Water Resource Surveillance for the Salton Sea in California By Adaptive Sequential Monitoring of Its Landsat Images

December 17, 2022

Abstract

Gradual loss of water resource in the Salton Sea has got much attention from researchers recently for its damage to the local environment and ecosystems for human beings, animals and plants. To monitor the water resource of the lake, researchers usually obtain certain water resource indices manually from databases such as the satellite images of the region. In this paper, we develop a new method to monitor the area of the Salton Sea automatically. By this method, the lake boundary is first segmented from each satellite image by an image segmentation procedure, and then its area is computed by a numerical algorithm. The sequence of lake areas computed from satellite images taken at different time points is then monitored by a control chart from the statistical process control literature. Because the lake area changes gradually over time, the control chart designed for detecting process mean drifts is used here for the water resource surveillance application.

Key Words: Control chart; Drift detection; Image segmentation; Local smoothing procedure; Satellite images; Statistical process control.

1 Introduction

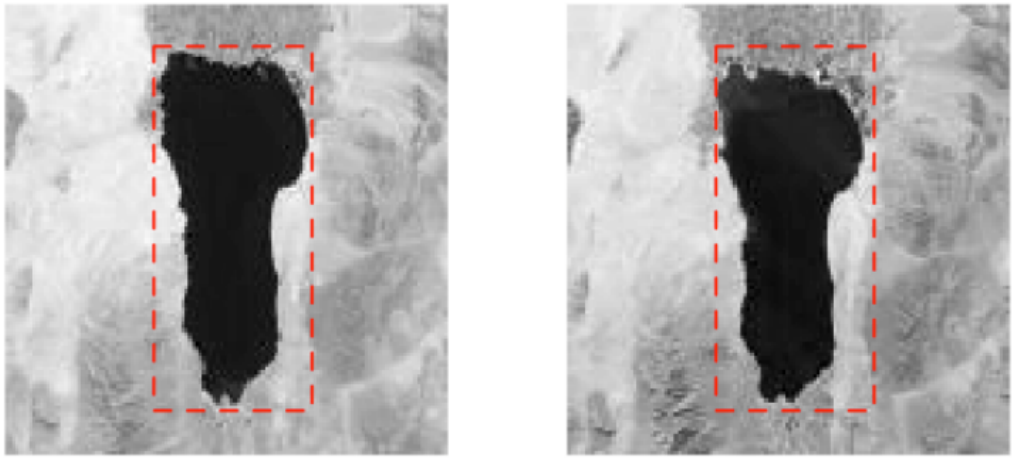
Salton Sea is the largest inland lake located at the southern end of California. See Figure 1 for a map of the lake and its surrounding areas. Because of the rapid municipal growth, increasing water resource demand, and fast evaporation due to the desert weather condition in the surrounding areas, the Salton Sea is in a serious danger of shrinkage (Cohen and Hyun 2006). A direct consequence of

the lake shrinkage is the damage of the local ecosystems for animals and plants (Shuford et al. 2002, 2004; Tiffany et al. 2007) and the worsening of the local environment for human health (Johnston et al. 2019; Jones and Fleck 2020). Therefore, effective sequential monitoring of the lake surface over time is important, which is the focus of this paper.

To monitor the lake surface of the Salton Sea, satellite images are routinely used. Since 1972, the Landsat project of NASA has launched 8 satellites to monitor the Earth surface. The current satellite (i.e., Landsat 8) can deliver an image of a given region (e.g., the Salton Sea region) every 16 days (USGS 2019). Therefore, there is a sequence of images for the Salton Sea region taken by the Landsat satellites at different times in the rich imagery database of the Landsat project of NASA. This sequence can be used for monitoring the Earth surface and resource in that region. Figure 2 presents two satellite images of the Salton Sea area taken on 04/13/1984 and 11/02/2011, respectively. By comparing the two images carefully, we can see some differences between the two images. For instance, the upper portion of the lake boundary moved downward quite dramatically after 27 years, and the lake became narrower as well.

Figure 1: Map of the Salton Sea area.

Figure 2: Two satellite images of the Salton Sea area taken on 04/13/1984 (left) and 11/02/2011 (right). To show the difference of the lake in the two images, a dashed rectangle of the same size is put around the lake in each image.



