**Stat**
The ISI's Journal for the Rapid
Dissemination of Statistics Research
(wileyonlinelibrary.com) DOI: 10.100X/sta.0000

# Metrics for SiZer Map Comparison

**Jan Hannig[a]*, Thomas C. M. Lee[b], Cheolwoo Park[c]**

Received 00 Month 2012; Accepted 00 Month 2012

**SiZer is a powerful visualization tool for uncovering real structures masked in noisy data. It produces a two-dimensional plot, the so-called SiZer map, to help the data analyst to carry out this task. Since its first proposal, many different extensions and improvements have been developed, including robust SiZer, quantile SiZer, and various SiZers for time series data, just to name a few. Given these many SiZer variants, one important question is, how can one evaluate the quality of a SiZer map produced by any one of these variants? The primary goal of this article aims to answer this question by proposing two metrics for quantifying the discrepancy between any two SiZer maps. With such metrics, one can systematically calculate the distance between a "true" SiZer map and a SiZer map produced by any one of the SiZer variants. Consequently, one can select a "best" SiZer variant for the problem at hand by selecting the variant that produces SiZer maps that are, on average, closest to the "true" SiZer map.** 

**Keywords:** Baddeley's delta metric, discrepancy measure, graphical data analysis, image metric, multiscale methods, oracle SiZer map

## 1. Introduction

The SiZer methodology of Chaudhuri & Marron (1999, 2000) and Hannig & Marron (2006) is a powerful visualization tool for exploring structures hidden in noisy data. It produces a graphical construct, termed the *SiZer map*, that aids the data analyst to isolate the significant structures from those spurious features that are due to sampling noise.

In the univariate setting, a SiZer map is a two-dimensional image that summarizes the results of a sequence of hypothesis tests. These tests are performed to test if locally the estimated slopes of the underlying regression function or density function are significantly increasing, decreasing, or neither. These tests are also done with slopes estimated at different scales (i.e., resolutions). The rationale behind is that, if all estimated slopes (at different scales) to the left of location $x$ are significantly decreasing while all estimated slopes to the right of $x$ are significantly increasing, then it provides strong evidence that there is a significant trough located at $x$ in the underlying function. A more formal description of SiZer is given in the next section.

[a]**Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, NC 27599-3260, U.S.A.**
[b]**Department of Statistics, University of California at Davis, CA 95616, U.S.A.**
[c]**Department of Statistics, University of Georgia, Athens, GA 30602-7952, U.S.A.**
*__Email: jan.hannig@unc.edu__

 
*Prepared using staauth.cls [Version: 2012/05/12 v1.00]*

Since the first proposal of SiZer, various extensions and improvements have been developed. For example, Hannig & Lee (2006) developed a *robust SiZer* which produces SiZer maps that are resistant to outliers, while Park et al. (2010) developed a *quantile SiZer* that targets various quantile structures. For SiZers devoted to time series or dependent data, see the proposals of Park et al. (2004), Rondonotti et al. (2007), Olsen et al. (2008b) and Park et al. (2009).

In higher dimensional settings, Godtliebsen et al. (2002) studied bivariate density estimation, Ganguli & Wand (2004) considered bivariate smoothing, Ganguli & Wand (2007) and Gonzalez-Manteiga et al. (2008) examined generalized additive models, and Holmstrom & Pasanen (2012) and Vaughan et al. (2012) applied the SiZer idea to images.

Other problem-specific SiZers have also been developed; e.g., Marron & de Una-Alvarez (2004) studied hazard rate estimation, Kim & Marron (2006) and Olsen et al. (2008a) considered changepoint and jump detection, and Park & Kang (2008) and Park et al. (2009) investigated the use of SiZer for multiple signal comparisons.

The Bayesian paradigm has also been a popular approach; see, for examples, Erästö & Holmström (2005), Godtliebsen & Oigard (2005), Oigard et al. (2006) and Erästö & Holmström (2007). Lastly, Marron & Zhang (2005) and Park et al. (2007) developed SiZers with, respectively, smoothing splines and wavelets, instead of local linear regression as in the original SiZer.

Given the existence of these many SiZers, a natural question to ask is, how do we evaluate their relative performances? To the best of our knowledge, there is no objective method available in the literature for handling this task. The primary goal of this article is to propose two new metrics for quantifying the distance between a "true" SiZer map and the SiZer map produced by any one of the SiZers mentioned above. With such metrics, one can systematically define the best SiZer as the one that produces SiZer maps that are, on average, closest to the "true" SiZer map (i.e., has the smallest distance). In other words, these metrics provide discrepancy measures for SiZer map comparison.

The rest of this article is organized as follows. Section 2 describes three existing SiZers that will be used in the later part of this article. The new SiZer map metrics, together with some theoretical properties, are presented in Section 3. Numerical examples illustrating the use of the proposed metrics are given in Section 4. Lastly, concluding remarks are provided in Section 5.

# 2. Various SiZer Proposals

In this section we provide the core ideas of the SiZer methodology proposed by Chaudhuri & Marron (1999), and describe two of its variants, the robust SiZer of Hannig & Lee (2006) and the spline SiZer of Marron & Zhang (2005). For the rest of this article the following nonparametric regression model is considered:

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i, \quad i = 1, \ldots, n, \quad \epsilon_i \sim \text{i. i. d. } N(0,1), \tag{1}$$

where $m(x)$ is a smooth regression function, $\{(X_i, Y_i)_{i=1}^n\}$ are observed data, and $\sigma^2(x) = \text{Var}(Y_i|X_i = x)$.

