Statistical Inferences and Visualization
Based on a Scale-Space Approach

by

Amy Vaughan

(Under the direction of Cheolwoo Park)

Abstract

SiZer (SIgnificant ZERo crossing of the derivatives) is a scale-space visualization tool for statistical inference that is originally developed by Chaudhuri and Marron (1999). It is an exploratory data analysis tool that uses local linear smoothing and convolutions with Gaussian kernel weights for inference and takes its views across a range of bandwidths. A number of authors also later go on to develop versions of this visualization tool that can account for two regression curves, instead of only one, and also dependent data. Motivated by these works, in this dissertation we will introduce a graphical method for the test of the equality of the mean of multiple time series based on SiZer. We will conduct a broad numerical study to demonstrate the sample performance of the proposed tool. In addition, we will investigate asymptotic properties of SiZer for the comparison of two time series.

As an extension of this original one dimensional SiZer, Godtliebsen et al. (2004) propose a two dimensional SiZer. This creates a tool that gives analysts the ability to look at images at a number of different resolutions, or bandwidths. We will introduce an inferential tool, called Spatial SiZer, that takes into account the detection of trends within datasets that have spatially dependent error structure. Also, we will compare several multiple testing adjustment procedures by a simulation study. Finally, we will display the performance of Spatial SiZer through several numerical studies for simulated and real datasets.

INDEX WORDS: Comparison of multiple curves, Image analysis, Local linear smoothing, Multiple testing adjustment, Scale-space, SiZer, Time series, Weak convergence

Statistical Inferences and Visualization Based on a Scale-Space Approach

by

Amy Vaughan

B.A., Murray State University, 1999

M.S., Georgia Southern University 2001

M.B.A., Georgia Southern University, 2001

M.S., The University of Georgia, 2006

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2009

© 2009
Amy Vaughan
All Rights Reserved

Statistical Inferences and Visualization

Based on a Scale-Space Approach

by

Amy Vaughan

Approved:

Major Professor: Cheolwoo Park

Committee: Nicole Lazar
Bill McCormick
Lynne Seymour
Lily Wang

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2009

Dedication

I would like to dedicate my dissertation to my parents. You are two very different, but extraordinarily strong people.

Acknowledgments

First and foremost, I would like to thank my major professor, Cheolwoo Park. The amount of time you gave in order to help me succeed is far above all customary expectations and for that I am forever grateful. For your patience and giving nature, I am indebted to you always.

I would also like to thank Professors Xiangrong Yin, Nicole Lazar, Lynne Seymour, William McCormick, Yehua Li, and Lily Wang for serving on my committee. I would like to acknowledge the technical and administrative support I received from Daphney, Loretta, Connie, Jimmy, and Jesse. I also have appreciated the teaching support and guidance that I received from Chandler Pike and Chris Franklin.

For moral support, I would like to thank my friends Ana, Ozgu, Jennifer, Johnathan, Archan, Ross, Ellen, Ayanna, Elizabeth, Jasper, Jinae, Jien, and others. I am also extremely grateful for the love and support that I have received from my parents and my sisters throughout all of these years of schooling.

Last but not least, I would like to thank my best friend Kelli. You have been there for me through everything and have been my emotional backbone too many times to count. You are the true definition of a friend and my life without you I hope to never know.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

SiZer, SIgnificant ZERo crossing of the derivatives, is developed by Chaudhuri and Marron (1999), as an exploratory data analysis tool. It provides a way to look at data, in one or multiple populations, so that one may be able to uncover underlying structure in the data, test it against underlying assumptions or potential models, and detect possible anomalies. SiZer is a more advanced version of a basic statistical graphic, such as a plot or chart, that can simultaneously look at data across a range of different bandwidths. It takes a nonparametric approach at smoothing curves and is a color coded tool that can provide still frame slides, or interactive movies to show the progression of information provided by varying levels of resolution in scale-space. Scale-space is proposed by Lindeberg (1994) as a formal theory which can handle image structures at different scales, by representing an image as a one-parameter family of smoothed images. The scale-space representation is parameterized by the size of the smoothing kernel used for suppressing fine-scale structures. SiZer assists in the determination of which features are "really present" in a dataset by constructing a so-called SiZer map, which visualizes statistical inference. It resolves the difficulty of gray areas, where the significance of certain features is debatable, and thus allows for informed inferences.

SiZer is based on scale-space ideas that provide kernel estimation across an extensive range of bandwidths. Looking at this entire range sidesteps a common statistical problem of attempting to find an optimal bandwidth for smoothing. Instead, statistical inference can be done and all the information that is available at each individual level of resolution can be detected. This way, the emphasis is changed from finding significant features in noisy data of the "true underlying curve" to finding them in the "curve at that given level of resolution".

SiZer helps us to detect which features in a smooth actually represent a trend and which ones are just sampling noise artifacts. It detects the precise location of local extrema, peaks and valleys. These significant features are determined by the zero crossings of the derivative and marked by color changes when this slope is deemed to be significantly increasing or decreasing.

The original SiZer investigates densities and single regression curve estimation under the assumption of independent errors. The regression estimation uses local linear smoothing (Fan and Gijbels, 1996) and convolutions with Gaussian kernel weights. SiZer constructs confidence intervals for its statistical inference and the original tool uses the quantile based on the idea of the number of independent blocks. This quantile is later improved by Hannig and Marron (2006) using advanced distributional theory.

Park et al. (2004) propose a dependent SiZer that does not assume independent errors and can detect which apparently significant features in the SiZer are attributable to the presence of dependence in the dataset. This dependent SiZer extends the methodology to time series data and uses an assumed autocovariance function when performing goodness of fit tests. Although this allows one to see how the data differ from the assumed model, it is often difficult to know the autocovariance structure that should be used as the true model. This leads Rondonotti et al. (2007) to propose a time series SiZer, that uses an estimated autocovariance function to detect significant features while still considering the dependence structure. Park et al. (2009) develop an improved version of the time series SiZer by using the extreme value theory. They also propose a new autocovariance estimator that does not use pilot bandwidths and residuals from an estimate, but instead uses a differenced time series to further decrease the spurious pixels in a SiZer map.

Park and Kang (2008) extend SiZer to look at two independent regression curves and thus shift the focus away from the derivative of a curve and onto the differences in two curves. When looking at SiZer maps, the need for causality of the creation of extrema states that when progressing to a higher level of smoothing, peaks and valleys should disappear monotonically. To aid in the development of the independent multiple regression curves case here, in this dissertation we will propose some asymptotic properties that deal with the need for causality of the creation of extrema and weak convergence of the empirical scale surface.

The problem of testing the equality of nonparametric regression curves with independent errors has been widely studied in the literature. Relevant work in this area includes Härdle and Marron (1990), Hall and Hart (1990), Delgado (1993), Kulasekera (1995), Bowman and Young (1996), Kulasekera and Wang (1995), Neumeyer and Dette (2003), Munk and Dette (1998), Dette and Neumeyer (2001), and Pardo-Fernandez et al. (2007). Koul and Stute (1998) and Li (2006) study fitting a regression function in the presence of long memory.

In this dissertation, we will develop a SiZer tool which is capable of comparing multiple time series. In order to compare multiple time series, a SiZer tool based on regression function estimation is needed. This is an extension of the existing SiZer for time series (Rondonotti et al., 2007, and Park et al., 2009) since they are applicable to only one time series. Moreover, this is also an advancement of Park and Kang (2008) since they consider only the independent case. This proposed tool provides insight to the differences between the curves by combining statistical inference with visualization. The method presented here not only keeps the advantages of the original SiZer tools, but also extends their usefulness to a broader range of scientific problems.

Our view of SiZer is also expanded from one to two-dimensions. The statistical inference of image analysis becomes difficult here since overlays are no longer possible. Godtliebsen et al. (2004) propose a two-dimensional version that replaces the focus on derivatives, or slopes of a curve, with the partial derivatives, or local slopes of a surface. However, this tool does not take into account the possible spatial dependence structure. In this dissertation, a SiZer tool is developed which can achieve this goal. Several multiple testing adjustment procedures are also considered, as are the performance of various levels of spatial dependence and bandwidths. This tool will aid statisticians in solving a much wider variety of tangible problems, such as functional Magnetic Resonance Imaging (fMRI) data and Satellite image data.

Chapter 2 reviews SiZer tools, including the original, dependent, and time series SiZer. Chapter 3 introduces the comparison of two or more regression curves, including the independent and dependent cases; simulations when the dependence structure is known and unknown; real data analysis; and asymptotic results. Chapter 4 covers a two-dimensional SiZer, which includes review of the independent case, introduction of the dependent case, and addresses multiple comparisons with a look at several different quantiles in order to compare the control of Type I error rates within the original and proposed SiZer cases. Chapter 5 discusses the ideas that have been proposed and the performance of the tools which are presented.

Chapter 2

Review of SiZer Tools

In this chapter, three SiZer tools are reviewed: original SiZer, dependent SiZer, and SiZer for time series.

2.1 Original SiZer

Curve estimation using nonparametric smoothing techniques is an effective tool for unmasking important structures from noisy data. The usual approach in the statistics literature, focuses on the "true underlying function," $f(x)$. Let $\hat{f}_h(x)$ be a kernel function estimator of $f(x)$ with a bandwidth h. A problem in nonparametric kernel estimation is that $E[\hat{f}_h(x)]$ is not necessarily equal to $f(x)$, so there is an inherent bias. This problem does not appear in classical parametric statistics, where one assumes a "correct" parametric model for $f(x)$ with parameters that can be unbiasedly estimated.

SiZer shifts the attention away from the "true underlying curve," $f(x)$, to the "true curves viewed at different scales of resolution", which is $E[\hat{f}_h(x)]$. Here, $E[\hat{f}_h(x)]$ is a "smoothed version" of the function $f(x)$ and can be viewed as the theoretical scale-space surface if $\hat{f}_h(x)$ is considered as the empirical scale-space surface. The empirical version here is by definition unbiased for the theoretical version. For a given function f (that is, underlying signal), various amounts of signal blurring (at least some is present in any real visual system) are represented by the convolution $f * K_h$ for different values of h. In fact, this family of convolutions (family of smooths) becomes the focus of the analysis, with the idea that this is all of the information that is available from a finite amount of data in the presence of noise at that resolution. This is very different from the classical statistical approach, where the focus is f. Confidence intervals are sought for the scale-space version $f_h'(x) \equiv E\hat{f}_h'(x)$. (For regression, this E is taken to be conditional on a set of predictor variables.) The center point of such intervals is automatically correct.

The methodology is motivated by "scale-space" ideas from computer vision (Lindeberg, 1994), with the idea that this contains all the information available in the data when working within that bandwidth. Instead of trying to find the optimum bandwidth for smoothing the data, a problem in classical statistics, SiZer focuses simultaneously on a wide range of values of the smoothing parameter, h. Different levels of smoothing may reveal different pieces of useful information. A large value of the smoothing parameter models "macroscopic or distant vision", where one can hope to resolve only large scale features. Similarly, a small value of the smoothing parameter models "microscopic vision" that can resolve small scale features. The smoothed version of the target curve is used to figure out which features visible in a smooth are "really there".

Local extrema (peaks and valleys) of the curve $\hat{f}_h(x)$ for fixed h are determined by the zero crossings of the derivative $[d\hat{f}_h(x)/dx]$. It is important that as one moves from lower to higher levels of smoothing, these structures should disappear monotonically in the scale-space surface. The smoothing method should not create "spurious structures" when going from a finer to a coarser scale. One-dimensional kernels should have the property that they do not increase the number of local extrema in any signal under convolution. Any structure that would appear at a higher level of smoothing than was previously viewed would be a falsely discovered local extremum.

SiZer can be used for both kernel density estimation and regression function estimation. The kernel density estimator based on univariate data $X_1, X_2, \ldots, X_n$ is

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where K is a kernel function, which is usually taken to be a smooth density symmetric around zero. The fact that the number of peaks in a kernel density estimate based on a Gaussian kernel $K(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$ decreases monotonically with the increase in the bandwidth, where the Gaussian kernel is the only kernel to always possess such a property, (for example, see the proof of Theorem 1 in Section 3.2) makes it especially necessary to use as the kernel here, because of the desire to not create spurious structures. The regression problem is based on data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$. In the regression case, use can be made of either the Priestley-Chao (Priestley and Chao, 1972) estimate; the Gasser-Müller (Gasser and Müller, 1984) estimate; or the local linear estimate (Wand and Jones, 1995), among others. Among these options SiZer utilizes the local linear smooths since they have some preferable properties, for example, boundary adjustment (Fan and Gijbels, 1996).

Suppose that the nonparametric regression model is given by

$$Y_i = f(X_i) + \sigma(X_i)\epsilon_i, \quad i = 1, \ldots, n$$

where ϵ_i's are independent errors with mean 0 and variance 1 and $\sigma^2(x) = Var(Y_i|X_i = x)$. In the local linear approach, the regression function is approximated by a series of local weighted least squares fits. That is, at a particular point x_0, the estimates are obtained by minimizing

$$\sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1(x_0 - X_i))]^2 K_h(x_0 - X_i)$$

over $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, where $K_h(\cdot) = K(\cdot/h)/h$. Using a Taylor expansion, it is easy to show that the solution of the regression function above provides estimates of a regression function and its first derivative at x_0 for different bandwidths; that is, $\hat{\beta}_0 \approx f_h(x_0) = K_h * f(x_0)$, and $\hat{\beta}_1 \approx f'_h(x_0) = K'_h * f(x_0)$ where * denotes the convolution. More specifically,

$$\hat{\boldsymbol{\beta}} = (X^T W X)^{-1} X^T W Y$$

where $Y = (Y_1, \ldots, Y_n)^T$, the design matrix of the local linear fit at x_0 is

$$X = \begin{pmatrix} 1 & (x_0 - X_1) \\ 1 & (x_0 - X_2) \\ \vdots & \vdots \\ 1 & (x_0 - X_n) \end{pmatrix},$$

and W =diag($K_h(x_0 - X_i)$). From this solution, the family of smooths parameterized by h can be constructed, in addition to the confidence intervals that underlie the SiZer analysis which determine the presence of significant structures. In order to achieve reasonable computational speed, fast binned implementation of the smoothers and the corresponding hypothesis tests are used. Binning allows for the repeated calculations of smoothers to become a rapidly computed discrete convolution when the data used are bin counts on an equally spaced grid. SiZer does allow for simple binning and linear binning to be done along a grid of 401 grid points where each value is replaced by the nearest grid point or midpoint of the bin. The number of grid points is originally chosen by Fan and Marron (1994), where they find that fewer than 400 grid points often results in distracting granularity in the image and greater than 400 grid points give negligible improvements in the resolution.

The confidence intervals used in SiZer are of the form

$$\hat{f}_h'(x_0) \pm q(h)\widehat{SD}(\hat{f}_h'(x_0))$$

where $q(h)$ is an appropriate Gaussian quantile, which is discussed below. The proposed estimate of SD is motivated by the fact that the derivative estimator is a weighted sum of the observed responses, and the conditional (given $X_1, \cdots, X_n$) weighted sample variances are used; that is,

$$Var(\hat{f}_h'(x)|X_1, \ldots, X_n) = Var(n^{-1}\sum_{i=1}^{n} W_h(x, X_i)Y_i|X_1, \ldots, X_n) = \sum_{i=1}^{n} \sigma^2(X_i)(W_h(x, X_i))^2$$

where

$$W_h(x, X_i) = \frac{\{\hat{s}_2(x;h)(x - X_i) - \hat{s}_1(x;h)\}K_h(x - X_i)}{\hat{s}_2(x;h)\hat{s}_0(x;h) - \hat{s}_1(x;h)^2}$$

and

$$\hat{s}_r(x;h) = \frac{1}{n}\sum_{i=1}^{n}(x - X_i)^r K_h(x - X_i).$$

SiZer visually displays the statistical significance of features over both location x and scale h, in a SiZer map. As one moves down the y axis from top to bottom, the scale decreases, that is, the bandwidth, h, gets smaller. As one moves along the x axis from left to right, the sequential value of the observation (e.g. time) increases. The SiZer map is either a gray-scale or a color map, reflecting statistical significance of the slope at (x, h) locations in scale-space. At each (x, h) location, the curve is blue (color map)/black (gray-scale) if it is significantly increasing (corresponding confidence interval > 0), red/white where it is decreasing (corresponding confidence interval < 0), and purple/intermediate gray when the curve cannot be concluded to be either decreasing or increasing (corresponding confidence interval contains 0). Finally, if there is not enough information in the data set to make statements about significance at this scale-space (x, h) location, then no conclusion can be drawn, so gray/darker shade of gray is used to indicate that the data are sparse. Sparse data means that the effective sample size in the window is less than 5, where the effective sample size is defined as

$$ESS(x, h) = \frac{\sum_{i=1}^{n} K_h(x - X_i)}{K_h(0)}.$$

The original SiZer uses an approximate quantile that provides simultaneous confidence limits based on the "number of independent blocks". This quantile is based on the fact that when x_1 and x_2 are sufficiently far apart, so that the kernel windows centered at x_1 and x_2 are essentially disjoint, the estimates $\hat{f}_h'(x_1)$ and $\hat{f}_h'(x_2)$ are essentially independent, but when x_1 and x_2 are close together, the estimates are highly correlated. The simultaneous confidence limit problem is then approximated by m independent confidence interval problems, where m reflects the number of independent blocks. Here m is defined as

$$m = m(h) = \frac{n}{avg_{x \in D_h} ESS(x, h)}$$

where D_h is the set of x locations where the data are dense, $D_h = \{x : ESS(x, h) \geq 5\}$.

This aforementioned quantile is later found to be part of the cause of many spurious pixels in the SiZer maps, highlighting significant sections where there in fact were none. To counteract this problem, Hannig and Marron (2006) develop some advanced distributional theory in order to bring the number of spuriously highlighted pixels down to the desired $\alpha 100\%$ of cases. They propose a row-wise and a global adjustment in order to reduce the number of false positives to $\alpha 100\%$ of the row or $\alpha 100\%$ of the map, respectively. The row-wise adjustment, they recommend, due to the significant loss of power for the global adjustment, uses the quantile

$$q(h) \equiv C_R = \Phi^{-1}\left(\left(1 - \frac{\alpha}{2}\right)^{1/(\theta g)}\right), \tag{2.1}$$

where θ, the cluster index, is defined as

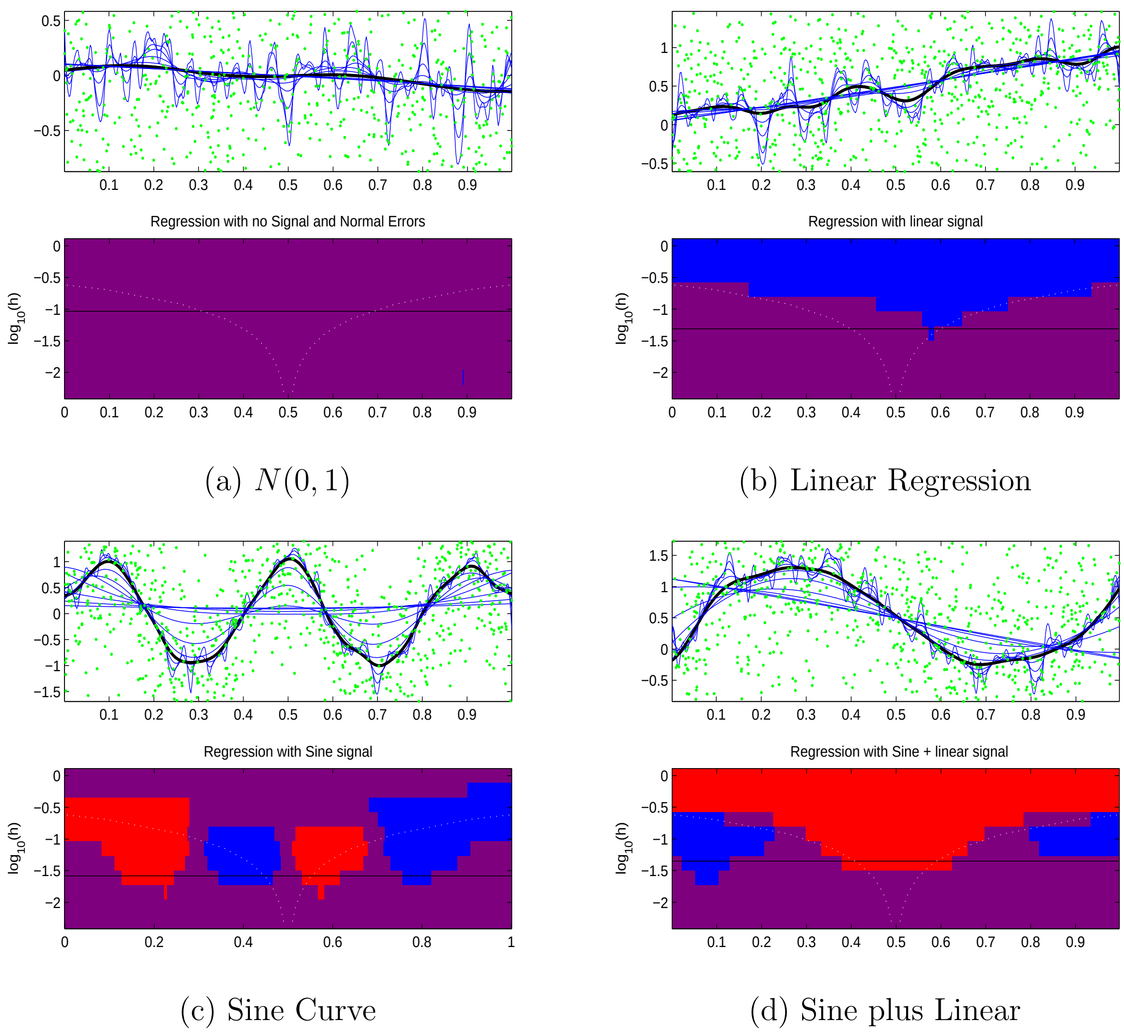
$$\theta = 2\Phi\left(\sqrt{3 \log g}\, \frac{\tilde{\Delta}}{2h}\right) - 1.$$

Here, $\tilde{\Delta}$ is the distance between the pixels of the SiZer map, g is the number of pixels on each row, h is the bandwidth used for the fixed row studied, and Φ is the standard normal distribution function. Suppose each T_i represents the hypothesis test statistic (modeled as a random variable) at each pixel location in the SiZer map. If the data contains no signal, then the probability that there is a spurious color on the jth row is

$$\begin{aligned}
& P[T_i < -C_R \;\text{ or }\; T_i > C_R \quad \text{for some } i = 1, \ldots, g] \\
\leq\; & P[\min(T_1, \cdots, T_g) < -C_R] + P[\max(T_1, \cdots, T_g) > C_R] \\
=\; & 2(1 - P[\max(T_1, \cdots, T_g) < C_R]) \\
\approx\; & 2(1 - \Phi(C_R)^{\theta g}) \\
=\; & \alpha
\end{aligned}$$

Thus, no more than about $\alpha 100\%$ of the rows will have spurious colors, as desired.

Figure 2.1 shows examples of SiZer plots. In each panel the top plot shows the original data points in green and smooths of the data at different bandwidths in blue. The individual thin blue curves in the top graph display the family of smooths; which are the kernel regression estimates of the curves viewed at different resolutions or bandwidths. Those smooths range from nearly raw data (small h, very wiggly thin line), to the limit as the window width goes to infinity (large h, nearly the simple least squares fit line). The solid black line is the optimal data-driven bandwidth as chosen by the method of Ruppert et al. (1995).



(a) $N(0,1)$ (b) Linear Regression

(c) Sine Curve (d) Sine plus Linear

Figure 2.1: Family of smooths plots (top panels) and SiZer maps (bottom panels) of some regression models.

The bottom plot in each panel is the SiZer plot for that data set, colored according to the description above. In each set of graphs, the horizontal locations in the graphs are the same in the top (family of smooths) and bottom (SiZer maps) panels, while the vertical locations in the SiZer maps correspond to the logarithm of bandwidths of the family of smooths. Where each curve in the top graph represents a different bandwidth, the range of these different bandwidths are plotted along the log scaled vertical axis in the SiZer maps below. The white dotted curves in the SiZer maps show effective window widths for each bandwidth, as intervals representing $\pm$ 2 bandwidths (that is, $\pm$ 2 standard deviations of the Gaussian kernel). Changes in color in the SiZer maps on the bottom are determined to occur when there is a zero crossing of the derivative, marked by a change in the sign of the slope. Again, the color scheme in the SiZer maps is blue (red) in locations where the curve is determined to be significantly increasing (decreasing), purple where the curve cannot be concluded to be either decreasing or increasing, and gray in regions where the data are too sparse to make statements about significance.

In Figure 2.1 (a), with no signal and normal errors, the SiZer map correctly identifies no significance (only noise) and marks the entire map as purple. In Figure 2.1 (b), the regression line shows in the smaller bandwidths as a constant increase. In Figure 2.1 (c), the sine curve has an estimate with a significantly increasing (blue) slope on the left, changes according to every turn in the sine curve, and then the slope ends by significantly decreasing (red) on the right edge of the graph. In Figure 2.1 (d), the sine curve is detected at moderate bandwidths; at the largest bandwidth, only the overall decrease of the linear trend can be detected.

2.2 Dependent SiZer

The dependent SiZer proposed by Park et al. (2004) compares the observed data with a specific null model being tested by using an assumed autocovariance function. This approach flags statistically significant differences between the data and a given null model. It uses a goodness of fit test to validate the null hypothesis without specifying an alternative hypothesis.

The original SiZer is not effective at differentiating between deterministic trends and natural variation by dependence in the time series. By adjusting the statistical inference with an autocovariance function, the dependent SiZer can account for this type of time series fluctuation and color the map appropriately. In this way, the dependent SiZer not only provides a goodness of fit test for an assumed model but also gives visual insight into how the data differ from the assumed model.

A time series model can be viewed in the regression setting as

$$Y_i = f(i) + \epsilon_i. \tag{2.2}$$

But the critical difference is that now the ϵ_i's are no longer independent, and thus

$$Cov(\epsilon_i, \epsilon_j) = \gamma(|i - j|).$$

Here, γ is an autocovariance function. In this case, the variance of the local linear estimator at i_0 is given by,

$$Var(\hat{\boldsymbol{\beta}}) = (X'WX)^{-1}(X'\Sigma X)(X'WX)^{-1}, \tag{2.3}$$

where, for the assumed correlation structure, Σ is the kernel weighted covariance matrix of the errors with generic element

$$\sigma_{ij} = \gamma(|i - j|)K_h(i - i_0)K_h(j - i_0). \tag{2.4}$$

When we take a look at time series models, they can contain autoregressive, moving average terms, or both. In combination, an autoregressive moving average model of

$$X_t = \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

has p autoregressive terms and q moving average terms. Here, $\phi_1, ..., \phi_p$ are the parameters for the autoregressive terms, $\theta_1, ..., \theta_q$ are the parameters for the moving average terms, and ϵ_t are the error terms.

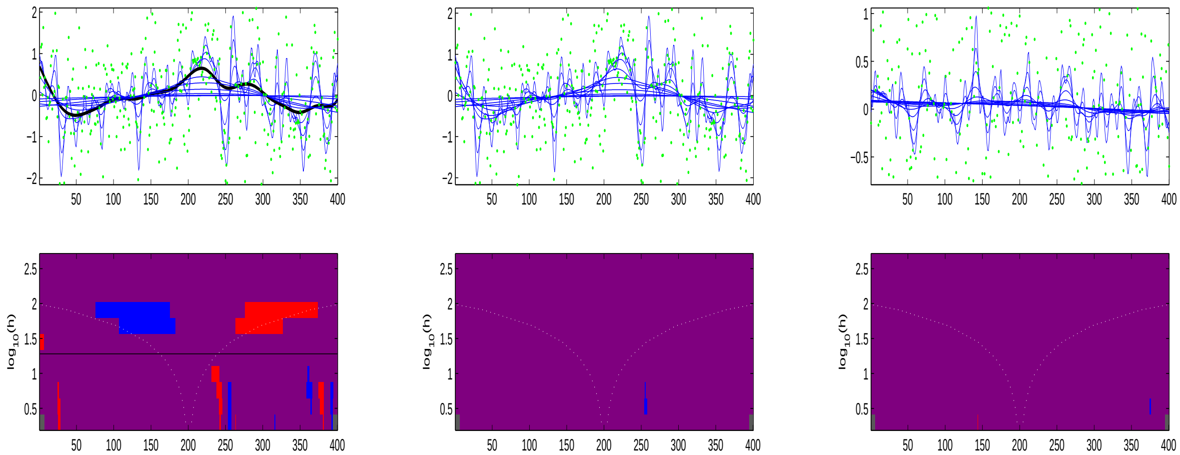


Figure 2.2: (a) Original SiZer plot of simulated AR(1) errors. The SiZer map in the lower panel flags some dependence artifacts as significant. (b) Dependent SiZer of simulated AR(1) errors when $\phi = 0.5$. (c) Dependent SiZer of simulated AR(1) errors when $\phi = -0.5$.

When the $\varepsilon_i's$ are drawn from AR(1), autoregressive dependency with lag of order 1 and a medium ϕ coefficient of 0.5 as an example, a careful look at the original SiZer map in Figure 2.2 (a) reveals some small unexpected colored regions that have mistakenly been identified as significant. In Figure 2.2 (b), when the dependence structure in the data is accounted for, almost all of these spuriously highlighted pixels have been corrected. In Figure 2.2 (c), when $\phi = -0.5$, the SiZer map reflects that the dependence structure has again been correctly detected with almost no spurious pixels present. This confirms that the dependent SiZer can successfully conduct a goodness of fit test for AR(1).

Figure 2.3 is given to point out the performance of the SiZer map at varying levels of dependence. In Figure 2.3 (a) when ϕ=0.50, the SiZer performs very well and has only a few highlighted pixels. In Figure 2.3 (b) when ϕ=0.90, again the SiZer correctly identifies almost the map in its entirety, with only a few spuriously highlighted pixels. In Figure 2.3 (c) when ϕ=0.95, the SiZer map does begin to show a weakness in identifying this level

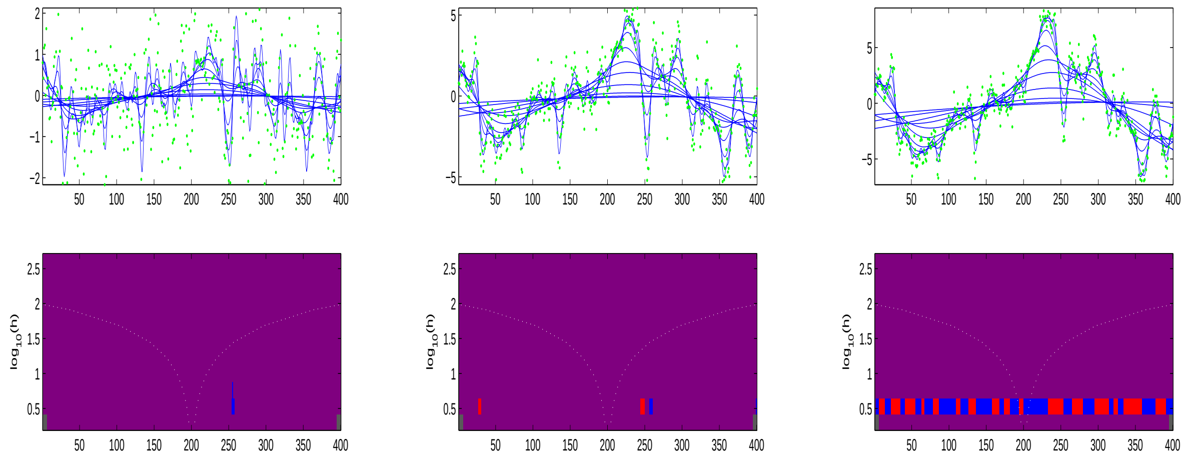


Figure 2.3: (a) Dependent SiZer of simulated AR(1) errors when $\phi = 0.50$. (b) Dependent SiZer of simulated AR(1) errors when $\phi = 0.90$. (c) Dependent SiZer of simulated AR(1) errors when $\phi = 0.95$.

of strong dependence in the map correctly. In the simulation, we use the given coefficient values, but in real data analysis we should estimate them. See Brockwell and Davis (2002) for the estimation.

2.3 SiZer for Time Series

Although the dependent SiZer can handle time series data, it assumes that the autocovariance function is known. Unlike in dependent SiZer, a new SiZer for time series, by Rondonotti et al. (2007), estimates γ by using the sample autocovariance function of the observed residuals from a "pilot smooth".

Let the same regression approach in (2.2) be used for SiZer for time series. The local linear fit approximates the regression function $f(i)$ and a sensible estimate of the variance is based on estimating γ by the sample autocovariance function of the observed residuals from a pilot smooth, using the pilot bandwidth h_p. A pilot bandwidth is used in order to estimate the function, then when a range of bandwidths is used to also estimate the function, graphs of the residuals between these two estimates can be obtained. One could take $h_p = h$, which means that h_p varies with h, but this would lead to a confounding of the different notions of scale and dependence structure. A small h_p assumes i.i.d. or weakly correlated errors, and a large one corresponds to strongly correlated errors. Because h and h_p are treated separately, another dimension needs to be added to the SiZer plot. This is approached through a series of SiZer plots, 11 total, 4 of which are chosen for viewing based on an Indicator of Residual component IR, which is a numerical ratio measure between the sum of the squared residuals, indexed by the pilot bandwidth h_p, to the sum of the squared residuals at the maximum pilot bandwidth. This residual measure represents the different trade-offs available between trend and dependence. When the pilot bandwidth is large, the dependence component of the data appears strongly in the residuals as noise and the IR is close to its maximum of 1 and when the pilot bandwidth is small, the dependence component of the data appears strongly in the pilot smooth as trend and the IR is close to its minimum of 0.

An example of these SiZer plots is shown in Figure 2.4. This is a generated MA(1), moving average of order 1 with a medium coefficient of $\theta = 0.5$, time series with signal $f(i) = \sin(6\pi i/n)$ where $n = 401$. The first panel in the first row shows the generated data. The second panel includes the data with pilot smooths. The last panel gives the IR for the 11 pilot bandwidths; of these, four are chosen to be displayed in the following rows of the plot. The chosen four always include the second plot, for which IR is close to 0%, and the plots that correspond to where IR is 25%, 50% and 75%. The smallest bandwidth is always excluded because it often contains too much noise. Further right on the top of Figure 2.4 is the bar diagram using this information and in this case, the second, fourth, sixth, and seventh bandwidths are selected. The series of plots in the second and third rows represent, respectively, the local linear fits and the residuals corresponding with the selected bandwidths.

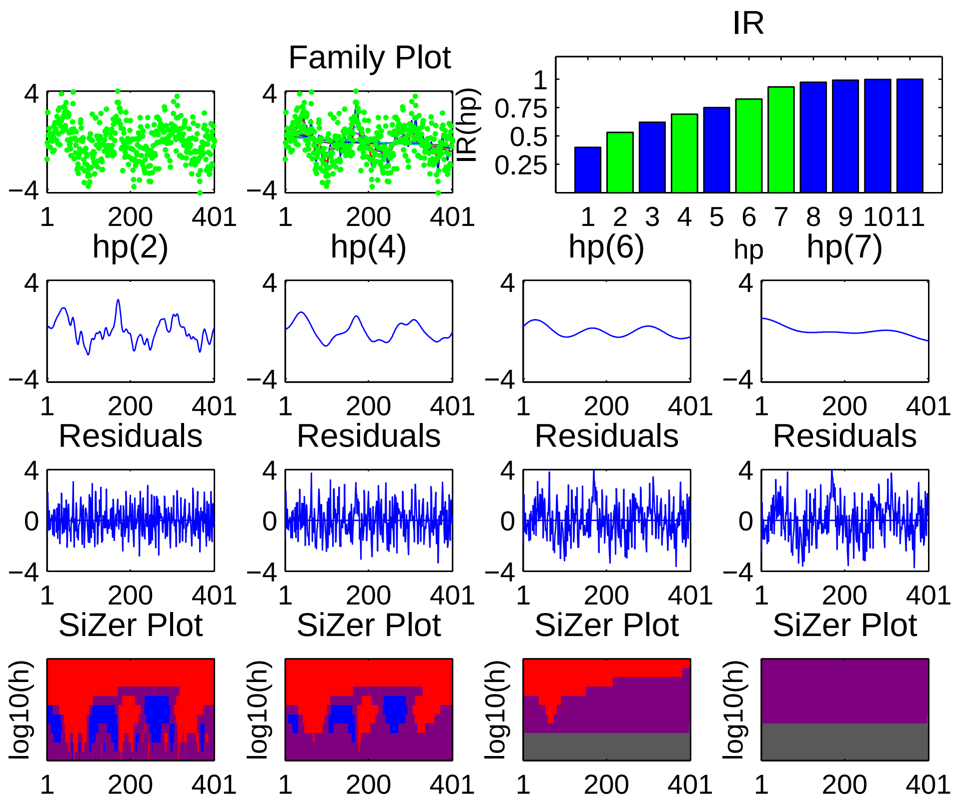


Figure 2.4: SiZer for time series: Sine plus MA(1)

The first SiZer map (corresponding to $h_p(2)$) shows significant features along the sine curve. Note that as we move to the other SiZer maps (that is, $h_p(4), h_p(6)$, and $h_p(7)$), an increasing amount of correlation appears in the error component, so that fewer features are significant at every level of resolution. Also, at the fine levels of resolution of the third and fourth maps there is less perceived useful information in the data, which means more data sparsity, thus more bottom lines of the SiZer plots are shaded gray. Since MA(1) is weakly correlated, it is reasonable to interpret the first or second SiZer map.

While this original SiZer for time series is useful, there is still room for improvement. The estimation of the quantile for the confidence interval relies on a heuristic idea rather than on theory, and the estimation of the autocovariance function is not accurate in some situations. Theoretical properties of the proposed method are also not provided. Park et al. (2009) aim to remedy these problems in a moderately correlated time series. They propose to estimate the quantile by extreme value theory and the autocovariance function based on differenced time series in scale-space. Weak convergence of the empirical scale-space surface to its theoretical counterpart is established in their paper under appropriate regularity conditions. These improvements should also help in reducing the number of spurious features that are flagged as significant.

In order to improve the quantile estimator, Park et al. (2009) extend the result of Hannig and Marron (2006). They use the same quantile as in (2.1), but now the cluster index is

$$\theta = 2\Phi\left(\sqrt{I \log g}\,\frac{\tilde{\Delta}}{h}\right) - 1, \tag{2.5}$$

where

$$I = \frac{\int \gamma(sh/\Delta)\, e^{-s^2/4}\, \frac{12-12s^2+s^4}{16}\, ds}{\int \gamma(sh/\Delta)\, e^{-s^2/4} \left\{1 - \frac{s^2}{2}\right\} ds}.$$

Park et al. (2009) also propose a new autocovariance estimator to fix spurious features when a time series has moderate correlation. Since the proposed estimator does not require a pilot bandwidth, there is no need to select bandwidths to display. The original SiZer for time series uses residuals obtained from pilot bandwidths to estimate an autocovariance function, with weights in a local linear estimate of f. This leads to an autocovariance estimate γ^* from the residuals which is not equal to the original γ, which is responsible for leftover spurious features.