In the literature, some individual satellite images have been used to study the characteristics (e.g., sulfide and oxygen concentration) of the Salton Sea, after the image data are processed manually (Ma et al. 2020; Tiffany et al. 2007). So far, we have not found any existing research to sequentially monitor the image sequence of the Salton Sea region automatically using a computer algorithm based on a properly designed statistical method, such as control charts in the statistical process control (SPC) literature (Qiu 2014). This paper aims to fill the gap by making the following two main contributions. First, a computer algorithm is developed to compute the lake area from a satellite image automatically based on an effective boundary curve estimation procedure. Second, a well-designed control chart is used for sequentially monitoring the sequence of lake areas computed from the satellite images, and a signal is given by the chart when a systematic change in the lake area is detected. It should be pointed out that, besides the Salton Sea application, the proposed method should be useful for online monitoring of the areas of other lakes or other Earth objects based on their satellite images as long as their boundary curves meet some regularity conditions given in Section 2.1 below.

The rest of the paper is organized as follows. In Section 2, the computer algorithm for automatic computation of the lake area from a satellite image is described in detail. Then, sequential monitoring of the lake areas computed from the satellite image sequence is presented in Section 3. Some remarks conclude the paper in Section 4.

2 Compute the Area of the Salton Sea From a Satellite Image

This section describes the proposed computer algorithm for computing the area of the Salton Sea automatically from a satellite image. The description is organized in two parts. In Section 2.1, the lake boundary curve is first approximated by a set of discrete points obtained by a boundary curve detection procedure, and then an estimate of the lake boundary curve is obtained by a local smoothing procedure in the polar coordinating system. In Section 2.2, lake area estimation based on the estimated lake boundary curve is discussed.

2.1 Estimation of the lake boundary curve

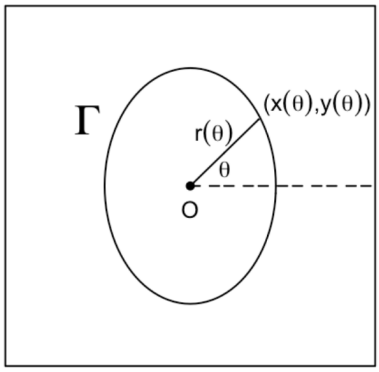
Boundary curve detection. To estimate the lake boundary curve automatically from a satellite image, there are two types of relevant image processing methods. One is related to edge detection, and the other is related to image segmentation (Gonzalez and Woods 2002; Qiu 2005). Edge detection methods aim to locate pixels at which the image intensity surface has jumps. Most edge detectors are based on estimation of the first-order or second-order directional derivatives of the image intensity function, and their detected edge pixels are usually scattered in the whole design space and may not form closed curves (e.g., Canny 1986; Sun and Qiu 2007). Image segmentation is mainly for locating the boundary of one or more objects in an observed image. Image segmentation methods are usually constructed based on the specific characteristics of observed images. The simplest method for image segmentation is by thresholding the intensity values of an observed image into two or more groups so that the observed image is divided into different segments accordingly. Other popular methods include the ones based on histograms of the image intensities, detected edge pixels, clustering, seeded region growing, and more (e.g., Adams and Bischof 1994; Guo et al. 2020).

From the satellite images shown in Figure 2, it can be seen that the boundary curve of the Salton Sea, denoted as Γ, is a closed curve and it can have the following expression in the polar coordinate system with respect to a given point of origin (or pole) O:

$$\begin{cases} x(\theta) = r(\theta)\cos(\theta) \\ y(\theta) = r(\theta)\sin(\theta), \end{cases} \tag{1}$$

where $(x(\theta), y(\theta))$ is a point on Γ, $\theta \in [0, 2\pi)$ is the angle formed by the line segment from the pole O to $(x(\theta), y(\theta))$ and the positive direction of the x-axis, $r(\theta) > 0$ is the Euclidean distance from $(x(\theta), y(\theta))$ to the pole. A demonstration of the boundary curve and its expression in the polar coordinate system is given in Figure 3. It should be pointed out that the boundary curve Γ may not have the expression (1) in some other applications when its shape is complicated. In such cases, one solution is to divide the related observed image into several sub-images first so that the part of the boundary curve in each sub-image would have an expression similar to (1), and then the estimates of different parts of Γ obtained from different sub-images are combined as an estimate of the entire Γ. Because this paper focuses on the Salton Sea application only, such more general cases will be left for our future research.