## 2.1. Original SiZer of Chaudhuri & Marron (1999)

The original SiZer applies local linear regression (e.g., Fan & Gijbels, 1996) to estimate $m(x)$ and its derivative $m'(x)$ locally at different scales by changing the values of the bandwidth $h$. Let $K(\cdot)$ be a kernel, taken as the standard normal density function, and write $K_h(\cdot) = K(\cdot/h)/h$. For a given pair of $(x, h)$, denote $(\hat{\beta}_0, \hat{\beta}_1)$ as the solution to the minimization problem

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0,\beta_1}{\arg\min} \sum_{i=1}^{n} [Y_i - \{\beta_0 + \beta_1(X_i - x)\}]^2 K_h(X_i - x).$$



*Prepared using staauth.cls*

Let $m_h(x) = K_h * m(x)$ and $m'_h(x) = K'_h * m(x)$ with $*$ denoting the convolution. Then the estimates for $m_h(x)$ and $m'_h(x)$ are given, respectively, as $\hat{m}_h(x) = \hat{\beta}_0$ and $\hat{m}'_h(x) = \hat{\beta}_1$. From this, one can construct a family of smooths that displays the estimated regression functions with different $h$ values; see the top panels of Figure 1 for examples.

A SiZer map is a two-dimensional plot summarizing the results of a sequence of hypothesis tests. The horizontal axis of the map represents the location $x$, while the vertical axis represents the scale $h$. A test is performed for each pair of $(x, h)$, with four possible outcomes: the slope $m'_h(x)$ at location $x$ with scale $h$ is zero, $m'_h(x)$ is increasing, $m'_h(x)$ is decreasing, and not enough data to perform a reliable test. This test can be achieved by constructing a confidence interval for $m'_h(x)$:

$$\hat{m}'_h(x) \pm q(h)\widehat{\mathrm{SE}}\{\hat{m}'_h(x)\}, \tag{2}$$

where $q(h)$ is an appropriate Gaussian quantile used for adjusting multiple testing (Hannig & Marron, 2006), and $\widehat{\mathrm{SE}}\{\hat{m}'_h(x)\}$ is an estimate of the standard error of $\hat{m}'_h(x)$ (Fan & Gijbels, 1996). The pixel at $(x, h)$ in the SiZer map is colored blue if the confidence interval $> 0$, which suggests that the regression curve is increasing at location $x$ and scale $h$. Similarly, the pixel is colored red if the confidence interval $< 0$ suggesting the regression curve at $(x, h)$ is decreasing, and purple if the confidence interval contains 0 where a significant slope cannot be concluded. If the data are too sparse to determine the significance at $(x, h)$, then the pixel is colored gray. A pixel is declared as sparse when the effective sample size, defined as $\mathrm{ESS}(x, h) = \sum_{i=1}^{n} K_h(X_i - x)/K_h(0)$, is less than 5.

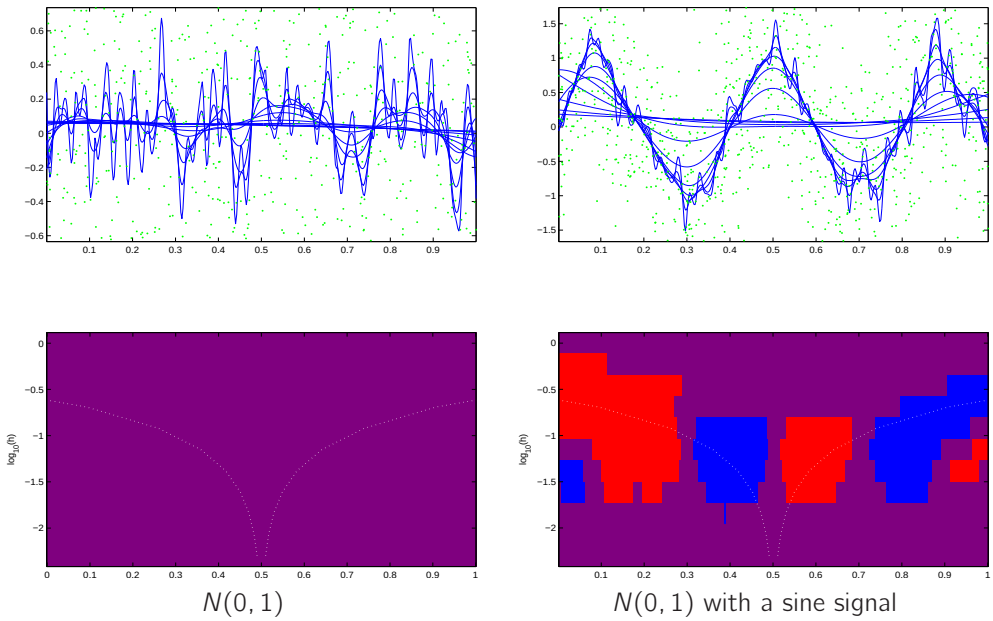



**Figure 1.** Plots for family of smooths (top panels) and SiZer maps (bottom panels).

Figure 1 displays two SiZer plots. In the top panels, the observed data points $(X_i, Y_i)$ are shown in green. The true regression function for the left panels is $m(x) = 0$ (i.e., no signal), while for the right is a sine curve. The individual thin blue curves display the family of smooths; i.e., the local linear estimates of the curves computed with different bandwidths. The bottom panels show the corresponding SiZer maps, where the horizontal axes show the location $x$, while the vertical axes indicate the logarithm of the bandwidths $h$ of the family of smooths. For the data set in the left side of Figure 1, where there is no signal but only i. i. d. standard normal errors, the corresponding SiZer map correctly identifies no significant structure and marks the entire map as purple. For the SiZer map for the data set in the right side, it can be seen that the sine curve estimate begins with a significantly increasing (blue) slope on the

left, then changes according to every turn in the sine curve, and lastly ends with a significantly decreasing (red) slope on the right edge of the plot.