To address this issue Park et al. (2009) do not estimate the covariance from the estimated residuals $\hat{\epsilon_i}$. Instead, they estimate the covariance structure directly from a (possibly several times) differenced time series to remove a trend in the data. One of the advantages of this approach is that the estimator of the covariance no longer depends on the pilot bandwidth. This is a big advantage because it is not necessary to interpret several SiZer maps at the same time nor to select some bandwidths for further investigation. Let e_i be the differenced time series, that is, $e = Ay$ where A is the difference matrix, e.g.,

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

if the first difference is used. A simple calculation shows for all i, j that

$$Cov(e_i, e_j) = \sum_{k=1}^{n} a_{i,k}a_{j,k}\gamma(0) + \sum_{k=1}^{n-1}(a_{i,k}a_{j,k+1} + a_{i,k+1}a_{j,k})\gamma(1) + ... + (a_{i,1}a_{j,n} + a_{i,n}a_{j,1})\gamma(n-1).$$

From this we can set

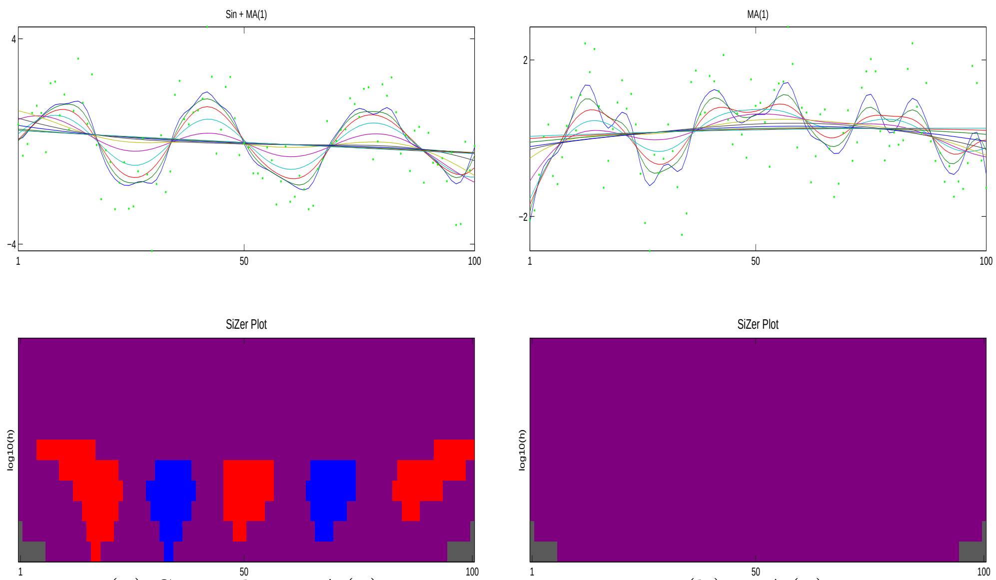
$$e_ie_j = \sum_{k=1}^{n} a_{i,k}a_{j,k}\gamma(0) + \sum_{k=1}^{n-1}(a_{i,k}a_{j,k+1} + a_{i,k+1}a_{j,k})\gamma(1) + ... + (a_{i,1}a_{j,n} + a_{i,n}a_{j,1})\gamma(n-1) + \delta_{ij},$$

where $E\delta_{ij} \approx 0$. Thus, there are n^2 equations and n variables. Estimating γ by least squares fails because the above equation does not lead to a stable solution. Therefore, it is necessary to regularize the problem. First, since $\gamma(0) \geq |\gamma(i)|$ for each i, one must consider only such solutions. Additionally, Park et al. (2009) regularize the least squares problem by introducing the penalty $\lambda \sum_{i=1}^{n-1} i\gamma(i)^2$. The weight i is motivated by the belief that the covariance $\gamma(i)$ should be decaying as i increases.

This leads to the following constrained ridge regression

$$\arg\min_{\gamma \in R} \left\{ \sum_{i,j} \left(e_i e_j - \sum_{k=1}^{n} a_{i,k} a_{j,k} \gamma(0) - \sum_{k=1}^{n-1} (a_{i,k} a_{j,k+1} + a_{i,k+1} a_{j,k}) \gamma(1) \right.\right.$$
$$\left.\left. - \cdots - (a_{i,1} a_{j,n} + a_{i,n} a_{j,1}) \gamma(n-1) \right)^2 + \lambda \sum_{i=1}^{n-1} i \gamma(i)^2 \right\},$$

where $R = \{\gamma : \gamma(0) \geq |\gamma(i)|, i = 1, \ldots, n-1\}$. They investigate several choices of λ and find that $\lambda = 1$ works well as long as the time series is weakly to only moderately dependent.



(a) Sine plus MA(1) (b) MA(1)

Figure 2.5: Fewer spurious features appear in the SiZer maps based on Park et al. (2009) new proposed quantile and autocovariance estimator.

Comparing the plot on the left in Figure 2.5 to the one in Figure 2.4, much improvement is seen. In Figure 2.4, the SiZer map flags the sine trend fairly well, however, there are some spurious features that are highlighted where the global downward trend, due to MA(1) with θ=0.5 dependence, is flagged as significant in the red color that appears at large resolutions. In Figure 2.5 (a), this dependence structure is correctly accounted for and the spurious pixels are no longer highlighted. In Figure 2.5 (b), the SiZer map also correctly colors the entire map purple with no spuriously highlighted pixels since there is no deterministic trend, just MA(1) with θ=0.5 errors.

Chapter 3

SiZer for the Comparison of Multiple Regression Curves

In this chapter, an important problem in statistical inference will be addressed, namely comparing two or more populations. Here, in the spirit of SiZer, the comparison will be done nonparametrically and using a scale-space approach. The statistical challenge in this problem is in testing whether there is any statistically significant differences of the population curves. Section 3.1 reviews Park and Kang (2008) for the comparison of two curves and Section 3.2 provides theoretical justification newly developed in this dissertation. Section 3.3 proposes a SiZer for the case of two time series and demonstrates the performance of SiZer both when the dependence structure is known and when it must be estimated. Also newly presented are two asymptotic properties to support the convergence of the empirical and theoretical scale-space surfaces. In Section 3.4, a SiZer for the analysis of more than two time series is newly proposed and its performance is evaluated.

3.1 Review of the Independent Case: Two Regression Curves

Suppose that there are two different samples and $n = n_1 + n_2$ independent observations from the following regression models:

$$Y_{ij} = f_i(X_{ij}) + \sigma_i(X_{ij})\epsilon_{ij}, \quad j = 1, \ldots, n_i, \ i = 1, 2, \tag{3.1}$$

where X_{ij}'s are covariates, the ϵ_{ij}'s are independently distributed random errors with mean 0 and variance 1, $f_i(X_{ij}) = E(Y_i|X_{ij})$ is the unknown regression function of the ith sample and $\sigma_i^2(X_{ij}) = Var(Y_i|X_{ij})$ is the conditional variance function of the ith sample ($i = 1, 2$).

Park and Kang (2008) expand SiZer to consider the nonparametric comparison of two regression curves f_1 and f_2. Within this context, SiZer represents the SIgnificant ZERo crossing of the *differences*, since now significance is determined by the difference of two smoothed functions. Hypothesis tests are for

$$H_0 : f_{1,h}(x_0) = f_{2,h}(x_0) \text{ vs. } H_1 : f_{1,h}(x_0) \neq f_{2,h}(x_0)$$

for a fixed point x_0. SiZer visually displays the significance of differences between two regression functions in families of smooths $\hat{f}_{i,h}(x), i = 1, 2$ over both location x and scale h, using a color map. It is based on confidence intervals for $\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)$. The formula for these confidence intervals is

$$\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x) \pm q(h) \cdot \widehat{SD}(\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)),$$

where $q(h)$ is an appropriate quantile, using the advanced theory developed by Hannig and Marron (2006) discussed previously. The quantile is as in (2.1) and here θ is the cluster index given by

$$\theta = 2\Phi\left(\sqrt{\log g}\,\frac{\tilde{\Delta}}{2h}\right) - 1.$$

For the estimation of the standard deviation, $\hat{f}_{i,h}(x)$ can be written as

$$\hat{f}_{i,h}(x) = \frac{1}{n_i}\sum_{j=1}^{n_i} W_{i,h}(x, X_{ij})Y_{ij}$$

where

$$W_{i,h}(x, X_{ij}) = \frac{\{\hat{s}_2(x; h) - \hat{s}_1(x; h)(x - X_{ij})\}K_h(x - X_{ij})}{\hat{s}_2(x; h)\hat{s}_0(x; h) - \hat{s}_1(x; h)^2}$$

and

$$\hat{s}_r(x; h) = \frac{1}{n_i}\sum_{j=1}^{n_i}(x - X_{ij})^r K_h(x - X_{ij}).$$

Then,

$$Var(\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)) = Var(\hat{f}_{1,h}(x)) + Var(\hat{f}_{2,h}(x)),$$

and

$$Var(\hat{f}_{i,h}(x)) = \frac{1}{n_i^2}\sum_{j=1}^{n_i} \sigma_i^2(X_{ij})(W_{i,h}(x, X_{ij}))^2.$$

To estimate $\sigma^2(x)$, we use a simple smooth of the residuals; for example,

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n} \hat{e}_i^2 K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)},$$

where $\hat{e}_i = Y_i - \hat{f}_h(X_i)$.

To demonstrate this SiZer for two regression curves, we present the following three examples. In these examples X_1 and X_2 are generated from a $U(0,1)$ distribution and we take $n_1 = 1000$ and $n_2 = 2000$. In Figure 3.1, graphs (a) through (f) show the original data series of the three examples. The first example, in Figures 3.1 (a) and (d), has the same constant mean 0 with independent $N(0,1)$ errors:

$$Y_{ij} = \varepsilon_{ij},\ j = 1, ..., n_i,\ i = 1, 2,$$

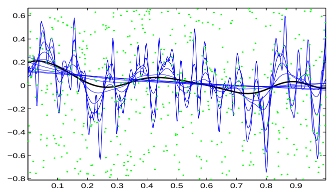
where $\varepsilon_{ij} \sim N(0,1)$ for $i = 1, 2$. In the top panel of (g), the thin curves display the family of smooths, which are the differences of the two local linear smooths, $\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)$. These differences are located around 0 because both samples have zero constant functions and the same noise distribution. The SiZer map in the lower panel correctly shows no significant difference and displays the solid color of purple. In the second example, Figures 3.1 (b) and (e), one sample has a mean of 2 and the other has mean 0 with both having error distribution $N(0,1)$:

$$Y_{1j} = 2 + \varepsilon_{1j}, \quad \text{and} \quad Y_{2j} = \varepsilon_{2j}$$

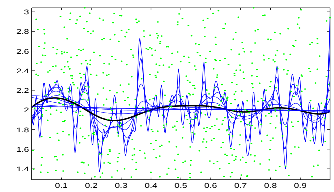
where again, $\varepsilon_{ij} \sim N(0,1)$ for $i = 1, 2$. The upper panel of plot (h) shows the difference of two smooths is approximately 2, which corresponds to the difference of the two true regression functions. The SiZer map shows positive differences, colored blue, across almost all scales since the mean of the first sample is greater than that of the second sample by 2. In the third example, the regression functions are:

$$Y_{1j} = \sin(6\pi X_{1j}) + \varepsilon_{1j} \quad \text{and} \quad Y_{2j} = \varepsilon_{2j}$$

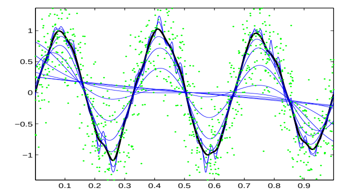
where $\varepsilon_{ij} \sim N(0, 0.25)$ for $i = 1, 2$. The difference of the two smooths clearly reveals the sine curves in the top panel of Figures 3.1 (c) and (f), and the SiZer map in plot (i) shows



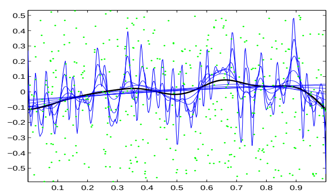
(a) $N(0,1)$



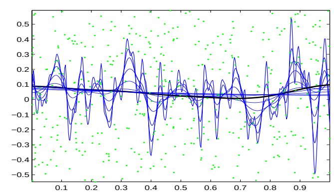
(b) $N(2,1)$



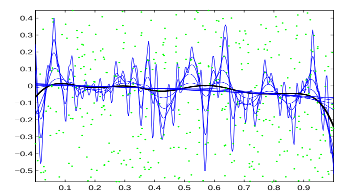
(c) Sine+$N(0,0.25)$



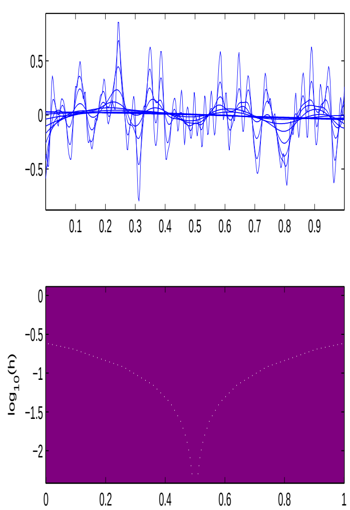
(d) $N(0,1)$



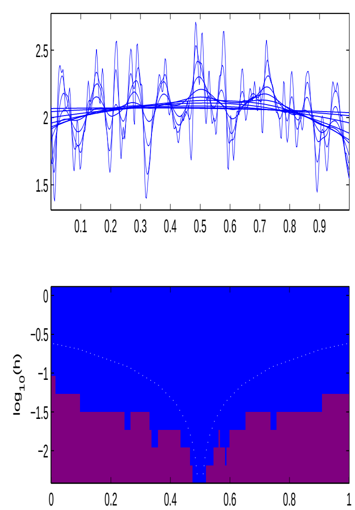
(e) $N(0,1)$



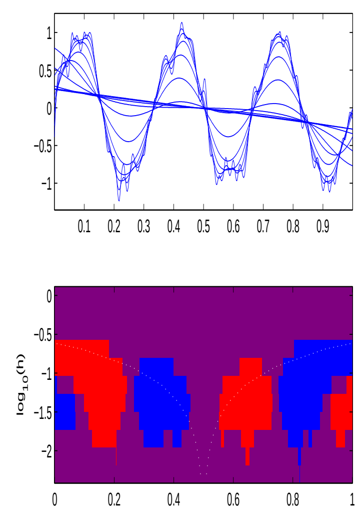
(f) $N(0,0.25)$



(g) SiZer plot



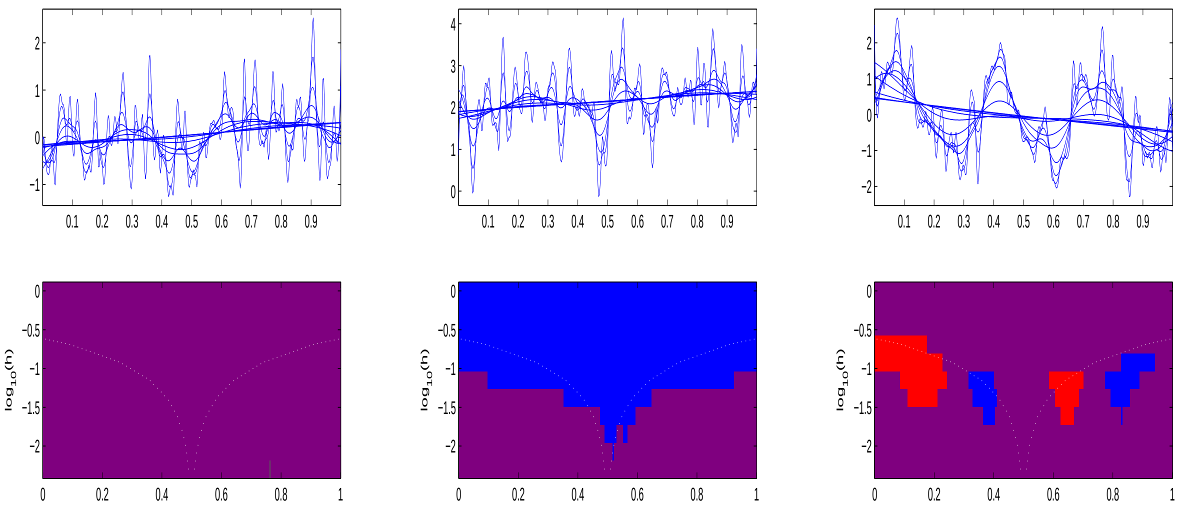
(h) SiZer plot



(i) SiZer plot

Figure 3.1: Original data from the first example ((a) and (d)), the second ((b) and (e)), and the third ((c) and (f)). SiZer plots for comparing two regression curves. The two samples are drawn from (g) normal errors with the same mean ((a) and (d)), (h) normal errors with different means ((b) and (e)), and (i) normal errors with a sine curve versus a constant mean ((c) and (f)).

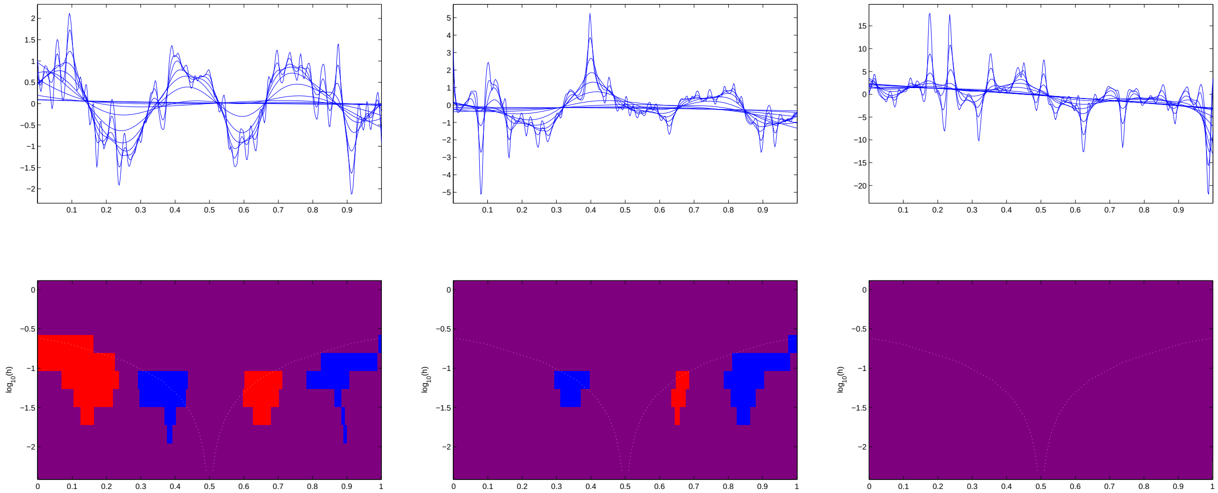
positive (blue) and negative (red) differences along the sine curve. These graphs show that SiZer can successfully detect differences between two regression curves in various settings.



(a) (b) (c)

Figure 3.2: SiZer plots for comparing two regression curves when the errors have very different variances. The two samples are drawn from (a) normal errors with the same mean, (b) normal errors with different means, and (c) normal errors with a sine curve versus a constant mean.

Figure 3.2 demonstrates the performance of the SiZer with the same regression curves viewed in Figure 3.1, now with very different error variances. In Figure 3.2 (a), there are the same constant means, but now with $\varepsilon_{1j} \sim N(0,1)$ and $\varepsilon_{2j} \sim N(0,16)$. The SiZer map again correctly identifies no significant difference as in Figure 3.1 (a). In Figure 3.2 (b), one sample has a mean of 2 and the other has a mean of 0, now with $\varepsilon_{1j} \sim N(0,1)$ and $\varepsilon_{2j} \sim N(0,16)$. The SiZer map again shows positive differences, colored blue, across almost all scales denoting only the greater mean of the first sample and not the noise of the second sample. In Figure 3.2 (c), one sample has a sine curve and the other has a constant mean of 0, now with $\varepsilon_{1j} \sim N(0,0.25)$ and $\varepsilon_{2j} \sim N(0,16)$. The SiZer map displays its ability to detect the trend and not the noise and shows the alternating color scheme along the sine curve.



(a) Sine+$t(3)$ and $t(4)$ (b) Sine+$t(2)$ and $t(3)$ (c) Sine+$t(1)$ and $t(2)$

Figure 3.3: The third example with various t distributions.

In order to see the effect of non-normal errors a redisplay of the third example with the sine curve trend is given now with various t distributions. In Figure 3.3 (a), although we use the t distributions with the degrees of freedom 3 and 4, SiZer map flags the sine curve as significant, which demonstrates robustness of the tool. However, as we decrease the degrees of freedom (equivalently, as the tail parts of the distribution get thicker), it is more challenging for SiZer to capture the trend as can be seen in Figures 3.3 (b) and (c).

3.2 Asymptotic Properties: Independent Case

This section discusses the development of the theoretical justification for the method proposed by Park and Kang (2008) described in 3.1. One of the main focuses in nonparametric curve estimation is that of structures, such as peaks and valleys. When performing this type of estimation, it is important that as one goes from lower to higher levels of smoothing within the scale-space surface, these structures should disappear monotonically. The smoothing method should not create spurious structures when going from a finer to a coarser scale.

There must be a "causality" for the creation of extrema, and new structures should not be produced with additional smoothing. In what follows, it is assumed that x varies in a subinterval J of $(-\infty, \infty)$ and h varies in a subinterval H of $(0,\infty)$. The data are binned to give new data values $(v_j, \tilde{Y}_{ij})$ where the v_j's $(j = 1, \ldots, m)$ are the midpoints of each bin and the $\tilde{Y}_{ij}$'s are the bin averages. Note that one can choose between the simple binning and linear binning (Fan and Marron, 1994) in running a SiZer code, but we use the simple binning in the proof for simplicity. Using the binned values,

$$\begin{aligned}
\hat{g}_h(x) \equiv \hat{f}_{1,h}(x) - \hat{f}_{2,h}(x) &= \frac{1}{mh}\sum_{j=1}^{m} \tilde{Y}_{1j} K\left(\frac{x-v_j}{h}\right) - \frac{1}{mh}\sum_{j=1}^{m} \tilde{Y}_{2j} K\left(\frac{x-v_j}{h}\right) \\
&= \frac{1}{mh}\sum_{j=1}^{m} (\tilde{Y}_{1j} - \tilde{Y}_{2j}) K\left(\frac{x-v_j}{h}\right) \\
&= \frac{1}{mh}\sum_{j=1}^{m} Z_j K\left(\frac{x-v_j}{h}\right)
\end{aligned}$$

where $Z_j = \tilde{Y}_{1j} - \tilde{Y}_{2j}$ and $K(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. Silverman (1981) proves that convolutions with Gaussian kernels have the number of their zero crossings of the derivative smooth is always a decreasing function of h. Presented here is a theorem parallel to that of Silverman's (1981) for the difference between two nonparametric regression problems.

Theorem 1. *Assume that the scale-space surface $\hat{f}_{i,h}(x)$ arises as a local linear regression problem, $\hat{g}_h(x)$ using $\tilde{Y}_{ij}$'s, binned values, and a Gaussian kernel K. Then for each fixed $h \in H$ and $i = 1, 2$, the number of zero crossings of $\hat{g}_h(x)$ will be a decreasing and right continuous function of h. Furthermore, the same result holds for $E\{\hat{g}_h(x)\}$ of the theoretical scale-space surface.*

Proof.

Let us denote the theoretical scale-space surfaces $E[\hat{g}_h(x)]$ by $g_h(x)$.

Note that for the Gaussian kernel $K(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$,

$$[\hat{g}_{h_1}(x)] * K(x/h_2) = \hat{g}_{\sqrt{h_1^2+h_2^2}}(x) \text{ and } [g_{h_1}(x)] * K(x/h_2) = g_{\sqrt{h_1^2+h_2^2}}(x)$$

for all $h_1, h_2 > 0$. Here $*$ denotes the usual convolution, and note that the fact that $K_{h_1} * K_{h_2}(x) = K_{\sqrt{h_1^2+h_2^2}}(x)$ is used. Now it follows from total positivity of the Gaussian kernel and the variation diminishing property of functions generated by convolutions with totally positive kernels [see Schoenberg (1950), Karlin (1968)] that the number of sign changes in $\hat{g}_h(x)$ will be a monotonically decreasing function of h. Suppose next that $\hat{g}_{h_0}(x)$ has $k \geq 0$ sign changes for some fixed $h_0 > 0$. Then arguing as in Silverman (1981), it is easy to see using the continuity of $\hat{g}_h(x)$ as a function of h and x that there exists $\epsilon > 0$ such that for all $h \in [h_0, h_0 + \epsilon)$, $\hat{g}_h(x)$ will have at least k sign changes. Hence the monotonic decrease in the number of sign changes as h increases implies that the number of sign changes in $\hat{g}_h(x)$ will be exactly equal to k for all $h \in [h_0, h_0 + \epsilon)$. An identical argument can be given for the number of sign changes in $g_h(x)$. This completes the proof of right continuity.

We now consider the statistical convergence of empirical scale-space surfaces to their theoretical counterparts. Consider regression problems based on independent (binned) observations $(v_1, \tilde{Y_{i1}}), (v_2, \tilde{Y_{i2}}), \cdots, (v_m, \tilde{Y_{im}})$ and assume that $\hat{f}_{i,h}(x)$ has the form $m^{-1} \sum_{i=1}^{m} \tilde{Y}_i W_m(h, x, v_i)$, where W_m is a smooth weight function that arises from the kernel function in usual kernel regression or kernel weighted local polynomial regression with bandwidth h. Thus, $\hat{g}_h(x) = m^{-1} \sum_{j=1}^{m} Z_j W_m(h, x, v_j)$ where $Z_j = \tilde{Y}_{1j} - \tilde{Y}_{2j}$, and $E\hat{g}_h(x) = m^{-1} \sum_{j=1}^{m} E(Z_j) W_m(h, x, v_j)$.

The following Theorem yields the weak convergence of the empirical scale-space surfaces under the i.i.d. setting. It is worth noting that the conditions assumed on the weight function in Theorem 2 are satisfied for many standard kernel regression estimates and kernel weighted local polynomial estimates for suitable distributions (X, Y).

Theorem 2. *Assume that*

$$E\{|Z_j - E(Z_j)|^{2+\rho}\} < \infty, \quad j = 1, \ldots, m$$

for some $\rho > 0$ *and* I *and* H *are compact subintervals of* $(-\infty, \infty)$ *and* $(0, \infty)$ *respectively. Assume that as* $m \to \infty$,

$$m^{-1} \sum_{j=1}^{m} Var(Z_j) W_m(h_1, x_1, v_j) W_m(h_2, x_2, v_j)$$

converges in probability to a covariance function $cov(h_1, x_1, h_2, x_2)$ *for all* (h_1, x_1) *and* (h_2, x_2) $\in H \times I$, *and*

$$m^{-(1+\rho/2)} \{\max_{1 \le j \le m} |W_m(h, x, v_j)|^{\rho}\} \sum_{j=1}^{m} \{W_m(h, x, v_j)\}^2 \to 0$$

in probability for all $(h, x) \in H \times I$. *Also, assume that as* h *varies in* H *and* x *varies in* I, $Var(Z_j)\{W_m(h, x, v_j)\}^2$ *will be uniformly dominated by a positive function* $M(v_j)$ *such that* $\sup_{j \ge 1} M(v_j) < \infty$. *Then as* $m \to \infty$, *the 2-parameter stochastic process*

$$m^{1/2}[\hat{g}_h(x) - E\hat{g}_h(x)]$$

with $(h, x) \in H \times I$ *converges weakly to a Gaussian process on* $H \times I$ *with zero mean and covariance function* $cov(h_1, x_1, h_2, x_2)$.

Proof.

First fix $(h_1, x_1), (h_2, x_2), \cdots, (h_l, x_l) \in H \times I$ and $t_1, t_2, \cdots, t_l \in (-\infty, \infty)$. If we let

$$m^{1/2} \sum_{j=1}^{l} t_j [\hat{g}_{h_j}(v_j) - E\hat{g}_{h_j}(v_j)] = V_m$$

which has zero mean, and variance

$$m^{-1} \sum_{j=1}^{l} \sum_{k=1}^{l} t_j t_k \sum_{p=1}^{m} Var(Z_p) W_m(h_1, x_j, v_p) W_m(h_2, x_k, v_p),$$

which converges in probability to $\sum_{j=1}^{l} \sum_{k=1}^{l} t_j t_k cov(h_j, x_j, h_k, x_k)$ as $m \to \infty$. Also, uniform boundedness of the $(2+\rho)$-th central moment of Z_p and the condition that

$$m^{-(1+\rho/2)} \{\max_{1 \le j \le m} |W_m(h, x, v_j)|^{\rho} \sum_{j=1}^{m} \{W_m(h, x, v_j)\}^2 \to 0$$

in probability as $m \to \infty$ together imply that Lyapunov's condition holds for V_m, and consequently its limiting distribution must be normal. Finally, it follows using the Cramer-Wold device that as $m \to \infty$, the joint limiting distribution of

$$m^{1/2}[\hat{g}_{h_i}(x_i) - E\hat{g}_{h_i}(x_i)] = U_m(h_i, x_i)$$

for $1 \leq i \leq l$ is multivariate normal with zero mean and $cov(h_j, x_j, h_k, x_k)$ as the (j, k)-th entry of the limiting variance covariance matrix for $1 \leq j, k \leq l$.

Next, fix $h_1 < h_2$ in H and $x_1 < x_2$ in I. Then the last condition assumed in the statement of the theorem implies that

$$\begin{aligned}
& E\{U_m(h_2, x_2) - U_m(h_2, x_1) - U_m(h_1, x_2) + U_m(h_1, x_1)\}^2 \\
= \;& m^{-2} \sum_{j=1}^{m} Var(Z_j)\{W_m(h_2, x_2, v_j) - W_m(h_2, x_1, v_j) - W_m(h_1, x_2, v_j) + W_m(h_1, x_1, v_j)\}^2 \\
\leq \;& C_2(h_2 - h_1)^2(x_2 - x_1)^2\{m^{-1} \sum_{j=1}^{m} M(v_j)\} \leq C_3(h_2 - h_1)^2(x_2 - x_1)^2
\end{aligned}$$

for some constants C_2 and $C_3 > 0$.

It now follows [see Bickel and Wichura (1971)] that the sequence of processes

$$m^{1/2}[\hat{g}_h(x) - E\hat{g}_h(x)]$$

on $H \times I$ will have the tightness property, and consequently the assertion in the theorem follows.

The following theorem involves the difference between the behavior of the empirical and the theoretical scale-space surfaces under the supremum norm on $H \times I$ and the uniform convergence of the empirical version to the theoretical one as the sample size grows. One more condition is needed:

Condition A. In the setup of Theorem 2, as h varies in H and x varies in I, both $Var(Z_j)\{W_m(h_1, x_1, v_j)\}^2$ and $Var(Z_j)\{W_m(h_2, x_2, v_j)\}^2$ are uniformly dominated by a positive function $M^*(v_j)$ such that $\sup_{j \geq 1} \; M^*(v_j) < \infty$, which will provide bounds for the variance of Z_j.

Theorem 3. *Assume Condition A as well as the setup of Theorem 2. Then as* $m \to \infty$,

$$\sup_{x \in I, h \in H} m^{1/2} |\hat{g}_h(x) - E\{\hat{g}_h(x)\}|$$

converges weakly to a random variable that has the same distribution as that of $\sup_{x \in I, h \in H} |Z(h, x)|$. *Here* $Z(h, x)$ *with* $h \in H$ *and* $x \in I$ *is a Gaussian process with zero mean and covariance function* $cov(h_1, x_1, h_2, x_2)$ *as defined in Theorem 2 so that*

$$Pr\{Z(h, x) \text{ is continuous for all } (h, x) \in H \times I\} = 1,$$

and consequently $Pr\{\sup_{x \in I, h \in H} |Z(h, x)| < \infty\}$*=1. It immediately follows from the preceding theorem that we have*

$$\sup_{x \in I, h \in H} |\hat{g}_h(x) - E\{\hat{g}_h(x)\}| = O_p(m^{-1/2}) \text{ as } m \to \infty.$$

Proof. For (h_1, x_1) and (h_2, x_2) in $H \times I$,

$$\begin{aligned} E\{Z(h_2, x_2) - Z(h_1, x_1)\}^2 &= cov(h_2, x_2, h_2, x_2) + cov(h_1, x_1, h_1, x_1) - 2cov(h_2, x_2, h_1, x_1) \\ &\leq C_4(h_2 - h_1)^2 + (x_2 - x_1)^2 \end{aligned}$$

for some constant $C_4 > 0$. This follows straight away from the fact that

$$E\{U_m(h_2, x_2) - U_m(h_1, x_1)\}^2 \leq C_4(h_2 - h_1)^2 + (x_2 - x_1)^2$$

for all $m \geq 1$ with some appropriate choice of C_4 if Condition A holds. Here U_m is as in the proof of Theorem 2. Next, consider the compact metric space $H \times I$ metrized by the pseudo metric

$$d\{(h_2, x_2), (h_1, x_1)\} = [E\{Z(h_2, x_2) - Z(h_1, x_1)\}^2]^{1/2},$$

which the canonical metric associated with the Gaussian process $Z(h, x)$. Let $N(\epsilon)$ be the smallest number of closed d-balls with radius $\epsilon > 0$ in this metric space that are required to cover $H \times I$. So, $\log\{N(\epsilon)\}$ is the usual metric entropy of $H \times I$ under the metric d. Note that for any $\epsilon >$ diameter$(H \times I)$, $N(\epsilon) = 1$ and $N(\epsilon) = O(\epsilon^{-2})$ for $0 < \epsilon \leq$ diameter$(H \times I)$. Hence, using an analogous argument in Adler (1990) $\int_0^\infty [log\{N(\epsilon)\}]^{1/2} d\epsilon < \infty$ is necessary so that the size of the metric space does not explode. This ensures the continuity of the sample paths of the process $V(h,x)$ as well as the finiteness of $\sup_{x\in I, h\in H} \ |V(h,x)|$ with probability one [see Adler (1990, pp. 104-107)]. The proof of the theorem is now complete in view of the weak convergence of the centered and normalized empirical scale-space process to the Gaussian process $V(h,x)$ on $H \times I$ established in Theorem 2.

3.3 SiZer for the Comparison of Two Time Series

This section addresses the problem of comparing two populations to the case where errors are not independent. Unlike in previous work, we do not assume the independence structure. Here, the goal is to discover meaningful structures in two populations by comparing two time series based on the differences of two kernel estimates.

A statistical challenge in this problem is testing whether there are any statistically significant differences between the trends of these time series. Suppose that we have two regularly spaced time series with the same length, and thus there are $2n$ observations from the following regression models:

$$Y_{ij} = f_i(j) + \epsilon_{ij}, \ j = 1, \ldots, n, \ i = 1, 2, \tag{3.2}$$

where the ϵ_{ij}'s are dependent random errors with mean 0, variance σ^2, $Cov(\epsilon_{ij}, \epsilon_{ik}) = \gamma_i(|j - k|)$ for all i=1, 2, $j, k = 1, \ldots, n$, f_i is the unknown regression function of the ith sample (i=1,2). It is assumed that ϵ_{1j} and ϵ_{2j} are independent of each other.

The main concern here is to develop a graphical device for testing the hypothesis of the equality of mean regression functions:

$$H_0 : f_1(x) = f_2(x)$$

when the errors are weakly correlated.

Again, confidence intervals for $f_{1,h}(x) - f_{2,h}(x)$ are of the form

$$\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x) \pm q(h) \cdot \widehat{SD}(\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)), \tag{3.3}$$

where $q(h)$ is a quantile which, for significance level α, defined as

$$q(h) = \Phi^{-1}\left(\left(1-\frac{\alpha}{2}\right)^{1/(\theta g)}\right), \tag{3.4}$$

where Φ is the standard normal distribution function and g is the number of bins. The "cluster index" θ is given by

$$\theta = 2\Phi\left(\sqrt{I\log g}\,\frac{\tilde{\Delta}}{h}\right) - 1$$

where now

$$I = \frac{\int(\gamma_1(sh/\Delta) + \frac{2-s^2}{8}\gamma_2(sh/\Delta))e^{-s^2/4}ds}{\int(\gamma_1(sh/\Delta) + \gamma_2(sh/\Delta))e^{-s^2/4}ds}. \tag{3.5}$$

The design points in time series are equidistant, and thus we can assume that without loss of generality that the ith point is in the location $i\Delta$ for some $\Delta > 0$. Here, Δ denotes the distance between design points. Let $\tilde{\Delta}$ denote the distance between the pixels of the SiZer map and $p = \tilde{\Delta}/\Delta$ denote the number of data points per SiZer column. Also, γ_1 and γ_2 are the autocovariance functions of the first and second time series, respectively. This quantile (3.4) is used in our implementation and it can be derived as follows.

SiZer uses the local linear smoother defined by

$$\sum_{j=1}^{n}\{Y_{ij} - (\beta_{i0} + \beta_{i1}(x_0 - j))\}^2 K_h(x_0 - j).$$

To color the pixels SiZer checks whether the difference of the estimates of the two regression functions

$$\begin{aligned}
\hat{\beta}_{i0} &= c_i^{-1}\left[\sum_{j=1}^{n} K_h(x-j)Y_{ij}\right]\left[\sum_{j=1}^{n}(x-j)^2\,K_h(x-j)\right] \\
&- c_i^{-1}\left[\sum_{j=1}^{n}(x-j)\,K_h(x-j)\right]\left[\sum_{j=1}^{n}(x-j)K_h(x-j)\,Y_{ij}\right], \qquad (3.6) \\
c_i &= \left[\sum_{j=1}^{n} K_h(x-j)\right]\left[\sum_{j=1}^{n}(x-j)^2\,K_h(x-j)\right] - \left[\sum_{j=1}^{n}(x-j)\,K_h(x-j)\right]^2,
\end{aligned}$$

for $i = 1, 2$, is significantly different from 0. Suppose that $T_1, ..., T_g$ are the test statistics in the SiZer map. Note that

$$T_k \approx \sum_{q=1}^{n} w_{kp-q}^{h}(Y_{1,q} - Y_{2,q}).$$

The form of the w^h_{kp-q} is given in the first term of (3.6). Note that w^h_{kp-q} is proportional to $K_{h/\Delta}(kp-q)$ and thus the weights w^h_q are proportional to the Gaussian kernel with standard deviation h/Δ.

Let γ_1 be the autocovariance function of the first time series and γ_2 be the autocovariance function of the second. The full joint distribution of $T_1, ..., T_g$ also depends on the correlation between the T_k's. This correlation is approximated by

$$\begin{aligned}
\rho_{j-i} =& corr(T_i, T_j) \\
=& \frac{\sum_q \sum_r w^h_{ip-q} w^h_{jp-r} (\gamma_1(q-r) + \gamma_2(q-r))}{\sum_q \sum_r w^h_q w^h_r (\gamma_1(q-r) + \gamma_2(q-r))} \\
\approx& \frac{\iint K_{h/\Delta}(ip-x)\, K_{h/\Delta}(jp-y)\, (\gamma_1(x-y) + \gamma_2(x-y))\, dxdy}{\iint K_{h/\Delta}(x)\, K_{h/\Delta}(y) (\gamma_1(x-y) + \gamma_2(x-y))\, dxdy} \\
=& \frac{\int (\gamma_1(s) + \gamma_2(s)) \int K_{h/\Delta}(ip-s-y)\, K_{h/\Delta}(jp-y)\, dy\, ds}{\int (\gamma_1(s) + \gamma_2(s)) \int K_{h/\Delta}(s+y)\, K_{h/\Delta}(y)\, dy\, ds} \\
=& \frac{\int (\gamma_1(s) + \gamma_2(s))\, e^{-(ip-jp-s)^2\Delta^2/(4h^2)}\, ds}{\int (\gamma_1(s) + \gamma_2(s)) e^{-s^2\Delta^2/(4h^2)}\, ds} \\
=& \frac{\int (\gamma_1(s) + \gamma_2(s))\, e^{-[(i-j)\tilde{\Delta} - s\Delta]^2/(4h^2)}\, ds}{\int (\gamma_1(s) + \gamma_2(s)) e^{-s^2\Delta^2/(4h^2)}\, ds}.
\end{aligned}$$

Here the third line follows by replacing the sums by integral approximations and the last step follows by observing that $p\Delta = \tilde{\Delta}$. Thus

$$\rho_{j,g} = \frac{\int (\gamma_1(s) + \gamma_2(s))\, e^{-(Cj/\sqrt{\log g} - s)^2/4}\, ds}{\int (\gamma_1(s) + \gamma_2(s))\, e^{-s^2/4}\, ds}.$$

Finally, since $\gamma_i(s)$ is an even function, we get by dominated convergence theorem

$$\lim_{g\to\infty} \log g (1 - \rho_{k,g}) = k^2 \frac{C^2 \int (\gamma_1(s) + \gamma_2(s))\, \frac{2-s^2}{8} e^{-s^2/4}\, ds}{\int (\gamma_1(s) + \gamma_2(s))\, e^{-s^2/4}\, ds}.$$

Therefore just as in Hannig and Marron (2006), we conclude that

$$P\left[\max_{i=1,\dots,g} T_i \le x\right] \approx \Phi(x)^{\theta g},$$

where the cluster index

$$\theta = 2\Phi\left(\sqrt{I \log g}\, \frac{\tilde{\Delta}}{h}\right) - 1$$

and

$$I = \frac{\int (\gamma_1(sh/\Delta) + \gamma_2(sh/\Delta)) \frac{2-s^2}{8} e^{-s^2/4} \, ds}{\int (\gamma_1(sh/\Delta) + \gamma_2(sh/\Delta)) \, e^{-s^2/4} \, ds}.$$

The local linear estimate $\hat{f}_{i,h}(x)$ can be written as

$$\hat{f}_{i,h}(x) = \frac{1}{n} \sum_{j=1}^{n} w_n(h, x, j) Y_{ij}. \tag{3.7}$$

where

$$w_n(h, x, j) = \frac{\{\hat{s}_2(x; h) - \hat{s}_1(x; h)(x - j)\} K_h(x - j)}{\hat{s}_2(x; h)\hat{s}_0(x; h) - \hat{s}_1(x; h)^2}$$

and

$$\hat{s}_r(x; h) = \frac{1}{n} \sum_{j=1}^{n} (x - j)^r K_h(x - j).$$

Then, by independence

$$Var(\hat{f}_{1,h}(x) - \hat{f}_{2,h}(x)) = Var(\hat{f}_{1,h}(x)) + Var(\hat{f}_{2,h}(x)),$$

and

$$Var(\hat{f}_{i,h}(x)) = \frac{\sigma_i^2}{n^2} \sum_{j=1}^{n} (w_n(h, x, j))^2 + \frac{2}{n^2} \sum \sum_{j<k} w_n(h, x, j) w_n(h, x, k) \gamma_i(k - j).$$

In order to construct the confidence interval in (3.3) the estimate of the autocovariance function γ_i in (3.5) is needed. These γ_i's are estimated as explained previously in Section 2.3.