Figure 3: Demonstration of the boundary curve Γ in the polar coordinate system.



For applications in which the boundary curve Γ has the expression (1), Qiu and Sun (2009) suggested an image segmentation approach to estimate Γ in the context of computing the gene expression levels based on an observed microarray image. They showed that their suggested method was more effective than some alternative image segmentation approaches mentioned above, and it could be shown by similar auguments to those in Qiu and Sun (2007) that their estimated Γ would be statistical consistent under some regularity conditions in the sense that the estimated Γ would converge to the true Γ when the image resolution increases. For these reasons, this method is adopted in the current paper and briefly described below.

Assume that an observed satellite image is described by the following model:

$$z_{ij} = f(x_i, y_j) + \epsilon_{ij}, \quad i = 1, 2, \cdots, n_x, j = 1, 2, \cdots, n_y,$$

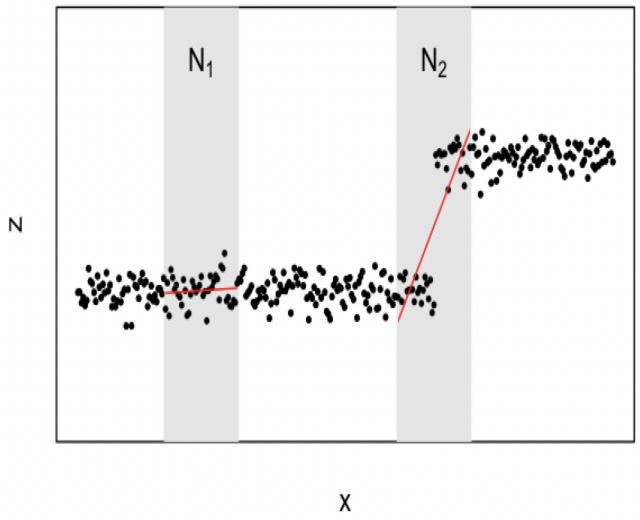
where z_{ij} is the observed image intensity at the (i,j)th pixel (x_i, y_j), $f(x_i, y_j)$ is the true image intensity at that pixel, and ϵ_{ij} is the pointwise noise with mean 0 and variance σ^2. Without loss of generality, it is further assumed that the pixels are equally spaced in the design region $\Omega = [0,1] \times [0,1]$. Namely, $(x_i, y_j) = (i/n_x, j/n_y)$, for each i and j. To estimate the boundary curve Γ, Qiu and Sun (2009) first considered the following local linear kernel (LLK) smoothing procedure in a neighborhood of a given point $(x, y) \in \Omega$:

$$\min_{a,b,c} \sum_{(x_i,y_j)\in\mathcal{N}(x,y)} \left[z_{ij} - a - b(x_j - x) - c(y_j - y)\right]^2 K\left((x_i - x)/h_x, (y_j - y)/h_y\right), \tag{2}$$

where $\mathcal{N}(x,y) = \{(x^{'}, y^{'}) : \sqrt{((x^{'} - x)/h_x)^2 + ((y^{'} - y)/h_y)^2} \leq 1\}$ is an elliptical neighborhood of the point (x,y) with the bandwidths h_x and h_y in the $x-$ and $y-$axis, respectively, and $K(u,v)$ is a density kernel function with the support $\{(u,v) : u^2 + v^2 \leq 1\}$. The solutions to (a,b,c) of the minimization problem (2) are denoted as $\widehat{a}(x,y)$, $\widehat{b}(x,y)$, $\widehat{c}(x,y)$, respectively. Then, the LLK

estimator of $f(x,y)$ is defined to be $\widehat{a}(x,y)$, and the LLK estimator of the gradient direction of $f(\cdot,\cdot)$ at (x,y) (i.e., $(f'_x(x,y), f'_y(x,y))'$) is defined to be $(\widehat{b}(x,y), \widehat{c}(x,y))'$.

Figure 4: Illustration of the local linear kernel smoothing procedure in one-dimensional cases, where N_1 and N_2 denote two neighborhoods of two given points and the two red lines are fitted in the two neighborhoods, respectively, for estimating the values of the mean function at the two given points.