### 2.2. Robust SiZer of Hannig & Lee (2006)

Hannig & Lee (2006) proposed a robust version of SiZer that is capable of identifying potential outliers and producing statistical inference with different degrees of robustness. When compared to the original SiZer, a major difference of the robust SiZer is that it uses the $M$-type local linear regression to obtain estimates for $m_h(x)$. To be more specific, $(\hat{\beta}_0, \hat{\beta}_1)$ are obtained as the joint minimizer of

$$\sum_{i=1}^{n} \rho_c \left[ \frac{Y_i - \{\beta_0 + \beta_1(X_i - x)\}}{\hat{\sigma}(X_i)} \right] K_h(X_i - x),$$

where $\hat{\sigma}(x)$ is a robust estimate of $\sigma(x)$ and $\rho_c(x)$ is the Huber loss function (Huber, 1981)

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

The tuning parameter $c$ controls the robustness of the estimates: smaller $c$ yields more robust estimates while $c = \infty$ essentially corresponds to the original SiZer with the difference being the use of a different estimator of variance. For the rest of this article the typical choice $c = 1.345$ is used in all the numerical experiments. This robust SiZer also provides an indicator for outliers in its SiZer maps, but we do not consider this feature in this article. The interpretation of a robust SiZer map remains the same as for the original SiZer map. For other details, such as the expression for the estimate of the standard error of the robust derivative estimates, see Hannig & Lee (2006).

### 2.3. Spline SiZer of Marron & Zhang (2005)

Marron & Zhang (2005) extended the kernel type estimation to the smoothing spline estimation in SiZer analysis. The spline SiZer estimates the regression function by minimizing

$$\sum_{i=1}^{n} \{Y_i - m(X_i)\}^2 + \lambda \int \left\{ m''(x) \right\}^2 dx,$$

where $\lambda$ is the smoothing spline tuning parameter that determines the smoothness of the regression estimate $\hat{m}_\lambda(x)$. In this spline SiZer, $\lambda$ is the scale parameter which plays the same role as the bandwidth $h$ in the original SiZer. The interpretation of a spline SiZer map also remains the same as for the original SiZer map. For other implementation details, such as the expressions for the first derivative estimate and its standard error, see Marron & Zhang (2005).

## 3. Two Metrics for SiZer Map Comparison

This section proposes two metrics for SiZer map comparison. We start by introducing some notation. For ease of presenting mathematical expressions in this section we assume that a SiZer map $S$ is encoded as follows

$$S(x) = \begin{cases} 1 & \text{if pixel } x \text{ is blue} \\ -1 & \text{if pixel } x \text{ is red} \\ 0 & \text{if pixel } x \text{ is purple or grey.} \end{cases} \tag{3}$$

The rationale for this re-coding is that blue color corresponds to significant evidence that the function is increasing, red color to significant evidence that the function is decreasing, while purple and gray colors do not provide significant evidence either way.

Next define the sets of pixels for the blue, purple and red regions as, respectively, $S_B = \{x : S(x) = 1\}$, $S_P = \{x : S(x) = 0\}$ and $S_R = \{x : S(x) = -1\}$.

When measuring the distance between two pixels $x_1 = (x_{11}, x_{12})$ and $x_2 = (x_{21}, x_{22})$ on a SiZer map, we choose to use the $l_1$ distance; i.e., $d(x_1, x_2) = |x_{11} - x_{21}| + |x_{12} - x_{22}|$. This choice is driven by the fact that the axes on the SiZer map correspond to two very different quantities: location and scale.

After defining the distance between two points, a distance between a pixel $x$ and a set of pixels $B$ is defined as the shortest distance from pixel $x$ to $B$; i.e.,

$$d(x, B) = \min\{d(x, b) : b \in B\}.$$

## 3.1. A Direct Extension of Baddeley (1992)

Baddeley (1992) proposed an error metric for quantifying the distance between two binary images. This metric uses the $p$-th order mean difference between thresholded distance transformations of the two images. It measures the discrepancies between the two images by considering the pattern of pixels. It satisfies the three axioms of a metric: minimality, symmetry, and triangle inequality, which generates a desired topology that defines continuity and convergence. Also, it is robust to noise.

Let $X$ denote a pixel set with $N$ pixels. A binary image is an indicator function $I_B(x)$ with the set $B$ being the collection of white pixels with the rest of the pixels being black. Then, Baddeley's metric $\Delta_b$ is defined for two binary images $B_1$ and $B_2$ as

$$\Delta_b(B_1, B_2) = \left[\frac{1}{N}\sum_{x \in X}\left|\omega\{d(x, B_1)\} - \omega\{d(x, B_2)\}\right|^p\right]^{1/p},$$

where $\omega(t)$ is a concave, continuous, and non-decreasing function for $t \geq 0$. Baddeley (1992) recommends the cutoff transformation

$$\omega(t) = \min(t, c)$$

for a fixed $c > 0$. In our analysis, we choose $c = 5$ and $p = 2$, which is a common choice as suggested by Baddeley (1992).