3.3.1 Two time series: Simulation

The first part of this subsection illustrates simulated examples when the dependence structure is known and the second part deals with cases when the dependence structure is unknown for two time series. Real data analysis on two time series is done in the third part of this subsection, and the final part of this subsection contains two asymptotic results.

Two time series: Simulation when the dependence structure is known

Here we explore the performance of SiZer when the autocovariance function is given in advance. This is an extension of Park et al. (2004) for two time series, and it is particularly useful when the dependence structure is known from previous studies.

In these simulations various combinations of error structures and mean regression functions are considered. Six simulated examples are provided and each example has sample size $n = 100$. The first example has the same constant mean 0 for both samples:

$$\text{(i)} \qquad Y_{ij} = \varepsilon_{ij}, \quad j = 1, \ldots, n, \ i = 1, 2.$$

For the second example, one time series has a sine curve as a regression function and the other has mean 0:

$$\text{(ii)} \qquad Y_{1j} = 4\sin(6\pi j/n) + \varepsilon_{1j}, \quad \text{and} \quad Y_{2j} = \varepsilon_{2j}.$$

The third example studies two different regression functions:

$$\text{(iii)} \qquad Y_{1j} = 4\sin(6\pi j/n) + 3j/n + \varepsilon_{1j}, \quad \text{and} \quad Y_{2j} = 4\sin(6\pi j/n) + \varepsilon_{2j}.$$

The fourth example has constant mean of 2 in the first model and constant mean of 0 in the second model:

$$\text{(iv)} \qquad Y_{1j} = 2 + \varepsilon_{1j}, \quad \text{and} \quad Y_{2j} = \varepsilon_{2j}.$$

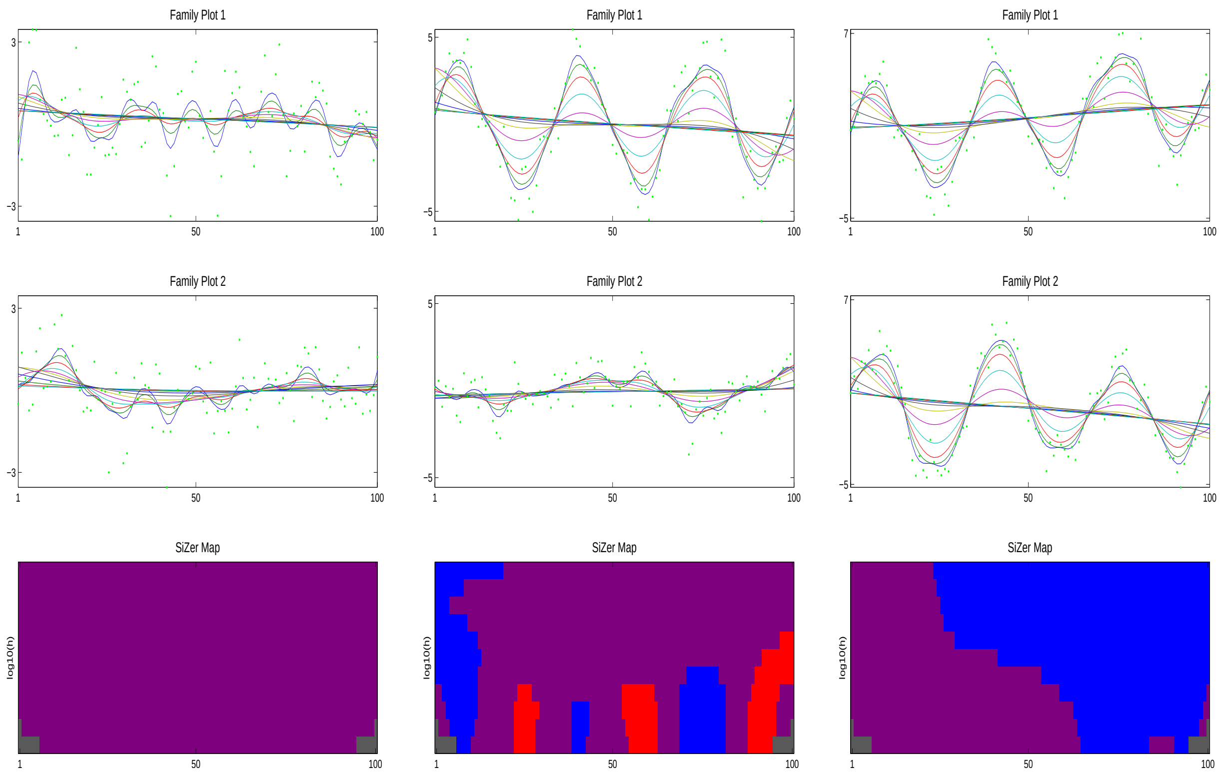
For the fifth example, one time series has an exponential trend as a regression function and the other has mean 0:

$$\text{(v)} \qquad Y_{1j} = \exp(j/n) + \varepsilon_{1j}, \quad \text{and} \quad Y_{2j} = \varepsilon_{2j}.$$

Finally, the sixth example studies two different regression functions:

$$\text{(vi)} \qquad Y_{1j} = 4\sin(6\pi j/n) + \exp(j/n) + \varepsilon_{1j}, \quad \text{and} \quad Y_{2j} = \exp(j/n) + \varepsilon_{2j}.$$

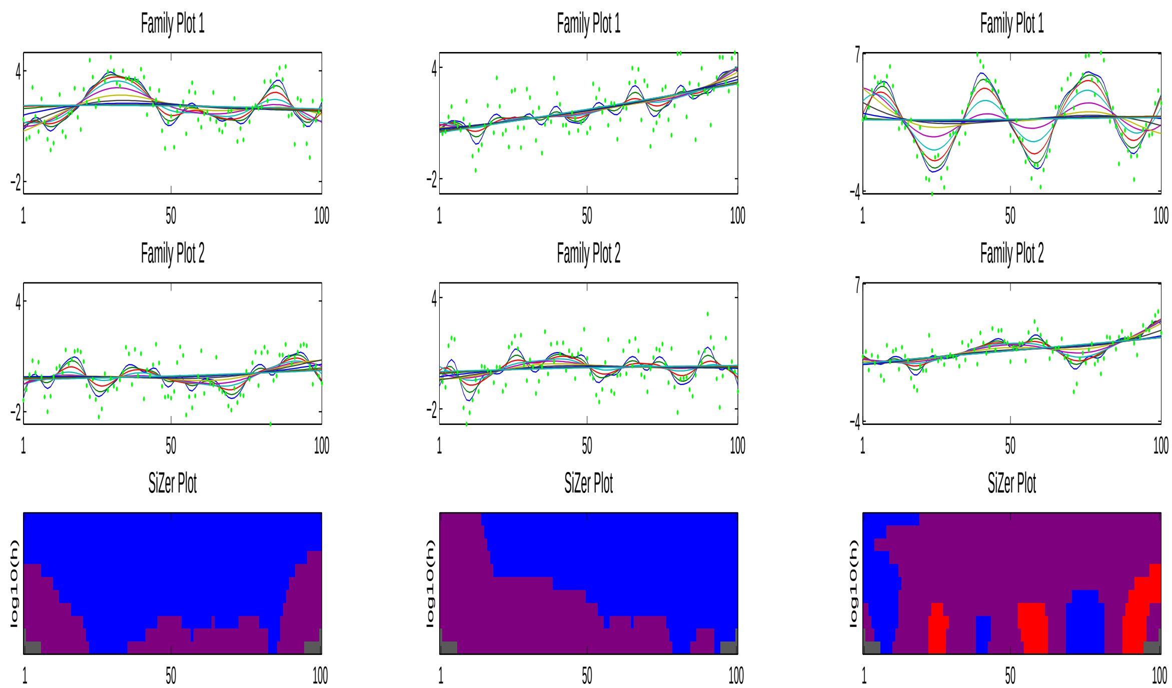
Three combinations of error structures are considered, MA(1) versus AR(1), MA(1) versus MA(5), and AR(1) versus MA(5) for situations where there are two short-term correlated datasets or where one dataset is correlated long-term. The ϕ for AR(1) is 0.5, the θ for MA(1) is 0.5, and the θs for MA(5) are 0.9, 0.8, 0.7, 0.6, and 0.5. The correct SiZer plots would show no significant difference for the first example, a sine trend for the second, and a linear trend for the third. The fourth example would have a linear trend for a constant value, an exponential trend for the fifth, and a sine trend for the sixth.



(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 3.4: Comparison of two time series with MA(1) and AR(1) errors. Autocovariance functions are given in advance.

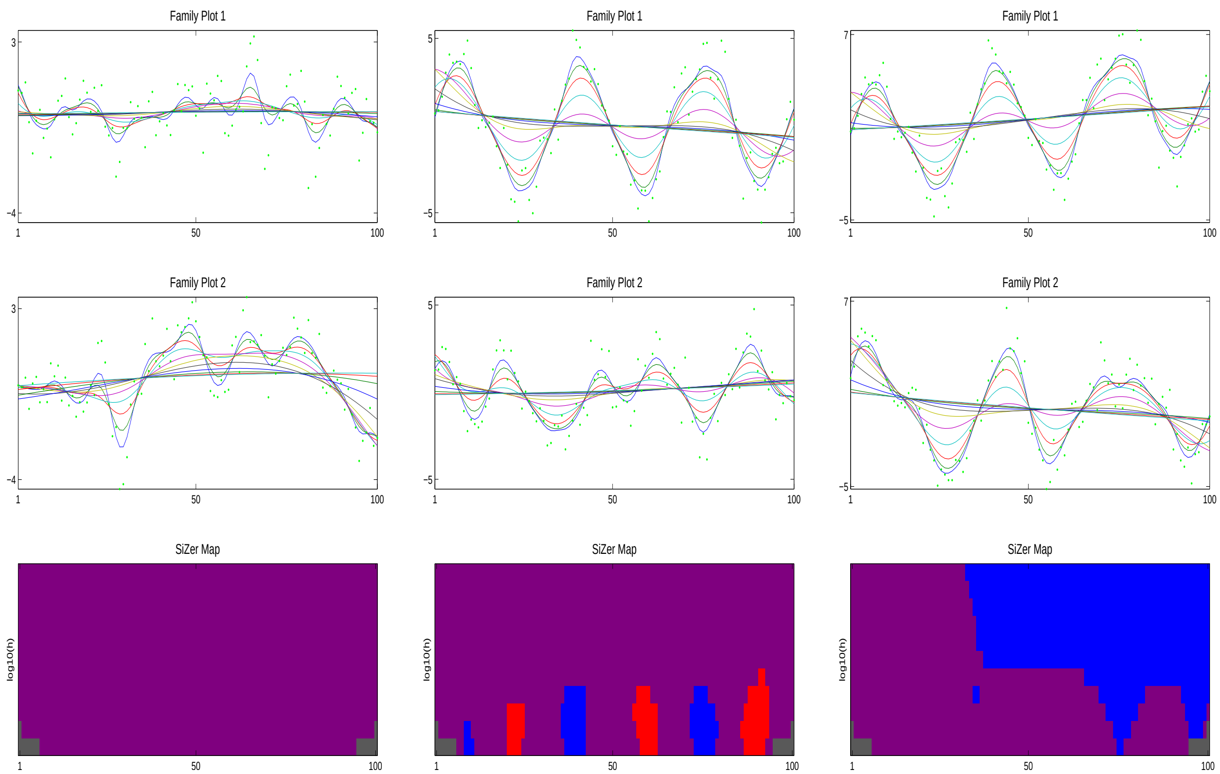
Figure 3.4 displays SiZer plots with MA(1) and AR(1) for the first three examples. In the top two panels, the green dots are actual data points and the thin colored curves display the family of smooths; that is, $\hat{f}_{i,h}(x)$ for $i = 1, 2$ and where h is the bandwidth. The SiZer maps in the third panels report the equality test of the two time series by investigating the confidence intervals in (3.3) at each (x, h). The horizontal locations in the SiZer map are the same x values as in the top panels, and the vertical locations in the SiZer plot correspond to the logarithm of bandwidths of the family of smooths shown as thin colored curves in the top panels. The SiZer map in Figure 3.4 (a) shows only purple, meaning no significant difference, as expected. When the two regression curves are different, the SiZer maps correctly capture the differences: the SiZer map in Figure 3.4 (b) shows positive (blue) and negative (red) differences along the sine curve, but also has some spurious pixels present; and the map in Figure 3.4 (c) flags a rough linear trend as significant. From these three simulations, SiZer has shown that for the comparison of two time series, it performs well in its ability to detect significant trends through assumed dependence structure for weakly correlated data.



(a) constant vs. zero mean (b) expon vs. zero mean (c) sine plus expon vs. expon

Figure 3.5: Additional comparisons of two time series with MA(1) and AR(1) errors. Autocovariance functions are given in advance.

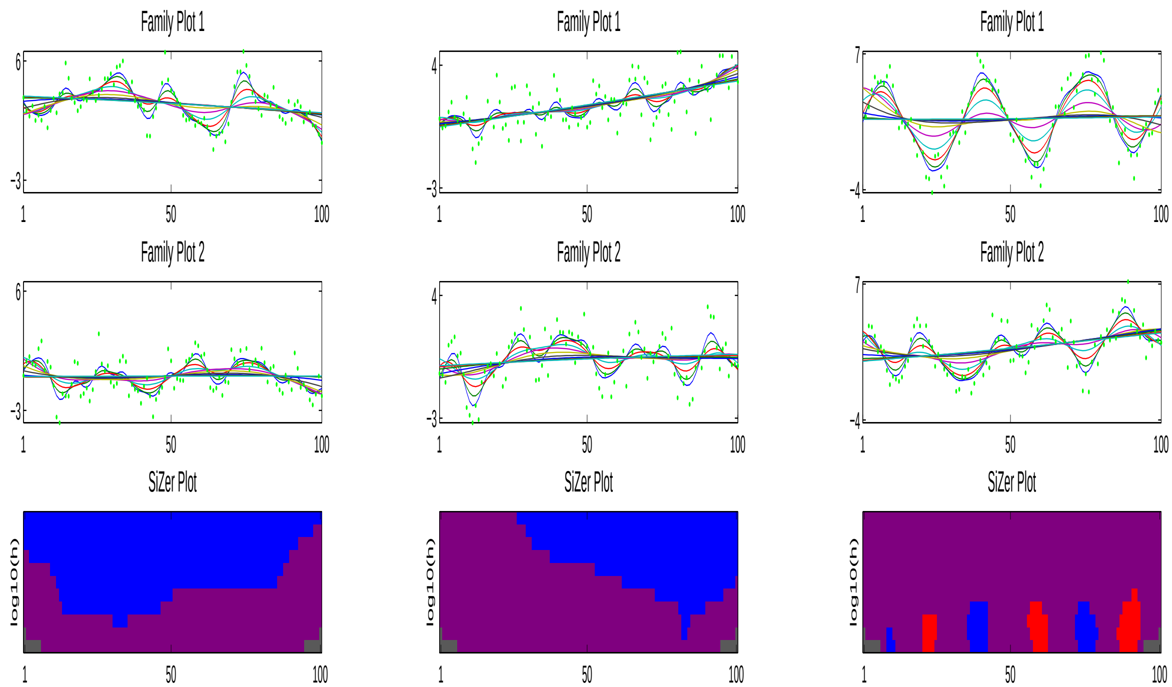
Figure 3.5 displays SiZer plots with the weak dependence structure of MA(1) and AR(1) for the last three examples. SiZer flags a roughly constant trend for Figure 3.5 (a) and thus the entire top of the map is blue to denote the positive difference between the constant and the mean of 0. In Figure 3.5 (b) the SiZer map also catches the important trend given by an exponential curve, and although it is not quite as clearly delineated as is desired, because the exponential trend at the beginning is indistinguishable from the natural variation, the presence of this difference is marked clearly. In Figure 3.5 (c), the SiZer plot detects a strong difference that is attributed to the sine trend from the upper family plot, with only a few spurious pixels in the upper left hand corner.



(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 3.6: Comparison of two time series with MA(1) and MA(5) errors. Autocovariance functions are given in advance.

Figure 3.6 displays SiZer plots with MA(1) and MA(5) for the first three examples. The results are similar to Figure 3.4, with (a) being correctly marked completely purple and (b) having the sine trend marked even more clearly here with no spurious pixels. In Figure 3.6 (c), once again, a general linear trend is detected, with a couple of spurious pixels near the center of the picture. Thus, SiZer performs reasonably well for cases in which either of the time series has a short or a long-term correlation, but it does need some improvement in clearly delineating the linear trend as the significant signal. Such improvement may include a more accurate estimation of an autocovariance function for the case of long-term correlation.

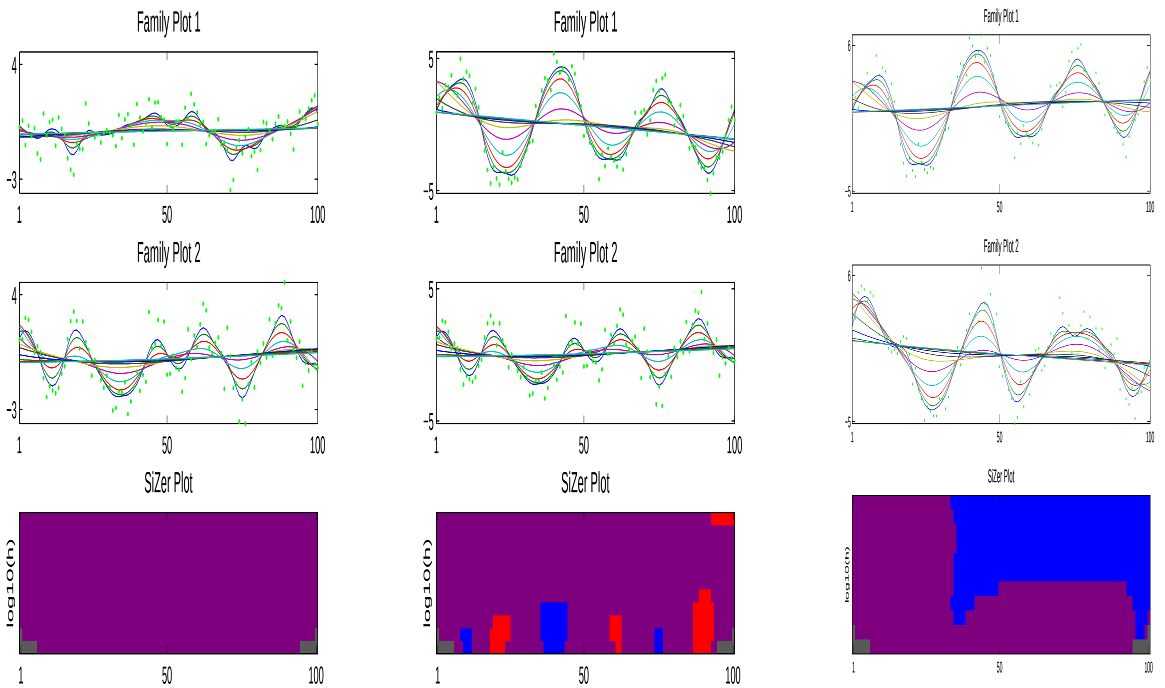


(a) constant vs. zero mean (b) expon vs. zero mean (c) sine plus expon vs. expon

Figure 3.7: Additional comparisons of two time series with MA(1) and MA(5) errors. Autocovariance functions are given in advance.

Figure 3.7 displays SiZer plots with MA(1) and MA(5) for the last three trend examples. Figure 3.7 (a) is correctly colored blue in the top portion of the SiZer map for the presence of the constant trend and Figure 3.7 (b) also shows that the rough exponential trend from the first regression model is identified by the SiZer map. In Figure 3.7 (c), the SiZer map captures all of the changes in the sine trend.

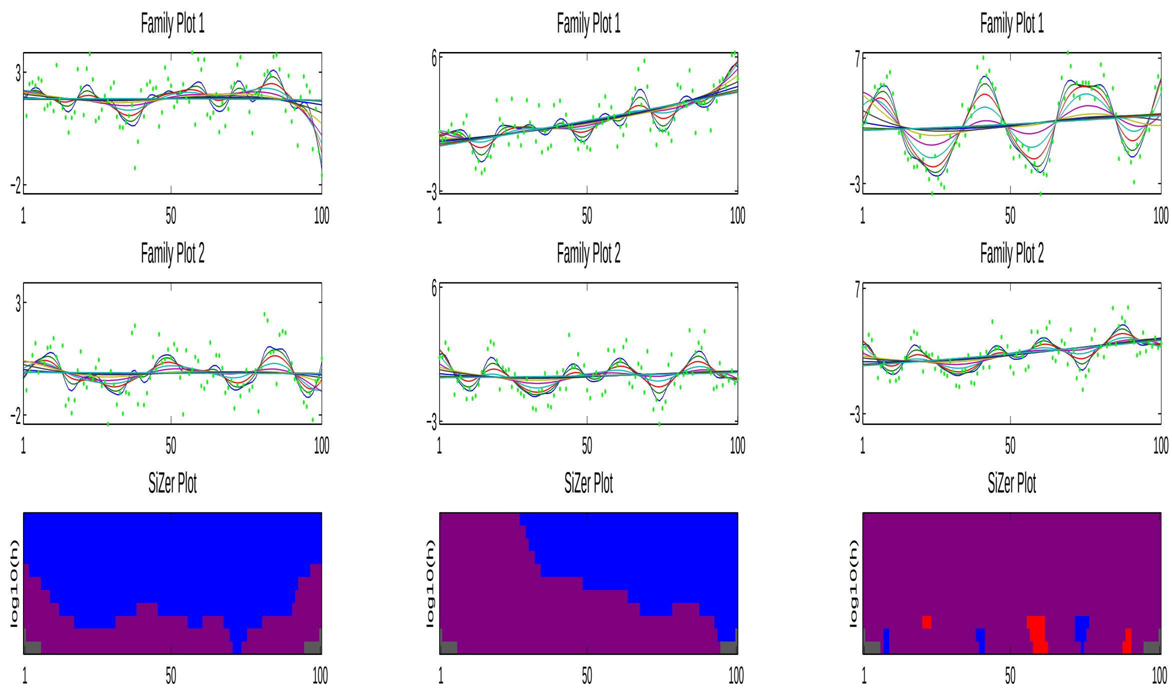
Figure 3.8 displays the first three SiZer plots that have AR(1) and MA(5) as their correlation structure. Even though the bottom family plot for (a) is very wiggly due to the MA(5) error, the SiZer map can correctly identify this as error structure and not a significant signal



(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 3.8: Comparison of two time series with AR(1) and MA(5) errors. Autocovariance functions are given in advance.

or trend. The SiZer map shows all purple as desired. In Figure 3.8 (b), it can again detect the difference between changes in the plot that should be attributed to the sine curve and that which is MA(5) error structure and highlights the sine structure in the bottom plot. There are a few spurious pixels in the top right hand corner that are undesirable, but all changes in the sine trend are correctly detected with no gaps in the signal differences. Finally, in Figure 3.8 (c), a rough linear trend is correctly mapped out in the bottom plot as the difference between the top plot which has a sine and a linear trend and the plot in the middle, which is a sine trend only.



(a) constant vs. zero mean (b) expon vs. zero mean (c) sine plus expon vs. expon

Figure 3.9: Additional comparisons of two time series with AR(1) and MA(5) errors. Autocovariance functions are given in advance.

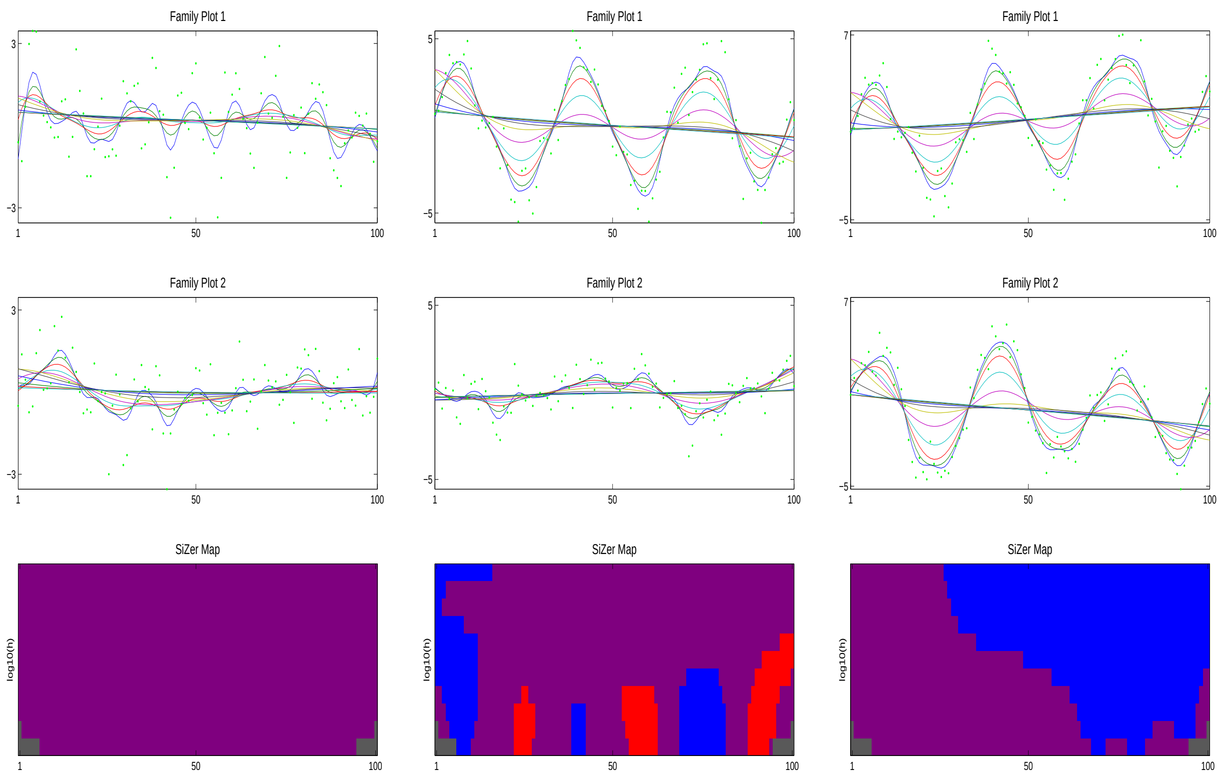
In Figure 3.9 (a), the positive difference from the constant trend is modeled roughly by the blue portion in the top of the SiZer map. In Figure 3.9 (b), the exponential trend is reflected accurately by the upper positive portion highlighted blue in the SiZer map. Finally, in Figure 3.9 (c), SiZer does an accurate job in finding the difference between the two family plots' regression functions with every change over in the sine trend detected and no spurious pixels highlighted.

These simulated examples have shown that SiZer is able to correctly differentiate between the actual signal and the dependence structure in the difference between two time series. Although there is still room for improvement when it comes to the clear changeover points in the trends which are close to linear, the SiZer maps do correctly detect various types of regression structure, even when both weak and strong dependence structure are present, and correctly identifies only true trend as significant.

Two time series: Simulation when the dependence structure is unknown

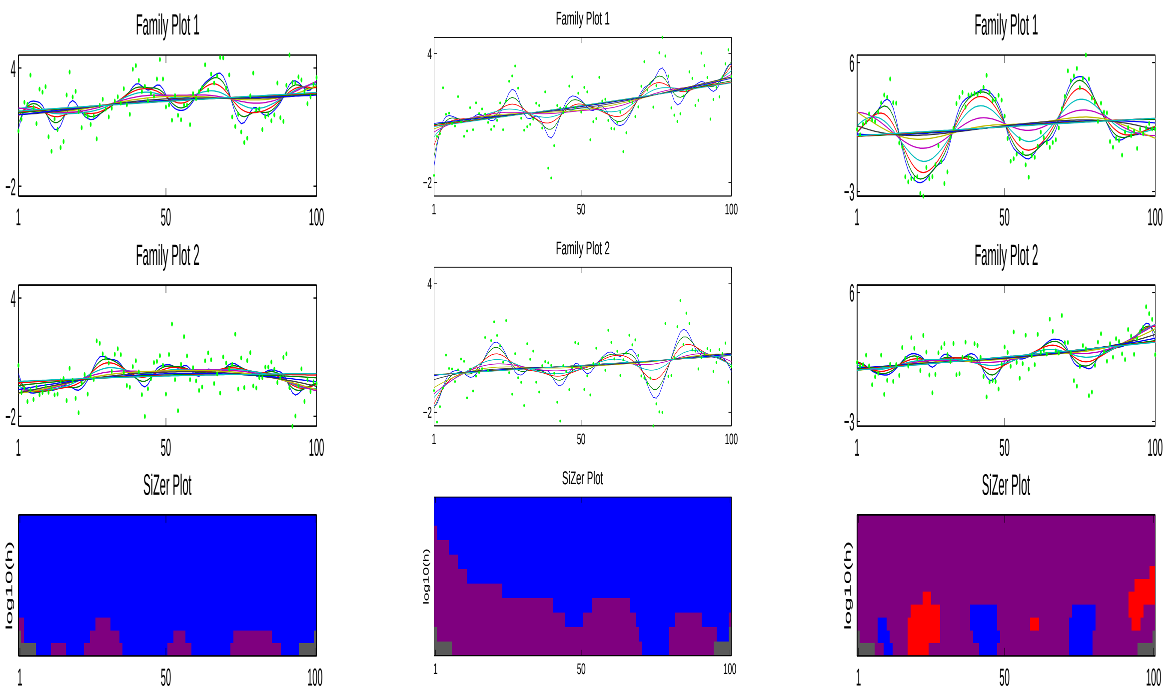
In the real world, one can rarely assume the true autocovariance structure in advance. Therefore, the performance of our new approach will be examined with the nonparametrically estimated autocovariance functions in Section 2.3 by repeating the simulation studies in the first part of Section 3.3.1. The results will be compared with those in the first part of Section 3.3.1 in order to assess the performance of the autocovariance function estimator, which does not require knowledge of the order of the autocovariance structure.

Figure 3.10 displays SiZer plots with MA(1) and AR(1) for the first three examples. In Figure 3.10 (a), there is no signal in either of the family plots and thus only the MA(1) and AR(1) autocorrelations are present. As one can see from the plots, similar to Figure 3.4, SiZer flags no trend for Figure 3.10 (a) and thus the entire map is purple. In Figure 3.10 (b) the SiZer map also catches all of the important trends given in a sine curve although it is not quite as clearly partitioned as is desired, with some spurious pixels in the upper left hand corner. Figure 3.10 (c), is very similar to Figure 3.4 which has the true autocovariance function, and it does detect a strong difference in the upper portion of the map that is attributed to the linear trend.



(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

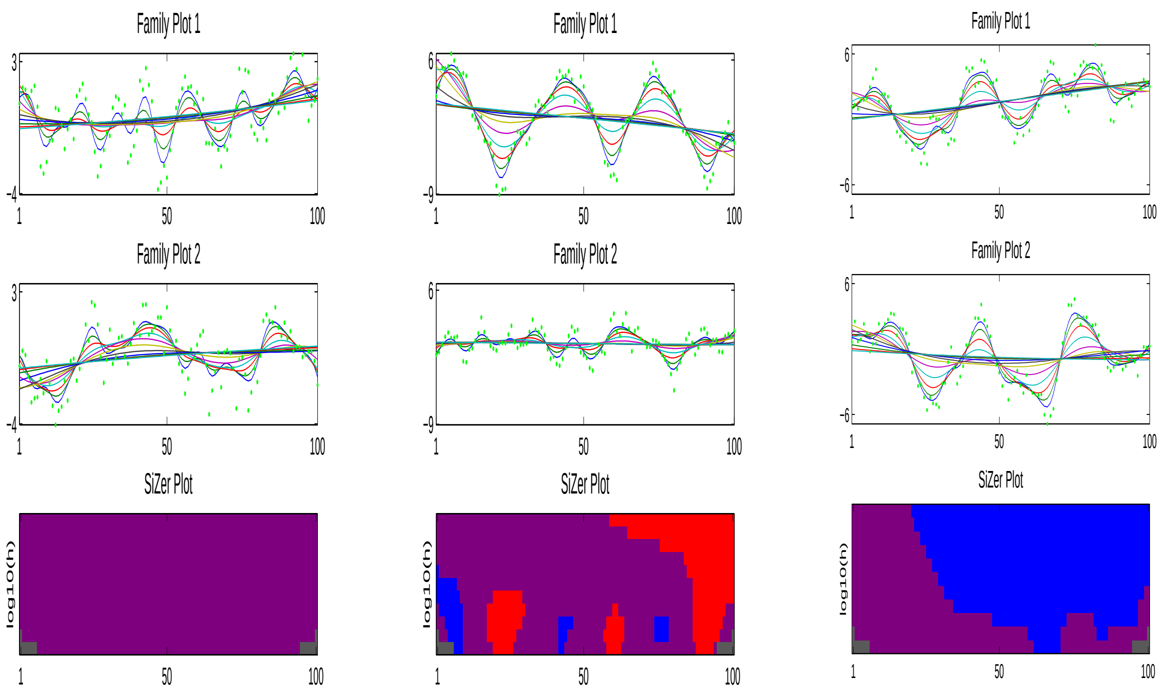
Figure 3.10: Comparison of two time series with MA(1) and AR(1). Autocovariance functions are estimated from the time series.



(a) constant vs. zero mean (b) expon vs. zero mean (c) sine plus expon vs. expon

Figure 3.11: Additional comparisons of two time series with MA(1) and AR(1) errors. Autocovariance functions are estimated from the time series.

Figure 3.11 displays SiZer plots with MA(1) and AR(1) for the last three examples. As one can see from the plots, similar to Figure 3.5, SiZer flags a roughly constant trend for Figure 3.11 (a) and thus the entire top of the map is blue to denote the positive difference between the constant and the mean of 0. In Figure 3.11 (b) the SiZer map also catches the important trend given by an exponential curve, and although it is not quite as clearly delineated as is desired, the presence of this difference is marked clearly. Figure 3.11 (c) is very similar to Figure 3.5 which has the true autocovariance function, and it does again detect a strong difference in the portion of the map that is attributed to the sine trend.

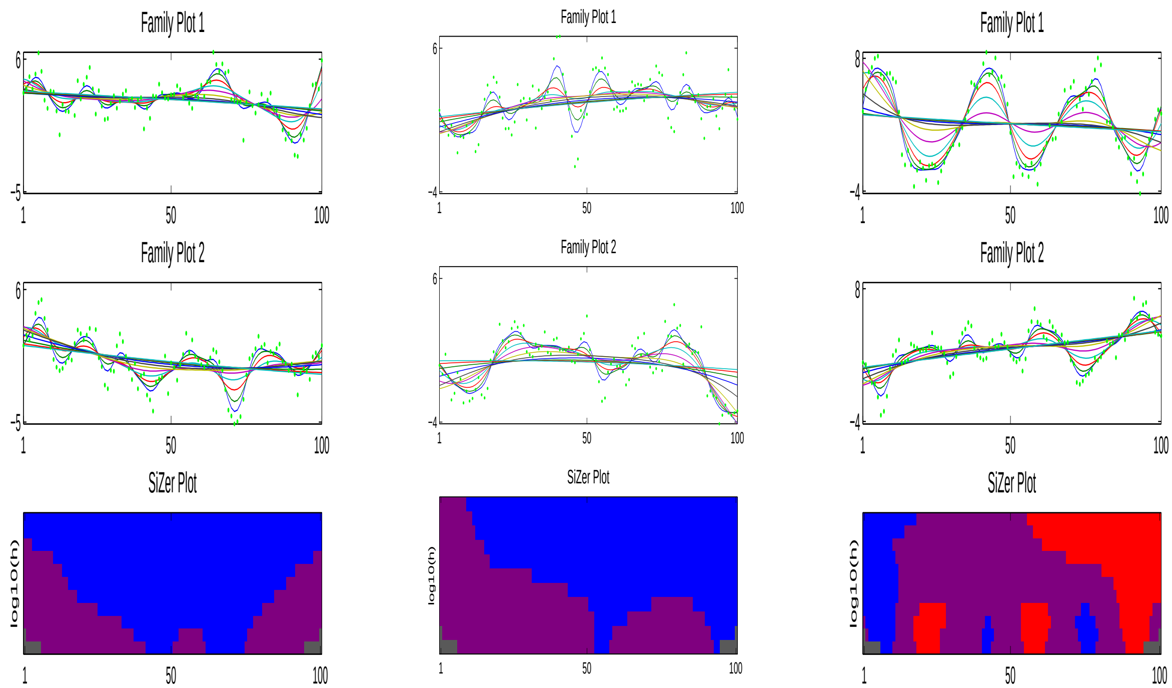


(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 3.12: Comparison of two time series with MA(1) and MA(5). Autocovariance functions are estimated from the time series.

Figure 3.12 displays SiZer plots with MA(1) and MA(5) for the same first three trend examples. Here again, almost the same conclusions can be made as from the plots in Figure 3.6. Figure 3.12 (a), is correctly colored purple for the presence of no trend. In Figure 3.12

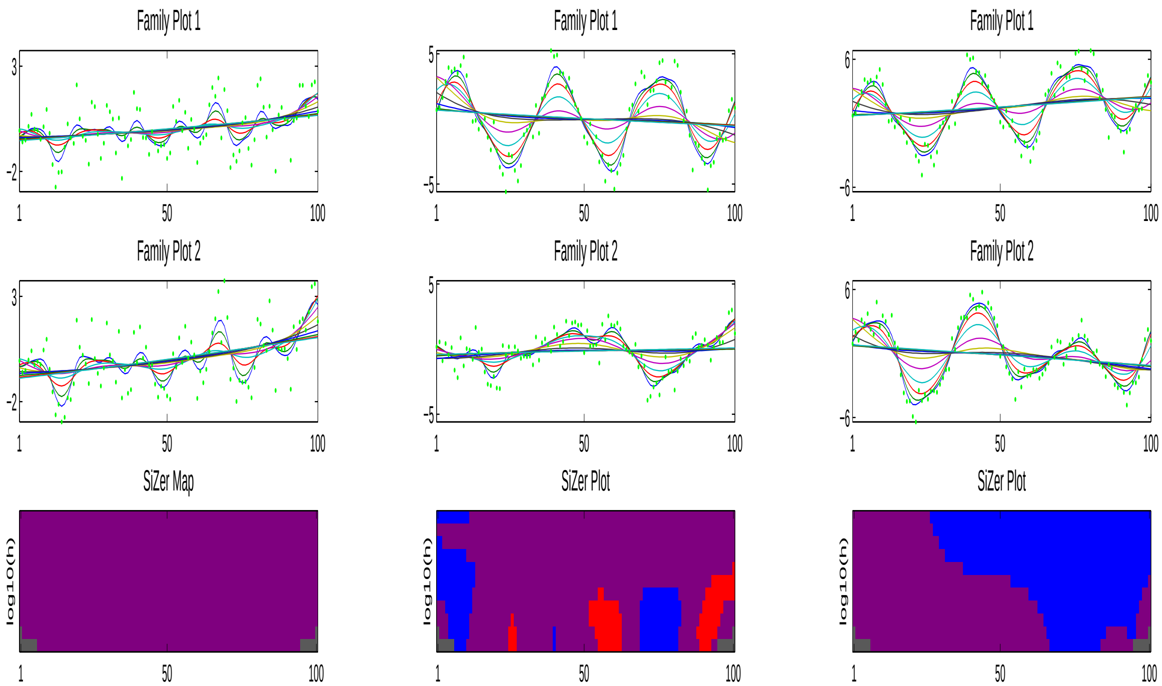
(b), the SiZer map captures all of the changes in the sine trend; the only misdiagnoses are in the top right hand corner where there are some spurious pixels. Figure 3.12 (c) also shows again that the positive linear trend from the first plot is identified by the SiZer map.



(a) constant vs. zero mean (b) expon vs. zero mean (c) sine plus expon vs. expon

Figure 3.13: Additional comparisons of two time series with MA(1) and MA(5) errors. Autocovariance functions are estimated from the time series.