By using the LLK smoothing procedure (2), a plane is fitted in the neighborhood $\mathcal{N}(x,y)$ of the given point $(x,y) \in \Omega$ to approximate the true image intensity surface in that neighborhood, and the LLK estimator $\widehat{a}(x,y)$ of $f(x,y)$ is actually a weighted average of the observed image intensities in the neighborhood with the weights with the weights determined by the kernel function $K(u,v)$. Figure 4 illustrates the LLK smoothing procedure in one-dimensional cases, where N_1 and N_2 denote two neighborhoods of two given points and the two red lines are fitted in the two neighborhoods, respectively, for estimating the values of the mean function at the two given points. From the plot, it can be seen that if the mean function has a jump at a given point (i.e., the case with N_2), the fitted line would be steep. If the mean function is continuous around a given point (i.e., the case with N_1), then the fitted line would be relatively flat. Similarly, in the current two-dimensional problem, if there is an edge curve in $\mathcal{N}(x,y)$, then the searched local plane would be steep. Otherwise, it would be relatively flat. Note that the steepness of a plane is measured by the magnitude of its gradient. So, the magnitude of the estimated gradient $(\widehat{b}(x,y), \widehat{c}(x,y))'$ can be used to judge whether the given point (x,y) is close to the boundary curve Γ or not. Now, let us consider a ray starting from the pole O in the direction of θ in the polar coordinate system. Then, any point on the ray has

the expression (r, θ), for $r \geq 0$, which corresponds to the point $(r\cos(\theta), r\sin(\theta))$ in the Cartesian coordinate system. Based on the above intuition, a natural boundary detection criterion is

$$M(r, \theta) = \widehat{b}^2(r\cos(\theta), r\sin(\theta)) + \widehat{c}^2(r\cos(\theta), r\sin(\theta)),$$

which is the squared magnitude of the estimated gradient mentioned above in the polar coordinate system. Let $\{\theta_l \in [0, 2\pi), l = 1, 2, \ldots, m\}$ be a sequence of equally spaced θ values at which Γ needs to be estimated, and $\widetilde{r}(\theta_l) = \max_{0 \leq r \leq R_{\theta_l}} M(r, \theta_l)$, where R_{θ_l} is the maximum length of the ray in the direction θ_l within the design space Ω. Then, the boundary curve Γ can be approximated by the following point set in the polar coordinate system:

$$\widetilde{\Gamma} = \{(\theta_l, \widetilde{r}(\theta_l)), l = 1, 2, \ldots, m\}.$$

Post-detection smoothing. The detected pointset $\widetilde{\Gamma}$ for approximating the boundary curve Γ is a set of discrete points scattered around the true boundary curve Γ. To remove some randomness in this pointset and obtain an estimate of Γ, we consider the following LLK smoothing procedure in the polar coordinate system for a given $\theta \in [0, 2\pi]$:

$$\min_{a,b} \sum_{\theta_l \in [\theta - h_\theta, \theta + h_\theta]} [\widetilde{r}(\theta_l) - a - b(\theta_l - \theta)]^2 K_\theta\left((\theta_l - \theta)/h_\theta\right), \tag{3}$$

where $h_\theta > 0$ is a bandwidth parameter, and K_θ is a density kernel function with the support $[-1, 1]$. The solution to a of the procedure (3) is denoted as $\widehat{r}(\theta)$. Then, the LLK estimate of Γ is $\widehat{\Gamma} = \{(\theta, \widehat{r}(\theta)) : \theta \in [0, 2\pi]\}$ in the polar coordinate system, or

$$\begin{cases} \widehat{x}(\theta) = \widehat{r}(\theta)\cos(\theta), \\ \widehat{y}(\theta) = \widehat{r}(\theta)\sin(\theta), \end{cases} \quad \text{for } \theta \in [0, 2\pi],$$

in the Cartesian coordinate system. See Figure 5 below for a demonstration.