Since SiZer maps take three basic values (blue, red and purple), we propose to compare these blue, purple and red areas separately. More precisely we define our first metric for comparing two SiZer maps $S_1$ and $S_2$ as

$$\Delta^B(S_1, S_2) = \frac{\Delta_b(S_{B,1}, S_{B,2}) + \Delta_b(S_{P,1}, S_{P,2}) + \Delta_b(S_{R,1}, S_{R,2})}{3}, \tag{4}$$

where $S_{B,i}$, $S_{P,i}$ and $S_{R,i}$ are the blue, purple and red regions of SiZer maps $S_i$, $i = 1, 2$, respectively. Note that, as the average of 3 metrics, $\Delta^B$ is again a metric; i.e., it satisfies minimality, symmetry, and triangle inequality.

## 3.2. An Improved SiZer Metric

One potential drawback with the above metric $\Delta^B$ in (4) is that it ignores the meaning and ordering of the SiZer colors. In particular replacing red color with purple is much lesser of an issue than replacing red with a blue, as the latter totally switches the conclusion. In this subsection we propose a second SiZer map metric that repairs this drawback.

To explain this second metric we first introduce the image metric of Wilson et al. (1997), which generalizes Baddeley's binary metric to gray scale images. This metric is based on the distance of a point $(x, y)$, where $x \in X$ is a pixel location and $y \in Y$ is a gray scale intensity, to the graph of the gray scale image $G_S = \{(x, y) : y \leq S(x)\}$. The simple form of the gray scale metric is

$$\Delta_g(S_1, S_2) = \left( \frac{1}{KN} \sum_x \sum_y \left| \omega[d'\{(x, y), G_{S_1}\}] - \omega[d'\{(x, y), G_{S_2}\}] \right|^p \right)^{1/p}, \tag{5}$$

where $d'\{(x_1, y_1), (x_2, y_2)\} = \max\{d(x_1, x_2), c|y_1 - y_2|\}$, $\omega(t) = \min(t, \tilde{c})$, $K$ is the number of intensity levels in the gray scale image, $N$ is the number of pixels in the image. The role of the function $\omega(t)$ is to speed up numerical calculations by limiting the number of distances and gray scale levels considered when computing the sum. This is not desirable in our situation as the SiZer map has only three levels. In what follows we therefore eliminate the effects of $\omega(t)$ by selecting $\tilde{c} > 2c$.

Let us now apply (5) to the SiZer maps encoded as in (3). First notice that

$$\begin{aligned} \omega[d'\{(x, -1), G_S\}] &= 0, \\ \omega[d'\{(x, 0), G_S\}] &= \min\{d(x, S_P \cup S_B), c\}, \\ \omega[d'\{(x, 1), G_S\}] &= \min[d(x, S_B), \max\{c, d(x, S_P \cup S_B)\}, 2c]. \end{aligned}$$

Consequently

$$\begin{aligned} \Delta_g(S_1, S_2) = \Big[ \frac{1}{3N} \sum_x \big| \min\{d(x, S_{P,1} \cup S_{B,1}), c\} - \min\{d(x, S_{P,2} \cup S_{B,2}), c\} \big|^p \\ + \big| \min[d(x, S_{B,1}), \max\{c, d(x, S_{P,1})\}, 2c] \\ - \min[d(x, S_{B,2}), \max\{c, d(x, S_{P,2})\}, 2c] \big|^p \Big]^{1/p}. \end{aligned}$$

The gray scale metric $\Delta_g$ in (5) has an inherent non-symmetry in the gray scale space due to the definition of $G_S$. In particular a different distance will be obtained if one switches the white-black direction in the gray scale; e.g., $\Delta_g(S_1, S_2) \neq \Delta_g(-S_1, -S_2)$. While this non-symmetry could be a desirable property for gray scale images where bright pixels are sometimes considered to be more important, it is highly undesirable for comparison of SiZer images. To explain the issue, suppose we modify all the SiZer maps by exchanging the blue and red colors, effectively changing the sign of the underlying data. It is intuitively clear that the distance between the modified SiZer maps should remain the same, which is not true for $\Delta_g$. Additionally, when defining a proper SiZer metric, it is advantageous to place more emphasis on the statistically significant regions (i.e., blue and red) in the SiZer maps. This is because conclusions based on SiZer maps are derived from these regions.