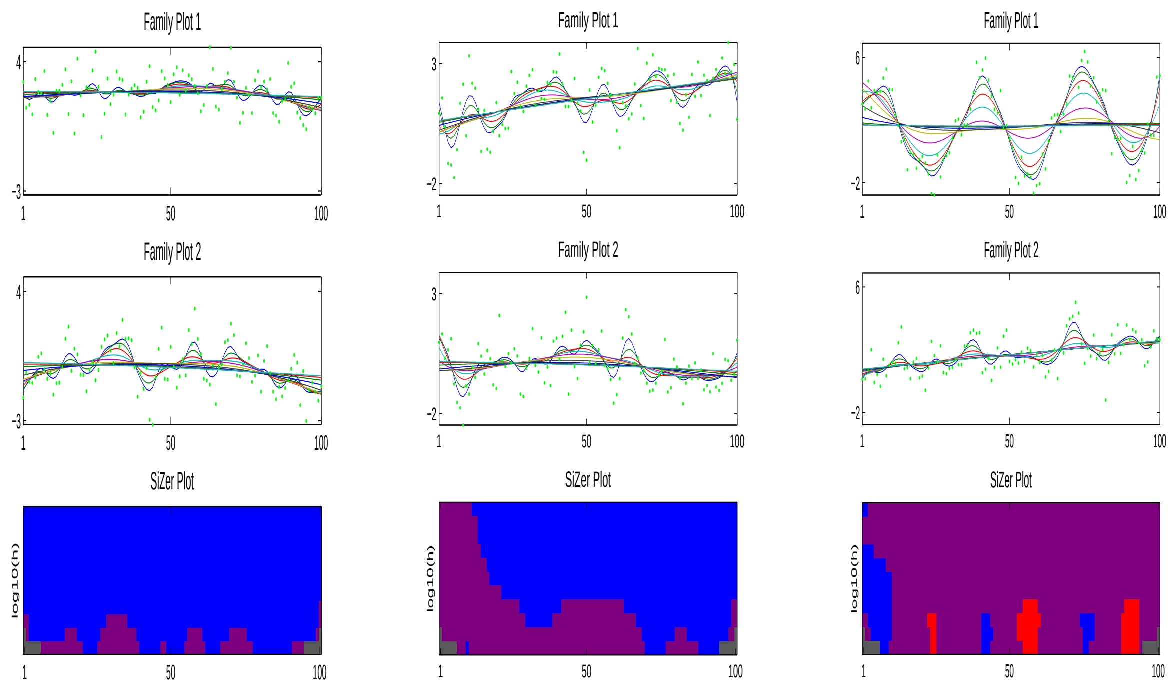
Figure 3.13 displays SiZer plots with MA(1) and MA(5) for the last three trend examples. Here again, almost the same conclusions can be made as from the plots in Figure 3.11. Figure 3.13 (a) is correctly colored blue in the top portion of the SiZer map for the presence of the constant trend and in Figure 3.13 (b), also shows again that the rough exponential trend from the first plot is identified by the SiZer map. In Figure 3.13 (c), the SiZer map captures all of the changes in the sine trend and the only misdiagnoses are in the top hand corners where there are some spurious pixels.



(a) zero mean vs. zero mean (b) sine vs. zero mean (c) sine plus linear vs. sine

Figure 3.14: Comparison of two time series with AR(1) and MA(5). Autocovariance functions are estimated from the time series.

Figure 3.14 presents plots that have AR(1) and MA(5) for the same first three trends. Here, it can be seen that Figure 3.14 (a) again, as in Figure 3.8 (a), correctly detects the fact that there is no signal in either plot, and colors the SiZer map completely purple. In Figure (b), it catches all changes within the sine trend very cleanly. For Figure 3.14 (c), like Figure 3.8 (c), it highlights the appropriate linear trend after the difference between the signals of the two plots is taken. With these three figures it can be seen that SiZer succeeds in capturing the important differences in two correlated time series while estimating the autocovariance function. In addition, it also does a very fair job of highlighting the differences in trend whether both time series have weak correlation or if one of the two time series has a stronger correlation.



(a) constant vs. zero mean (b) expon vs. zero mean (c) sine plus expon vs. expon

Figure 3.15: Additional comparisons of two time series with AR(1) and MA(5) errors. Autocovariance functions are estimated from the time series.

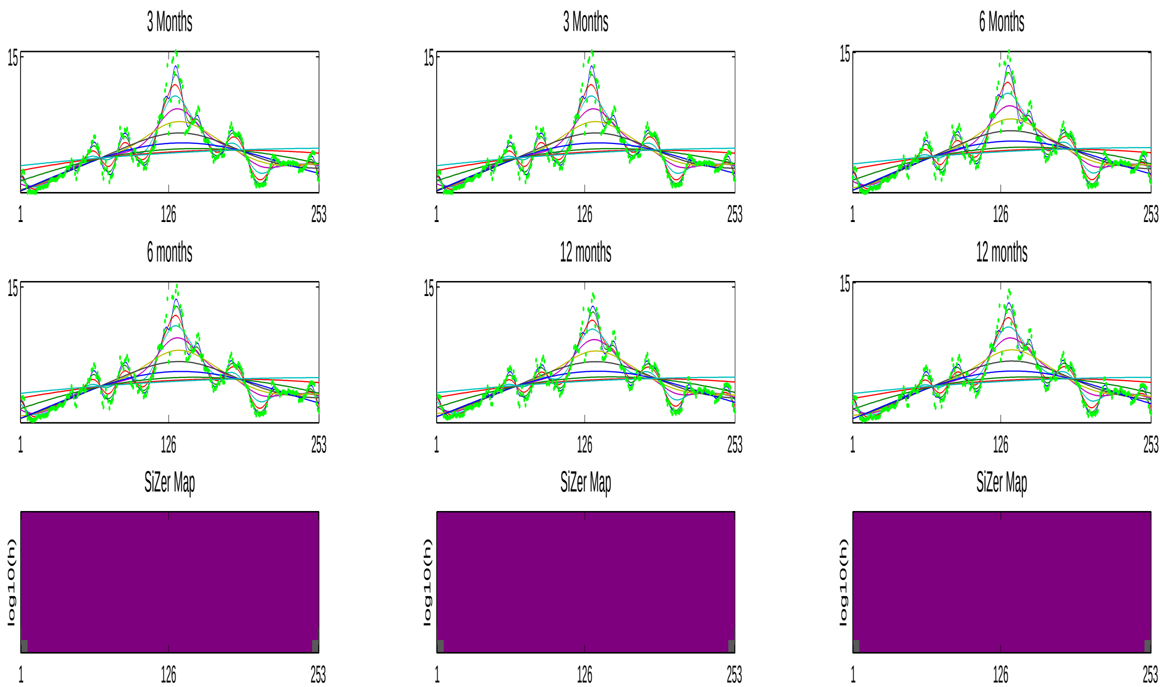
From Figure 3.15 (a), the positive difference from the constant trend is modeled roughly by the blue portion in the top of the SiZer map. From the first plot in Figure 3.15 (b), the exponential trend is reflected accurately by the upper positive portion highlighted blue in the SiZer map below. Finally, in Figure 3.15 (c), SiZer does an excellent job in estimating the difference between the two family plots' regression functions with a sine trend and only a few spurious pixels. These simulated examples have shown that while there may still be some need for improvement when it comes to cleanly delineating the difference between two time series, the SiZer maps do correctly detect various types of dependence structure and correctly identify only true trend as significant.

3.3.2 Two time series: Real data analysis

This subsection is devoted to illustrating our procedure applied to two time series with real data.

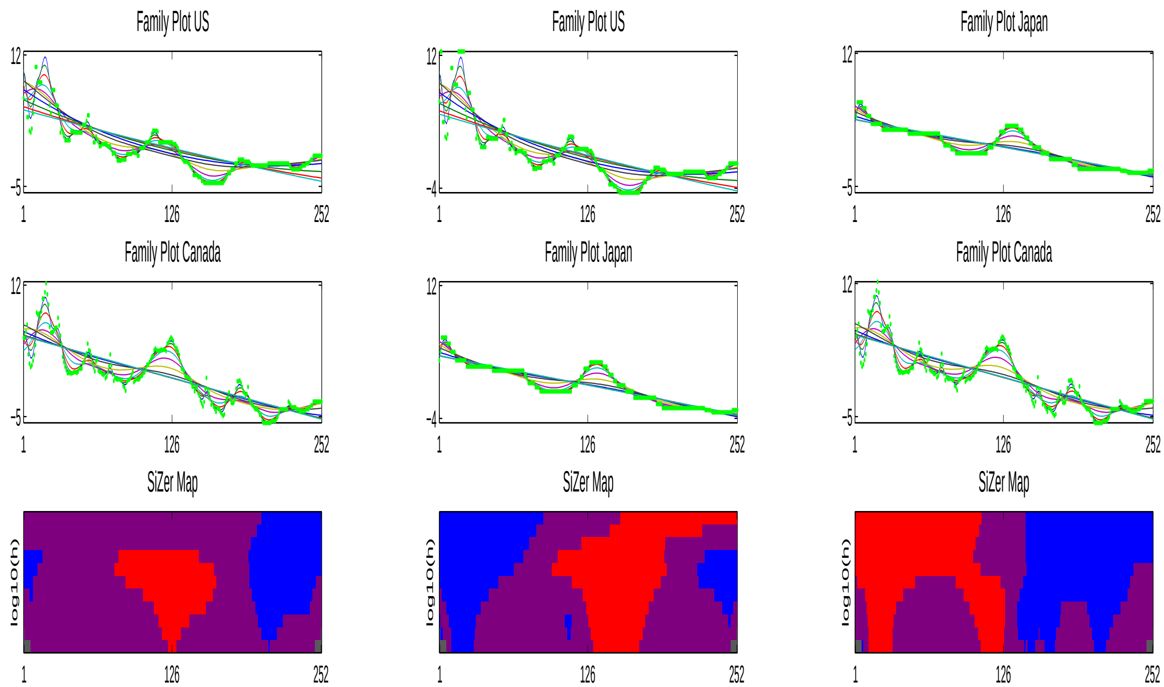
Example 1. This first example involves the weekly yields of the 3-month, 6-month, and 12-month Treasury bills. The data set was taken from July 1959 to August 2001 and can be seen in sources such as Fan and Yao (2003). In order to reduce computational burden in estimating the autocovariance function, we have taken the average of every 2 consecutive months and used that as our data, causing no change in trend. The almost identical structure can be seen in the family plots of all 3 time periods. In Figures 3.16 (a), (b), and (c) the almost identical trend of the yields is accurately detected by the SiZer maps, all of which are purple, indicating no significant difference between any pair of the time periods. Therefore, we conclude that there is no statistical difference in terms of their mean functions in the three time series.

Example 2. This example displays the monthly long-term interest rates for US, Canada and Japan from January 1980 to December 2000 (Christiansen and Pigott, 1997). Before plotting, the global mean has been taken out for each country so that all of the data are centered in



(a) M3 vs. M6 (b) M3 vs. M12 (c) M6 vs. M12

Figure 3.16: Comparison of the yields of the 3-month, 6-month, and 12-month Treasury bills measured as the bi-monthly average from July 1959 to August 2001.



(a) US vs. Canada (b) US vs. Japan (c) Japan vs. Canada

Figure 3.17: Comparison of the trends for long term target interest rates for US, Canada, and Japan from January 1980 to December 2000.

order to compare the relative trends of interest rates between the countries. In Figure 3.17 (a), we see that the long-term interest rates for the US and Canada moved quite closely together from approximately 1993-1995, despite different business cycle positions at those times. This is indicated by the purple marking of no significance between the red and final blue highlighted portions. Also confirmed in the SiZer map are the events of the fall of the Canadian rates to just below the US rates for the first time in over a decade around 1996, indicated by the final blue difference section in the map.

In Figure 3.17 (b), in the period from 1982 to mid 1984, US rates rise as the Japanese rates fall, believed by Christiansen and Pigott (1997) to be caused in part by the effects of US fiscal expansion in raising the demand for domestic savings relative to its supply. This is indicated by the blue highlighted pixels to the left of the graph indicating this early 1980's time period. In Figures 3.17 (a) and (b) there are significant divergences in the interest rates in the late 1980's: as US rates begin to fall back, rates in Canada and Japan are increasing. In both plots, the larger values of Canada and Japan cause a significant negative difference, denoted red in the middle of both plots. This short-term similarity between Canada and Japan can be seen in Figure 3.17 (c); however the graph is clearly dominated by the more rapid descent of the Canadian rates through the overall decrease of both countries.

3.3.3 Two Time Series: Asymptotic results

In this subsection, the statistical convergence of the difference between the empirical and the theoretical scale-space surfaces is proposed; this provides theoretical justification of SiZer for the comparison of two time series in scale-space. Chaudhuri and Marron (2000) address this issue for one independent sample and Park et al. (2009) extend it to single correlated data. In Section 3.2, we gave newly presented asymptotic properties two independent samples. Here, the results are extended to the case of comparing two time series.

The first theorem provides the weak convergence of the empirical scale-space surfaces and their differences with their theoretical counterparts. The second theorem states the behavior of the difference between the empirical and the theoretical scale-space surfaces under the supremum norm and the uniform convergence of the empirical version to the theoretical one.

Let I and H be compact subintervals of [0,∞) and (0,∞), respectively. Let

$$\hat{g}_h(x) = \frac{1}{n}\sum_{j=1}^{n} Z_j w_n(h, x, j)$$

where $Z_j = Y_{1j} - Y_{2j}$. The following set of assumptions are needed for the following theorems.

(A.1) The errors $(\varepsilon_{i1}, \varepsilon_{i2}, ...)$ in (3.1) are stationary, ϕ-mixing with the mixing function $\phi(j)$ satisfying $\sum_{j=1}^{\infty} \phi(j)^{1/2} < \infty$. To define ϕ-mixing, let $\{X_i; -\infty < i < \infty\}$ be a stationary sequence of random variables. If $M_{-\infty}^{k}$ and M_{k+j}^{∞} are the sequences generated by $\{X_i; i \leq k\}$ and $\{X_i; i \geq k+j\}$, respectively and $E_1 \in M_{-\infty}^{k}$ and $E_2 \in M_{k+j}^{\infty}$. Then if there exists a sequence $\phi(1), \phi(2), ...$ such that

$$|P(E_2|E_1) - P(E_2)| \leq \phi(j), \quad \phi(j) \geq 0,$$

where $1 \geq \phi(1) \geq \phi(2) \geq \cdots$, $j \geq 1$, $(-\infty < k < \infty)$, and $\lim_{j\to\infty} \phi(j) = 0$, then $\{X_i; -\infty < i < \infty\}$ is called ϕ-mixing. That is, in ϕ-mixing sequences, the lag-i covariance $\gamma(i) = Cov(X_k, X_{k+i}) \to 0$ as i increases. Intuitively, $X_1, X_2, ..., X_n$ is ϕ-mixing if X_i and X_{i+j} become essentially independent as j becomes large.

(A.2) The errors have a bounded moment $E\{|\varepsilon_{ij}|^{2+\rho}\} < \infty$ for some $\rho > 0$.

(A.3) For an integer $n \geq 0$, as $n \to \infty$

$$\frac{1}{n}\left[\sum_{j=1}^{n}\sum_{k=1}^{n}(\gamma_1(|j-k|) + \gamma_2(|j-k|))w_n(h_1, x_1, j)w_n(h_2, x_2, k)\right]$$

converges to a covariance function $cov(h_1, x_1, h_2, x_2)$ for all (h_1, x_1) and $(h_2, x_2) \in H \times I$. This assumption is fair to assume for short-term to moderate dependence, which is what is needed for Q_n that is defined below.

(A.4) $n^{-(1+\rho/2)}\{\max_{1\leq j\leq n} |w_n(h, x, j)|^{\rho}\} \sum_{j=1}^{n} w_n(h, x, j)\}^2 \to 0$ for all $(h, x) \in H \times I$.

(A.5) $w_n(h, x, j)w_n(h, x, k)$ are uniformly dominated by a positive finite number M.

(A.6)

$$\left\{\frac{\partial w_n(h,x,j)}{\partial x}\right\}\left\{\frac{\partial w_n(h,x,k)}{\partial x}\right\}, \ \left\{\frac{\partial w_n(h,x,j)}{\partial h}\right\}\left\{\frac{\partial w_n(h,x,k)}{\partial h}\right\}$$

and

$$\left\{\frac{\partial w_n(h,x,j)}{\partial x}\right\}\left\{\frac{\partial w_n(h,x,k)}{\partial h}\right\}$$

are uniformly dominated by a positive finite number M^*.

Theorem 4. *Suppose that assumptions (A.1)-(A.5) are satisfied. Define*

$$U_n(h,x) = n^{1/2}[\hat{g}_h(x) - E\{\hat{g}_h(x)\}], \quad (h,x) \in H \times I.$$

As $n \to \infty$, $U_n(h,x)$ converges to a Gaussian process on $H \times I$ with zero mean and covariance function $cov(h_1, x_1, h_2, x_2)$.

Proof. It is enough to show that all the finite dimensional distributions of the process converge weakly to the normal distribution, and that the process satisfies the tightness condition.

Fix $(h_1, x_1), (h_2, x_2), ..., (h_l, x_l) \in H \times I$ and $(t_1, ..., t_l) \in (-\infty, \infty)$. Define

$$\begin{aligned} Q_n &= n^{1/2}\sum_{j=1}^{l} t_j[\hat{g}_{h_j}(x_j) - E\{\hat{g}_{h_j}(x_j)\}] \\ &= n^{-1/2}\sum_{p=1}^{n}(\varepsilon_{1p} - \varepsilon_{2p})\sum_{j=1}^{l} t_j w_n(h_j, x_j, p). \end{aligned}$$

Then $E(Q_n)$=0 and

$$\begin{aligned} Var(Q_n) &= \frac{1}{n}\sum_{j=1}^{l}\sum_{k=1}^{l} t_j t_k \left[\sum_{p=1}^{n}\sum_{q=1}^{n}(\gamma_1(|p-q|) + \gamma_2(|p-q|))w_n(h_j, x_j, p)w_n(h_k, x_k, q)\right] \\ &\to \sum_{j=1}^{l}\sum_{k=1}^{l} t_j t_k cov(h_j, x_j, h_k, x_k) \qquad (3.8) \end{aligned}$$

as $n \to \infty$ by assumption (A.3).

Assumptions (A.2) and (A.4) imply that Lyapunov's and hence Lindeberg's condition hold for the terms in Q_n. This and assumption (A.1) verify the conditions of the main theorem in Utev (1990), which states that for a sequence of series of ϕ-mixing random variables $X_{1,n}, ..., X_{k_n,n}$ with zero means and finite variances and for $\phi_n(\cdot)$ that is a mixing coefficient corresponding to the n^{th} series, $S_n = X_{1,n} + \ldots + X_{k_n,n}$, $\sigma_n^2 = E(S_n^2)$, the distribution of S_n/σ_n tends weakly to normality. This allows us to conclude that Q_n converges in distribution to a normal random variable with variance given by (3.8). By the Cramer-Wold device, the limiting distribution of $U_n(h_j, x_j)$ $(j = 1, ..., l)$ is the multivariate normal distribution with zero mean and $cov(h_j, x_j, h_k, x_k)$ as the (j, k)th entry of the limiting variance-covariance matrix.

Now, fix $h_1 < h_2$ in H and $x_1 < x_2$ in I. Then, by Bickel and Wichura (1971) the second moment of the increment of U_n is defined by

$$\begin{aligned} &E\{U_n(h_2, x_2) - U_n(h_2, x_1) - U_n(h_1, x_2) + U_n(h_1, x_1)\}^2 \\ &= \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} (\gamma_1(|k-j|) + \gamma_2(|k-j|)) D_j D_k \end{aligned} \tag{3.9}$$

where

$$D_j = w_n(h_2, x_2, j) - w_n(h_2, x_1, j) - w_n(h_1, x_2, j) + w_n(h_1, x_1, j).$$

Then, by assumption (A.5), (3.9) is bounded by

$$C_1(x_2 - x_1)^2 (h_2 - h_1)^2 \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} (\gamma_1(|j-k|) + \gamma_2(|j-k|)),$$

which is again bounded by $C_2(x_2 - x_1)^2(h_2 - h_1)^2$, since conditions (A.1) and (A.2) imply that $\sup_n n^{-1} \sum_{j=1}^{n} \sum_{k=1}^{n} (\gamma_1(|j-k|) + \gamma_2(|j-k|)) < \infty$ by use of Doukhan (1994). Doukhan states that if we let $\mathbf{X} = (\mathbf{X_t})_{\mathbf{t} \in \mathbf{Z}}$ be a real valued stationary centered at expectation and mixing random process, the central limit theorem problem is to provide explicit sufficient conditions on $\mathbf{X}$ for a central limit theorem to hold. Set $S_n = X_1 + ... + X_n$, then convergence of $\frac{\sigma_n^2}{n}$ holds if the sequence $\mathbf{X}$ is ϕ-mixing, $\sum_{n=0}^{\infty} \phi_n^{\frac{\delta+1}{2+\delta}} < \infty$ and $E|X_1|^{2+\delta} < \infty$ for some $\delta > 0$ and $\sigma_n^2 = E|S_n|^2$. Then the tightness property of the sequence of processes

$$n^{1/2}[\hat{g}_h(x) - E\{\hat{g}_h(x)\}]$$

on $H \times I$ is implied by Theorem 3 in Bickel and Wichura (1971).

Bickel and Wichura (1971) state that if we let $T_1, \cdots, T_q$ be subsets of [0,1], each of which contains 0 and 1, then $T = T_1 \times \cdots \times T_q$. Then suppose that each X_n vanishes along the lower boundary of T, and that there exist constants $\beta_0 > 1, \gamma_0 > 0$ and a finite nonnegative measure μ on T with continuous marginals such that (X_n, μ) is said to satisfy condition (β_0, γ_0) for each n if

$$P\{\min\{|X(B)|, |X(C)|\} \geq \lambda\} \leq \lambda^{-\gamma}(\mu(B \cup C))^{\beta}$$

for all $\lambda > 0$ and every pair of neighboring blocks B and C in T. Then the tightness condition is in force. Together with the finite dimensional convergence property, this implies that the theorem holds.

Theorem 5. *Suppose that assumptions (A.1)-(A.6) are satisfied. As* $n \to \infty$

$$\sup_{x \in I, h \in H} n^{1/2} |\hat{g}_h(x) - E\{\hat{g}_h(x)\}|$$

converges weakly to a random variable that has the same distribution as that of $\sup_{x \in I, h \in H} |G(h,x)|$. $G(h,x)$ *is a Gaussian process with zero mean and covariance function* $cov(h_1, x_1, h_2, x_2)$ *so that*

$$P\{G(h,x) \textit{ is continuous for all } (h,x) \in H \times I\} = 1,$$

and consequently $P\{\sup_{x \in I, h \in H} |G(h,x)| < \infty\}$*=1.*

Proof. Let D_j^* be

$$D_j^* = w_n(h_2, x_2, j) - w_n(h_1, x_1, j).$$

Then,

$$\begin{aligned} E\{U_n(h_2, x_2) - U_n(h_1, x_1)\}^2 &= \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} (\gamma_1(|j-k|) + \gamma_2(|j-k|)) D_j^* D_k^* \\ &\leq C_3\{(h_2 - h_1)^2 + (x_2 - x_1)^2\}. \end{aligned}$$

The rest of the proof proceeds as in Theorem 3 in Section 3.2, by defining the pseudo metric d here by $d\{(h_2, x_2), (h_1, x_1)\} = [E\{G(h_2, x_2) - G(h_1, x_1)\}^2]^{1/2}$.

3.4 SiZer for the Comparison of More than Two Time Series

This section is devoted to testing the equality of k time series. The model is

$$Y_{ij} = f_i(j) + \epsilon_{ij}, \quad i = 1, \ldots, k, \; j = 1, \ldots, n.$$

Consider testing the scale-space version of the hypotheses:

$$H_0 \; : \; f_{1,h}(x) = f_{2,h}(x) = \cdots = f_{k,h}(x) \quad vs. \quad H_1 \; : \; not \; H_0. \tag{3.10}$$

The extension of the approach in Subsection 3.3 is not straightforward because the pairwise comparison would not be sufficient for this testing problem. Therefore, we instead propose to compare two sets of residual time series under the null and alternative hypotheses, respectively. First individual local linear estimates are fit under the alternative hypotheses in (3.10), then the set of residuals found from taking the individual datapoints from their respective group estimates are computed, composing the first set of residuals. Next, all of the datasets are combined together and one common local linear estimate is approximated, and the residuals found from every data value and this common function are computed, composing the second set of residuals. If the null hypothesis is true, then the residuals from the combined estimate and the residuals from the subdivided estimates should be roughly the same. In this way, the comparison of multiple time series is converted into the comparison of two time series. A similar idea is used in Park and Kang (2008) for the independent case.

For obtaining the residuals, one could use a pilot bandwidth h_p to estimate the mean function, that is different from the bandwidth h used for constructing the SiZer map. For simplicity, however, $h_p = h$ is taken in this analysis.

We summarize our procedure as follows:

1. Using k response sets, $Y_{1j}, Y_{2j}, \ldots, Y_{kj}$, create an estimated function that is a fit to its own data. Thus, $\hat{f}_{1,h}, \, \hat{f}_{2,h}, \ldots, \hat{f}_{k,h}$ are obtained where

$$\hat{f}_{i,h}(x) = \frac{1}{n} \sum_{j=1}^{n} w_n(h, x, j) Y_{ij}$$

for $i = 1, 2, \ldots, k$.

2. Compute the residuals for these k sets, take each and compute $Y_{ij} - \hat{f}_{i,h}(j)$ as the estimate of the errors ϵ_{ij}'s from the ith population.

3. Obtain $\hat{f}_h(\cdot)$, the local linear estimator of the common scale-space regression function $f_h(\cdot)$ under H_0, which has the form:

$$\hat{f}_h(x) = \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} w_n(h, x, j) Y_{ij}, \tag{3.11}$$

   where $w_n(h, x, j)$'s are the local linear weights.

4. Compute the residuals for this one combined set, let $Y_{ij} - \hat{f}_h(j)$ be the estimate of the errors $\tilde{\epsilon}_{ij}$'s under the null hypothesis in (3.10).

The idea is that if H_0 is true, $Y_{ij} - \hat{f}_{i,h}(j)$ and $Y_{ij} - \hat{f}_h(j)$ would be similar time series. Hence, the equality of k time series can be verified by comparing these two sets of residual time series with the tool proposed in Section 3.3 to see whether or not their means are actually equal.

3.4.1 More than two time series: Simulation

In this subsection, some simulated examples are presented to compare three different time series. In each example, three time series are generated from either $N(0, 1)$, MA(1) with $\theta = 0.5$, MA(5) with $\theta's = 0.9, 0.8, 0.7, 0.6, 0.5$, or AR(1) with $\phi = 0.5$, with the length $n = 100$. In the first example, the mean regression functions are all zero. Therefore, each graph should demonstrate that there is no signal and if the time series structure can be estimated accurately, the behavior of residuals under each of the three estimation functions should be the same as that of the commonly estimated function. These two mean functions of residuals should thus leave a difference of nothing and the lack of trend would leave us with three SiZer maps of no significant trend.

In Figure 3.18, the first column of graphs shows the generated time series and their family of smooths. Here, the dependence structures of the time series are known; the case where the

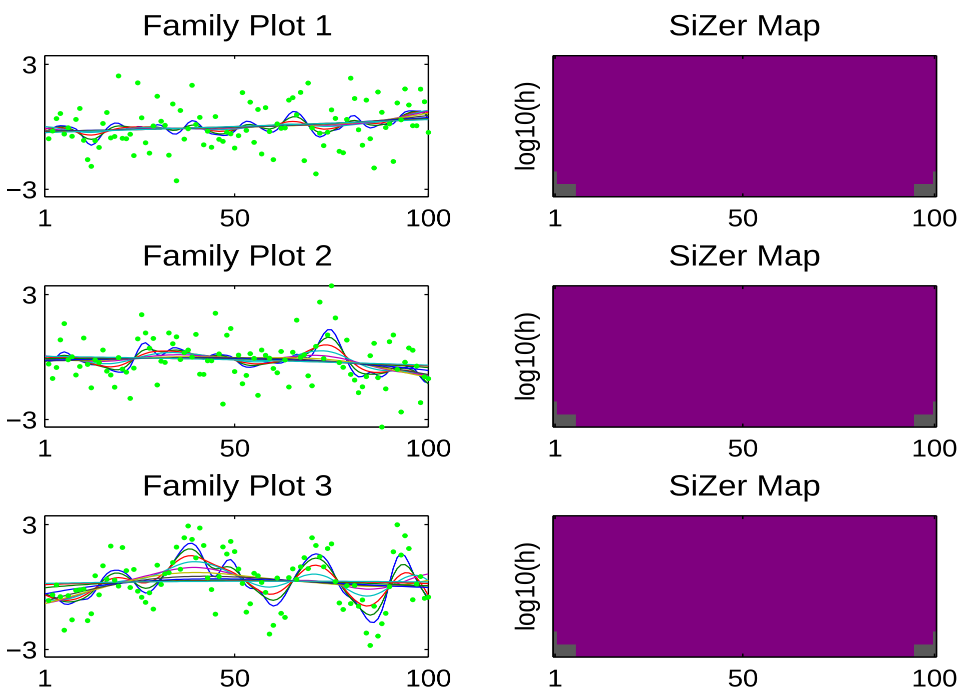


Figure 3.18: SiZer plots for comparing three time series with the same zero mean. Family Plot 1 has no signal and Normal errors. Family Plot 2 has no signal and MA(1) errors. Family Plot 3 has no signal and AR(1) errors.

time series structures are estimated is given in the next subsection. The second column shows the SiZer maps constructed by comparing two sets of residual time series. In other words, for $i = 1, 2, 3$, the ith row of the second column corresponds to the SiZer map comparing $Y_{ij} - \hat{f}_{i,h}(j)$ and $Y_{ij} - \hat{f}_h(j)$. All purple colors indicate that the mean functions of the residual time series are indeed similar according to SiZer analysis.

In the second example, the error structures remain the same as in the first, but the sine curve $f_1(x) = \sin(6\pi x)$ is added to the first sample. The last two samples remain with no signal, only the correlated error structure.

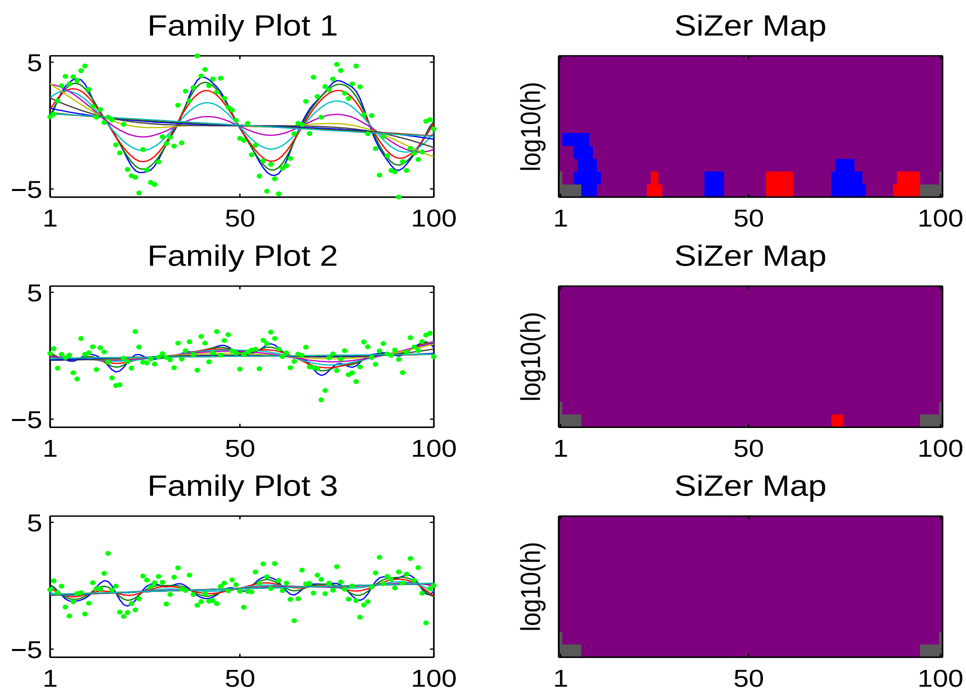


Figure 3.19: SiZer plots for comparing three time series with different mean functions. Family Plot 1 has a Sine signal and Normal errors. Family Plot 2 has no signal and MA(1) errors. Family Plot 3 has no signal and AR(1) errors.

Figure 3.19 shows some significant features in the SiZer maps, which implies the differences of the mean functions. The first one shows far more significant trends since it is different from the others, and the trend clearly suggests the presence of the sine curve that is inserted into the first sample.

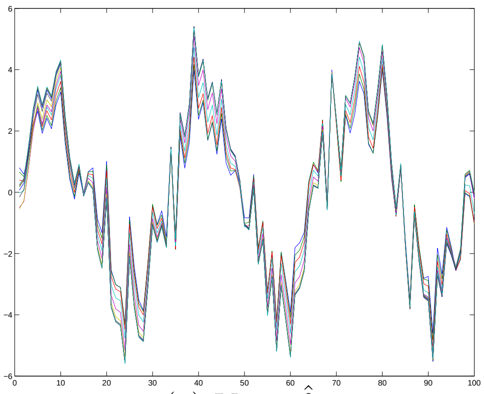
For illustration purposes, Figure 3.20 displays two sets of residuals in the SiZer maps in Figure 3.19. For example, we compare the two sets of residuals in Figures 3.20 (a) and (b), and create the SiZer map in the first row in Figure 3.19. Multiple lines in the plots correspond to different bandwidths. In Figures 3.20 (a) and (b), they look similar for large bandwidths, but show some differences for small bandwidths, whose significance appear in the first SiZer map of Figure 3.19. The rest of the plots look very similar each other, and their corresponding SiZer maps show only purple in Figure 3.19.

In Figure 3.21 we can see that the SiZer maps flag some significant differences between the mean functions of the family plots, where the top plot has a sine and linear signal and the bottom two are again both just correlated noise and have no signal. These SiZer maps show that unlike what the null hypothesis proposes, these three functions are not equal and should not be combined into one single function because their residuals are not equal to those residuals where each set has its own representative function.

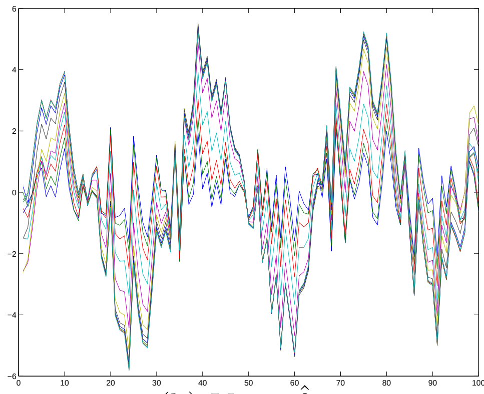
The next simulation examines the situation where all three samples have different dependent structures. In these three examples, the first family plot will have AR(1) errors, plot 2 will have MA(1) errors, and the third and final plots will have MA(5) errors. This will also allow a look at the SiZer maps' behavior and ability to detect differences between multiple time series when weak and strongly correlated errors are present and mixed together.

In the first sample here, Figure 3.22 correctly shows that all SiZer maps are purple, that there is no difference between the mean of the two residual sets. These functions could all be combined because each one of them has an equal signal, that of no signal at all. This detection is correctly identified despite the strong correlation of MA(5) errors in the bottom family plot.

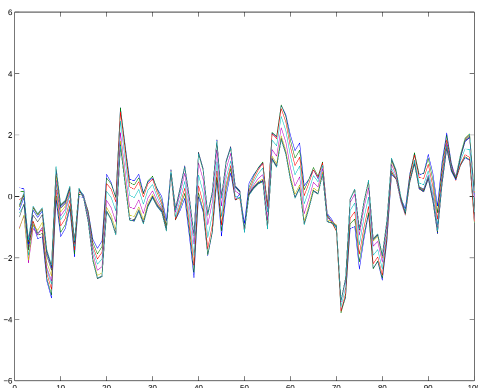
In Figure 3.23, when the sine signal is present in the first sample and no signal is in the last two samples, the first SiZer plot correctly identifies the difference between the two



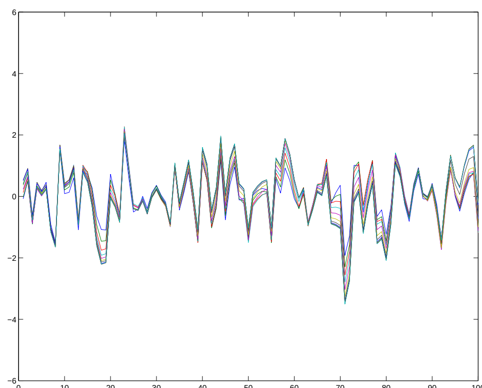
(a) $Y_{1j} - \hat{f}_h$



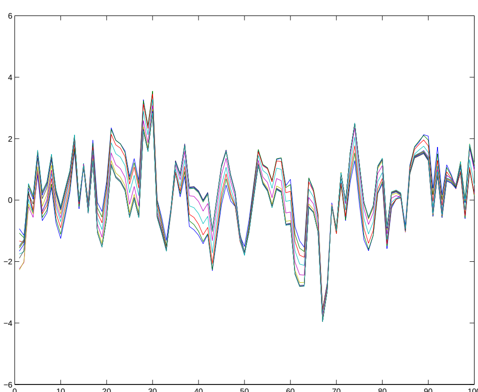
(b) $Y_{1j} - \hat{f}_{1,h}$



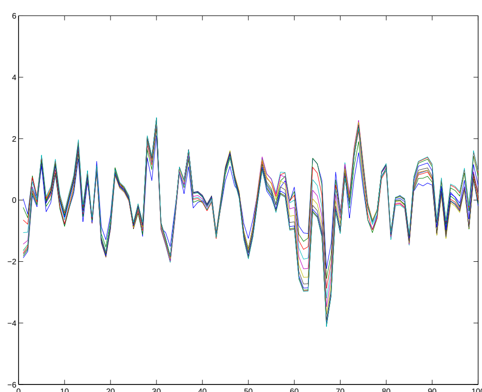
(c) $Y_{2j} - \hat{f}_h$



(d) $Y_{2j} - \hat{f}_{2,h}$



(e) $Y_{3j} - \hat{f}_h$



(f) $Y_{3j} - \hat{f}_{3,h}$

Figure 3.20: Two sets of residual plots in the SiZer maps in Figure 3.19.

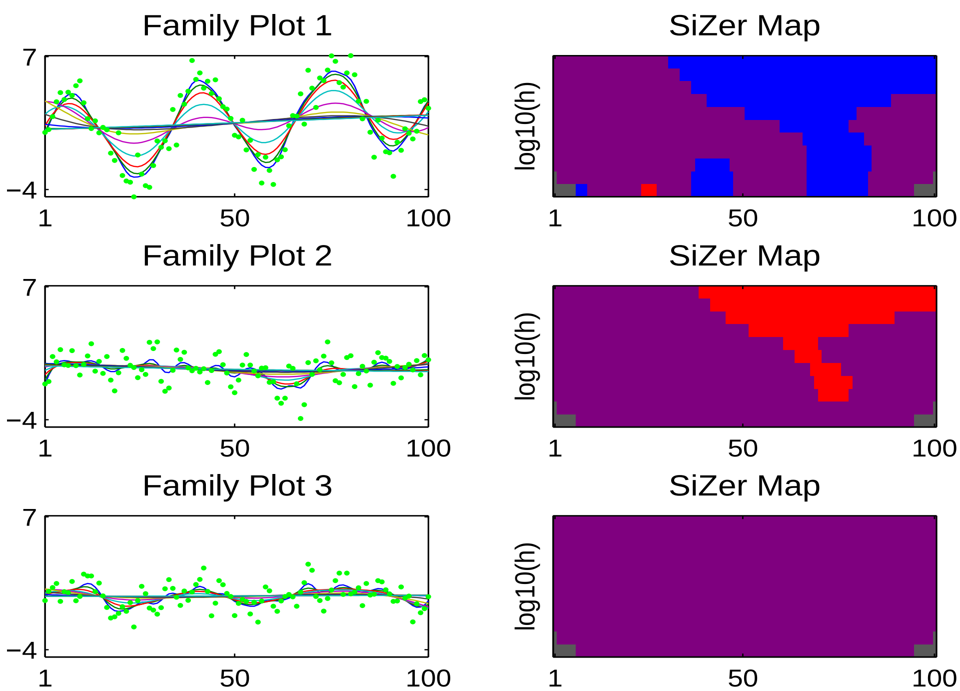


Figure 3.21: SiZer plots for comparing three time series with different mean functions. Family Plot 1 has a Sine plus Linear signal and Normal errors. Family Plot 2 has no signal and MA(1) errors. Family Plot 3 has no signal and AR(1) errors.