Similar to the LLK procedure (2), the LLK estimate $\widehat{r}(\theta)$ obtained in (3) is a weighted average of $\{\widetilde{r}(\theta_l), l = 1, 2, \ldots, m\}$ in the neighborhood $[\theta - h_\theta, \theta + h_\theta]$. Because the boundary curve Γ is a closed curve, we assume that $r(0) = r(2\pi)$ and that $r(\theta)$ is a periodic function with a period of 2π. This periodic property has been used in (3). Namely, when computing the LLK estimate $\widehat{r}(\theta)$ in the boundary regions $[0, h_\theta]$ and $[2\pi - h_\theta, 2\pi]$, the values of $\widetilde{r}(\theta_l)$ for $\theta_l \in [2\pi - h_\theta, 2\pi]$ are also used as $\widetilde{r}(2\pi - \theta_l)$, and the values of $\widetilde{r}(\theta_l)$ for $\theta_l \in [0, h_\theta]$ are also used as $\widetilde{r}(2\pi + \theta_l)$.

Bandwidth selection. In the minimization problems (2) and (3), there are three bandwidth parameters h_x, h_y and h_θ to choose in advance. In this paper, we use the cross-validation (CV)

procedures described below. For choosing h_x and h_y, we first define

$$\Omega_c = \{(x,y) \in \Omega : \sqrt{[(x - \widetilde{r}(\theta_l)\sin(\theta_l))/h_x]^2 + [(y - \widetilde{r}(\theta_l)\cos(\theta_l))/h_y]^2} > 1, l = 1, 2, \cdots, m\}.$$

Then, Ω_c denotes all points in the design space Ω that their distance to any detected boundary points in $\widetilde{\Gamma}$ is at least h_x in the x-axis direction and at least h_y in the y-axis direction. Then, h_x and h_y are chosen by minimizing the following CV score:

$$CV(h_x, h_y) = \sum_{(x_i,y_j)\in\Omega_c} \left[z_{ij} - \widehat{f}_{-i,-j}(x_i, y_j)\right]^2 /(N_c - 1),$$

where N_c denotes the number of pixels in Ω_c, and $\widehat{f}_{-i,-j}$ is the estimator of $f(x_i, y_j)$ by (2) after the (i,j)-th observation is excluded from computation. For choosing h_θ, we define the following CV score in a similar way:

$$CV(h_\theta) = \sum_{l=1}^{m} \left[\widetilde{r}(\theta_l) - \widehat{r}_{-l}(\theta_l)\right]^2 /m,$$

where $\widehat{r}_{-l}(\theta_l)$ is the estimator of $r(\theta_l)$ by (3) after the point $(\theta_l, \widetilde{r}(\theta_l))$ is excluded. Then, the three parameters are chosen by minimizing the respective CV scores.

2.2 Lake area estimation

After the lake boundary curve Γ is estimated by $\widehat{\Gamma}$, the lake area can be computed as follows. For simplicity, let us assume that $[0, 2\pi)$ is partitioned by the sequence $\{\theta_l, l = 1, 2, \ldots, m\}$. Then, the lake area can be approximated by the summation of the areas of all neighboring triangles formed by the pole O of the polar coordinate system and the consecutive rays from O to $\widehat{\Gamma}$ at the directions $\{\theta_l, l = 1, 2, \ldots, m\}$. See Figure 5 for a demonstration. For the triangle formed by the rays at the directions θ_{l-1} and θ_l, for $l = 2, , 3 \ldots, m$, it can be checked that its area is $\widehat{r}(\theta_l)\widehat{r}(\theta_{l-1})\sin(\theta_l - \theta_{l-1})/2$. For the triangle formed by the rays at the directions θ_m and θ_1, its area is $\widehat{r}(\theta_m)\widehat{r}(\theta_1)\sin(\theta_m - \theta_1)/2$. Thus, the lake area can be estimated by

$$\widehat{S} = \sum_{l=2}^{m} \widehat{r}(\theta_l)\widehat{r}(\theta_{l-1})\sin(\theta_l - \theta_{l-1})/2 + \widehat{r}(\theta_m)\widehat{r}(\theta_1)\sin(\theta_m - \theta_1)/2.$$

For the satellite image observed on 04/13/1984 that is shown in the left panel of Figure 2, it has 128×128 pixels. The entire image covers a region of the size 75×75 square kilometers. The design space of the image is then standardized into $[0, 1] \times [0, 1]$. When m is chosen to be 200 (i.e., $\{\theta_l = 2\pi l/200, l = 1, 2, \ldots, 200\}$) and the pole O is chosen to be at (0.5,0.6) in the design space, the

Figure 5: Illustration of lake area estimation. Left panel: Dot points denote the detected boundary curve $\widetilde{\Gamma}$. Right panel: Dot points denote the estimated boundary curve $\widehat{\Gamma}$ after post-detection smoothing. In each plot, the thin curve denotes the true boundary curve Γ. The lake area is estimated by the summation of the areas of all consecutive triangles from the pole O to the detected boundary points in $\widehat{\Gamma}$ as shown in the right panel.