These considerations lead us to define a new SiZer metric:

$$\Delta^S(S_1, S_2) = \left[ \frac{1}{2N} \sum_x \left| d'\{(x, 1), G_{S_1}\} - d'\{(x, 1), G_{S_2}\} \right|^p + \left| d'\{(x, 1), G_{-S_1}\} - d'\{(x, 1), G_{-S_2}\} \right|^p \right]^{1/p}. \tag{6}$$

Notice that we sum over values $y = 1$ (i.e., blue) for both the original maps $(S_1, S_2)$ and the inverted maps $(-S_1, -S_2)$. By symmetry the last two terms correspond to summing over $y = -1$ (i.e., red) in the original map with the red-blue

direction inverted. After some algebra, (6) simplifies to

$$\Delta^S(S_1, S_2) = \left(\frac{1}{2N}\sum_x \Big|\min[d(x, S_{R,1}), \max\{c, d(x, S_{P,1})\}, 2c] - \min[d(x, S_{R,2}), \max\{c, d(x, S_{P,2})\}, 2c]\Big|^p + \Big|\min[d(x, S_{B,1}), \max\{c, d(x, S_{P,1})\}, 2c] - \min[d(x, S_{B,2}), \max\{c, d(x, S_{P,2})\}, 2c]\Big|^p\right)^{1/p}. \tag{7}$$

**Proposition 1** *The quantity $\Delta^S(S_1, S_2)$ is a metric; i.e., it satisfies minimality, symmetry, and triangle inequality.*

**Proof:** The minimality and symmetry are straightforward. To prove triangle inequality notice that

$$\begin{aligned}\Delta^S(S_1, S_3) &= \frac{1}{(2N)^{1/p}}\left(\sum_x \Big|[d'\{(x,1), G_{S_1}\} - d'\{(x,1), G_{S_2}\}] + [d'\{(x,1), G_{S_2}\} - d'\{(x,1), G_{S_3}\}]\Big|^p \right.\\ &\qquad \left. + \sum_x \Big|[d'\{(x,1), G_{-S_1}\} - d'\{(x,1), G_{-S_2})\}] + [d'\{(x,1), G_{-S_2})\} - d'\{(x,1), G_{-S_3})\}]\Big|^p\right)^{1/p}\\ &\le \Delta^S(S_1, S_2) + \Delta^S(S_2, S_3).\end{aligned}$$

The second inequality follows by the triangle inequality for $l_p$ norms.

# 4. Illustrative Examples

Two sets of experiments were conducted to illustrate the sample behaviors of the two SiZer map metrics proposed in the previous section. In particular these two metrics were applied to calculate the distances between various SiZer maps obtained by the original, robust and spline SiZers reviewed in Section 2. For all experiments we used $n = 50$, with $X_i$, $i = 1, \ldots, n$ being equally spaced in $[0, 1]$. For the original and robust SiZers, 41 bandwidth values of $h$ were selected from the interval $[1/(n-1), 1]$, while for the spline SiZer, $\lambda = nh^4$ were used (Silverman, 1984).

For benchmark comparisons, we define an *oracle SiZer map* as the one that is produced by applying the original SiZer to the true regression function $m(x)$ and true variance function $\sigma^2(x)$ in (1). In other words, the input data are $\{X_i, m(X_i)\}_{i=1}^n$, and the true variance function $\sigma^2(x)$ is plugged into the confidence interval in (2).

## 4.1. Experiment 1

In this first experiment we compare different oracle SiZer maps obtained from a sine function with different phase shifts indexed by a constant $a$: $m_a(x) = 2\sin\{2\pi(x + a)\}$. The noise variance used was a constant $\sigma(x) = 0.5$. For convenience denote the oracle SiZer map that corresponds to $m_a(x)$ as $S(a)$. Figure 2 displays $S(0)$, $S(0.1)$, $S(0.3)$ and $S(0.5)$; one can see that the red and blue regions in $S(0)$ and $S(0.5)$ are completely opposite and hence the distance between these two SiZer maps should be relatively large.

Both $\Delta^B$ and $\Delta^S$ were applied to calculate the distances between $S(0)$ and $S(a)$ for $a \in (0, 2]$. These calculated distances are plotted in Figure 3. For both metrics, such calculated distances, when viewed as a function of $a$, show a periodic behavior due to the sine wave. They reach their maximum around $a = 0.5$ and 1.5 where opposite maps are compared, and attain their minimum (zero) at $a = 1$ and 2 when identical maps are compared. Note that for $\Delta^B$, due to the drawback discussed in the beginning of Section 3.2, its maximum value does not occur at $a = 0.5$ or 1.5. The second and improved metric $\Delta^S$, does not have this issue and retains the monotonicity as $a$ increases from 0 to 0.5.

Therefore for this sine wave example, in general both $\Delta^B$ and $\Delta^S$ are returning reasonable distance values, with $\Delta^S$ being preferred.

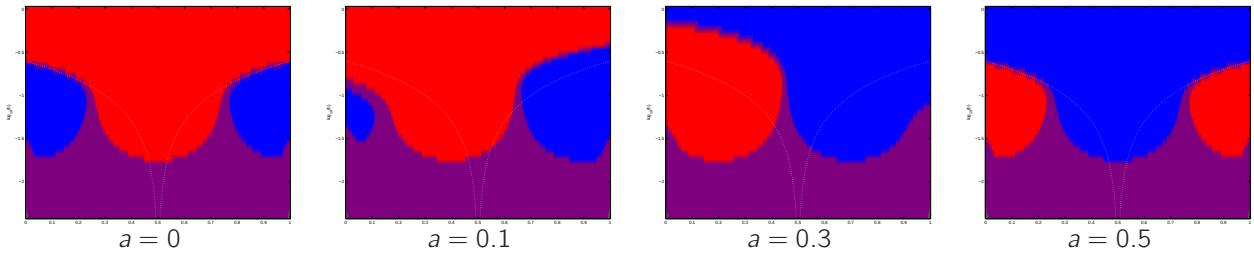
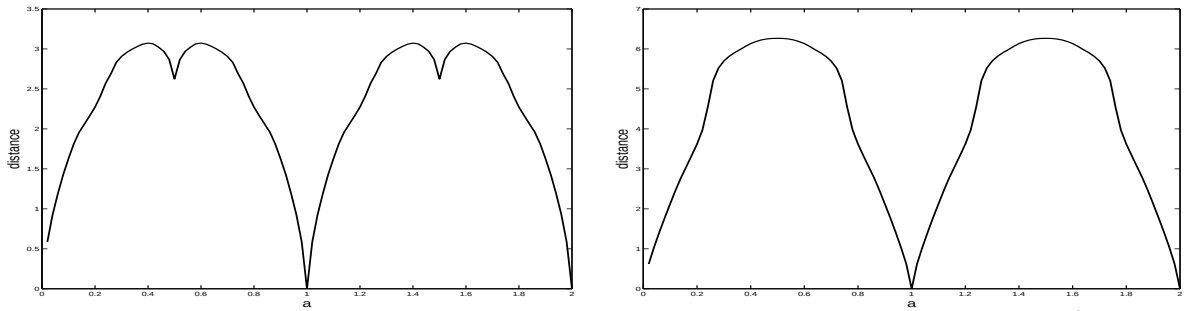