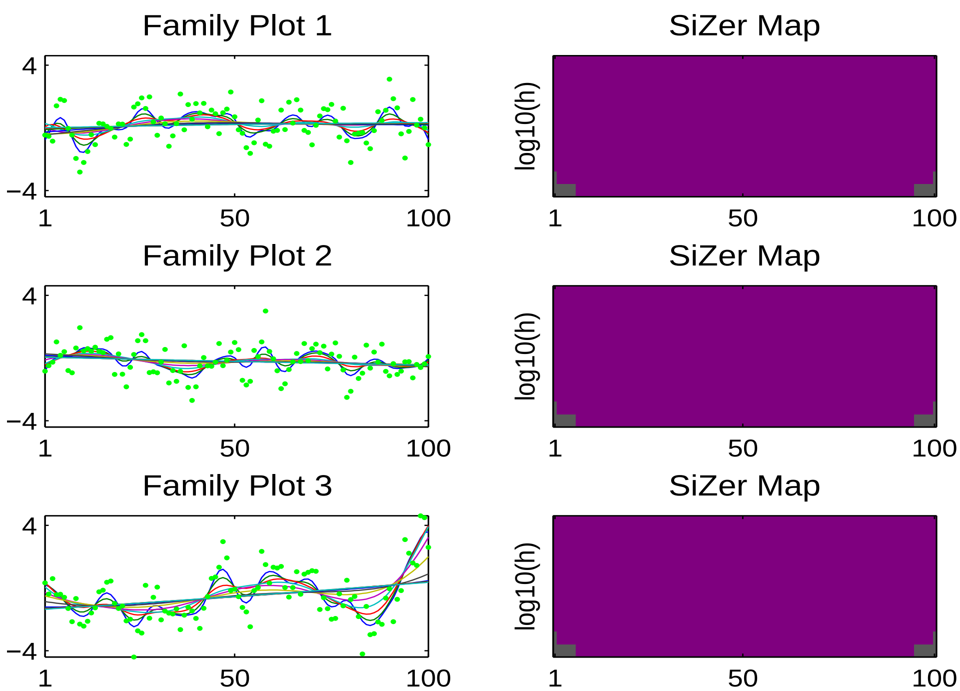


Figure 3.22: SiZer plots for comparing three time series with the same mean function and different dependent error structures. Family Plot 1 has no signal and AR(1) errors. Family Plot 2 has no signal and MA(1) errors. Family Plot 3 has no signal and MA(5) errors.

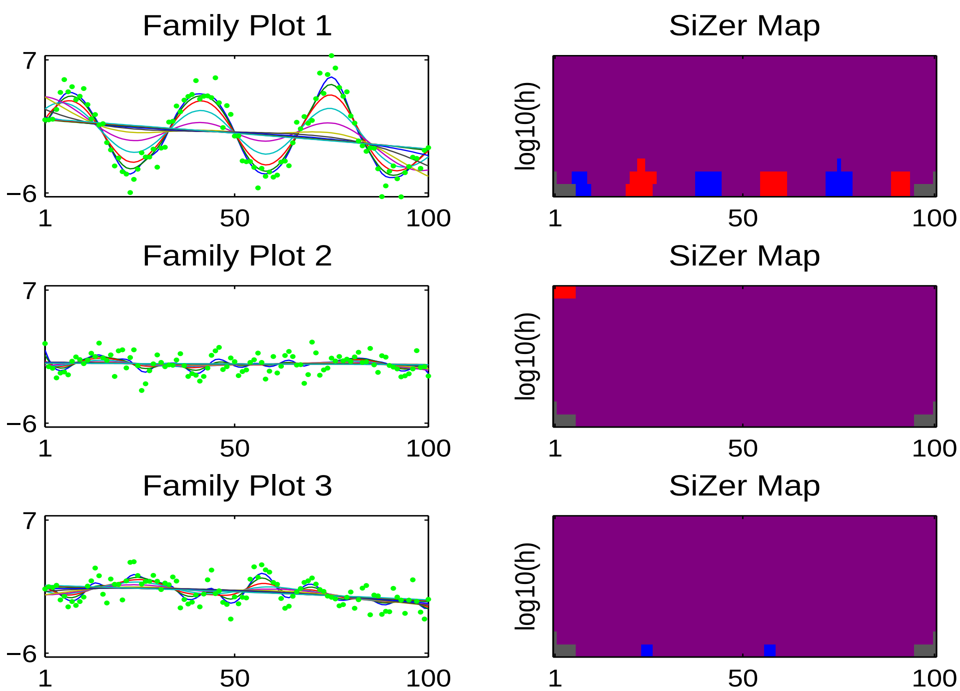


Figure 3.23: SiZer plots for comparing three time series with different mean functions. Family Plot 1 has a Sine signal and AR(1) errors. Family Plot 2 has no signal and MA(1) errors. Family Plot 3 has no signal and MA(5) errors.

residual sets when this individual function has a sine trend. These highlighted portions in this first and the few in the second and last SiZer map show that there are significant differences between these three datasets, besides that of their time series correlation structure.

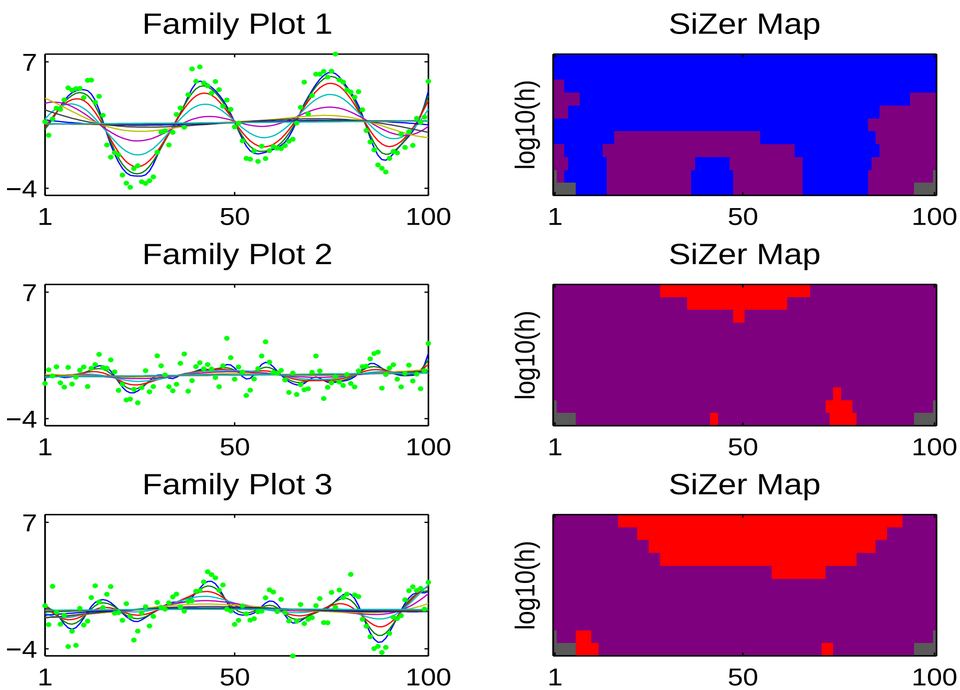


Figure 3.24: SiZer plots for comparing three time series with different mean functions. Family Plot 1 has a Sine + Linear signal and AR(1) errors. Family Plot 2 has no signal and MA(1) errors. Family Plot 3 has no signal and MA(5) errors.

In Figure 3.24, all three SiZer maps indicate that there is a significant difference between the two residual sets. Due to the fact that a sine plus linear signal is added to the first plot, highlighted pixels can be seen in all of the plots, indicating that it detects the difference between the residuals of the single function and when each dataset gets its own individual function. It has denoted that because of this first signal, these two sets of residuals are not equal, and thus the null hypothesis should be rejected.

3.4.2 More than two time series: Real data analysis

This subsection is devoted to illustrating the procedure for more than two time series applied to real data.

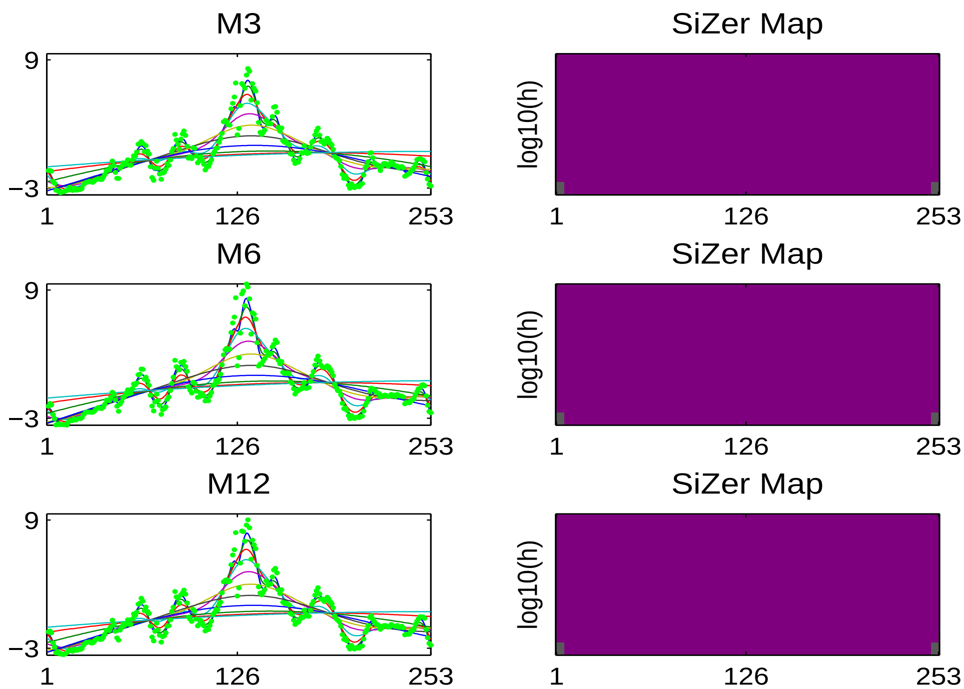


Figure 3.25: Multiple comparison of the trends for the yields of the 3-month, 6-month, and 12-month Treasury bills measured as the bi-monthly average from July 1959 to August 2001.

Instead of doing pairwise comparison for the real dataset examples used earlier, the previous approach is taken again to look at multiple comparison of time series, now with real datasets. In Figure 3.25, almost identical returns for 3, 6, and 12 month treasury bills are depicted in all of the SiZer maps, and all are colored purple as a result. This mimics the earlier decision that none of the pairs of yields are significantly different from one another, and here again there is no significant difference between the two sets of residuals. This indicates that, in line with the null hypothesis, these three time series have equal mean functions and could be combined together into one single common function of Treasury bills using (3.11) that does not indicate the span of months.

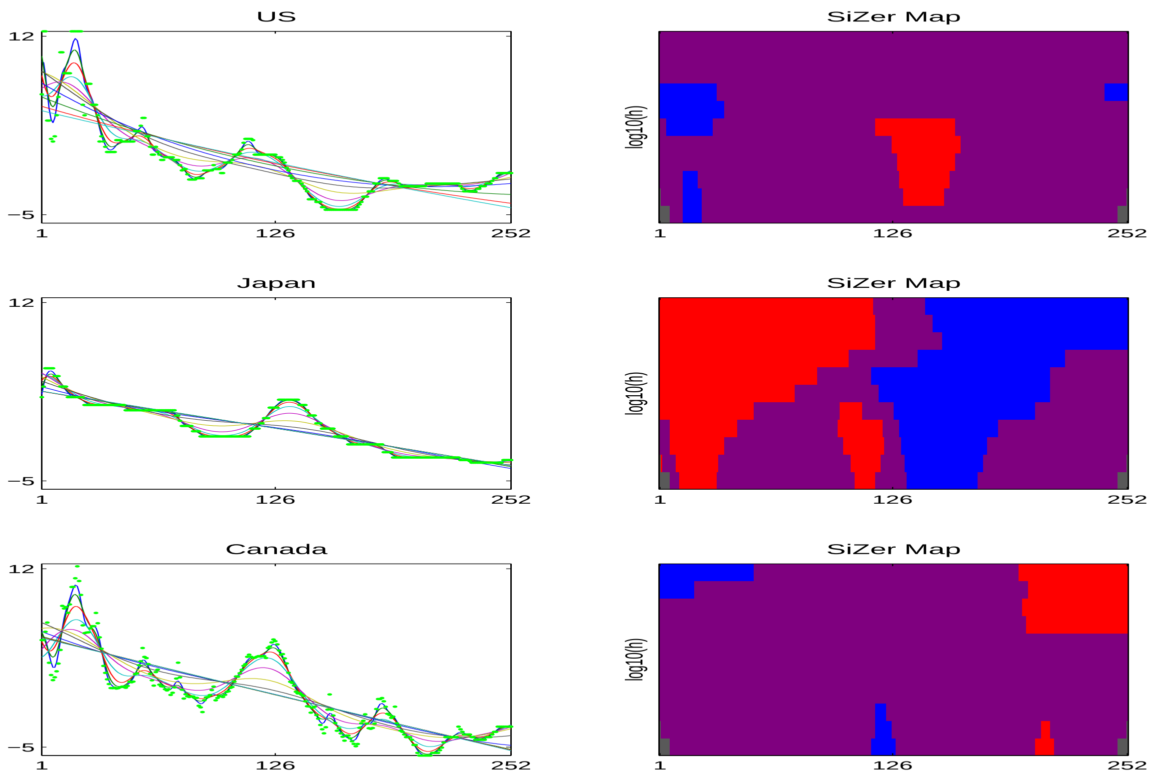


Figure 3.26: Multiple comparison of the trends for long term target interest rates for US, Canada, and Japan from January 1980 to December 2000.

Also we compare the three interest rate trends for the three countries looked at previously in Figure 3.17. In Figure 3.26, there are differences that occur within each SiZer map, denoting that there are significant differences present. Seen before in Figure 3.17, there exist pairwise differences between all of the countries. The presence of these differences is also correctly detected when we compare each set of residuals from each country's individual estimated function to the residuals from the overall estimation.

Chapter 4

Two Dimensional SiZer

In this chapter, we propose to extend the original one-dimensional SiZer to two dimensions, thereby allowing us to take spatial correlation structure into account in image analysis. In one dimension, the scale-space is viewed as an overlay of curves. In two dimensions, overlays are no longer possible, so Godtliebsen et al. (2004) propose a movie version that shows the progression through various bandwidths instead. The most challenging part becomes the statistical inference, which previously was based on where the derivatives of a curve had statistically significant increases and decreases. In two dimensions, the derivative, or slope, is replaced by partial derivatives, or gradients, and thus a new approach is required. Godtliebsen et al. (2002) and Duong et al. (2008) study multivariate kernel estimation in a scale-space. In this chapter we will focus on a regression setting.

4.1 Review of the Two Dimensional SiZer with Independent Errors

Godtliebsen et al. (2004) propose what they refer to as S^3, Significance in Scale-Space. They combine scale-space, statistical inference, and visualization methodology using the full scale-space (that is, all levels of resolution of the image) to try and separate out important underlying structures from spurious noise artifacts.

The statistical model that underlies S^3 is

$$Y_{i,j} = s(i, j) + \epsilon_{i,j},$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, m$ index pixel locations, s represents the underlying non-random signal (thought of as a smooth, deterministic function evaluated at a rectangular grid), and the $\epsilon_{i,j}$'s are the noise, assumed to be independent random variables with $E(\epsilon_{i,j}) = 0$ and $Var(\epsilon_{i,j}) = \sigma^2_{i,j}$. Note that the variance can be different at each location. The gray level scale-space slices are simply Gaussian smooths; that is, discrete two-dimensional convolutions of a spherically symmetric Gaussian density, with the data, denoted

$$\hat{s}_h(i,j) = \sum_{i'=1}^{n} \sum_{j'=1}^{m} Y_{i',j'} K_h(i-i', j-j'),$$

or in matrix notation

$$\underline{\hat{s}}_h = \underline{K}_h * \underline{Y}$$

where * denotes bivariate discrete convolution, and

$$K_h(i,j) = K_h(i) K_h(j),$$

for $i = (1-n), \ldots, (n-1)$ and $j = (1-m), \ldots, (m-1)$, where

$$K_h(i) = \frac{\exp(-(i/h)^2/2)}{\sum_{i'=1-n}^{n-1} \exp(-(i'/h)^2/2)}.$$

Using $\hat{s}_h$ above can cause the severe boundary effects due to a result of averaging in zeros from outside the image. To overcome this problem they subtract the mean of the $Y_{i,j}$ before smoothing. Therefore, the estimate becomes

$$\underline{\hat{s}}_h = A(\underline{Y}) + \underline{K}_h * (\underline{Y} - A(\underline{Y})),$$

where A is the matrix operator which returns the constant matrix whose common entries are the average of the entries of its matrix argument; that is, each

$$A(\underline{Y})_{i,j} = \frac{1}{nm} \sum_{i'=1}^{n} \sum_{j'=1}^{m} Y_{i',j'}.$$

Another consideration is again the number of points that should be inside each kernel window, where the effective sample size here is given as

$$\underline{ESS} = (\underline{K}_h * \underline{1})/(K_h(0,0)).$$

Here $\underline{1}$ is the n by m matrix having a one in each entry and the denominator of $K_h(0,0)$ is the rescaling that assigns value one to the pixel in the center, and appropriately down-weights the values assigned to other pixels. When h is large, ESS(i,j) is large and when h is small, ESS(i,j) is small. In boundary regions, the boundary effect yields an appropriately small value of ESS(i,j). To make S^3 inferences simultaneously across location, Godtliebsen et al. (2004) use an average effective sample size of

$$ESS_2 = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} ESS(i,j) \right) /(nm).$$

In their paper, they point out that data sparsity issues need much more attention because of possible large regions with no data. Therefore, data sparsity issues need much more attention. The idea of an effective sample size in image analysis is to take a kernel weighted count of the number of points in each window and give us a guideline for when the data is too sparse for inference. Because there are nm independent data points, the smoothing process can be viewed as averaging in groups of size ESS_2. Therefore, the number of independent averages is approximately

$$\ell = \frac{nm}{(ESS_2)}. \tag{4.1}$$

In estimating the noise level, or the variance of $\epsilon_{i,j}$, one must decide if it is reasonable to assume that $\sigma^2_{i,j}$ is constant. In the case of heteroscedasticity, Godtliebsen et al. (2004) estimate $\sigma^2_{i,j}$ by smoothing the squared residuals and subtracting the mean of the squared residuals out to take into account any boundary issues:

$$\underline{\hat{\sigma}}^2_h = \underline{ESSQ} \cdot \{A(C_S(\underline{Y} - \underline{\hat{s}}_h)) + \underline{K}_h * [C_S(\underline{Y} - \underline{\hat{s}}_h) - A(C_S(\underline{Y} - \underline{\hat{s}}_h))]\}$$

where $\underline{ESSQ}$ is the matrix with

$$\frac{ESS(i,j)}{ESS(i,j) - 1}$$

in entry (i,j), the operator $\cdot$ denotes element by element matrix multiplication, C_S is the matrix operator which squares all entries of its matrix argument, and A is the matrix operator which returns the constant matrix whose common entries are the average of the entries of its matrix argument. In the homoscedastic case, these estimates are pooled to estimate the common σ^2. Since interior points have a more stable $\sigma_h^2(i,j)$, an ESS weighted average is used,

$$\hat{\sigma}_h^2 = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} ESS(i,j) \hat{\sigma}_h^2(i,j) \right) / \left(\sum_{i=1}^{n} \sum_{j=1}^{m} ESS(i,j) \right).$$

The gradient of the underlying signal s at any given (i,j) location is

$$G(s) = [(s_1)^2 + (s_2)^2]^{1/2},$$

where s_1 is the partial derivative in the vertical direction (indexed by i) and s_2 is the partial derivative in the horizontal direction (indexed by j). The corresponding estimate of the gradient is

$$\hat{G}_h(s) = [(\hat{s}_{h,1})^2 + (\hat{s}_{h,2})^2]^{1/2},$$

where the partial derivatives are estimated by

$$\underline{\hat{s}}_{h,1} = \underline{K}_{h,1} * \underline{Y},$$

$$\underline{\hat{s}}_{h,2} = \underline{K}_{h,2} * \underline{Y},$$

where

$$K_{h,1}(i,j) = K_h'(i) K_h(j),$$

$$K_{h,2}(i,j) = K_h(i) K_h'(j),$$

and

$$K_h'(i) = (-i/h) K_h(i).$$

Note that they use the Nadaraya-Watson (e.g. see Nadaraya (1964)) type estimator rather than computing a numerical derivative for simplicity. Two dimensional derivatives using the local linear estimator can be developed, but we suggest it as future work.

The gradient version of S^3 flags pixels with arrows as significant when $\hat{G}_h(s)$ is higher than the noise level, rejecting a null hypothesis of the form

$$H_0 : G_h(s) = 0.$$

The null distribution of this test is based on the bivariate Gaussian distribution

$$\begin{pmatrix} \hat{s}_{h,1} \\ \hat{s}_{h,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_2^2 \end{pmatrix}\right),$$

which is exact if the noise terms $\epsilon_{i,j}$ have a Gaussian distribution or follow the Central Limit Theorem. Godtliebsen et al. (2004) use the independent case in their paper, therefore $\sigma_{12}^2 \approx 0$, and the approximate distribution is

$$\frac{\hat{s}_{h,1}^2}{\sigma_1^2} + \frac{\hat{s}_{h,2}^2}{\sigma_2^2} \sim \chi_2^2,$$

so the null hypothesis is rejected for pixels where

$$\frac{\hat{s}_{h,1}^2}{\hat{\sigma}_1^2} + \frac{\hat{s}_{h,2}^2}{\hat{\sigma}_2^2} > q_{\chi_2^2}(\alpha').$$

Here α is the nominal level that is commonly used and α' is the value of the significance level that is used to make the inference simultaneous across all pixels. The estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ for the independent case can be found in Godtliebsen et al. (2004). Doing an approximation based on the number of independent blocks, ℓ in (4.1), they propose using

$$q_{\chi_2^2}(\alpha') = -2\log(1 - (1-\alpha)^{1/\ell}).$$

as the quantile for determining significance. This quantile is calculated based on the equation $\left(P(\chi_2^2 \leq q_{\chi_2^2}(\alpha'))\right)^l = 1 - \alpha$.

Their approach to highlighting significant features in an image is to use arrows to point in the direction of a significant gradient. If the gradient, that is, the local slope at each pixel in the image is found to be statistically significant, they use green arrows, overlaid on the scale-space gray level image, to point in that direction. A statistically significant extreme is therefore surrounded by a ring of significant gradients pointing towards the peak or valley. In addition to the MPEG movie versions of the image that can be viewed progressing through many different levels of resolution, their paper chooses different slices of the scale-space that represent several scales. The images begin with a substantial noise component still present, which can lead to not very clearly defined outlines of the images and few arrows that appear to indicate statistically significant features. These images end their progression with a blurring of the outlines of the images and many structures marked as significant, which is not particularly useful, since features of interest are typically not visible at this scale. Some slice choice in between the beginning and end of the movie is therefore obviously typically the optimal choice of resolution.

4.2 Incorporation of the Variogram

This section provides a brief introduction of spatial statistics. See Cressie (1991) for more details. Spatial data are distinguished by observations that are obtained at spatial locations. A random field, or random process, is denoted by $Z(\tilde{s})$ indexed by spatial location $\tilde{s}$ in d-dimensional Euclidean space. The process $Z(\tilde{s})$ may exhibit spatial correlation, similar to time series where the correlation between different time points must be taken into account. Looking at

$$\{Z(\tilde{s}) : \ \tilde{s} \in D\},$$

D is a fixed subset of $\mathcal{R}^d$. Observing $\{Z(\tilde{s}_1), ..., Z(\tilde{s}_n)\}$ at known spatial locations $\{\tilde{s}_1, ..., \tilde{s}_n\}$, the most common way to model the correlation structure is with the variogram, $2\tilde{\gamma} = Var[Z(\tilde{s}_1 - Z(\tilde{s}_2)]$.

A random process $Z(\cdot)$ which has $E\left(Z(\tilde{s})\right) = \mu$ and $cov\left(Z(\tilde{s}_1), Z(\tilde{s}_2)\right) = C(\tilde{s}_1 - \tilde{s}_2)$ is defined to be second-order, or weakly, stationary because the mean is constant over the spatial domain and the covariance depends on the separation between points but not on their absolute location. These conditions imply that the mean is constant across all locations and that the function $C(\cdot)$, the covariogram, is a stationary covariance function. Additionally, if $C(\cdot)$ is a function only of $||\tilde{s}_1 - \tilde{s}_2||$, then $C(\cdot)$ is called isotropic. This implies that the process is uniform in all directions and it is only the distance between two spatial locations that is of importance. Likewise, if $\{Z(\tilde{s}) : \ \tilde{s} \in D\}$ satisfies $E\left(Z(\tilde{s})\right) = \mu$ and $Var\left(Z(\tilde{s}_1) - Z(\tilde{s}_2)\right) = 2\tilde{\gamma}(\tilde{s}_1 - \tilde{s}_2)$ then $Z(\cdot)$ is said to be intrinsically stationary, which is similar to weak stationarity, but is instead concerned that the variance of the differences of Z at pairs of locations only depends on the distance between their locations. If $2\tilde{\gamma}(\tilde{s}_1 - \tilde{s}_2)$ is also a function only of $||\tilde{s}_1 - \tilde{s}_2||$, then $2\tilde{\gamma}(\cdot)$ is called isotropic.

If we replace locations, $\tilde{s}_1$ and $\tilde{s}_2$ with u and $u+v$, we have a covariogram $Cov(Z(u), Z(u+v)) = C(v)$, where the function depends only on v, the distance between the two locations. In time series, this is often referred to as the autocovariance function. Having a stationary process is beneficial because it is more general and the process behaves the same at any location, but it also makes it easier to carry out spatial prediction, or kriging. Again, if a spatial process is intrinsically stationary, it has a constant mean and the variance of the differences of Z at pairs of locations only depends on v, the displacement between locations. In this case there is what is called a semivariogram

$$\tilde{\gamma}(v) = \frac{1}{2} Var\left(Z(u+v) - Z(u)\right).$$

The semivariogram is often the most preferred tool for characterizing spatial processes. Related to this is the variogram, which is a function of the spatial dependence of variance. It is given with the following equation

$$2\tilde{\gamma}(v) = Var\left(Z(u+v) - Z(u)\right).$$

Let (i', j') be the location $(u+v)$, and (i'', j'') be the location u and let $Y_{i,j} \equiv Z(u)$, then it can be said that

$$2\tilde{\gamma}(v) = Var\left(Z(u+v) - Z(u)\right) = Var\left(Y_{i',j'} - Y_{i'',j''}\right)$$

and if the lag v only depends on the difference between the two locations, then

$$2\tilde{\gamma}\left(d((i', j'), (i'', j''))\right) = Var\left(Y_{i',j'} - Y_{i'',j''}\right),$$

where d is the value of the Euclidean distance between the coordinates (i', j') and (i'', j''). To mimic the true variogram, it is necessary to calculate the empirical variogram in data analysis. Because some estimators of the variogram, such as the method of moments, do not yield a valid variogram model, empirical variograms are often approximated by model functions such as exponential and Matern. The exponential variogram is given as

$$2\tilde{\gamma}(v) = c_0 + c_1(1 - \exp(-v/b))$$

and the Matern variogram is given as

$$2\tilde{\gamma}(v) = c_0 + c_1 \left(1 - \frac{(v/2b)^{\nu}}{2\Gamma(\nu)}\right) B_{\nu}(v/b)$$

where c_0 is the nugget, the height of the jump of the discontinuity at the origin; c_1 is the sill, the limit of the variogram tending to infinity lag distances; ν is a parameter that controls the smoothness of the random field; b represents the range, or the distance in which the difference of the variogram from the sill becomes negligible; and B_{ν} is a modified Bessel function of the second kind. If $\nu = 0.5$, it is a special case of exponential variogram and if $\nu \to \infty$ then the Matern converges to the Gaussian variogram.

4.3 Review of Multiple Comparison Procedures

Multiple hypothesis testing is a difficult process that attempts to perform individual inferences on null hypotheses, while maintaining an acceptable Type I error rate control. This can be daunting if the data include numerous locations, such as when testing for the presence of a signal in spatial data. One choice is whether it is desirable to test each voxel location separately, or focus on testing clusters of data for signal presence. Historically, the approach has been to test each location separately and then adjust the level of the test to the multiplicity of locations in order to control what is known as the familywise error rate (FWER). A classic way of controlling for the FWER has been done by the Bonferroni correction, which is known to be conservative. Another concern is the significant loss of power that typically comes along with this and other multiple hypothesis testing procedures. Because of its difficulty and loss of power, many practitioners decide to circumvent doing controls for multiple testing altogether. In multiple testing problems, it is also often difficult to identify a threshold that will control a measure of false positives across the entire image. In this section, we explore some newer procedures that perform multiple tests and compare their performances.

The most commonly used correction is the Bonferroni, which is a standard fixed threshold procedure where the p-values are assessed according to

$$p_{(i)} \leq T.$$

Here $p_{(i)}$ is an ordered p-value and T will typically correspond to a threshold test by which any test statistic that leads to a p-value which is less than the given threshold will be declared significant. This threshold for Bonferroni is

$$T = \alpha/N$$

where N is the total number of locations tested, or hypotheses performed, and α is again the nominal level α. Bonferroni normally overcorrects and because it is so tight, it is too conservative, declaring few individual locations to be significant.

The first method to be looked at as a Bonferroni alternative is Holm's method, which, similar to Bonferroni, makes no assumptions on the dependence of the tests. Holm (1979) compares the smallest p-value, $p_{(1)}$, to α/N and if one rejects the corresponding hypothesis, then move on to the next smallest p-value, $p_{(2)}$, which is compared to $\alpha/(N-1)$ and so on. This is referred to a step-down method and it starts at i=1 and stops the first time the inequality below is violated. It then rejects all hypotheses that have smaller p-values. The inequality for comparison is

$$p_{(i)} \leq \alpha \frac{1}{N-i+1}.$$

Although it is nice that Bonferroni and Holm do not make assumptions on dependence since this makes them flexible and hence more applicable in a wide variety of situations, they also do not make use of the spatial structure that may exist in certain datasets.

The second criterion is control of FDR (False Discovery Rate), which involves the proportion of errors among the rejected hypotheses, not just the probability of getting even a single false positive, such as in the FWER. Benjamini and Hochberg (1995) propose FDR as a way to look at the number of errors committed in independent multiple-comparison problems and also only be concerned about the probability of a false rejection given that a rejection has occurred. Again, the ordered p-values, $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(N)}$ and H_i their corresponding H_0's are used. Let k be the largest i for which

$$p_{(i)} \leq \alpha \frac{i}{N},$$

then all H_i, $i = 1, \ldots, k$ are rejected. This will control the FDR, the proportion of rejected H_0's which are erroneously rejected, at rate α.

The next procedure, proposed by Pavlicova et al. (2003), is p-value adaptive thresholding (PAT), an adaptation of FDR. This procedure also uses N tested hypotheses and their ordered p-values. PAT, however, has two steps,

$$(1) N_0 = \max\{i : p_{(i)} \leq \frac{\alpha}{N - i + 1}\}$$

$$(2) k = \max\{i : p_{(i)} \leq \frac{(i - N_0 + 1)\alpha}{N - N_0 + 1} \ ; i = N_0, \ldots, N\}.$$

The PAT procedure rejects those null hypotheses whose p-values are $p_{(1)} \leq ... \leq p_{(k)}$. In step (1), the number of hypotheses is reduced using the Holm's procedure and then FDR is applied to the remaining hypotheses in step (2). Note that if N_0 is set equal to 1, the PAT procedure reduces to the original FDR.

Another adaptation to the FDR procedure is presented by Benjamini and Yekutieli (2001), which shall here be called dFDR, since it can be applied to data that has positively dependent test statistics. Their proposal is to replace α with $\frac{\alpha}{(\sum_{i=1}^{N} \frac{1}{i})}$ in

$$k = \max\{i : p_{(i)} \leq \frac{i}{N} \alpha\}.$$

Let $H_{0(i)}$ be the hypothesis corresponding to $p_{(i)}$. Then one would reject the null hypotheses $H_{0(1)}, \ldots, H_{0(k)}$ according to their ordered p-values. This procedure obviously increases the range of problems that can be evaluated by a FDR-type procedure. Now, better assessments can be performed on data that possess positive dependency instead of possibly avoiding the issue of multiple testing altogether.

In addition to these four thresholding procedures, the quantile that uses independent blocks is tested here. This was previously used in several versions of SiZer and it will be used for Original 2-*d* SiZer and a new SiZer that allows for dependence. In the next section, the use of these quantiles will be incorporated into our new version of SiZer, the Spatial SiZer, which takes the dependent structure of an image into account.

4.4 Spatial SiZer

In this section, we introduce a dependent two-dimensional SiZer which accounts for the case when the $\epsilon_{i,j}$'s are not assumed to be independent random variables. This Spatial SiZer provides the benefit of performing statistical analysis on spatially correlated data. Most commonly, data with spatial dependence structure come from images that are difficult to model because of the additional dimension and the usually high level of noise. This tool can circumvent these problems and provide informative and understandable analysis with visualizations for a vast range of statistical problems. One of those types of problems can include PET analysis where additional tests cannot be run on a patient because of the safety levels of radiation the patient can be exposed to. SiZer can take these types of images that are noisy and determine whether a vague feature is in fact a significant structure.

In this Spatial SiZer, we too use the green arrows to point in the direction where there are significant gradients on the image so that a statistically significant local slope is surrounded by a ring of arrows pointing towards its peak. Again, views are taken of the images at various levels of signal blurring, which is represented via the convolution of $s * K_h$ for different values of h.

As in the independent version, the signal estimate used is

$$\hat{s}_h(i,j) = \sum_{i'=1}^{n} \sum_{j'=1}^{m} Y_{i',j'} K_h(i - i', j - j'),$$

but now the $\epsilon_{i,j}$'s can have dependent structure and $2Cov(\epsilon_{i,j}, \epsilon_{i',j'}) = \sigma^2_{i,j} + \sigma^2_{i',j'} - 2\tilde{\gamma}(d((i,j),(i',j')))$ where d is the Euclidean distance and $\tilde{\gamma}$ is the semivariogram. Then, the variance for the partial derivative in the vertical direction, indexed by 1, is

$$
\begin{aligned}
\hat{\sigma}^2_1 &= Var(\hat{s}_{h,1}(i,j)) \\
&= Cov\left(\sum_{i'=1}^{n}\sum_{j'=1}^{m} Y_{i',j'}K_{h,1}(i-i',j-j'), \sum_{i''=1}^{n}\sum_{j''=1}^{m} Y_{i'',j''}K_{h,1}(i-i'',j-j'')\right) \\
&= \sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,1}(i-i',j-j')K_{h,1}(i-i'',j-j'')Cov(Y_{i',j'}, Y_{i'',j''}) \\
&= \sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,1}(i-i',j-j')K_{h,1}(i-i'',j-j'') \times \\
& \quad \left[\frac{\sigma^2_{i',j'} + \sigma^2_{i'',j''}}{2} - \tilde{\gamma}\left(d((i',j'),(i'',j''))\right)\right] \\
&= \left(\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,1}(i-i'',j-j'')\right)\left(\sum_{i'=1}^{n}\sum_{j'=1}^{m} K_{h,1}(i-i',j-j')(\sigma^2_{i',j'})\right) \\
&- \sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,1}(i-i',j-j')K_{h,1}(i-i'',j-j'')\left[\tilde{\gamma}\left(d((i',j'),(i'',j''))\right)\right]
\end{aligned}
$$

Similarly, the variance in the horizontal direction is

$$
\begin{aligned}
\hat{\sigma}^2_2 &= Var(\hat{s}_{h,2}(i,j)) = \left(\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,2}(i-i'',j-j'')\right)\left(\sum_{i'=1}^{n}\sum_{j'=1}^{m} K_{h,2}(i-i',j-j')(\sigma^2_{i',j'})\right) \\
&- \sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,2}(i-i',j-j')K_{h,2}(i-i'',j-j'')\left[\tilde{\gamma}\left(d((i',j'),(i'',j''))\right)\right]
\end{aligned}
$$

and the covariance between the partial derivatives in the horizontal and vertical directions is

$$\begin{aligned}\hat{\sigma}^2_{12} &= Cov(\hat{s}_{h,1}, \hat{s}_{h,2}) \\ &= \frac{1}{2}\left(\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,1}(i-i'', j-j'')\right)\left(\sum_{i'=1}^{n}\sum_{j'=1}^{m} K_{h,2}(i-i', j-j')(\sigma^2_{i',j'})\right) \\ &+ \frac{1}{2}\left(\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,2}(i-i'', j-j'')\right)\left(\sum_{i'=1}^{n}\sum_{j'=1}^{m} K_{h,1}(i-i', j-j')(\sigma^2_{i',j'})\right) \\ &- \sum_{i'=1}^{n}\sum_{j'=1}^{m}\sum_{i''=1}^{n}\sum_{j''=1}^{m} K_{h,1}(i-i', j-j')K_{h,2}(i-i'', j-j'')\left[\tilde{\gamma}\left(d((i',j'),(i'',j''))\right)\right].\end{aligned}$$

Here, $\sigma^2_{i,j}$ can be estimated using the formulas introduced in Section 4.1, and $\tilde{\gamma}$ can be estimated using a parametric model as explained in Section 4.2. In our analysis, we assume that the variance of $\epsilon_{i,j}$ is a constant but this assumption can be released to the case of heteroscedasticity. Once these pieces are computed, they are put together to form the covariance matrix

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}^2_1 & \hat{\sigma}^2_{12} \\ \hat{\sigma}^2_{12} & \hat{\sigma}^2_2 \end{pmatrix}.$$

We no longer assume that there are independent errors, that is $\sigma^2_{12} \neq 0$. This leads to the need for a new test statistic and resulting distribution, that can be seen as

$$\begin{pmatrix} \hat{\sigma}^2_1 & \hat{\sigma}^2_{12} \\ \hat{\sigma}^2_{12} & \hat{\sigma}^2_2 \end{pmatrix}^{-1/2} \begin{pmatrix} \hat{s}_{h,1} \\ \hat{s}_{h,2} \end{pmatrix} = \begin{pmatrix} \hat{t}_{h,1} \\ \hat{t}_{h,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

The sum of the squares of these two test statistics results in

$$\hat{t}^2_{h,1} + \hat{t}^2_{h,2} \sim \chi^2_2(\alpha').$$

Thus, the null hypothesis

$$H_0 : G_h(s) = 0.$$

is rejected for those pixels which have values

$$\hat{t}^2_{h,1} + \hat{t}^2_{h,2} > q_{\chi^2_2}(\alpha').$$

This $q_{\chi_2^2}(\alpha')$ represents the appropriate quantile from the χ_2^2 distribution and it is chosen via one of the six thresholding techniques described in Section 4.3. The nominal level of α used in all of the examples within this dissertation is α=0.05 and the α' is chosen so that the inference is made simultaneously across all pixel values, and therefore, all hypothesis tests at the overall level α.

For pixels whose null hypothesis is rejected, we determine that pixel location to have a gradient significantly higher than those of its surrounding area. To denote this significance, an arrow is drawn in that gradient direction using the corresponding vertical and horizontal direction vector

$$\begin{pmatrix} \hat{s}_{h,1} \\ \hat{s}_{h,2} \end{pmatrix}.$$

4.5 Numerical Study

The first part of this section illustrates simulated examples with a signal involving cosine and either an exponential or Matern dependence structure. The second part deals with real data analysis on datasets that were also previously analyzed by Godtliebsen et al. (2004).

4.5.1 Simulation

In the simulation, we generate a signal of image sizes of 20×20 from the following equation for $i = 1, ..., n$ and $j = 1, ..., m$

$$s(i,j) = 3\left[\cos\left(\frac{180 \times 10}{\pi}\left(i - \frac{n}{2}\right)\right) \cdot \cos\left(\frac{180 \times 10}{\pi}\left(j - \frac{m}{2}\right)\right)\right]_+ . \tag{4.2}$$

Here, n is the length in the vertical direction, m is the length in the horizontal direction, $(\cdot)_+$ indicates the positive part of the function (the negative pieces are set to zero).

Then, an error with mean zero and with either an exponential or Matern covariance function is generated in order to construct a covariance matrix for the error field. The exponential covariance function is

$$\Sigma = a \times \exp\left(\frac{-d}{b}\right) \tag{4.3}$$

where d represents a Euclidean distance, a represents the level of covariance, and b is the spatial range, or how long the correlation lasts. We show a's of 0.01 and 0.1 in the tables of results. Values for b are 3, 10, and 30. The bandwidths at which the data are viewed include h=1.5, 2, 3, 4. All of the thresholding quantiles mentioned in the previous section are also used, and are represented by q=1, Holm's FWER; q=2, FDR; q=3, PAT; q=4, dFDR; q=5, independent blocks with the Spatial SiZer; and q=6, independent blocks with the Original SiZer.