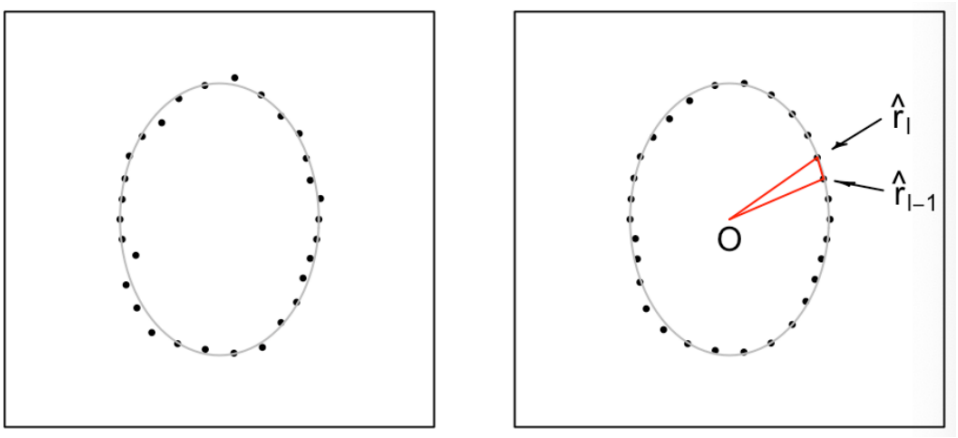
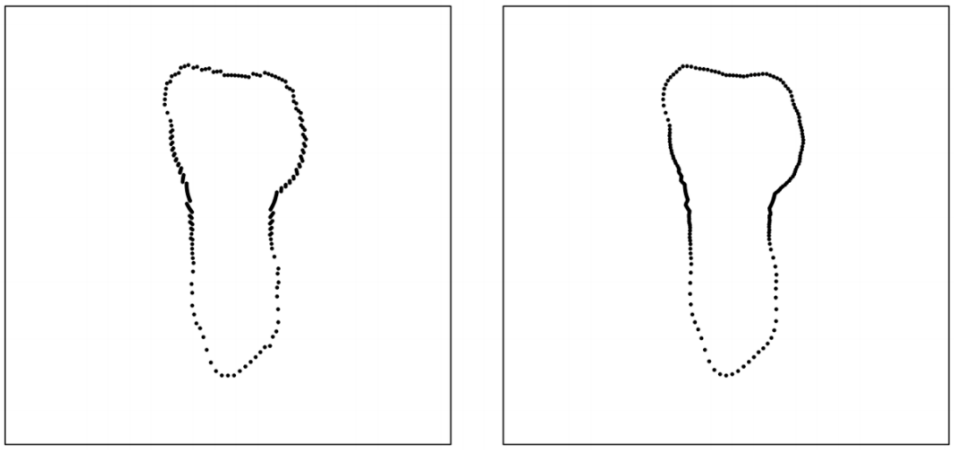


Table 1: Calculated areas of the Salton Sea in the standardized unit from the satellite image observed on 04/13/1984 when m changes among 50, 100, 200, 500 and 1000. In the calculation, the estimated boundary curves with and without using post-smoothing are both considered

	$m = 50$	100	200	500	1,000
Without Post-Smoothing	0.1661	0.1686	0.1692	0.1697	0.1697
With Post-Smoothing	0.1645	0.1681	0.1691	0.1697	0.1697

detected boundary curve by the procedure described above is shown in Figure 6. The estimated lake area is 0.1692 in the standardized unit, or $0.1692 \times 75^2 = 951.75$ square kilometers. In this example, we tried several different values of m, and the calculated lake area values are shown in Table 1 in the standardized unit. From the table, it can be seen that i) the calculated lake area will stabilize when m gets larger and larger and its value would not change much after $m \geq 500$, and ii) the calculated lake areas with and without considering post-smoothing of the detected boundary curve are almost the same when $m \geq 500$ in this example. So, based on these results, we will choose $m = 500$ in all subsequent analyses.