**Figure 2.** Oracle SiZer maps for $m_a(x) = 2\sin\{2\pi(x+a)\}$ for different values of $a$.




(a) calculated distances from $\Delta^B$ (b) calculated distances from $\Delta^S$

**Figure 3.** In (a): the $\Delta^B$ distances between the oracle SiZer maps of $m_0(x) = 2\sin\{2\pi(x)\}$ and $m_a(x) = 2\sin\{2\pi(x+a)\}$ for $a \in (0, 2]$. In (b): same as (a) but for $\Delta^S$.

## 4.2. Experiment 2

In the second experiment we are interested in calculating the distances between the oracle SiZer map and the SiZer maps obtained by the original, robust and spline SiZers reviewed in Section 2. The regression function is a polynomial $m_{\text{poly}}(x) = 1 - 48x + 218x^2 - 315x^3 + 145x^4$ and three variance functions are considered:

$$\begin{aligned} \text{constant:} \quad & \sigma_1(x) = 0.238, \\ \text{linear:} \quad & \sigma_2(x) = 0.369(2.5 - 2x), \\ \text{quadratic:} \quad & \sigma_3(x) = 0.135(-4x^2 + 4x + 0.5). \end{aligned}$$

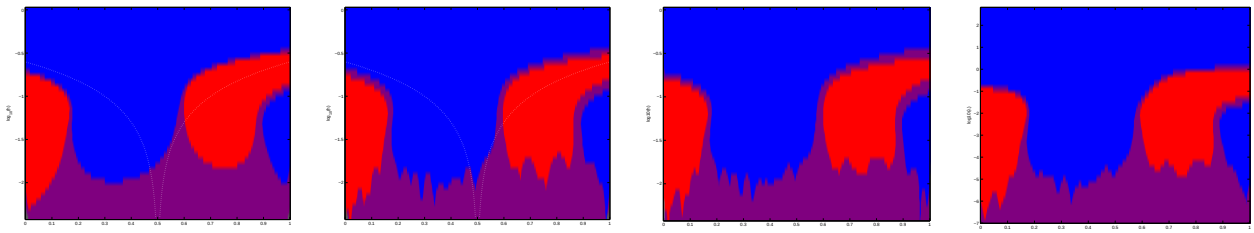
The multiplicative constants at the front are used to ensure the signal-to-noise ratio are the same for all variance functions.

For $m_{\text{poly}}(x)$ with $\sigma_1(x)$, a typical data set was first generated and the original, robust and spline SiZers were then applied to this data set to obtain their corresponding SiZer maps. These three SiZer maps, together with the oracle SiZer map, are displayed in Figure 4. Also provided in Figure 4 are the $\Delta^B$ and $\Delta^S$ values between the oracle SiZer map and any one of the three "non-oracle" SiZer maps. Visually the original and robust SiZer maps are quite similar, and their $\Delta^B$ and $\Delta^S$ values reflect this. The spline SiZer map seems to be more deviated from the oracle SiZer map and its $\Delta^B$ and $\Delta^S$ values are larger. For this particular example both $\Delta^B$ and $\Delta^S$ produced reasonable distance values that agree with a visual comparison of the SiZer maps.

The same procedure was performed for $\sigma_2(x)$ and $\sigma_3(x)$. Figures 5 and 6 summarize the results in a similar manner as in Figure 4.

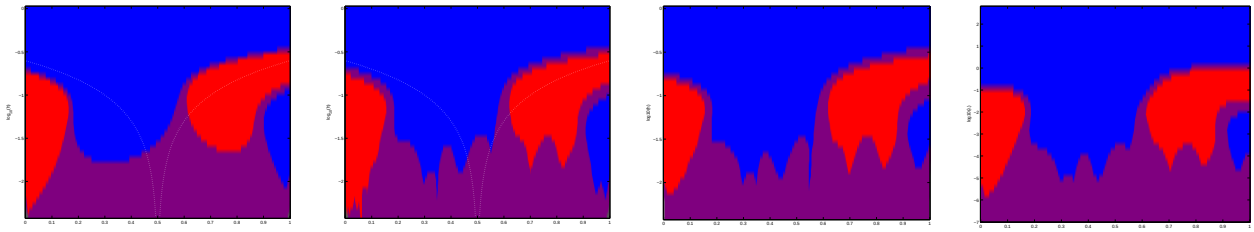
Lastly the whole set of experiments were repeated again, with 12 outliers randomly added to each simulated noisy data sets. These outliers were independently generated by a zero mean normal distribution with a standard deviation ten times larger than the corresponding "clean noise" standard deviation. The results are summarized in Figures 7 to 9 in a similar manner as before.