In addition to the exponential covariance, a Matern covariance function is used:

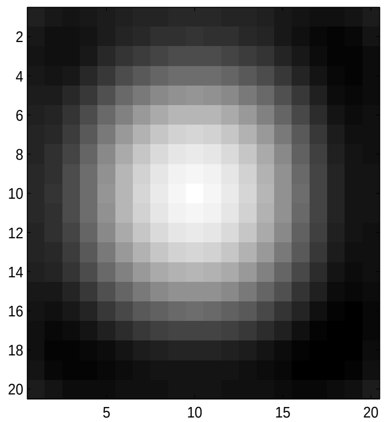
$$\Sigma = (a(d/b)^{\nu}) \times besselK\left(\nu, \frac{d}{b}\right) \tag{4.4}$$

where d is again a Euclidean distance; $besselK$ is a modified Bessel function of the second kind; a represents the sill parameter, the height at which the variogram flattens out and the range is obtained; b is the spatial range parameter, the distance at which the difference of the variogram from the sill gets negligible; and ν is a parameter representing smoothness. The Matern simulation uses the same values for a and b that are used for the exponential covariance function, and the values for ν include 0.5, 1, 2.

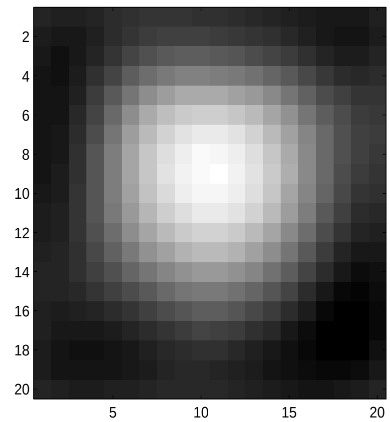
For example, Figure 4.1 (a) has the signal that is generated with no error added to the image. The next three figures, (b)-(d), have the simulated data, where now error has been added to the signal to create the new images with either exponential or Matern covariance functions.

The tables at the end of Chapter 4 show the effectiveness of the multiple comparison procedures using the various thresholding procedures and the Spatial versus Original SiZer. To evaluate the effectiveness of these procedures, Type I error, Type II error, and the proportion of pixels identified correctly, are all used. The performance of the cases given include the Matern covariance function and the exponential covariance function, which is a special case of the Matern, when ν=0.5.

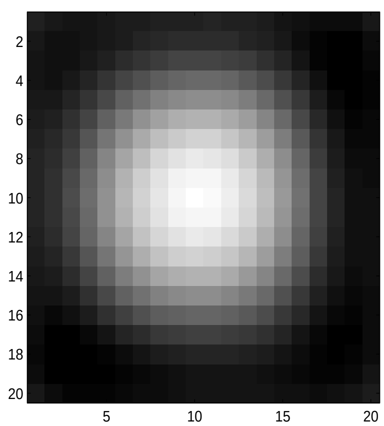
Table 4.1 shows Type I errors for the Matern covariance structure and $a = 0.01$. Here when q=5, the independent blocks quantile with Spatial SiZer is by far the best performer



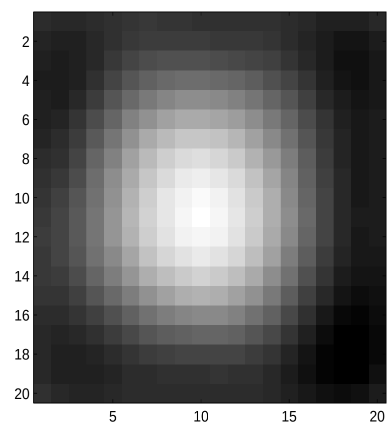
(a) No Error



(b) Exponential Error, a=0.1, b=3



(c) Matern Error, a=0.01, b=3, ν=1



(d) Matern Error, a=0.1, b=3, ν=1

Figure 4.1: Original Signal (a) and simulated data, which consists of the signal + error (b-d)

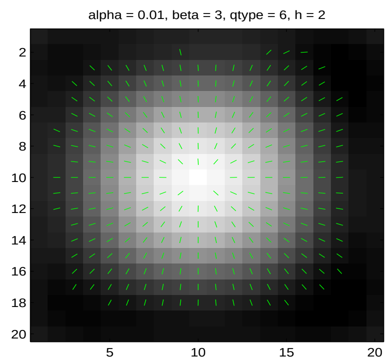
in every scenario. No matter what the level of b or ν or h, q=5 is always the lowest level of Type I error and it is drastically lower than the Original SiZer, where the differences within the same combination of parameters can be very commonly as different as 0.0000 for the Spatial and 0.2237 for the Original. In Table 4.2, with a=0.1 and a Matern covariance function, there exists the same type of performance, with q=5 having the lowest Type I error rate for every combination of the parameters, and when the error rate is 0 it can also tie with other quantiles within the Spatial SiZer. Again, the performance of the Original SiZer is dismal, with its Type I error rate ranging between 0.1988 and 0.2374. In Table 4.3 with the exponential covariance function, the same type of performance is exhibited. The lowest values for the Type I error rate are when q=5 for all outcomes, regardless of b or h. The performance of the Original SiZer, q=6, is also very disappointing at all combinations of a, b and h.

In Tables 4.4, 4.5, and 4.6 looking at the Type II error performance, often q=2, Holm's FWER with the Spatial SiZer, does the best. If q=6, the Original SiZer with independent blocks does fairly well, but as a caveat, as has been seen, the Original SiZer always has a large number of spurious pixels in it, this means that it is quite prone to easily declaring pixels to be significant. Thus, the cases where it would declare something to not be significant when it actually is, are expected to be very few. In Tables 4.5 and 4.8, the Spatial SiZer has higher type II errors and lower correct proportions when a=0.1. This occurs more specifically when q=5, but, that quantile and the Spatial SiZer still have significantly achieved the lowest Type I error. Since we are more concerned about controlling for Type I error, we prefer $q = 5$ even though the method sometimes produces high Type II error (equivalently low power). This means that the method $q = 5$ may miss important features in the image, which shows the limitation of the proposed method. We propose the increase of power as future work. This can be achieved by improving a quantile that can account for spatial dependence in the data.

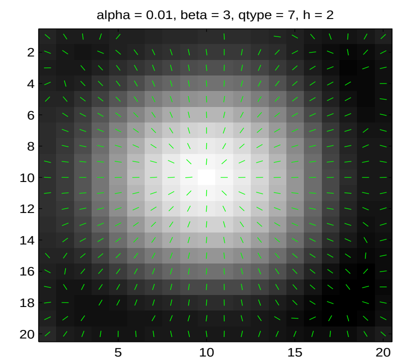
In Table 4.7, at the smaller bandwidths, where h=1.5 or 2, the independent blocks quantile with the Spatial SiZer, q=5, is consistently the best across the board when it comes to correctly identifying pixels' statistical significance. This proportion identified correct includes those that were significant and deemed significant and those insignificant and deemed insignificant. When moving into the larger bandwidths, other quantiles begin to perform better. At h=3 with the smaller value of b, q=4, PAT (p-value adaptive thresholding) performs the best within the Spatial SiZer and q=5 performs the best at larger values of b=10 or 30. At the largest value of h, now various quantiles, including q=4 are performing the best with the Spatial SiZer, and at all values within the table, the Original SiZer, q=6, is always outperformed. In Table 4.9, when a=0.01, we see the same pattern as at the smaller bandwidths, where h=1.5 or 2, the independent blocks quantile with the Spatial SiZer, q=5, is consistently the best across the board when it comes to correctly identifying pixels' statistical significance. Then when moving into the larger bandwidths, other quantiles begin to perform better. However, when a=0.1 in Table 4.9, the independent blocks quantile with the Spatial SiZer, q=5, does not perform well and another quantile should be chosen.

These results imply that if the proportion of correctly identified pixels is one's main priority, instead of reducing Type I error, the independent blocks quantile with the Spatial SiZer, q=5, may be preferred for smaller resolutions and at the larger bandwidths, various methods should be evaluated to see which performs best under the present scenario. As previously mentioned, however, the quantile which takes spatial correlation into account will improve the performance even more. We propose its development and theoretical justification as future work.

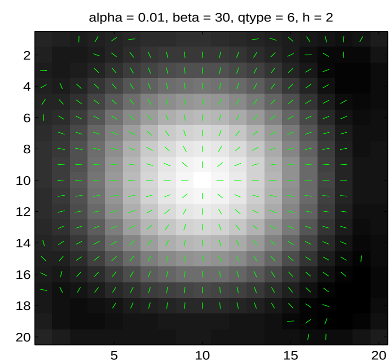
Figure 4.2 presents plots that have the signal mentioned above in equation (4.2), along with error generated with the exponential covariance function (4.3). In this Figure, all of the graphs in the left panel are using q=5, which is the independent blocks quantile in the Spatial SiZer and the right panel are graphs with q=6, the independent blocks quantile with the Original SiZer. In Figures 4.2 (a) and (c), a=0.01 and the midrange view of bandwidth h=2 are used. Both graphs seem to have a few spurious pixels out around the edges, but the number of these pixels is quite small in comparison to those of the Original SiZer in (b) and



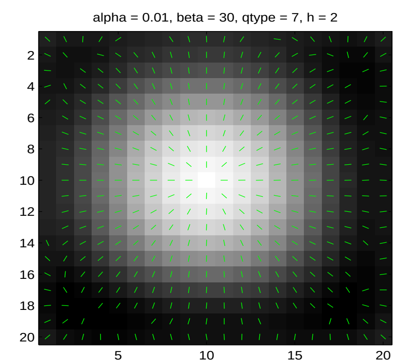
(a) a=0.01, b=3, h = 2, q=5



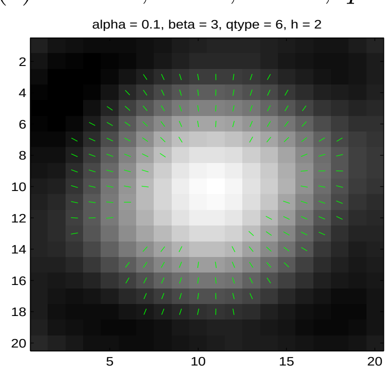
(b) a=0.01, b=3, h = 2, q=6



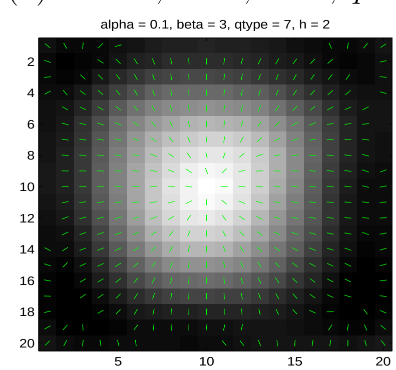
(c) a=0.01, b=30, h = 2, q=5



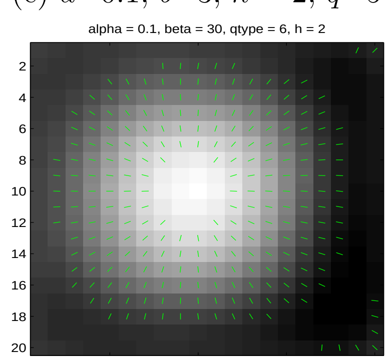
(d) a=0.01, b=30, h =2, q=6



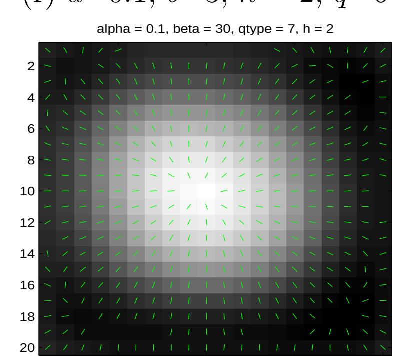
(e) a=0.1, b=3, h = 2, q=5



(f) a=0.1, b=3, h = 2, q=6



(g) a=0.1, b=30, h = 2, q=5



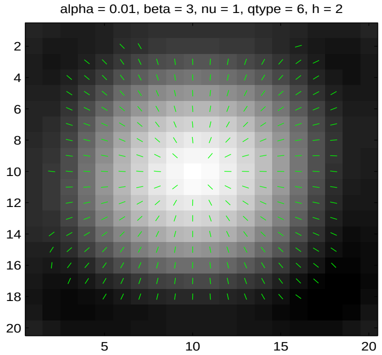
(h) a=0.1, b=30, h = 2, q=6

Figure 4.2: SiZer maps with the simulated signal and exponential covariance function.

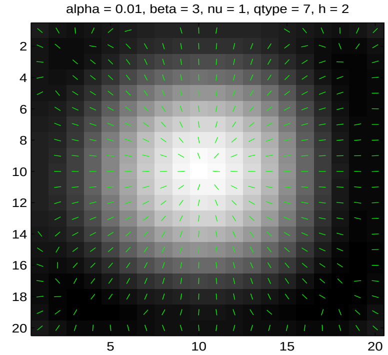
(d). Figures 4.2 (e) and (g) use a=0.1, which is the higher of the two a's used and thus denotes that there is a higher level of correlation here. When there is a higher degree of dependence or correlation present, it is harder to determine what should be declared actual trend and what should be attributed to the error structure. Thus, there are fewer pixels highlighted as significant than in Figures 4.2 (a) and (c) as an increasing amount of variation is attributed to the error structure and not real trend.

In Figure 4.2 (e), we see that although the Spatial SiZer clearly focuses correctly on the center, there is a corner where pixels seem to be missing where one would expect significant pixels to be. On the other hand, the Original SiZer in Figure 4.2 (f), albeit it encompasses the signal in the center completely, has highlighted a large number of spurious pixels. Because Figures 4.2 (b),(d),(f), and (h) have a number of spuriously highlighted pixels in the corners of the Figures and the range that is declared significant is also falsely enlarged, this shows that the Spatial SiZer does a much better job at accounting for the level of dependent structure than the Original SiZer.

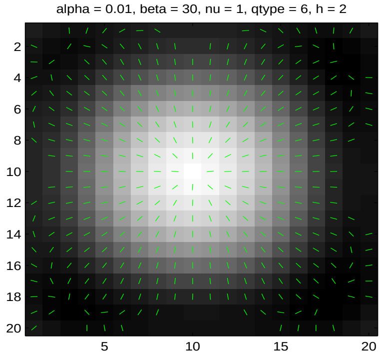
Figure 4.3 presents plots that also have the signal mentioned above in equation (4.2), now with error generated with the Matern covariance function (4.4). Again in this Figure, all of the graphs in the left panel are using q=5, which is the independent blocks quantile in the Spatial SiZer and the right panel are graphs with q=6, the independent blocks quantile with the Original SiZer. In Figures 4.3 (a) and (c), a=0.01 and the midrange view at bandwidth h=2. Similar to those with the exponential covariance function in Figure 4.2, both graphs seem to have a few spurious pixels out around the edges, especially when b=30, but the number of these pixels is still smaller than for the Original SiZer in (b) and (d). Figures 4.3 (e) and (g) have an a=0.1, which indicates the higher level of correlation here, therefore it is harder to detect trend. Thus, there are fewer pixels highlighted as significant than in Figures 4.3 (a) and (c). Figure 4.3 (g) does have a few spurious pixels in its corners, but in the comparative figure, Figure 4.3 (h), there is a much larger number of spuriously highlighted pixels in the corners of the figure and around the circumference of the signal in the center,



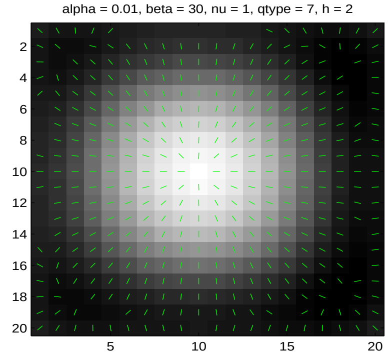
(a) a=0.01, b=3, ν=1, h = 2, q=5



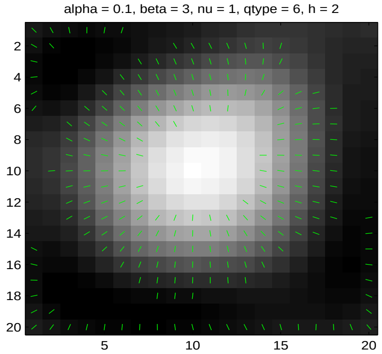
(b) a=0.01, b=3, ν=1, h = 2, q=6



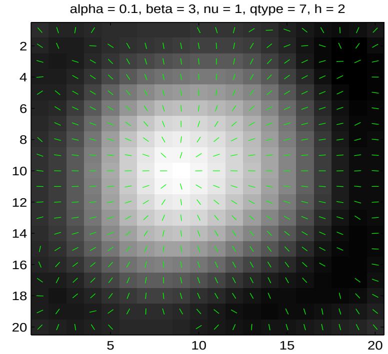
(c) a=0.01, b=30, ν=1, h = 2, q=5



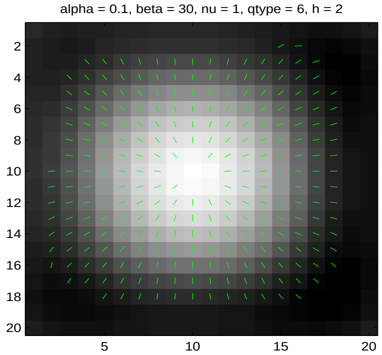
(g) a=0.01, b=30, ν=1, h = 2, q=6



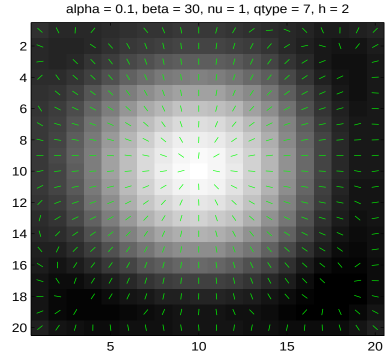
(e) a=0.1, b=3, ν=1, h = 2, q=5



(f) a=0.1, b=3, ν=1, h = 2, q=6



(g) a=0.1, b=30, ν=1, h = 2, q=5



(h) a=0.1, b=30, ν=1, h = 2, q=6

Figure 4.3: SiZer maps with the simulated signal and Matern covariance function.

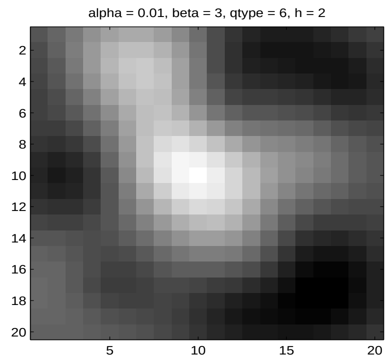
showing again that the Spatial SiZer does a much better job at accounting for the dependent structure than does the Original SiZer.

In Figure 4.4 are the SiZer maps of the cases where the Original SiZer, q=6, and Spatial SiZer, q=5 are used to assess a situation where there is in fact no signal present, merely error. In these maps, the Spatial SiZer's performance is exemplary, with not a single pixel erroneously highlighted. In the Original SiZer, however, there are numerous green arrows, denoting that there is a significant trend present, where in fact there is not. Looking at both a=0.01 and a=0.1 and both extremes of the b, where it is 3 or 30, in all circumstances, the Spatial SiZer can correctly differentiate between the dependent error structure and true significant trend, while the Original SiZer cannot. In Figure 4.5, the Spatial SiZer with the Matern covariance function has identical performance to that of the exponential in Figure 4.4 with no pixels highlighted as significant. Also in Table 4.10 are the baseline references for Type I error for when there is no signal. The table uses an exponential covariance, which also coincides for the case where Matern covariance has ν=0.5. Again the Spatial SiZer with independent blocks, q=5, has a greatly reduced level of Type I errors when compared to the Original SiZer with independent blocks, q=6.

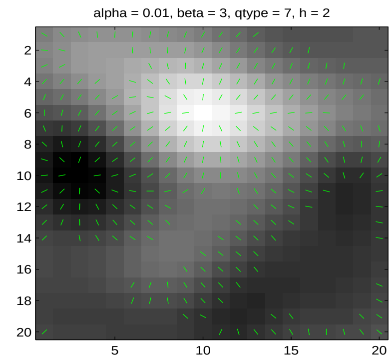
4.5.2 Real data analysis

This subsection will focus on three datasets that were analyzed by Godtliebsen et al. (2004) method S^3, Significance in Scale-Space, discussed in Section 4.1. In all of these results, the data are fit using a Matern variogram with a weighted least squares method. A Matern model is used since it has a general form. The weights used for the fitting are those of Cressie (1991), which apportions more weight to lags that are closer to zero, and the values for a, b, and ν are given below for each analysis. Presented in Figure 4.6 are the original three datasets.

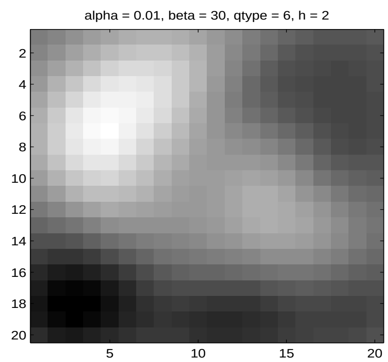
In this first example, Figures 4.7 and 4.8 are SiZer plots created from raw data of confocal microscopy, courtesy of Havard Rue. These data have a low noise level, but the original 80×80 image has a dim elliptical figure in its bottom left hand corner, thus still making it difficult



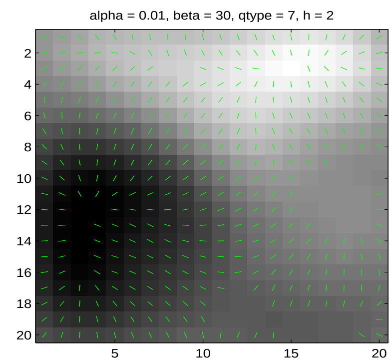
(a) a=0.01, b=3, h = 2, q=5



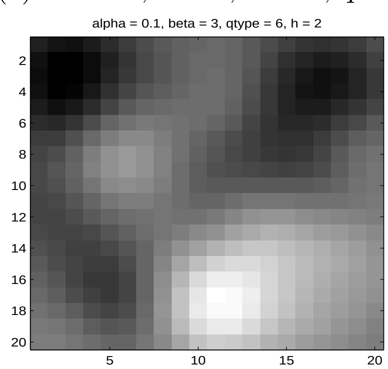
(b) a=0.01, b=3, h = 2, q=6



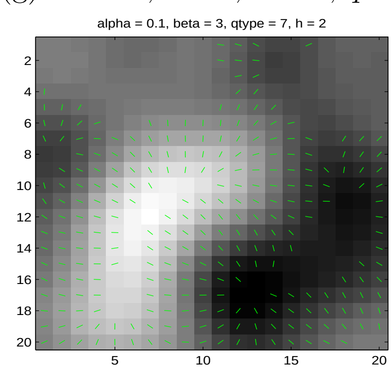
(c) a=0.01, b=30, h = 2, q=5



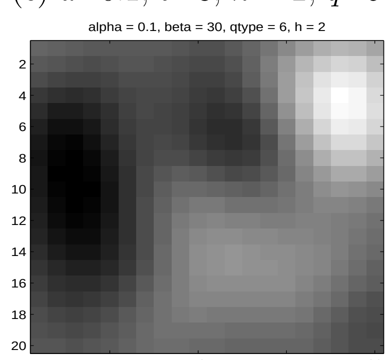
(g) a=0.01, b=30, h = 2, q=6



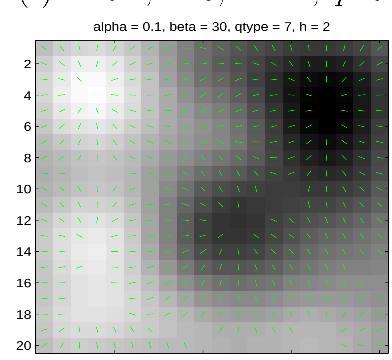
(e) a=0.1, b=3, h = 2, q=5



(f) a=0.1, b=3, h = 2, q=6

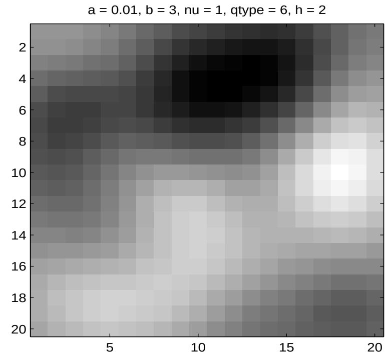


(g) a=0.1, b=30, h = 2, q=5

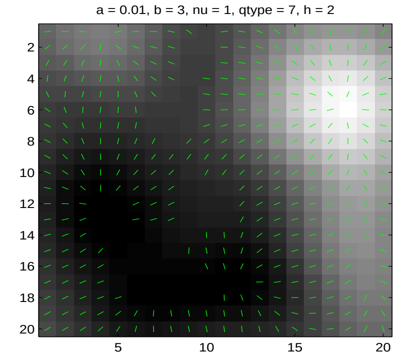


(h) a=0.1, b=30, h = 2, q=6

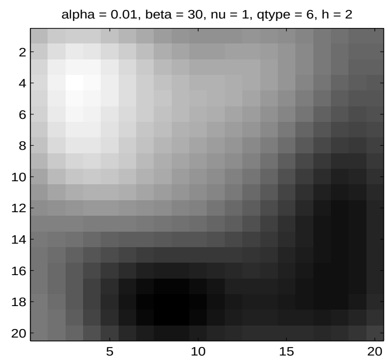
Figure 4.4: SiZer maps with no signal and exponential covariance function.



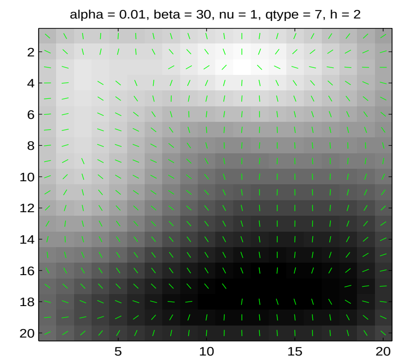
(a) a=0.01, b=3, h = 2, q=5



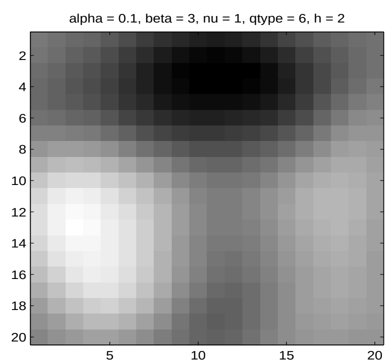
(b) a=0.01, b=3, h = 2, q=6



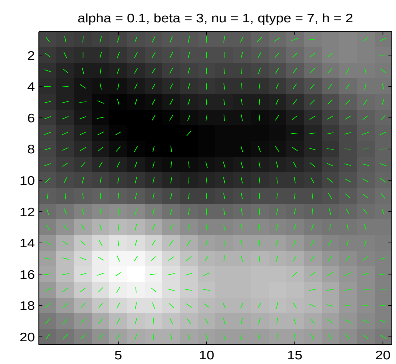
(c) a=0.01, b=30, h = 2, q=5



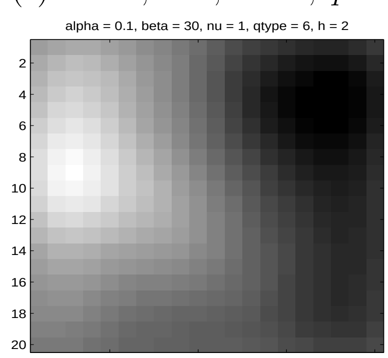
(g) a=0.01, b=30, h = 2, q=6



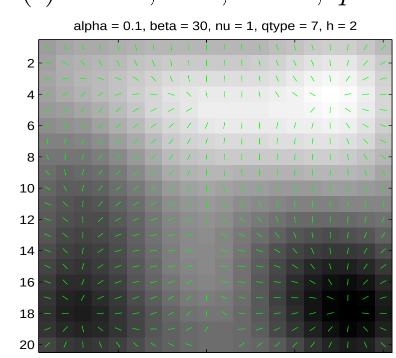
(e) a=0.1, b=3, h = 2, q=5



(f) a=0.1, b=3, h = 2, q=6

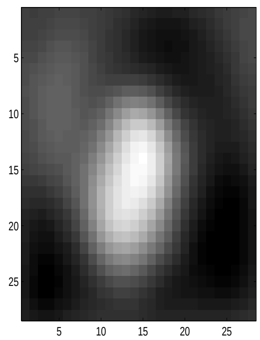


(g) a=0.1, b=30, h = 2, q=5

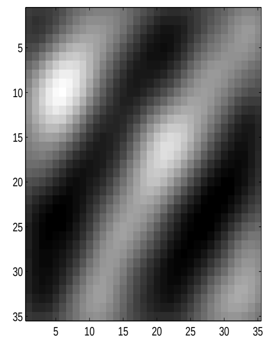


(h) a=0.1, b=30, h = 2, q=6

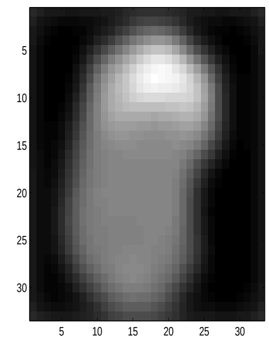
Figure 4.5: SiZer maps with no signal and Matern covariance function.



(a) Confocal dataset



(b) Gamma Camera dataset

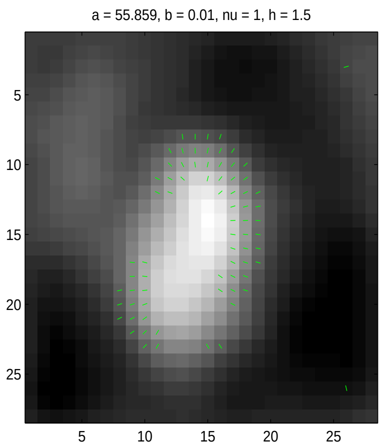


(c) MRI dataset

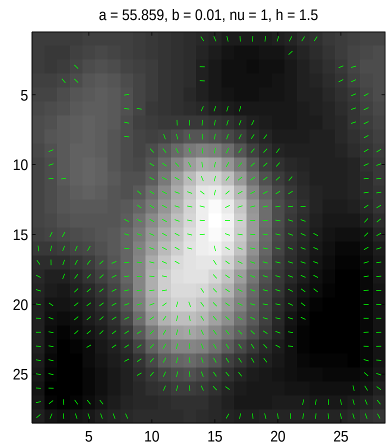
Figure 4.6: Original real data.

to distinguish whether the object is a true feature or noise similar to that of other nearby gray areas. The parameter estimates here in the Spatial SiZer are a=55.859, b=0.01, and ν=1. The Spatial SiZer also clearly denotes the object to be significantly different from its neighboring gray areas by surrounding the ellipsoid with green arrows. The Original SiZer, although its arrows do converge to the center of the object, marking its peak as the significant focal point, has a number of surrounding spurious pixels. The smaller bandwidths, especially 1.5 and 2 in Figure 4.7 (a) and (c), do a good job of marking the oval shape of the object and nothing else extraneous.

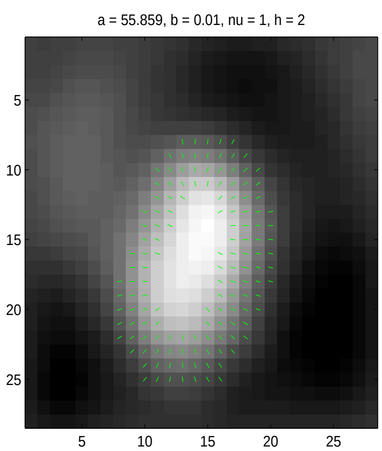
Figures 4.9 and 4.10 are SiZer plots from a gamma camera image of a phantom designed to reflect structure expected from cancerous bones. The gray levels show radiation counts and the radioactive isotope accumulates in regions with bone cancer, so the bright spots on these ribs indicate cancerous regions. Because the image is so noisy, it is especially difficult to identify which brighter spots actually depict cancer. At the center of these figures is the most questionable spot. Using Original SiZer, although the ribbed structures are clearly marked as visible, as the bandwidth h increases, the image becomes oversmoothed and the arrows



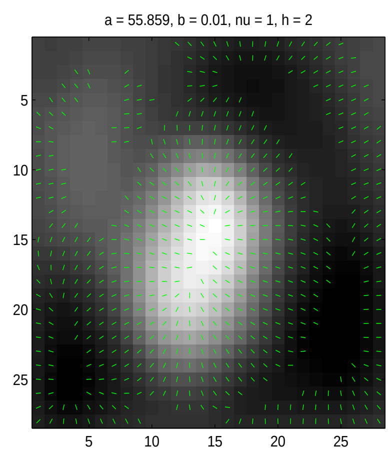
(a) $h = 1.5$, Spatial



(b) $h = 1.5$, Original

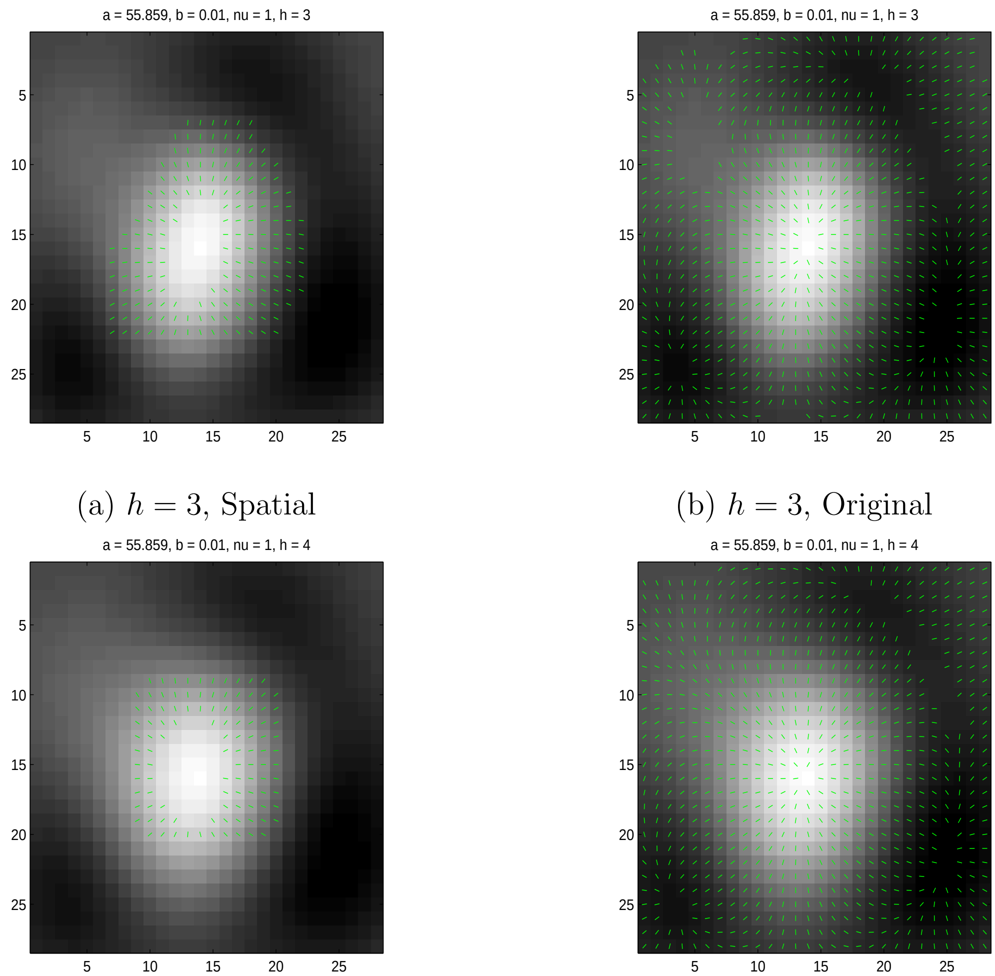


(c) $h = 2$, Spatial



(d) $h = 2$, Original

Figure 4.7: Confocal SiZer maps using the quantile with independent blocks in the Spatial SiZer (left panels) and the Original SiZer (right panels), h=1.5, 2.



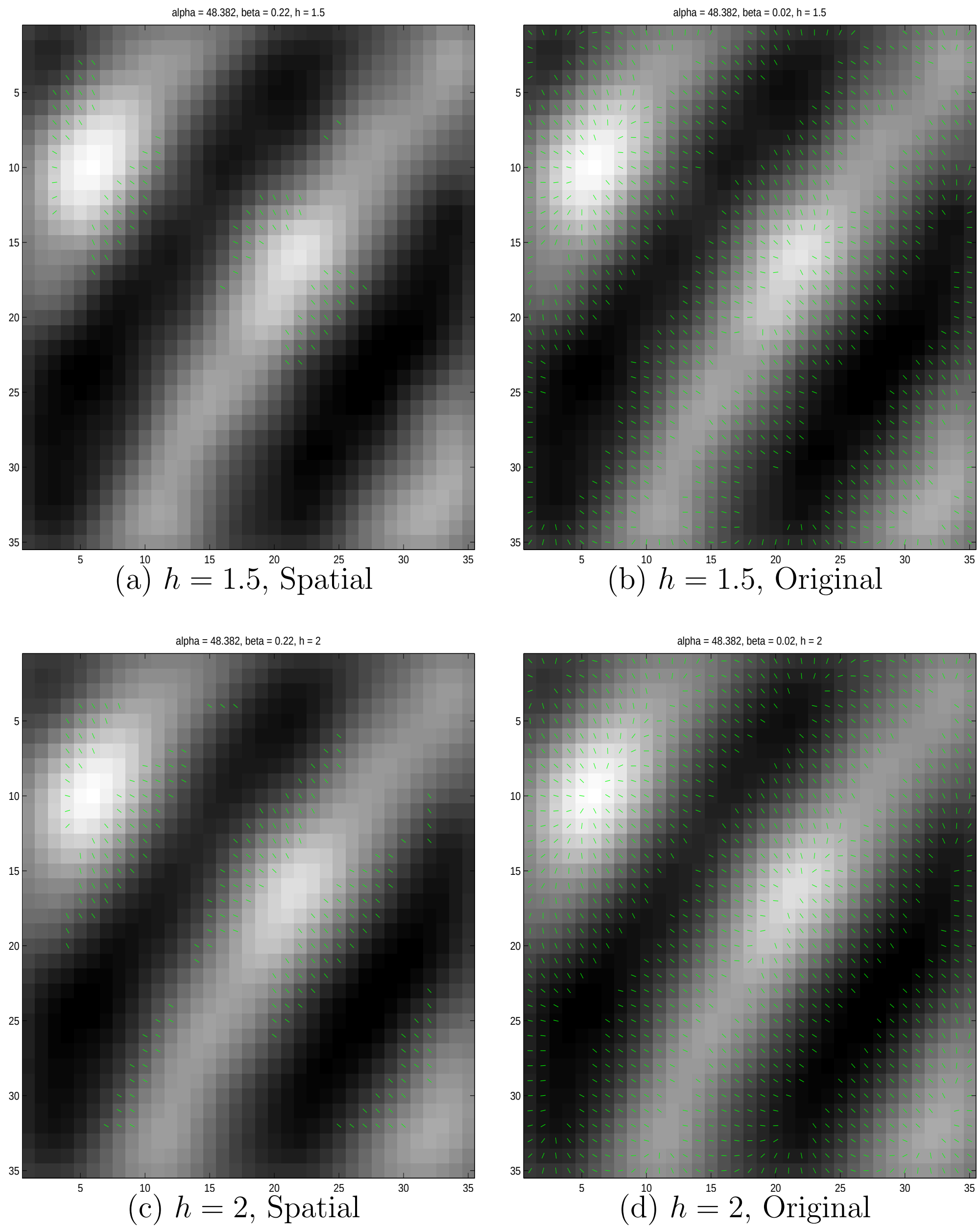
(a) $h = 3$, Spatial

(b) $h = 3$, Original

(c) $h = 4$, Spatial

(d) $h = 4$, Original

Figure 4.8: Confocal SiZer maps using the quantile with independent blocks in the Spatial SiZer (left panels) and the Original SiZer (right panels), h=3, 4.