Figure 6: Left panel: detected boundary curve $\widetilde{\Gamma}$ from the satellite image of the Salton Sea observed on 04/13/1984 when $m = 200$. Right panel: estimated boundary curve $\widehat{\Gamma}$ after post-detection smoothing.



3 Sequential Monitoring of the Areas of Salton Sea Over Time

A major statistical tool to monitor sequential processes is control charts in the SPC literature (cf., Qiu 2014), which aim to decide whether the sequential process under monitoring is in-control (IC) or out-of-control (OC) each time after a new batch of process observations is collected. In the Salton Sea example, its area can only change gradually over time and would not have step mean shifts in a normal circumstance. So, for monitoring the lake area over time, instead of using a control chart designed for detecting step mean shifts, it should be more appropriate to consider a chart designed for detecting mean drifts. In the SPC literature, there has been much existing discussion on detecting mean drifts. For example, Gan (1991) and Gan (1992) investigated the performance of the conventional EWMA and CUSUM chart under linear drift, respectively. Fahmy and Elsayed (2006) proposed a statistic based on the deviation between the target mean and the expected mean of the process to detect linear trends in the process mean. Zhou et al. (2010) suggested a control chart for monitoring a process with patterned mean or variance changes based on the combination of the generalized likelihood ratio statistic and EWMA. Zou et al (2009) compared several different control charts for detecting linear drifts, including the conventional CUSUM and EWMA charts, the generalized EWMA chart by Han and Tsung (2004), and two control charts based on the generalized likelihood ratio statistic under the linear drift and step shift alternatives. In this paper, the GLR-L chart suggested in Zou et al. (2009) for monitoring a process with linear

mean drift will be used for the following reasons. First, the GLR-L chart has the advantage that the size of the mean change does not need to be pre-specified in order to design the chart properly. The other methods mentioned above do not have this property. Second, according to Zou et al. (2009), the GLR-L chart has the best overall performance among all the charts considered in their numerical comparisons. Third, like other Change-Point Detection (CPD) charts, the GLR-L chart can provide good estimators of the drift time and the IC and OC distribution parameters at the same time when it gives a signal of process mean drift. However, the original GLR-L chart in Zou et al. (2009) was designed for cases with equally spaced observation times. In the current problem, however, the observation times are unequally spaced because some observed images cannot be used because of poor quality due mainly to bad weather. Thus, we need to modify the original GLR-L chart to accomodate unequally spaced observation times as follows.

Let the lake area computed at time t_n be $\widehat{S}_{t_n}$, for $n \geq 1$. Assume that $\widehat{S}_{t_n} \sim N(\mu_0, \sigma_0^2)$ if the process is in-control (IC) at t_n, where μ_0 and σ_0^2 are the IC mean and variance. If the process is out of control (OC) at time t_n and the related shift is a linear drift starting at $t_\tau < t_n$, then it is assumed that $\widehat{S}_{t_n} \sim N(\mu_0 + \gamma(t_n - t_\tau)), \sigma_0^2)$, where γ is the slope of the drift. Let $e_{t_n} = (\widehat{S}_{t_n} - \mu_0)/\sigma_0$ be the standardized value of $\widehat{S}_{t_n}$, for $n \geq 1$. Then, we can obtain the following logarithm of the likelihood ratio

$$R(t_\tau, \gamma | e_{t_i}, i = 1, \cdots, n) = 2 \ln \prod_{i=\tau+1}^{n} \left(\exp^{-[e_{t_i} - (t_i - t_\tau)\gamma]^2/2} \right) \Big/ \left(\exp^{-e_{t_i}^2/2} \right).$$

Then, the GLR-L charting statistic is

$$\begin{aligned} R_n &= \max_{0 \leq \tau < n, \gamma} R(t_\tau, \gamma | e_{t_i}, i = 1, \cdots, n) \\ &= \max_{0 \leq \tau < n} \max_{\gamma} R(t_\tau, \gamma | e_{t_i}, i = 1, \cdots, n) \\ &= \max_{0 \leq \tau < n} \left[\sum_{i=\tau+1}^{n} (t_i - t_\tau) e_{t_i} \right]^2 \Big/ \sum_{i=\tau+1}^{n} (t_i - t_\tau)^2. \end{aligned} \tag{4}$$

The GLR-L chart gives a signal if

$$R_n > h, \tag{5}$$

where $h > 0$ is a control limit.