Overall from Figures 5 to 9, the $\Delta^B$ and $\Delta^S$ distances are seemed to be in agreement with human judgment. In particular, they suggest that, the robust SiZer maps are nearly as good as the original SiZer maps for cases without outliers, while these robust SiZer maps could be substantially better for cases with outliers.



oracle SiZer map | original SiZer map ($\Delta^B = 0.63, \Delta^S = 0.66$) | robust SiZer map ($\Delta^B = 0.60, \Delta^S = 0.60$) | spline SiZer map ($\Delta^B = 0.70, \Delta^S = 0.80$)

**Figure 4.** Different SiZer maps for $m_{\text{poly}}(x)$ with variance function $\sigma_1(x)$. The numbers below the original SiZer map are the $\Delta^B$ and $\Delta^S$ distances between the oracle and the original SiZer maps. Similarly for the numbers below the robust and spline SiZer maps.



oracle SiZer map | original SiZer map ($\Delta^B = 0.69, \Delta^S = 0.77$) | robust SiZer map ($\Delta^B = 0.77, \Delta^S = 0.86$) | spline SiZer map ($\Delta^B = 0.88, \Delta^S = 1.02$)

**Figure 5.** Similar to Figure 4 but for variance function $\sigma_2(x)$.

# 5. Concluding Remarks

In this article two metrics for SiZer map comparison are proposed. The first metric, $\Delta^B$, is a direct extension of the binary image metric proposed by Baddeley (1992). It is simple, but suffers from a minor issue as it ignores the meaning and ordering of the SiZer colors. To overcome this issue, a second metric $\Delta^S$ is proposed. Results from numerical experiments suggest that both $\Delta^B$ and $\Delta^S$ produce distances that are generally agree with human visual judgment, with $\Delta^S$ being preferred. Codes for calculating both metrics are available at the authors' website. Overall $\Delta^S$ is our recommended choice for SiZer map comparison.