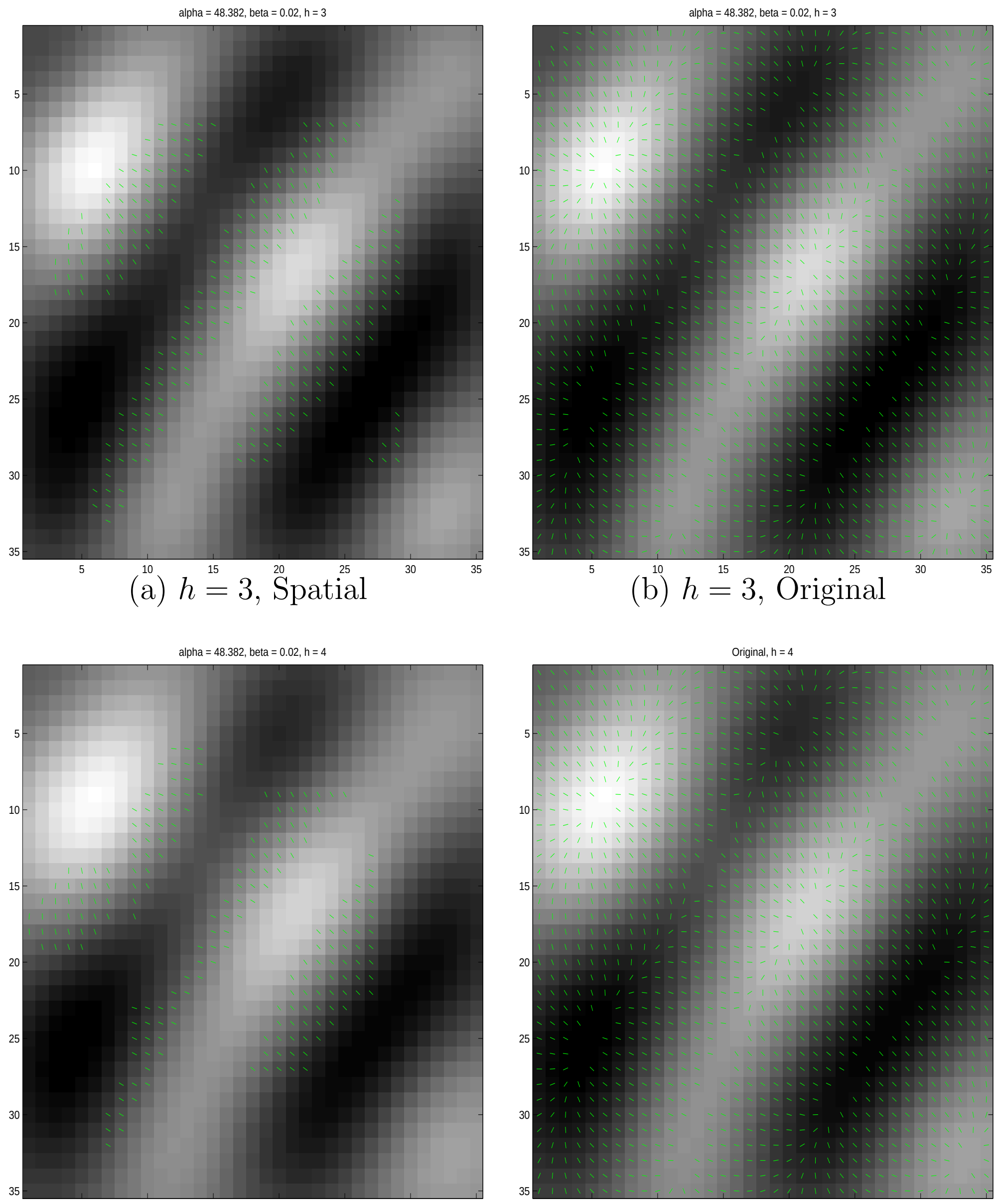
(a) $h = 1.5$, Spatial

(b) $h = 1.5$, Original

(c) $h = 2$, Spatial

(d) $h = 2$, Original

Figure 4.9: Gamma camera data SiZer maps using the quantile with independent blocks in the Spatial SiZer (left panels) and the Original SiZer (right panels), h=1.5, 2.



(a) $h = 3$, Spatial (b) $h = 3$, Original

(c) $h = 4$, Spatial (d) $h = 4$, Original

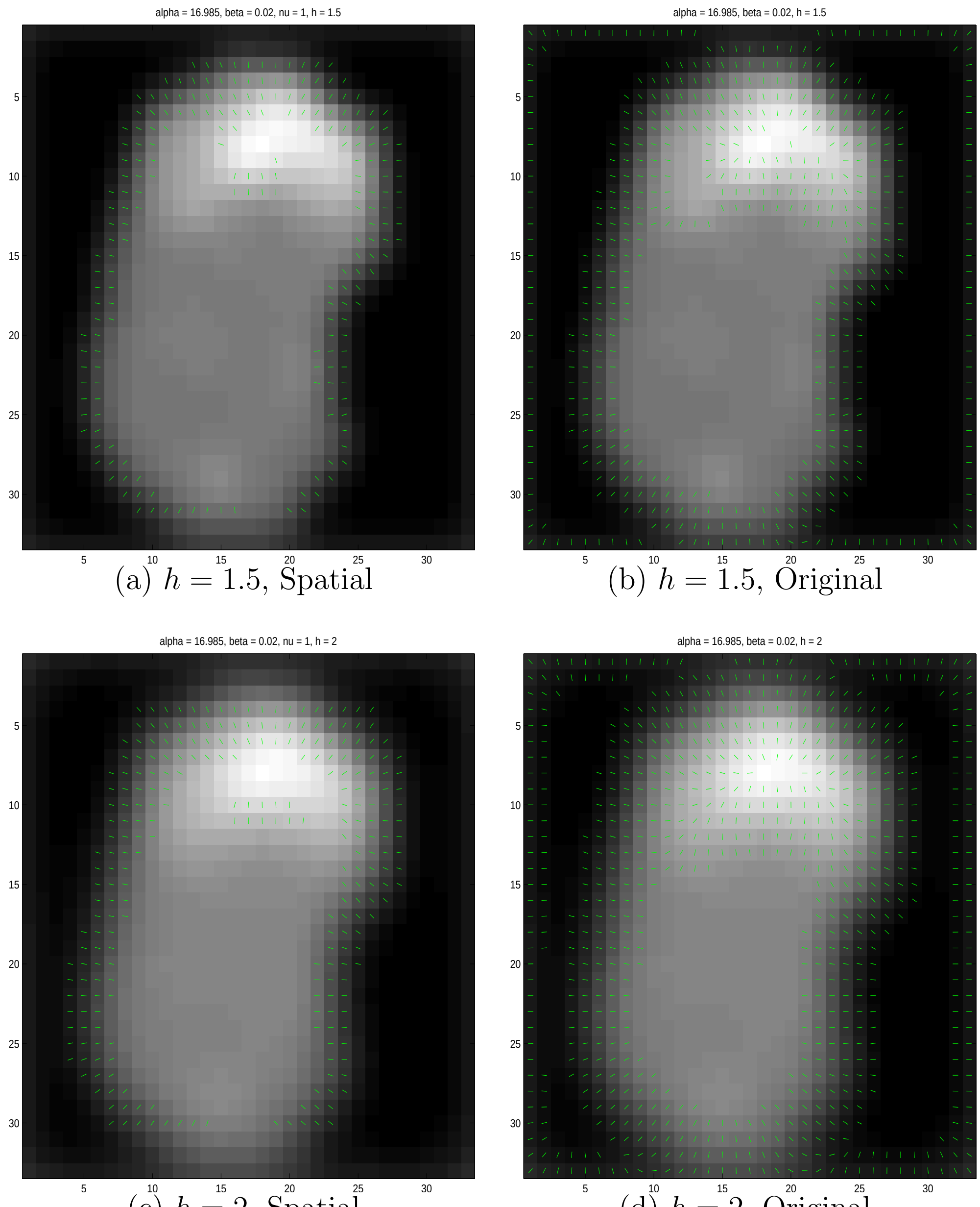
Figure 4.10: Gamma camera data SiZer maps using the quantile with independent blocks in the Spatial SiZer (left panels) and the Original SiZer (right panels), h=3, 4.

can barely be separated enough to mark the different ribs. For the parameter estimates in our Spatial SiZer, a=48.382, b=0.02, and ν=1 are used. Although at finer scales, smaller h's, there is a large amount of noise present, it can still be seen that the arrows of the Spatial SiZer point inward toward the center of the bright features. This indicates that the highlighted spots we see in these bones are indeed to be considered cancerous. These Figures also offer a great look at the tradeoff between different levels of resolution and the amount of detail that is detectable at either the macroscopic or microscopic view of an image.

Figures 4.11 and 4.12 are data derived from a time series of perfusion MR images. Here for the parameter estimates in our Spatial SiZer, a=16.985, b=0.02, and ν=1 are used. The Spatial SiZer highlights the borderline of the image rather cleanly and the most prominent feature of interest is in the top half of the image where the lighter regions are present. These data, although evaluated in Godtliebsen et al. (2004) were originally presented in Chu et al. (1998). There are a couple of other bright pixels in the image, but they are known to be non-Gaussian sampling artifacts and are correctly not highlighted as significant in the Spatial SiZer here. The Spatial SiZer clearly outperforms the Original again because of the greatly reduced number of spuriously highlighted pixels. The smaller bandwidths also seem to delineate the outline of the object more precisely and at these small scales, structures that are observed are deemed actually there in the dataset.

4.6 Future Work

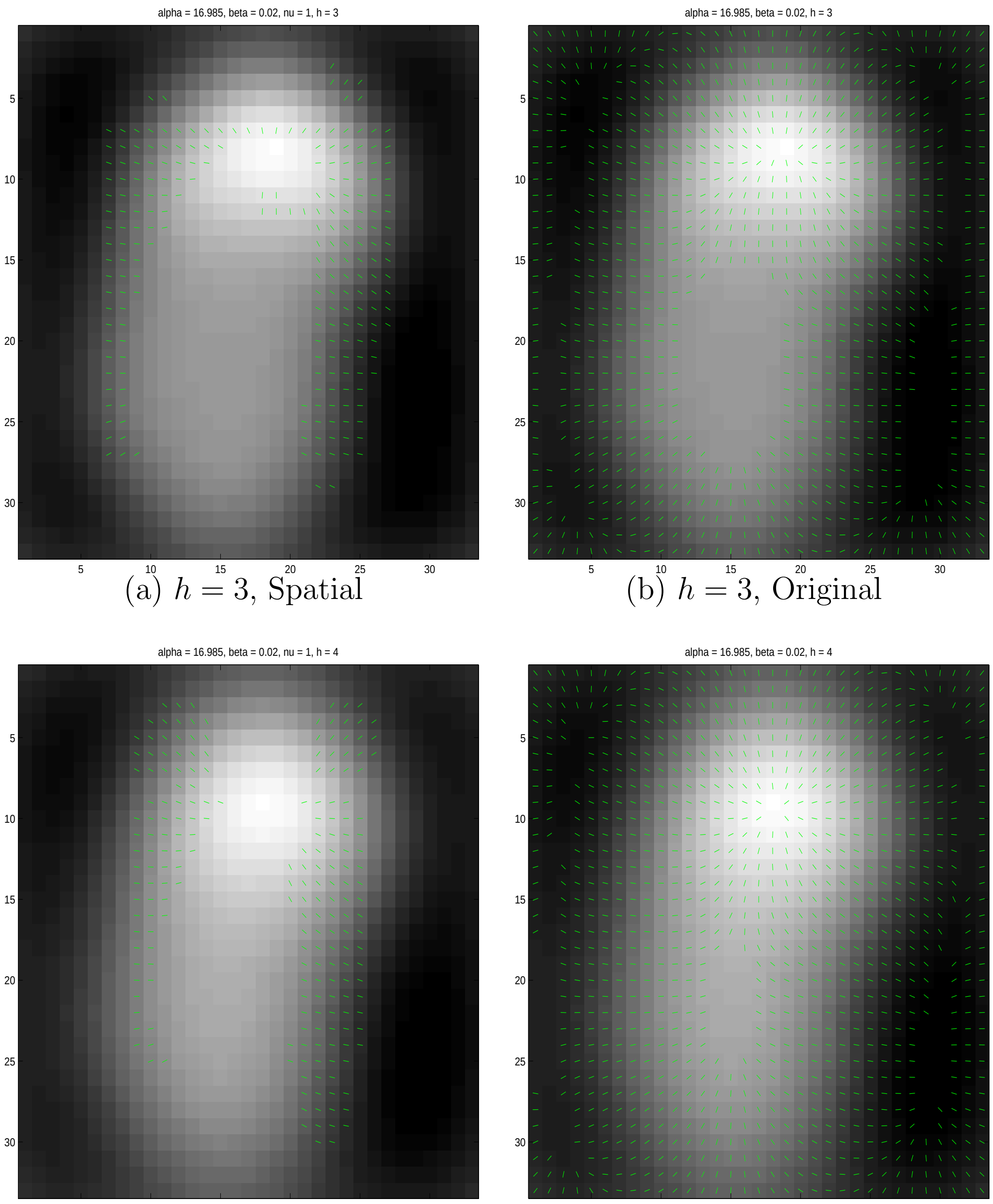
As discussed in Section 2.1, the idea of using independent blocks to aid in approximating a quantile is recently updated by Hannig and Marron (2006) to the idea of estimating a quantile using advanced distributional theory that provides row-wise or global adjustments to reduce the number of spuriously highlighted pixels. We will propose an updated version of the two-dimensional SiZer which incorporates this improvement, using a form of the global adjustment. Although the independent blocks quantile performs quite well in our



(a) $h = 1.5$, Spatial (b) $h = 1.5$, Original

(c) $h = 2$, Spatial (d) $h = 2$, Original

Figure 4.11: MRI data SiZer maps using the quantile with independent blocks in the Spatial SiZer (left panels) and the Original SiZer (right panels), h=1.5, 2.



(a) $h = 3$, Spatial (b) $h = 3$, Original

(c) $h = 4$, Spatial (d) $h = 4$, Original

Figure 4.12: MRI data SiZer maps using the quantile with independent blocks in the Spatial SiZer (left panels) and the Original SiZer (right panels), h=3, 4.

Spatial SiZer, we are optimistic that what spurious pixels do occur may be removed with this advancement.

In addition to the arrows, Godtliebsen et al. (2004) have versions of their 2-dimensional SiZer that use streamlines and curvature. The streamlines are curves that indicate the gradient direction. These interpret the structure of a surface by indicating the physical path that a drop of water would take in flowing downhill. Where significant structures are deemed to occur, the streamlines are found to run together. Curvature is indicated by using different colored dots overlaid on the image. The color of the dots is based on the eigenvalues of the Hessian matrix. At a specific location, whether one or both of its eigenvalues are positive or negative, a different array of colors is used to mark the significance of the points. These alternate versions are proposed in case analysts feel they have better intuition interpreting one of the methods over the others and also a specific method may show particular features better than the other two. We also propose incorporating these types of visualizations into our Spatial SiZer in the future.

Similar to what was introduced in Section 3.4, with the concept of the comparison of multiple time series, we would like to be able to test the equality of multiple images. This would compare k images at multiple locations and resolutions. An ANOVA type test statistic might be developed for the comparison and if some differences are found among the images, multiple pairwise comparisons could then be performed. For the comparison of two images one can obtain a difference image by subtracting two images and apply a modified Spatial SiZer to the difference image. Here, the modified Spatial SiZer estimates and tests the mean of images rather than the partial derivatives. In this case one needs to reestimate the variogram and recalculate a quantile.

As we are mainly interested in difference images, the pixel intensities take on both positive and negative values and tests will be conducted to determine the statistical significance of the signs with spatial correlation taken into account. Different colors will then be used to indicate areas of positive and negative pixel values. Smoothing amounts to local weighted averaging of the pixel intensities, which means that a new SiZer will analyze the images in different resolutions. At the smallest scale smoothing, differences in individual pixels or average differences in very small neighborhoods of pixels are considered. Raising the smoothing level corresponds to analyzing the average differences in increasingly large neighborhoods of pixels. Then, the changes in isolated pixels will tend to be smoothed out and what should remain are just large scale mean changes.

A very challenging extension will be to expand SiZer into more than two dimensions. The statistical inference part of Spatial SiZer extends in a straightforward way, but the visualization will require some creative ideas. Scott (1992) provides a good discussion of a number of interesting possibilities in this direction for density estimation.

Table 4.1: Type I errors and (standard errors) of a signal with Matern covariance, a=0.01

		a=**0.01**								
		b=**3**			b=**10**			b=**30**		
h	q	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**
1.5	1	0.1700 (0.0094)	0.1737 (0.0094)	0.1617 (0.0090)	0.1702 (0.0093)	0.2117 (0.0115)	0.2200 (0.0119)	0.2141 (0.0117)	0.2329 (0.0126)	0.2250 (0.0122)
	2	0.2251 (0.0012)	0.2299 (0.0011)	0.2193 (0.0015)	0.2275 (0.0010)	0.2558 (0.0005)	0.2603 (0.0002)	0.2567 (0.0005)	0.2672 (0.0002)	0.2601 (0.0001)
	3	0.2194 (0.0011)	0.2235 (0.0012)	0.2118 (0.0016)	0.2207 (0.0010)	0.2536 (0.0006)	0.2589 (0.0003)	0.2549 (0.0006)	0.2667 (0.0002)	0.2599 (0.0000)
	4	0.2105 (0.0012)	0.2144 (0.0012)	0.2024 (0.0016)	0.2113 (0.0011)	0.2494 (0.0007)	0.2560 (0.0004)	0.2507 (0.0006)	0.2657 (0.0002)	0.2590 (0.0002)
	5	0.0464 (0.0010)	0.0391 (0.0011)	0.0385 (0.0011)	0.0496 (0.0009)	0.0711 (0.0014)	0.0911 (0.0018)	0.0719 (0.0015)	0.1743 (0.0009)	0.0840 (0.0007)
	6	0.1985 (0.0009)	0.2010 (0.0010)	0.2008 (0.0012)	0.2001 (0.0009)	0.1996 (0.0006)	0.1997 (0.0006)	0.1997 (0.0005)	0.1996 (0.0004)	0.1986 (0.0002)
2	1	0.1171 (0.0066)	0.1126 (0.0065)	0.0905 (0.0051)	0.1164 (0.0065)	0.1684 (0.0091)	0.1785 (0.0097)	0.1769 (0.0093)	0.2122 (0.0115)	0.1834 (0.0099)
	2	0.1963 (0.0015)	0.1961 (0.0016)	0.1723 (0.0020)	0.1997 (0.0013)	0.2313 (0.0006)	0.2361 (0.0004)	0.2347 (0.0005)	0.2544 (0.0003)	0.2332 (0.0002)
	3	0.1766 (0.0019)	0.1743 (0.0020)	0.1459 (0.0023)	0.1799 (0.0017)	0.2233 (0.0010)	0.2293 (0.0007)	0.2272 (0.0008)	0.2530 (0.0003)	0.2299 (0.0002)
	4	0.1660 (0.0019)	0.1648 (0.0019)	0.1377 (0.0020)	0.1687 (0.0018)	0.2151 (0.0012)	0.2224 (0.0008)	0.2207 (0.0008)	0.2495 (0.0003)	0.2271 (0.0002)
	5	0.0011 (0.0003)	0.0012 (0.0003)	0.0012 (0.0003)	0.0013 (0.0003)	0.0328 (0.0010)	0.0504 (0.0012)	0.0414 (0.0010)	0.1594 (0.0007)	0.0464 (0.0004)
	6	0.2212 (0.0012)	0.2216 (0.0012)	0.2242 (0.0013)	0.2215 (0.0012)	0.2224 (0.0011)	0.2204 (0.0012)	0.2207 (0.0010)	0.2207 (0.0007)	0.2218 (0.0005)
3	1	0.0485 (0.0027)	0.0422 (0.0025)	0.0102 (0.0011)	0.0502 (0.0030)	0.1001 (0.0051)	0.1032 (0.0053)	0.1189 (0.0060)	0.1934 (0.0096)	0.1056 (0.0053)
	2	0.1414 (0.0018)	0.1287 (0.0018)	0.0782 (0.0021)	0.1425 (0.0017)	0.1805 (0.0009)	0.1839 (0.0008)	0.1943 (0.0009)	0.2378 (0.0004)	0.1862 (0.0003)
	3	0.1113 (0.0019)	0.1020 (0.0019)	0.0552 (0.0019)	0.1148 (0.0019)	0.1655 (0.0011)	0.1701 (0.0010)	0.1818 (0.0011)	0.2318 (0.0004)	0.1699 (0.0004)
	4	0.0956 (0.0018)	0.0827 (0.0017)	0.0317 (0.0016)	0.0993 (0.0017)	0.1452 (0.0012)	0.1478 (0.0010)	0.1616 (0.0011)	0.2262 (0.0004)	0.1458 (0.0004)
	5	0.0001 (0.0001)	0.0001 (0.0001)	0.0000 (0.0000)	0.0001 (0.0001)	0.0145 (0.0007)	0.0161 (0.0006)	0.0273 (0.0007)	0.1364 (0.0008)	0.0175 (0.0003)
	6	0.2177 (0.0013)	0.2190 (0.0010)	0.2199 (0.0012)	0.2183 (0.0011)	0.2164 (0.0009)	0.2175 (0.0009)	0.2153 (0.0010)	0.2106 (0.0007)	0.2091 (0.0004)
4	1	0.0175 (0.0012)	0.0105 (0.0011)	0.2339 (0.0132)	0.0196 (0.0017)	0.0325 (0.0024)	0.0279 (0.0021)	0.0515 (0.0035)	0.1283 (0.0073)	0.0238 (0.0019)
	2	0.0873 (0.0012)	0.0599 (0.0013)	0.0194 (0.0010)	0.0905 (0.0011)	0.0935 (0.0011)	0.0844 (0.0012)	0.1159 (0.0009)	0.1929 (0.0005)	0.0821 (0.0008)
	3	0.0654 (0.0011)	0.0518 (0.0012)	0.2651 (0.0049)	0.0690 (0.0012)	0.0861 (0.0012)	0.0774 (0.0014)	0.1073 (0.0010)	0.1883 (0.0005)	0.0774 (0.0007)
	4	0.0430 (0.0009)	0.0284 (0.0011)	0.2642 (0.0053)	0.0449 (0.0012)	0.0597 (0.0012)	0.0511 (0.0013)	0.0842 (0.0011)	0.1719 (0.0006)	0.0455 (0.0005)
	5	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0003 (0.0001)	0.0001 (0.0001)	0.0032 (0.0004)	0.0613 (0.0009)	0.0000 (0.0000)
	6	0.2210 (0.0006)	0.2215 (0.0005)	0.2237 (0.0006)	0.2217 (0.0006)	0.2215 (0.0006)	0.2211 (0.0005)	0.2197 (0.0005)	0.2196 (0.0004)	0.2188 (0.0003)

Table 4.2: Type I errors and (standard errors) of a signal with Matern covariance, a=0.1

		a=**0.1**								
		b=**3**			b=**10**			b=**30**		
h	q	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**
1.5	1	0.0294 (0.0019)	0.1534 (0.0093)	0.1407 (0.0089)	0.1575 (0.0089)	0.1474 (0.0082)	0.1484 (0.0082)	0.1524 (0.0083)	0.1867 (0.0101)	0.1477 (0.0081)
	2	0.0492 (0.0012)	0.2003 (0.0041)	0.1832 (0.0057)	0.2087 (0.0020)	0.2069 (0.0015)	0.2074 (0.0015)	0.2080 (0.0014)	0.2407 (0.0009)	0.2050 (0.0006)
	3	0.0484 (0.0012)	0.1987 (0.0041)	0.1828 (0.0053)	0.2058 (0.0020)	0.1983 (0.0015)	0.1973 (0.0014)	0.2000 (0.0015)	0.2357 (0.0010)	0.1963 (0.0006)
	4	0.0394 (0.0012)	0.1878 (0.0041)	0.1713 (0.0048)	0.1935 (0.0020)	0.1880 (0.0014)	0.1873 (0.0013)	0.1899 (0.0014)	0.2282 (0.0010)	0.1876 (0.0005)
	5	0.0085 (0.0005)	0.1195 (0.0036)	0.1034 (0.0041)	0.1023 (0.0021)	0.0500 (0.0014)	0.0479 (0.0018)	0.0563 (0.0015)	0.0476 (0.0013)	0.0341 (0.0006)
	6	0.1988 (0.0017)	0.2124 (0.0016)	0.2174 (0.0019)	0.2057 (0.0016)	0.2085 (0.0015)	0.2025 (0.0015)	0.2024 (0.0015)	0.2015 (0.0010)	0.1994 (0.0008)
2	1	0.0047 (0.0021)	0.0657 (0.0055)	0.0596 (0.0054)	0.1215 (0.0084)	0.0864 (0.0051)	0.0823 (0.0052)	0.0944 (0.0054)	0.1220 (0.0066)	0.0656 (0.0036)
	2	0.0016 (0.0003)	0.0858 (0.0052)	0.0781 (0.0048)	0.1641 (0.0061)	0.1545 (0.0020)	0.1516 (0.0020)	0.1658 (0.0015)	0.2046 (0.0011)	0.1371 (0.0004)
	3	0.0068 (0.0019)	0.0851 (0.0052)	0.0758 (0.0049)	0.1624 (0.0060)	0.1410 (0.0020)	0.1344 (0.0022)	0.1500 (0.0015)	0.1874 (0.0014)	0.1173 (0.0004)
	4	0.0087 (0.0023)	0.0780 (0.0050)	0.0694 (0.0044)	0.1524 (0.0058)	0.1253 (0.0019)	0.1207 (0.0020)	0.1356 (0.0015)	0.1746 (0.0015)	0.1042 (0.0005)
	5	0.0000 (0.0000)	0.0413 (0.0038)	0.0370 (0.0032)	0.0843 (0.0044)	0.0092 (0.0008)	0.0076 (0.0009)	0.0120 (0.0009)	0.0051 (0.0008)	0.0004 (0.0001)
	6	0.2097 (0.0017)	0.2149 (0.0020)	0.2247 (0.0018)	0.2143 (0.0018)	0.2196 (0.0019)	0.2214 (0.0015)	0.2194 (0.0014)	0.2223 (0.0013)	0.2202 (0.0011)
3	1	0.0000 (0.0000)	0.0104 (0.0030)	0.0011 (0.0011)	0.0082 (0.0030)	0.0734 (0.0108)	0.1134 (0.0131)	0.0274 (0.0024)	0.0293 (0.0022)	0.2298 (0.0135)
	2	0.0002 (0.0001)	0.0104 (0.0029)	0.0011 (0.0011)	0.0104 (0.0034)	0.0750 (0.0022)	0.0583 (0.0022)	0.0931 (0.0021)	0.1118 (0.0017)	0.0335 (0.0008)
	3	0.0001 (0.0001)	0.0103 (0.0029)	0.0011 (0.0011)	0.0102 (0.0034)	0.1186 (0.0086)	0.1548 (0.0106)	0.0892 (0.0021)	0.0887 (0.0018)	0.2607 (0.0057)
	4	0.0000 (0.0000)	0.0102 (0.0029)	0.0015 (0.0011)	0.0099 (0.0033)	0.1034 (0.0106)	0.1440 (0.0125)	0.0572 (0.0018)	0.0643 (0.0019)	0.1719 (0.0132)
	5	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0063 (0.0022)	0.0001 (0.0001)	0.0000 (0.0000)	0.0001 (0.0001)	0.0000 (0.0000)	0.0000 0.0000
	6	0.2170 (0.0018)	0.2219 (0.0015)	0.2302 (0.0017)	0.2222 (0.0015)	0.2249 (0.0013)	0.2245 (0.0011)	0.2227 (0.0011)	0.2205 (0.0009)	0.2179 (0.0008)
4	1	0.0000 (0.0000)	0.0063 (0.0028)	0.0048 (0.0024)	0.0000 (0.0000)	0.2302 (0.0128)	0.2420 (0.0131)	0.2266 (0.0119)	0.2203 (0.0135)	0.2420 (0.0131)
	2	0.0000 (0.0000)	0.0071 (0.0026)	0.0048 (0.0024)	0.0000 (0.0000)	0.2608 (0.0043)	0.2750 (0.0000)	0.1027 (0.0042)	0.0200 (0.0015)	0.2750 (0.0000)
	3	0.0000 (0.0000)	0.0071 (0.0026)	0.0048 (0.0024)	0.0000 (0.0000)	0.2605 (0.0044)	0.2750 (0.0000)	0.2567 (0.0048)	0.2498 (0.0072)	0.2750 (0.0000)
	4	0.0055 (0.0027)	0.0071 (0.0026)	0.0048 (0.0024)	0.0000 (0.0000)	0.2609 (0.0045)	0.2750 (0.0000)	0.2559 (0.0053)	0.2329 (0.0097)	0.2750 (0.0000)
	5	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0066 (0.0019)	0.0000 (0.0000)	0.0085 (0.0021)	0.0000 (0.0000)	0.0000 (0.0000)
	6	0.2243 (0.0011)	0.2277 (0.0011)	0.2374 (0.0013)	0.2291 (0.0011)	0.2285 (0.0010)	0.2308 (0.0010)	0.2273 (0.0009)	0.2233 (0.0007)	0.2220 (0.0007)

Table 4.3: Type I errors and (standard errors) of a signal with exponential covariance, a=0.01 and a=0.1

		a=**0.01**			a=**0.1**		
h	q	b=**3**	b=**10**	b=**30**	b=**3**	b=**10**	b=**30**
1.5	1	0.2014 (0.0013)	0.2165 (0.0014)	0.2491 (0.0006)	0.0607 (0.0059)	0.1756 (0.0014)	0.1759 (0.0012)
	2	0.2319 (0.0011)	0.2439 (0.0010)	0.2593 (0.0003)	0.0774 (0.0061)	0.2058 (0.0017)	0.2120 (0.0015)
	3	0.2271 (0.0011)	0.2390 (0.0011)	0.2579 (0.0004)	0.0763 (0.0061)	0.2007 (0.0017)	0.2038 (0.0013)
	4	0.2179 (0.0012)	0.2308 (0.0012)	0.2546 (0.0005)	0.0686 (0.0060)	0.1899 (0.0016)	0.1935 (0.0013)
	5	0.0458 (0.0011)	0.0492 (0.0012)	0.0847 (0.0016)	0.0331 (0.0053)	0.0897 (0.0016)	0.0509 (0.0013)
	6	0.1999 (0.0009)	0.1981 (0.0008)	0.1993 (0.0005)	0.2002 (0.0017)	0.2034 (0.0017)	0.2027 (0.0015)
2	1	0.1435 (0.0023)	0.1631 (0.0021)	0.2106 (0.0008)	0.0046 (0.0015)	0.1510 (0.0026)	0.1049 (0.0015)
	2	0.2060 (0.0015)	0.2181 (0.0013)	0.2393 (0.0005)	0.0065 (0.0016)	0.1870 (0.0029)	0.1662 (0.0017)
	3	0.1888 (0.0019)	0.2049 (0.0015)	0.2336 (0.0006)	0.0063 (0.0016)	0.1832 (0.0029)	0.1483 (0.0018)
	4	0.1778 (0.0020)	0.1943 (0.0017)	0.2273 (0.0006)	0.0051 (0.0015)	0.1692 (0.0027)	0.1344 (0.0017)
	5	0.0016 (0.0003)	0.0093 (0.0007)	0.0575 (0.0012)	0.0035 (0.0012)	0.0668 (0.0027)	0.0073 (0.0008)
	6	0.2222 (0.0012)	0.2240 (0.0011)	0.2226 (0.0011)	0.2117 (0.0017)	0.2178 (0.0014)	0.2219 (0.0016)
3	1	0.0731 (0.0016)	0.0842 (0.0015)	0.1527 (0.0009)	0.0000 (0.0000)	0.0648 (0.0062)	0.0177 (0.0013)
	2	0.1533 (0.0016)	0.1605 (0.0013)	0.2064 (0.0006)	0.0004 (0.0002)	0.0747 (0.0073)	0.0855 (0.0023)
	3	0.1297 (0.0018)	0.1407 (0.0017)	0.1955 (0.0007)	0.0003 (0.0001)	0.0744 (0.0073)	0.0748 (0.0023)
	4	0.1111 (0.0017)	0.1198 (0.0016)	0.1777 (0.0007)	0.0000 (0.0000)	0.0687 (0.0066)	0.0448 (0.0019)
	5	0.0009 (0.0002)	0.0036 (0.0004)	0.0443 (0.0009)	0.0000 (0.0000)	0.0436 (0.0042)	0.0000 (0.0000)
	6	0.2181 (0.0012)	0.2166 (0.0011)	0.2154 (0.0009)	0.2177 (0.0017)	0.2222 (0.0013)	0.2220 (0.0012)
4	1	0.0336 (0.0009)	0.0293 (0.0009)	0.0819 (0.0012)	0.0000 (0.0000)	0.0008 (0.0006)	0.2750 (0.0000)
	2	0.0950 (0.0012)	0.0815 (0.0013)	0.1350 (0.0007)	0.0000 (0.0000)	0.0009 (0.0007)	0.0761 (0.0053)
	3	0.0778 (0.0012)	0.0734 (0.0013)	0.1258 (0.0008)	0.0000 (0.0000)	0.0009 (0.0007)	0.2750 (0.0000)
	4	0.0576 (0.0012)	0.0484 (0.0011)	0.1048 (0.0010)	0.0000 (0.0000)	0.0009 (0.0007)	0.2750 (0.0000)
	5	0.0006 (0.0001)	0.0004 (0.0001)	0.0143 (0.0006)	0.0000 (0.0000)	0.0005 (0.0004)	0.0027 (0.0000)
	6	0.2208 (0.0005)	0.2219 (0.0005)	0.2202 (0.0005)	0.2244 (0.0011)	0.2285 (0.0009)	0.2251 (0.0009)

Table 4.4: Type II errors and (standard errors) of a signal with Matern covariance, a=0.01

		a=**0.01**								
		b=**3**			b=**10**			b=**30**		
h	q	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**
1.5	1	0.0111 (0.0007)	0.0087 (0.0005)	0.0123 (0.0009)	0.0098 (0.0005)	0.0055 (0.0003)	0.0055 (0.0003)	0.0050 (0.0003)	0.0052 (0.0003)	0.0062 (0.0004)
	2	0.0069 (0.0003)	0.0054 (0.0002)	0.0074 (0.0003)	0.0063 (0.0002)	0.0044 (0.0001)	0.0046 (0.0001)	0.0044 (0.0001)	0.0043 (0.0001)	0.0050 (0.0000)
	3	0.0079 (0.0003)	0.0062 (0.0002)	0.0086 (0.0003)	0.0071 (0.0002)	0.0049 (0.0001)	0.0049 (0.0001)	0.0046 (0.0001)	0.0045 (0.0001)	0.0051 (0.0000)
	4	0.0091 (0.0003)	0.0072 (0.0002)	0.0104 (0.0004)	0.0081 (0.0002)	0.0055 (0.0001)	0.0054 (0.0001)	0.0051 (0.0001)	0.0048 (0.0001)	0.0053 (0.0001)
	5	0.0634 (0.0008)	0.0550 (0.0008)	0.0700 (0.0008)	0.0637 (0.0008)	0.0130 (0.0003)	0.0092 (0.0002)	0.0125 (0.0003)	0.0075 (0.0000)	0.0076 (0.0000)
	6	0.0231 (0.0005)	0.0222 (0.0005)	0.0216 (0.0005)	0.0209 (0.0005)	0.0172 (0.0003)	0.0171 (0.0004)	0.0175 (0.0004)	0.0143 (0.0002)	0.0142 (0.0001)
2	1	0.0217 (0.0013)	0.0199 (0.0011)	0.0292 (0.0019)	0.0204 (0.0012)	0.0100 (0.0006)	0.0087 (0.0005)	0.0094 (0.0005)	0.0066 (0.0004)	0.0072 (0.0004)
	2	0.0153 (0.0003)	0.0138 (0.0003)	0.0189 (0.0004)	0.0146 (0.0003)	0.0083 (0.0001)	0.0076 (0.0001)	0.0079 (0.0002)	0.0071 (0.0001)	0.0075 (0.0000)
	3	0.0179 (0.0004)	0.0170 (0.0004)	0.0236 (0.0005)	0.0170 (0.0003)	0.0088 (0.0002)	0.0081 (0.0001)	0.0084 (0.0002)	0.0073 (0.0001)	0.0075 (0.0000)
	4	0.0194 (0.0004)	0.0183 (0.0004)	0.0251 (0.0005)	0.0184 (0.0003)	0.0094 (0.0002)	0.0086 (0.0002)	0.0090 (0.0002)	0.0074 (0.0001)	0.0075 (0.0000)
	5	0.0939 (0.0012)	0.0964 (0.0015)	0.1258 (0.0011)	0.0905 (0.0012)	0.0420 (0.0005)	0.0370 (0.0005)	0.0384 (0.0005)	0.0174 (0.0003)	0.0366 (0.0002)
	6	0.0183 (0.0004)	0.0172 (0.0003)	0.0182 (0.0004)	0.0172 (0.0003)	0.0158 (0.0002)	0.0156 (0.0001)	0.0159 (0.0002)	0.0151 (0.0000)	0.0150 (0.0000)
3	1	0.0397 (0.0022)	0.0422 (0.0024)	0.0671 (0.0039)	0.0395 (0.0021)	0.0302 (0.0017)	0.0297 (0.0016)	0.0286 (0.0016)	0.0158 (0.0009)	0.0305 (0.0017)
	2	0.0297 (0.0004)	0.0322 (0.0004)	0.0450 (0.0005)	0.0297 (0.0004)	0.0251 (0.0004)	0.0245 (0.0004)	0.0229 (0.0003)	0.0155 (0.0002)	0.0269 (0.0004)
	3	0.0353 (0.0004)	0.0367 (0.0004)	0.0514 (0.0006)	0.0351 (0.0004)	0.0280 (0.0003)	0.0272 (0.0003)	0.0260 (0.0003)	0.0166 (0.0001)	0.0310 (0.0002)
	4	0.0379 (0.0004)	0.0397 (0.0004)	0.0595 (0.0006)	0.0377 (0.0004)	0.0308 (0.0003)	0.0303 (0.0002)	0.0290 (0.0003)	0.0174 (0.0001)	0.0324 (0.0001)
	5	0.1619 (0.0007)	0.1815 (0.0010)	0.2352 (0.0010)	0.1598 (0.0009)	0.1261 (0.0008)	0.1278 (0.0009)	0.1014 (0.0007)	0.0518 (0.0003)	0.1218 (0.0004)
	6	0.0163 (0.0003)	0.0161 (0.0003)	0.0152 (0.0003)	0.0155 (0.0003)	0.0157 (0.0003)	0.0158 (0.0003)	0.0164 (0.0003)	0.0162 (0.0002)	0.0166 (0.0002)
4	1	0.0793 (0.0041)	0.1091 (0.0056)	0.0067 (0.0033)	0.0781 (0.0038)	0.0973 (0.0050)	0.1077 (0.0054)	0.0789 (0.0040)	0.0454 (0.0023)	0.1059 (0.0053)
	2	0.0461 (0.0004)	0.0572 (0.0005)	0.0847 (0.0007)	0.0454 (0.0003)	0.0479 (0.0004)	0.0526 (0.0004)	0.0410 (0.0003)	0.0274 (0.0002)	0.0507 (0.0002)
	3	0.0530 (0.0005)	0.0608 (0.0005)	0.0031 (0.0015)	0.0522 (0.0004)	0.0515 (0.0004)	0.0560 (0.0005)	0.0443 (0.0003)	0.0296 (0.0003)	0.0551 (0.0004)
	4	0.0654 (0.0005)	0.0819 (0.0008)	0.0046 (0.0023)	0.0646 (0.0006)	0.0759 (0.0006)	0.0840 (0.0007)	0.0621 (0.0005)	0.0390 (0.0002)	0.0850 (0.0003)
	5	0.2962 (0.0005)	0.3319 (0.0008)	0.4339 (0.0016)	0.2950 (0.0006)	0.3340 (0.0011)	0.3612 (0.0009)	0.3017 (0.0014)	0.1904 (0.0005)	0.3632 (0.0004)
	6	0.0043 (0.0001)	0.0045 (0.0002)	0.0048 (0.0002)	0.0045 (0.0002)	0.0041 (0.0002)	0.0044 (0.0001)	0.0038 (0.0001)	0.0033 (0.0001)	0.0030 (0.0001)