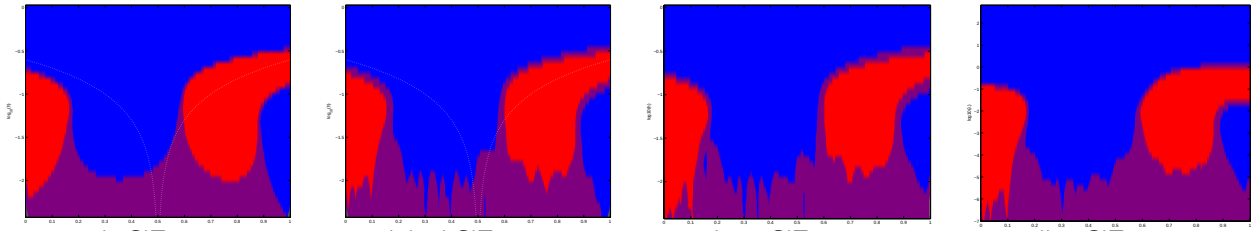



**Figure 6.** Similar to Figure 4 but for variance function $\sigma_3(x)$.

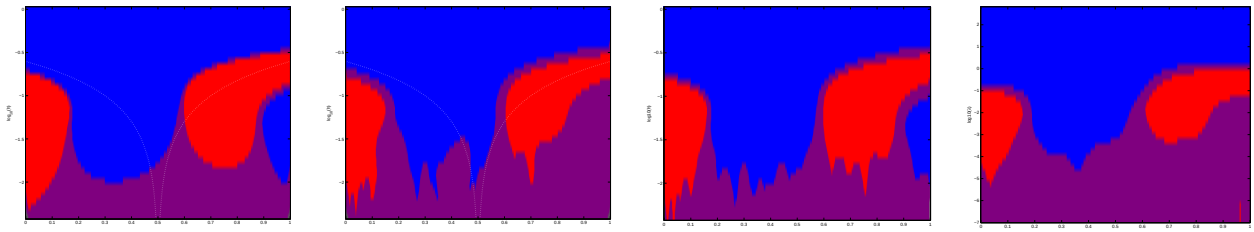



**Figure 7.** Similar to Figure 4 but with outliers.

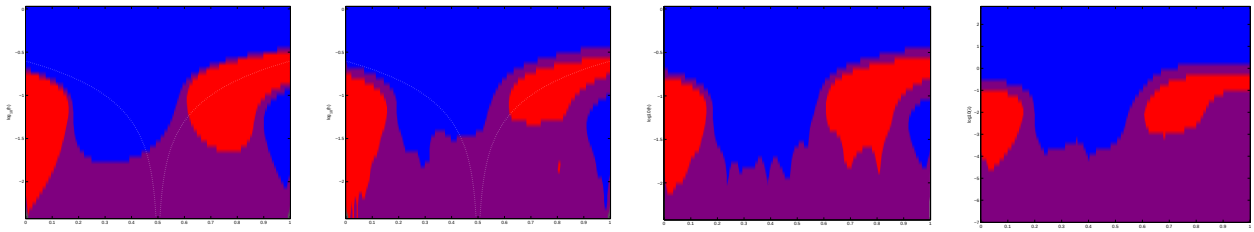



**Figure 8.** Similar to Figure 4 but for variance function $\sigma_2(x)$ and with outliers.

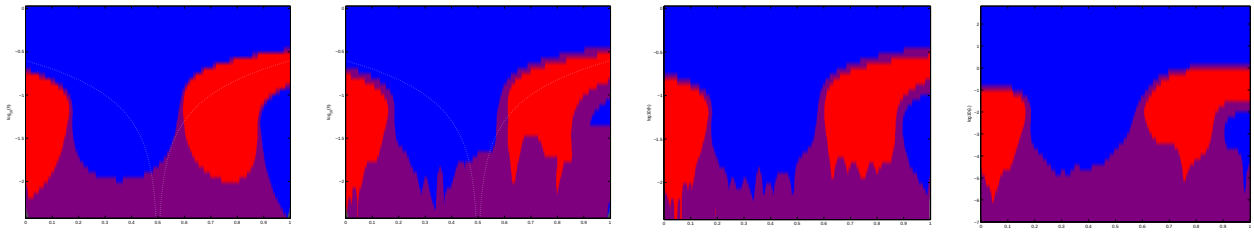



**Figure 9.** Similar to Figure 4 but for variance function $\sigma_3(x)$ and with outliers.

# Acknowledgements

The work of Hannig was supported in part by the National Science Foundation under Grants 1007543 and 1016441. The work of Lee was supported in part by the National Science Foundation under Grants 1007520, 1209226 and 1209232. The authors thank anonymous referees and the associate editor for helpful comments.


# References


Baddeley, AJ (1992), 'An error metric for binary images,' in Forstner, W & Ruwiedel, S (eds.), *Robust Computer Vision: Quality of Vision Algorithms*, Karlsruhe: Wichmann, pp. 59–78.

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

Chaudhuri, P & Marron, JS (2000), 'Scale space view of curve estimation,' *The Annals of Statistics*, **28**, pp. 408–428.

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

Erästö, P & Holmström, L (2007), 'Bayesian analysis of features in a scatter plot with dependent observations and errors in predictors,' *Journal of Statistical Computation and Simulation*, **77**, pp. 421–434.

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

Ganguli, B & Wand, MP (2004), 'Feature significance in geostatistics,' *Journal of Computational and Graphical Statistics*, **13**, pp. 954–973.

Ganguli, B & Wand, MP (2007), 'Feature significance in generalized additive models,' *Statistics and Computing*, **17**, pp. 179–192.

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

Godtliebsen, F & Oigard, TA (2005), 'A visual display device for significant features in complicated signals,' *Computational Statistics and Data Analysis*, **48**, pp. 317–343.

Gonzalez-Manteiga, W, Martinez-Miranda, MD & Raya-Miranda, R (2008), 'SiZer Map for inference with additive models,' *Statistics and Computing*, **18**, pp. 297–312.

Hannig, J & Lee, TCM (2006), 'Robust SiZer for exploration of regression structures and outlier detection,' *Journal of Computational and Graphical Statistics*, **15**, pp. 101–117.

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

Holmstrom, L & Pasanen, L (2012), 'Bayesian Scale Space Analysis of Differences in Images,' *Technometrics*, **54**, pp. 16–29.

Huber, PJ (1981), *Robust Statistics*, John Wiley & Sons, New York.

Kim, CS & Marron, JS (2006), 'SiZer for jump detection,' *Journal of Nonparametric Statistics*, **18**, pp. 13–20.

Marron, JS & de Una-Alvarez, J (2004), 'SiZer for length biased, censored density and hazard estimation,' *Journal of Statistical Planning and Inference*, **121**, pp. 149–161.

Marron, JS & Zhang, JT (2005), 'SiZer for smoothing splines,' *Computational Statistics*, **20**, pp. 481–502.

Oigard, TA, Rue, H & Godtliebsen, F (2006), 'Bayesian multiscale analysis for time series data,' *Computational Statistics and Data Analysis*, **51**, pp. 1719–1730.

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

Olsen, LR, Sorbye, SH & Godtliebsen, F (2008b), 'A scale-space approach for detecting non-stationarities in time series,' *Scandinavian Journal of Statistics*, **35**, pp. 119–138.

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

Park, C, Hannig, J & Kang, KH (2009), 'Improved SiZer for time series,' *Statistica Sinica*, **19**, pp. 1511 – 1530.

Park, C & Kang, KH (2008), 'SiZer analysis for the comparison of regression curves,' *Computational Statistics and Data Analysis*, **52**, pp. 3954–3970.

Park, C, Lee, TCM & Hannig, J (2010), 'Multiscale exploratory analysis of regression quantiles using quantile SiZer,' *Journal of Computational and Graphical Statistics*, **19**, pp. 497–513.

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

Park, C, Vaughan, A, Hannig, J & Kang, KH (2009), 'SiZer analysis for the comparison of time series,' *Journal of Statistical Planning and Inference*, **139**, pp. 3974–3988.

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

Silverman, B (1984), 'Spline smoothing: the equivalent kernel method,' *The Annals of Statistics*, **12**, pp. 898–916.

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

Wilson, DL, Baddeley, AJ & Owens, RA (1997), 'A new metric for grey-scale image comparison,' *International Journal of Computer Vision*, **24**, pp. 5–17.