Table 4.5: Type II errors and (standard errors) of a signal with Matern covariance, a=0.1

		a=0.1								
		b=3			**b=10**			**b=30**		
h	q	**ν=0.5**	**ν=1.0**	**ν=2.0**	**ν=0.5**	**ν=1.0**	**ν=2.0**	**ν=0.5**	**ν=1.0**	**ν=2.0**
1.5	1	0.1004 (0.0064)	0.0874 (0.0058)	0.1192 (0.0126)	0.0592 (0.0037)	0.0291 (0.0018)	0.0199 (0.0013)	0.0268 (0.0018)	0.0059 (0.0004)	0.0190 (0.0003)
	2	0.0681 (0.0023)	0.0549 (0.0018)	0.0833 (0.0068)	0.0381 (0.0010)	0.0174 (0.0006)	0.0118 (0.0005)	0.0163 (0.0006)	0.0050 (0.0002)	0.0079 (0.0002)
	3	0.0697 (0.0023)	0.0576 (0.0019)	0.0828 (0.0067)	0.0405 (0.0011)	0.0200 (0.0006)	0.0138 (0.0005)	0.0190 (0.0006)	0.0054 (0.0002)	0.0095 (0.0003)
	4	0.0956 (0.0030)	0.0771 (0.0031)	0.1062 (0.0070)	0.0522 (0.0013)	0.0238 (0.0007)	0.0160 (0.0006)	0.0227 (0.0007)	0.0058 (0.0002)	0.0115 (0.0003)
	5	0.4825 (0.0072)	0.3985 (0.0092)	0.5310 (0.0101)	0.2532 (0.0054)	0.1200 (0.0015)	0.0938 (0.0014)	0.1123 (0.0015)	0.0296 (0.0006)	0.0886 (0.0005)
	6	0.0411 (0.0013)	0.0297 (0.0008)	0.0250 (0.0008)	0.0297 (0.0008)	0.0231 (0.0006)	0.0219 (0.0005)	0.0245 (0.0007)	0.0200 (0.0004)	0.0176 (0.0004)
2	1	0.1509 (0.0089)	0.1062 (0.0063)	0.1954 (0.0138)	0.0786 (0.0045)	0.0551 (0.0030)	0.0470 (0.0025)	0.0470 (0.0030)	0.0163 (0.0010)	0.0554 (0.0006)
	2	0.1420 (0.0021)	0.0729 (0.0022)	0.1448 (0.0076)	0.0513 (0.0010)	0.0362 (0.0008)	0.0292 (0.0007)	0.0306 (0.0007)	0.0116 (0.0002)	0.0312 (0.0003)
	3	0.1370 (0.0028)	0.0754 (0.0027)	0.1680 (0.0099)	0.0537 (0.0010)	0.0419 (0.0009)	0.0347 (0.0009)	0.0357 (0.0008)	0.0134 (0.0003)	0.0388 (0.0005)
	4	0.1538 (0.0041)	0.1013 (0.0029)	0.1977 (0.0087)	0.0692 (0.0012)	0.0484 (0.0010)	0.0400 (0.0010)	0.0414 (0.0009)	0.0148 (0.0003)	0.0436 (0.0005)
	5	0.5329 (0.0075)	0.4737 (0.0094)	0.6456 (0.0047)	0.3243 (0.0067)	0.1919 (0.0019)	0.1731 (0.0016)	0.1652 (0.0017)	0.0841 (0.0017)	0.1800 (0.0008)
	6	0.0277 (0.0009)	0.0231 (0.0006)	0.0192 (0.0007)	0.0217 (0.0007)	0.0188 (0.0004)	0.0175 (0.0004)	0.0194 (0.0004)	0.0168 (0.0003)	0.0160 (0.0002)
3	1	0.2802 (0.0155)	0.1926 (0.0140)	0.3479 (0.0215)	0.1531 (0.0107)	0.0802 (0.0071)	0.0631 (0.0074)	0.0846 (0.0048)	0.0511 (0.0029)	0.0078 (0.0031)
	2	0.2902 (0.0017)	0.1676 (0.0065)	0.2993 (0.0099)	0.1331 (0.0033)	0.0723 (0.0009)	0.0776 (0.0011)	0.0577 (0.0008)	0.0356 (0.0005)	0.0783 (0.0004)
	3	0.2962 (0.0020)	0.1910 (0.0100)	0.3309 (0.0108)	0.1421 (0.0068)	0.0550 (0.0030)	0.0411 (0.0036)	0.0601 (0.0007)	0.0396 (0.0005)	0.0048 (0.0019)
	4	0.3086 (0.0020)	0.1946 (0.0089)	0.3634 (0.0100)	0.1585 (0.0065)	0.0702 (0.0044)	0.0521 (0.0050)	0.0756 (0.0008)	0.0446 (0.0005)	0.0419 (0.0054)
	5	0.5574 (0.0042)	0.5384 (0.0074)	0.6377 (0.0014)	0.4762 (0.0065)	0.3852 (0.0043)	0.4135 (0.0045)	0.2880 (0.0020)	0.2135 (0.0008)	0.4104 (0.0011)
	6	0.0171 (0.0005)	0.0148 (0.0005)	0.0132 (0.0005)	0.0146 (0.0004)	0.0145 (0.0004)	0.0132 (0.0004)	0.0144 (0.0004)	0.0152 (0.0004)	0.0150 (0.0003)
4	1	0.4723 (0.0242)	0.4574 (0.0260)	0.5049 (0.0261)	0.3882 (0.0204)	0.0158 (0.0053)	0.0001 (0.0001)	0.0152 (0.0053)	0.0148 (0.0053)	0.0002 (0.0002)
	2	0.5040 (0.0040)	0.4908 (0.0114)	0.5270 (0.0084)	0.3798 (0.0061)	0.0123 (0.0037)	0.0000 (0.0000)	0.0838 (0.0020)	0.0860 (0.0009)	0.0000 (0.0000)
	3	0.5232 (0.0034)	0.5128 (0.0107)	0.5313 (0.0083)	0.4195 (0.0058)	0.0132 (0.0040)	0.0000 (0.0000)	0.0105 (0.0028)	0.0080 (0.0023)	0.0000 (0.0000)
	4	0.5138 (0.0033)	0.5134 (0.0110)	0.5529 (0.0072)	0.4075 (0.0058)	0.0160 (0.0051)	0.0000 (0.0000)	0.0148 (0.0041)	0.0220 (0.0051)	0.0000 (0.0000)
	5	0.6336 (0.0030)	0.6387 (0.0056)	0.6342 (0.0023)	0.5964 (0.0041)	0.6658 (0.0030)	0.7245 (0.0003)	0.5248 (0.0027)	0.4588 (0.0011)	0.7250 (0.0000)
	6	0.0075 (0.0003)	0.0070 (0.0003)	0.0072 (0.0004)	0.0069 (0.0003)	0.0063 (0.0003)	0.0065 (0.0003)	0.0058 (0.0003)	0.0054 (0.0002)	0.0047 (0.0002)

Table 4.6: Type II errors and (standard errors) of a signal with exponential covariance, a=0.01 and a=0.1

		a=**0.01**			a=**0.1**		
h	q	b=**3**	b=**10**	b=**30**	b=**3**	b=**10**	b=**30**
1.5	1	0.0102 (0.0004)	0.0072 (0.0002)	0.0063 (0.0001)	0.0918 (0.0024)	0.0551 (0.0014)	0.0244 (0.0007)
	2	0.0058 (0.0002)	0.0047 (0.0002)	0.0044 (0.0001)	0.0536 (0.0018)	0.0308 (0.0009)	0.0128 (0.0004)
	3	0.0063 (0.0002)	0.0052 (0.0002)	0.0048 (0.0001)	0.0558 (0.0018)	0.0343 (0.0010)	0.0149 (0.0005)
	4	0.0074 (0.0003)	0.0059 (0.0002)	0.0054 (0.0001)	0.0720 (0.0023)	0.0420 (0.0012)	0.0183 (0.0005)
	5	0.0529 (0.0008)	0.0263 (0.0006)	0.0101 (0.0002)	0.3759 (0.0065)	0.1974 (0.0042)	0.0958 (0.0012)
	6	0.0217 (0.0005)	0.0193 (0.0005)	0.0167 (0.0003	0.0371 (0.0011)	0.0274 (0.0007)	0.0237 (0.0006)
2	1	0.0214 (0.0004)	0.0154 (0.0003)	0.0090 (0.0002)	0.1540 (0.0028)	0.0777 (0.0013)	0.0459 (0.0009)
	2	0.0134 (0.0003)	0.0102 (0.0002)	0.0076 (0.0001)	0.1222 (0.0023)	0.0440 (0.0010)	0.0259 (0.0006)
	3	0.0159 (0.0003)	0.0115 (0.0002)	0.0079 (0.0001)	0.1234 (0.0024)	0.0469 (0.0010)	0.0311 (0.0006)
	4	0.0172 (0.0003)	0.0124 (0.0002)	0.0082 (0.0001)	0.1405 (0.0028)	0.0607 (0.0012)	0.0348 (0.0007)
	5	0.0805 (0.0014)	0.0578 (0.0010)	0.0339 (0.0004)	0.4188 (0.0081)	0.2536 (0.0051)	0.1476 (0.0014)
	6	0.0180 (0.0003)	0.0165 (0.0003)	0.0159 (0.0002)	0.0258 (0.0008)	0.0209 (0.0005)	0.0192 (0.0004)
3	1	0.0408 (0.0004)	0.0381 (0.0003)	0.0305 (0.0003)	0.2942 (0.0017)	0.1210 (0.0040)	0.0883 (0.0010)
	2	0.0288 (0.0004)	0.0275 (0.0004)	0.0195 (0.0003)	0.2718 (0.0013)	0.0851 (0.0040)	0.0541 (0.0007)
	3	0.0328 (0.0003)	0.0311 (0.0003)	0.0224 (0.0003)	0.2748 (0.0016)	0.0872 (0.0041)	0.0581 (0.0007)
	4	0.0362 (0.0009)	0.0344 (0.0012)	0.0261 (0.0006)	0.2873 (0.0030)	0.1079 (0.0053)	0.0722 (0.0020)
	5	0.1453 (0.0016)	0.1450 (0.0017)	0.0839 (0.0009)	0.4796 (0.0072)	0.3695 (0.0089)	0.2607 (0.0029)
	6	0.0168 (0.0006)	0.0161 (0.0009)	0.0160 (0.0006)	0.0217 (0.0050)	0.0179 (0.0036)	0.0146 (0.0004)
4	1	0.0897 (0.0008)	0.1106 (0.0009)	0.0762 (0.0005)	0.4905 (0.0021)	0.3643 (0.0064)	0.0000 0.0000
	2	0.0443 (0.0004)	0.0504 (0.0004)	0.0375 (0.0002)	0.4626 (0.0022)	0.3200 (0.0054)	0.0842 (0.0023)
	3	0.0506 (0.0005)	0.0548 (0.0004)	0.0403 (0.0003)	0.4849 (0.0022)	0.3459 (0.0068)	0.0000 0.0000
	4	0.0607 (0.0005)	0.0760 (0.0007)	0.0547 (0.0004)	0.4811 (0.0021)	0.3456 (0.0056)	0.0000 0.0000
	5	0.2886 (0.0008)	0.3206 (0.0008)	0.2709 (0.0011)	0.5932 (0.0024)	0.5462 (0.0051)	0.5163 (0.0023)
	6	0.0043 (0.0001)	0.0042 (0.0001)	0.0038 (0.0001)	0.0068 (0.0003)	0.0064 (0.0003)	0.0055 (0.0002)

Table 4.7: Proportion identified correctly and (standard error) of a signal with Matern covariance, a=0.01

		a=**0.01**								
		b=**3**			b=**10**			b=**30**		
h	q	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**
1.5	1	0.6989 (0.0378)	0.6977 (0.0380)	0.7061 (0.0381)	0.7000 (0.0381)	0.6628 (0.0360)	0.6546 (0.0355)	0.6609 (0.0358)	0.6419 (0.0348)	0.6489 (0.0352)
	2	0.7680 (0.0012)	0.7648 (0.0011)	0.7733 (0.0014)	0.7663 (0.0010)	0.7399 (0.0005)	0.7352 (0.0003)	0.7390 (0.0005)	0.7286 (0.0002)	0.7349 (0.0001)
	3	0.7727 (0.0011)	0.7704 (0.0012)	0.7796 (0.0015)	0.7722 (0.0010)	0.7416 (0.0006)	0.7362 (0.0003)	0.7406 (0.0005)	0.7289 (0.0002)	0.7350 (0.0001)
	4	0.7804 (0.0013)	0.7785 (0.0012)	0.7872 (0.0015)	0.7807 (0.0011)	0.7451 (0.0007)	0.7385 (0.0004)	0.7442 (0.0006)	0.7295 (0.0002)	0.7357 (0.0002)
	5	0.8903 (0.0011)	0.9059 (0.0008)	0.8915 (0.0010)	0.8868 (0.0008)	0.9160 (0.0013)	0.8997 (0.0017)	0.9157 (0.0015)	0.8182 (0.0009)	0.9084 (0.0007)
	6	0.7784 (0.0012)	0.7769 (0.0012)	0.7776 (0.0014)	0.7790 (0.0011)	0.7832 (0.0008)	0.7832 (0.0008)	0.7828 (0.0008)	0.7861 (0.0005)	0.7872 (0.0003)
2	1	0.8037 (0.0731)	0.8080 (0.0732)	0.8218 (0.0734)	0.8051 (0.0731)	0.7632 (0.0724)	0.7554 (0.0722)	0.7562 (0.0724)	0.7241 (0.0717)	0.7518 (0.0721)
	2	0.7885 (0.0015)	0.7901 (0.0016)	0.8089 (0.0020)	0.7857 (0.0013)	0.7604 (0.0006)	0.7563 (0.0005)	0.7575 (0.0006)	0.7385 (0.0003)	0.7594 (0.0002)
	3	0.8056 (0.0018)	0.8088 (0.0019)	0.8306 (0.0021)	0.8031 (0.0016)	0.7679 (0.0010)	0.7626 (0.0007)	0.7644 (0.0008)	0.7397 (0.0003)	0.7626 (0.0002)
	4	0.8146 (0.0019)	0.8169 (0.0019)	0.8372 (0.0019)	0.8129 (0.0017)	0.7756 (0.0012)	0.7690 (0.0008)	0.7703 (0.0009)	0.7431 (0.0003)	0.7654 (0.0002)
	5	0.9050 (0.0012)	0.9025 (0.0014)	0.8731 (0.0011)	0.9082 (0.0012)	0.9252 (0.0008)	0.9126 (0.0010)	0.9202 (0.0009)	0.8232 (0.0008)	0.9170 (0.0004)
	6	0.7606 (0.0011)	0.7609 (0.0013)	0.7576 (0.0013)	0.7613 (0.0011)	0.7618 (0.0011)	0.7639 (0.0012)	0.7636 (0.0010)	0.7643 (0.0007)	0.7631 (0.0005)
3	1	0.7918 (0.0430)	0.7956 (0.0433)	0.8036 (0.0434)	0.7915 (0.0429)	0.7518 (0.0408)	0.7487 (0.0406)	0.7346 (0.0398)	0.6743 (0.0366)	0.7461 (0.0404)
	2	0.8289 (0.0017)	0.8391 (0.0017)	0.8769 (0.0019)	0.8278 (0.0015)	0.7944 (0.0010)	0.7917 (0.0009)	0.7829 (0.0008)	0.7467 (0.0005)	0.7869 (0.0005)
	3	0.8534 (0.0017)	0.8614 (0.0017)	0.8935 (0.0016)	0.8501 (0.0017)	0.8066 (0.0011)	0.8028 (0.0010)	0.7923 (0.0010)	0.7516 (0.0005)	0.7991 (0.0005)
	4	0.8665 (0.0017)	0.8776 (0.0015)	0.9088 (0.0014)	0.8630 (0.0015)	0.8240 (0.0012)	0.8219 (0.0011)	0.8094 (0.0011)	0.7565 (0.0004)	0.8218 (0.0004)
	5	0.8380 (0.0007)	0.8184 (0.0010)	0.7648 (0.0010)	0.8402 (0.0009)	0.8595 (0.0008)	0.8561 (0.0009)	0.8714 (0.0006)	0.8118 (0.0007)	0.8608 (0.0006)
	6	0.7660 (0.0013)	0.7649 (0.0011)	0.7650 (0.0014)	0.7662 (0.0012)	0.7680 (0.0010)	0.7668 (0.0011)	0.7683 (0.0011)	0.7733 (0.0007)	0.7743 (0.0005)
4	1	0.8011 (0.0388)	0.7787 (0.0377)	0.6582 (0.0318)	0.8006 (0.0388)	0.7673 (0.0372)	0.7618 (0.0369)	0.7666 (0.0370)	0.7225 (0.0350)	0.7675 (0.0372)
	2	0.8667 (0.0011)	0.8829 (0.0011)	0.8960 (0.0010)	0.8642 (0.0010)	0.8587 (0.0010)	0.8630 (0.0010)	0.8431 (0.0008)	0.7797 (0.0005)	0.8673 (0.0008)
	3	0.8817 (0.0009)	0.8875 (0.0010)	0.7318 (0.0034)	0.8788 (0.0011)	0.8625 (0.0010)	0.8666 (0.0011)	0.8483 (0.0009)	0.7822 (0.0005)	0.8675 (0.0007)
	4	0.8917 (0.0008)	0.8897 (0.0009)	0.7312 (0.0030)	0.8906 (0.0009)	0.8645 (0.0010)	0.8650 (0.0011)	0.8537 (0.0009)	0.7892 (0.0006)	0.8696 (0.0004)
	5	0.7038 (0.0005)	0.6681 (0.0008)	0.5661 (0.0016)	0.7050 (0.0006)	0.6658 (0.0010)	0.6388 (0.0009)	0.6951 (0.0011)	0.7483 (0.0008)	0.6368 (0.0004)
	6	0.7747 (0.0006)	0.7741 (0.0006)	0.7715 (0.0006)	0.7739 (0.0006)	0.7745 (0.0006)	0.7745 (0.0006)	0.7765 (0.0004)	0.7771 (0.0004)	0.7782 (0.0003)

Table 4.8: Proportion identified correctly and (standard error) of a signal with Matern covariance, a=0.1

		a=**0.1**								
		b=**3**			b=**10**			b=**30**		
h	q	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**	ν=**0.5**	ν=**1.0**	ν=**2.0**
1.5	1	0.7502 (0.0407)	0.6392 (0.0347)	0.6201 (0.0322)	0.6634 (0.0358)	0.7035 (0.0381)	0.7117 (0.0386)	0.7008 (0.0379)	0.6874 (0.0373)	0.7155 (0.0388)
	2	0.8828 (0.0025)	0.7448 (0.0040)	0.7336 (0.0051)	0.7532 (0.0022)	0.7757 (0.0016)	0.7809 (0.0014)	0.7757 (0.0016)	0.7543 (0.0010)	0.7871 (0.0005)
	3	0.8819 (0.0025)	0.7438 (0.0040)	0.7344 (0.0056)	0.7537 (0.0022)	0.7817 (0.0016)	0.7889 (0.0013)	0.7810 (0.0017)	0.7590 (0.0010)	0.7943 (0.0005)
	4	0.8650 (0.0031)	0.7351 (0.0039)	0.7225 (0.0061)	0.7543 (0.0023)	0.7882 (0.0015)	0.7967 (0.0012)	0.7875 (0.0017)	0.7661 (0.0010)	0.8009 (0.0004)
	5	0.5090 (0.0072)	0.4820 (0.0091)	0.3656 (0.0097)	0.6445 (0.0062)	0.8300 (0.0017)	0.8583 (0.0016)	0.8314 (0.0017)	0.9229 (0.0011)	0.8773 (0.0007)
	6	0.7602 (0.0021)	0.7579 (0.0019)	0.7577 (0.0022)	0.7646 (0.0018)	0.7684 (0.0017)	0.7757 (0.0016)	0.7732 (0.0019)	0.7786 (0.0013)	0.7830 (0.0011)
2	1	0.7875 (0.0727)	0.7691 (0.0727)	0.6902 (0.0719)	0.7415 (0.0724)	0.8010 (0.0731)	0.8132 (0.0732)	0.8005 (0.0729)	0.8030 (0.0732)	0.8265 (0.0735)
	2	0.8565 (0.0021)	0.8413 (0.0047)	0.7771 (0.0064)	0.7846 (0.0057)	0.8093 (0.0018)	0.8192 (0.0016)	0.8036 (0.0015)	0.7839 (0.0012)	0.8317 (0.0005)
	3	0.8563 (0.0021)	0.8396 (0.0049)	0.7563 (0.0083)	0.7839 (0.0057)	0.8171 (0.0018)	0.8310 (0.0017)	0.8143 (0.0014)	0.7993 (0.0014)	0.8439 (0.0006)
	4	0.8375 (0.0027)	0.8207 (0.0045)	0.7329 (0.0073)	0.7784 (0.0053)	0.8263 (0.0018)	0.8393 (0.0015)	0.8230 (0.0014)	0.8106 (0.0014)	0.8523 (0.0008)
	5	0.4671 (0.0075)	0.4850 (0.0077)	0.3175 (0.0034)	0.5915 (0.0068)	0.7989 (0.0019)	0.8193 (0.0014)	0.8229 (0.0017)	0.9109 (0.0015)	0.8197 (0.0007)
	6	0.7627 (0.0017)	0.7621 (0.0019)	0.7562 (0.0019)	0.7638 (0.0016)	0.7616 (0.0018)	0.7615 (0.0015)	0.7612 (0.0014)	0.7609 (0.0013)	0.7638 (0.0011)
3	1	0.5998 (0.0325)	0.6781 (0.0381)	0.5310 (0.0302)	0.7185 (0.0394)	0.7306 (0.0406)	0.7067 (0.0399)	0.7691 (0.0412)	0.8009 (0.0433)	0.6451 (0.0349)
	2	0.7097 (0.0017)	0.8221 (0.0077)	0.6997 (0.0098)	0.8566 (0.0028)	0.8526 (0.0020)	0.8642 (0.0019)	0.8492 (0.0018)	0.8526 (0.0014)	0.8882 (0.0008)
	3	0.7037 (0.0020)	0.7988 (0.0107)	0.6680 (0.0107)	0.8477 (0.0064)	0.8264 (0.0057)	0.8041 (0.0071)	0.8507 (0.0018)	0.8717 (0.0016)	0.7346 (0.0038)
	4	0.6914 (0.0020)	0.7952 (0.0096)	0.6352 (0.0097)	0.8317 (0.0061)	0.8264 (0.0064)	0.8039 (0.0076)	0.8672 (0.0019)	0.8911 (0.0016)	0.7863 (0.0079)
	5	0.4426 (0.0042)	0.4617 (0.0074)	0.3623 (0.0014)	0.5175 (0.0052)	0.6148 (0.0043)	0.5865 (0.0045)	0.7119 (0.0021)	0.7865 (0.0008)	0.5896 (0.0011)
	6	0.7659 (0.0017)	0.7634 (0.0015)	0.7567 (0.0018)	0.7632 (0.0014)	0.7606 (0.0013)	0.7623 (0.0011)	0.7630 (0.0012)	0.7643 (0.0011)	0.7672 (0.0010)
4	1	0.4281 (0.0207)	0.4347 (0.0222)	0.3896 (0.0197)	0.5109 (0.0253)	0.6515 (0.0316)	0.6542 (0.0316)	0.6564 (0.0320)	0.6629 (0.0325)	0.6542 (0.0316)
	2	0.4961 (0.0040)	0.5021 (0.0108)	0.4682 (0.0076)	0.6202 (0.0061)	0.7270 (0.0008)	0.7250 (0.0000)	0.8135 (0.0027)	0.8941 (0.0009)	0.7250 (0.0000)
	3	0.4768 (0.0034)	0.4801 (0.0100)	0.4639 (0.0074)	0.5806 (0.0058)	0.7264 (0.0006)	0.7250 (0.0000)	0.7329 (0.0021)	0.7422 (0.0049)	0.7250 (0.0000)
	4	0.4808 (0.0031)	0.4795 (0.0103)	0.4423 (0.0060)	0.5925 (0.0058)	0.7231 (0.0009)	0.7250 (0.0000)	0.7293 (0.0013)	0.7452 (0.0047)	0.7250 (0.0000)
	5	0.3664 (0.0030)	0.3613 (0.0056)	0.3658 (0.0023)	0.4036 (0.0041)	0.3276 (0.0020)	0.2756 (0.0003)	0.4667 (0.0017)	0.5412 (0.0011)	0.2750 (0.0000)
	6	0.7682 (0.0012)	0.7654 (0.0012)	0.7554 (0.0014)	0.7640 (0.0012)	0.7652 (0.0011)	0.7628 (0.0011)	0.7669 (0.0009)	0.7714 (0.0007)	0.7734 (0.0007)

Table 4.9: Proportion identified correctly and (standard error) of a signal with exponential covariance, a=0.01 and a=0.1

		a=0.01			**a=0.1**		
h	q	**b=3**	**b=10**	**b=30**	**b=3**	**b=10**	**b=30**
1.5	1	0.7884 (0.0013)	0.7763 (0.0014)	0.7446 (0.0006)	0.8476 (0.0053)	0.7694 (0.0022)	0.7997 (0.0013)
	2	0.7623 (0.0011)	0.7514 (0.0010)	0.7363 (0.0004)	0.8690 (0.0056)	0.7634 (0.0021)	0.7752 (0.0015)
	3	0.7666 (0.0011)	0.7559 (0.0011)	0.7373 (0.0004)	0.8679 (0.0056)	0.7651 (0.0021)	0.7813 (0.0014)
	4	0.7747 (0.0012)	0.7634 (0.0012)	0.7401 (0.0005)	0.8595 (0.0054)	0.7682 (0.0022)	0.7883 (0.0013)
	5	0.9014 (0.0011)	0.9245 (0.0011)	0.9052 (0.0015)	0.5911 (0.0061)	0.7130 (0.0047)	0.8534 (0.0014)
	6	0.7785 (0.0011)	0.7827 (0.0011)	0.7841 (0.0007)	0.7627 (0.0020)	0.7693 (0.0020)	0.7736 (0.0018)
2	1	0.8352 (0.0022)	0.8215 (0.0021)	0.7805 (0.0009)	0.8415 (0.0020)	0.7713 (0.0029)	0.8492 (0.0015)
	2	0.7806 (0.0015)	0.7718 (0.0013)	0.7531 (0.0005)	0.8713 (0.0014)	0.7690 (0.0029)	0.8079 (0.0016)
	3	0.7953 (0.0018)	0.7836 (0.0016)	0.7585 (0.0006)	0.8703 (0.0014)	0.7699 (0.0029)	0.8206 (0.0016)
	4	0.8050 (0.0020)	0.7933 (0.0018)	0.7645 (0.0006)	0.8545 (0.0019)	0.7702 (0.0027)	0.8308 (0.0016)
	5	0.9180 (0.0013)	0.9329 (0.0010)	0.9086 (0.0010)	0.5778 (0.0075)	0.6796 (0.0063)	0.8452 (0.0014)
	6	0.7598 (0.0012)	0.7595 (0.0011)	0.7615 (0.0011)	0.7625 (0.0016)	0.7613 (0.0014)	0.7590 (0.0015)
3	1	0.8860 (0.0019)	0.8776 (0.0018)	0.8167 (0.0012)	0.7056 (0.0017)	0.8140 (0.0031)	0.8936 (0.0020)
	2	0.8179 (0.0015)	0.8120 (0.0013)	0.7742 (0.0007)	0.7278 (0.0013)	0.8402 (0.0041)	0.8605 (0.0018)
	3	0.8376 (0.0017)	0.8283 (0.0017)	0.7822 (0.0008)	0.7251 (0.0016)	0.8386 (0.0040)	0.8673 (0.0019)
	4	0.8537 (0.0016)	0.8469 (0.0016)	0.7968 (0.0007)	0.7150 (0.0016)	0.8263 (0.0034)	0.8849 (0.0017)
	5	0.8528 (0.0009)	0.8502 (0.0009)	0.8712 (0.0007)	0.5151 (0.0055)	0.5828 (0.0052)	0.7369 (0.0013)
	6	0.7657 (0.0013)	0.7681 (0.0012)	0.7692 (0.0011)	0.7656 (0.0016)	0.7635 (0.0014)	0.7633 (0.0013)
4	1	0.8767 (0.0009)	0.8602 (0.0008)	0.8420 (0.0011)	0.5095 (0.0021)	0.6349 (0.0061)	0.7250 (0.0000)
	2	0.8607 (0.0012)	0.8682 (0.0012)	0.8276 (0.0007)	0.5375 (0.0022)	0.6792 (0.0051)	0.8397 (0.0032)
	3	0.8716 (0.0011)	0.8719 (0.0012)	0.8339 (0.0008)	0.5151 (0.0022)	0.6532 (0.0065)	0.7250 (0.0000)
	4	0.8817 (0.0012)	0.8756 (0.0009)	0.8406 (0.0009)	0.5190 (0.0021)	0.6536 (0.0053)	0.7250 (0.0000)
	5	0.7092 (0.0008)	0.6770 (0.0008)	0.7137 (0.0007)	0.4053 (0.0024)	0.4512 (0.0048)	0.4786 (0.0023)
	6	0.7749 (0.0005)	0.7739 (0.0005)	0.7761 (0.0005)	0.7688 (0.0011)	0.7652 (0.0010)	0.7694 (0.0008)

Table 4.10: Type I Errors and (standard errors) of no signal with exponential covariance, a=0.01, 0.1

		a=**0.01**	**0.1**	**0.01**	**0.1**	**0.01**	**0.1**
h	q	b=**3**		b=**10**		b=**30**	
1.5	1	0.5093 (0.0411)	0.6139 (0.0387)	0.6988 (0.0314)	0.6715 (0.0338)	0.6743 (0.0323)	0.6717 (0.0323)
	2	0.4910 (0.0407)	0.6007 (0.0389)	0.6954 (0.0313)	0.6666 (0.0339)	0.6714 (0.0321)	0.6757 (0.0317)
	3	0.5115 (0.0409)	0.6160 (0.0384)	0.7013 (0.0310)	0.6731 (0.0336)	0.6770 (0.0319)	0.6743 (0.0319)
	4	0.6142 (0.0387)	0.7028 (0.0343)	0.7759 (0.0235)	0.7489 (0.0275)	0.7438 (0.0262)	0.7364 (0.0273)
	5	0.0089 (0.0009)	0.0085 (0.0010)	0.0066 (0.0009)	0.0057 (0.0008)	0.0065 (0.0011)	0.0065 (0.0011)
	6	0.5746 (0.0092)	0.5787 (0.0089)	0.7350 (0.0080)	0.7300 (0.0085)	0.7706 (0.0080)	0.7703 (0.0078)
2	1	0.8153 (0.0034)	0.8162 (0.0038)	0.7611 (0.0037)	0.7625 (0.0037)	0.7445 (0.0036)	0.7462 (0.0037)
	2	0.8157 (0.0034)	0.8164 (0.0038)	0.7613 (0.0037)	0.7626 (0.0037)	0.7440 (0.0036)	0.7462 (0.0037)
	3	0.8157 (0.0034)	0.8164 (0.0038)	0.7613 (0.0037)	0.7626 (0.0037)	0.7440 (0.0036)	0.7462 (0.0037)
	4	0.8078 (0.0088)	0.8085 (0.0090)	0.7539 (0.0085)	0.7551 (0.0085)	0.7360 (0.0082)	0.7386 (0.0083)
	5	0.0074 (0.0073)	0.0075 (0.0074)	0.0088 (0.0088)	0.0085 (0.0085)	0.0094 (0.0094)	0.0087 (0.0087)
	6	0.6841 (0.0098)	0.6869 (0.0101)	0.8367 (0.0077)	0.8245 (0.0082)	0.8590 (0.0066)	0.8624 (0.0071)
3	1	0.7371 (0.0190)	0.7546 (0.0141)	0.5072 (0.0341)	0.5104 (0.0343)	0.7067 (0.0053)	0.7071 (0.0055)
	2	0.7373 (0.0189)	0.7546 (0.0141)	0.5072 (0.0341)	0.5104 (0.0343)	0.7067 (0.0053)	0.7071 (0.0055)
	3	0.7371 (0.0190)	0.7546 (0.0141)	0.5072 (0.0341)	0.5104 (0.0343)	0.7067 (0.0053)	0.7071 (0.0055)
	4	0.7371 (0.0190)	0.7546 (0.0141)	0.5840 (0.0295)	0.5684 (0.0314)	0.7067 (0.0053)	0.7071 (0.0055)
	5	0.0009 (0.0004)	0.0004 (0.0002)	0.0021 (0.0002)	0.0020 (0.0002)	0.0000 (0.0000)	0.0000 (0.0000)
	6	0.7727 (0.0116)	0.7839 (0.0117)	0.9062 (0.0059)	0.8961 (0.0070)	0.9227 (0.0050)	0.9227 (0.0052)
4	1	0.7298 (0.0133)	0.7387 (0.0077)	0.7240 (0.0071)	0.7309 (0.0074)	0.6934 (0.0097)	0.6897 (0.0097)
	2	0.7298 (0.0133)	0.7387 (0.0077)	0.7240 (0.0071)	0.7309 (0.0074)	0.6934 (0.0097)	0.6897 (0.0097)
	3	0.7298 (0.0133)	0.7387 (0.0077)	0.7240 (0.0071)	0.7309 (0.0074)	0.6934 (0.0097)	0.6897 (0.0097)
	4	0.7434 (0.0084)	0.7387 (0.0077)	0.7240 (0.0071)	0.7309 (0.0074)	0.6934 (0.0097)	0.6897 (0.0097)
	5	0.0004 (0.0001)	0.0003 (0.0001)	0.0000 (0.0000)	0.0000 (0.0000)	0.0001 (0.0001)	0.0001 (0.0001)
	6	0.8025 (0.0132)	0.8195 (0.0129)	0.9269 (0.0052)	0.9186 (0.0061)	0.9401 (0.0049)	0.9403 (0.0043)

Chapter 5

Discussion

In this dissertation, SiZer, Significant ZERo Crossings of the derivative has been taken from its original form to some new advancements that will make it applicable to a wider range of scientific problems. To aid in the advancement provided by Park and Kang (2008), the first offering here was to derive and prove asymptotic properties. Park and Kang (2008) provided a SiZer that could be based on differences instead of derivatives. This allows for the scale-space analysis to be performed on two independent regression curves. To enhance the foundation of their proposal, we have provided here three asymptotic properties that address the need for causality of the creation of extrema and weak convergence of the empirical scale-space surface.

Also proposed is a new SiZer tool that can compare multiple time series. This device provides inference for not only two independent regression curves, but two or more regression curves with a dependent error structure. This approach combines the ideas behind the work of authors that worked with time series and the SiZer for two regression curves. The SiZer that is proposed here works with time series and takes the difference between two time series to denote any meaningful differences between the two. The performance of this proposed instrument was found to have a few spurious pixels on occasion, but as a whole performed well with different regression models, various types of dependent structure, and with known or estimated autocovariance functions. To support this development, asymptotic properties have also been newly given for the weak convergence of the empirical and the theoretical scale-space surfaces in the case when comparing two time series.

Another SiZer was also proposed that can evaluate more than two time series. This takes an approach where residuals are calculated under the null hypothesis, that supposes each individual times series can be approximated by the same local linear estimator. Using this common function, residuals are then calculated. Under the alternative hypothesis, we compute a different estimator for each. Then, a second group of residuals are computed between each data value and the set function from which it came. These two groups of residuals are then compared as two sets of time series with the tool that was proposed previously. Again the performance of this tool is demonstrated with simulation and real data analysis, with various types of dependence correlation structure. The performance of this SiZer shows that it is able to detect whether or not there is significant difference present between the sets of residuals and thus whether there should be individual function estimators.

The final version of SiZer proposed is that of a two-dimensional SiZer that can analyze not only independent data, but data that possesses spatial correlation structure. Now, instead of focusing on derivatives, or differences between curves, we look at the partial derivative or gradient at any specific location. This two-dimensional SiZer is also an advancement upon one that could analyze independent data only. The accounting of this covariance has also led us to propose a new covariance matrix, and resulting test statistic. We investigate its performance with a number of multiple comparison procedures and find that once the spatial correlation is taken into account, the independent blocks quantile, used previously by a number of other authors, performs quite well at the task of reducing Type I error rates. This Spatial SiZer was found to perform extremely well in simulated and real datasets with greatly reduced Type I error rates and few spuriously highlighted pixels in comparison with the Original SiZer, which did not account for the dependence structure in a dataset.

We believe that these new proposals have assisted in strengthening the foundation of proposals by other authors by providing some accompanying asymptotic properties to their work. Also, the ability to address two or more time series and two-dimensional images with spatial structure is an advancement that will allow numerous scientific problems to be analyzed by the visualization techniques provided by SiZer that could not have previously been.

References

Adler, R. J. (1990). An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. *IMS.* Hayward, California.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 1, 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing Under Dependency. *The Annals of Statistics*, 29, 4, 1165–1199.

Bickel, P. J. and Wichura, M. J. (1971). Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. *The Annals of Mathematical Statistics*, 42, 1656–1670.

Bowman, A. and Young, S. (1996). Graphical Comparison of Nonparametric Curves. *Applied Statistics*, 45, 83–98.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting.* Springer-Verlag, New York.

Chaudhuri, P. and Marron, J.S. (1999). SiZer for Exploration of Structures in Curves. *Journal of the American Statistical Association*, 94, 807–823.

Chaudhuri, P. and Marron, J.S. (2000). Scale Space View of Curve Estimation. *The Annals of Statistics*, 28, 408–428.

Christiansen, H. and Pigott, C. (1997). Long-Term Interest Rates in Globalised Markets. *Economic Department Working Papers*, 175, 1–44.

Chu, C. K., Glad, I. K., Godtliebsen, F., and Marron, J. S. (1998). Edge Preserving Smoothers for Image Processing. *Journal of the American Statistical Association*, 93, 526–556.

Cressie, Noel A.C. (1991). *Statistics for Spatial Data.* Wiley Series in Probability and Mathematical Statistics, New York.

Delgado, M. A. (1993). Testing the Equality of Nonparametric Regression Curve. *Statistics & Probability Letters*, 17, 199–204.

Dette, H. and Neumeyer, N. (2001). Nonparametric Analysis of Covariance. *The Annals of Statistics*, 29, 1361–1400.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics 85. Springer, New York.

Duong, T., Cowling, A., Koch, I., Wand, M. P., (2008). Feature Significance for Multivariate Kernel Density Estimation. *Computational Statistics and Data Analysis*, 52, 4225–4242.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall, London.

Fan, J., and Marron, J. S. (1994). Fast Implementations of Nonparametric Curve Estimators, *Journal of Computational and Graphical Statistics*, 3, 35–56.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer.

Gasser, T. and Müller, H.G. (1984). Estimating Regression Functions and Their Derivatives by the Kernel Method. *Scandinavian Journal of Statistics*, 11, 171–185.

Godtliebsen, F., Marron, J.S., Chaudhuri, P. (2002). Significance in Scale Space for Bivariate Density Estimation. *Journal of Computational and Graphical Statistics*, 11, 1–21.

Godtliebsen, F., Marron, J.S., Chaudhuri, P. (2004). Statistical Significance of Features in Digital Images. *Image and Vision Computing*, 22, 1093–1104.

Hall, P. and Hart, J. D. (1990). Bootstrap Test for Difference Between Means in Nonparametric Regression. *Journal of the American Statistical Association*, 85, 1039–1049.

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

Härdle W. and Marron, J. S. (1990). Semiparametric Comparison of Regression Curves. *The Annals of Statistics*, 13, 63–89.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6, 65–70.

Karlin, S. (1968). *Total Positivity, Vol. I.* Stanford: Stanford University Press.

Koul, H. L. and Stute, W. (1998). Regression Model Fitting with Long Memory Errors. *Journal of Statistical Planning and Inferences*, 71, 35–56.

Kulasekera, K. B. (1995). Comparison of Regression Curves using Quasi-Residuals. *Journal of the American Statistical Association*, 90, 1085–1093.

Kulasekera, K. B. and Wang, J. (1995). Smoothing Parameter Selection for Power Optimality in Testing of Regression Curves. *Journal of the American Statistical Association*, 92, 500–511.

Li, F. (2006). Testing for the Equality of Two Nonparametric Regression Curves with Long Memory Errors. *Communications in Statistics-Simulation and Computation*, 35, 621–643.

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

Munk, A. and Dette, H. (1998). Nonparametric Comparison of Several Regression Functions: Exact and Asymptotic Theory. *The Annals of Statistics*, 26, 2339–2368.

Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability and its Applications*, 9, 141–142

Neumeyer, N. and Dette, H. (2003). Nonparametric Comparison of Regression Curves: An Empirical Process Approach. *The Annals of Statistics*, 31, 880–920.

Pardo-Fernandez, J. C., Keilegom, I. V., and Gonzalez-Manteiga, W. (2007). Testing for the Equality of k Regression Curves. *Statistica Sinica*, 17, 1115–1137.

Park, C., Hannig, J., and Kang, K. (2009). Improved SiZer for Time Series. *Statistica Sinica*, 19, 1511–1530.

Park, C. and Kang, K. (2008). SiZer Analysis for the Comparison of Regression Curves. *Computational Statistics and Data Analysis*, 52, 3954–3970.

Park, C., Marron, J.S., and Rondonotti, V. (2004). Dependent SiZer: Goodness of Fit Tests for Time Series Models. *Journal of Applied Statistics*, 31, 999–1017.

Pavlicova, M., Cressie, N., Santner, T. J., and Algaze, A. (2003). Using Enhanced FDR to Find Activation in FMRI images. *NeuroImage*, 19, S919.

Priestley, M.B. and Chao, M.T. (1972). Non-parametric Function Fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34, 385–392.

Rondonotti, V., Marron J.S., and Park, C. (2007). SiZer for Time Series: A New Approach to the Analysis of Trends. *Electronic Journal of Statistics*, 1, 268–289.

Ruppert, D., Sheather, S.J., and Wand, M.P. (1995). An Effective Bandwidth Selector for Local Least Squares Regression. *Journal of the American Statistical Association*, 90, 1257–1270.

Schoenberg, I. J. (1950). On Pólya Frequency Functions. VariationDiminishing Integral Operators of the Convolution Type. *Acta Scientiarum Mathematicarum*, 12, 97-106.

Scott, D. W. (1992). *Multivariate Density Estimation, Theory, Practice and Visualization*. New York: Wiley.

Silverman, B.W. (1981). Using Kernel Density Estimates to Investigate Multimodality. *Journal of the Royal Statistical Society*, B, 43, 97–99.

Utev, S. A. (1990). The Central Limit Theorem for φ-Mixing Arrays of Random Variables. *Theory of Probability and Its Applications*, 35, 110–117.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman & Hall/CRC.