To measure the performance of a control chart, such as the one defined in (4)-(5), when observation times are unequally spaced, we often use the IC average time to signal, denoted as ATS_0 and defined to be the average time from the beginning of online process monitoring to the signal time

by the chart in cases when the process under monitoring is IC, and the OC average time to signal, denoted as ATS_1 and defined to be the average time from the occurrence of a shift in the process distribution to the signal time of the chart after the process becomes OC. Usually, the value of ATS_0 is pre-specified, similar to the type-I error probability in the hypothesis testing setup. Then, a control chart performs better to detect a given shift if its ATS_1 value is smaller. See Qiu and Xiang (2014) for a related discussion.

We are now ready to monitor the Salton Sea area over time by sequentially monitoring its Landsat images, using the boundary curve estimation, lake area estimation, and sequential process monitoring procedures discussed in the previous parts of the paper. For this application, we consider the Landsat images of the Salton Sea area observed during a time period from April 13, 1984 to November 2, 2011. As discussed in Section 1, the Landsat satellite collects an image of this area every 16 days. However, some images collected during this time period cannot be used because of their poor quality due to bad weather, as pointed out earlier. After these images are excluded, a total of 334 images are included in this study, the lake areas computed as described in Section 2 are shown in the upper panel of Figure 7. From the plot, the first 200 observations of the lake area until February 10, 2002 (i.e., observations before the vertical line in the plot) are stable overall, although there are some ups and downs in the observations that reveal certain degree of data autocorrelation. In addition, the California government passed the Water Transfer Agreement in 2003 (cf., Taylor 2018), which would reduce Salton Sea inflow and thus affect the lake area. For these reasons, the first 200 observations of the lake area until February 10, 2002 are used as the IC data, and the remaining ones are used for process online monitoring. From the description of the GLR-L chart, it requires the process distribution to be normal and the process observations at different time points to be independent of each other. However, by applying the Shapiro and rank von Neumann ratio tests through the R functions **shapiro.test()** and **serialCorrelationTest()** for testing the normality and autocorrelation of the IC data, respectively, it is found that the normality assumption cannot be rejected, but the serial data correlation is significant, and the p-values of the two tests are 0.7456 and 2.2×10^{-16}, respectively. Therefore, we consider decorrelating the entire dataset (i.e., the data with all 334 observations) by the data decorrelation procedure discussed in Qiu et al. (2020). More specifically, the IC mean, variance and data correlation are first estimated from the IC data. Then, all 334 observations are decorrelated recursively based on the Cholesky decomposition of the estimated covariance matrices. The decorrelated data are shown in the lower panel of Figure

7. After data decorrelation, the R functions **shapiro.test()** and **serialCorrelationTest()** are applied to the IC data again, and the p-values are respectively 0.9503 and 0.2391. Thus, the serial data correlation seems to be mostly removed by the data decorrelation procedure.

Next, we apply the GLR-L chart defined in (4)-(5) to this dataset. The only parameter to choose for the GLR-L chart is the control limit h which is determined as follows. First, we fix the ATS_0 value to be 200×16 days, where the unit of 16 days is the time interval between two consecutive images collected by the Landsat satellite and 200 is the commonly used ATS_0 value in the statistical process control literature. The control limit h is then searched by a bootstrap procedure as follows. We first randomly select 1,000 observations from the 200 IC data points with replacement and 1,000 time intervals from the 199 time intervals between consecutive data points in the IC data with replacement. Then, these randomly selected observations and time intervals are used to form a sequential process for online process monitoring. The GLR-L chart with a given h value is then applied to this sequence, and a signal time can be recorded. The above bootstrap re-sampling procedure is repeated for 10,000 times, and the average of the 10,000 signal times is used for estimating the ATS_0 value. Then the h value is searched such that the estimated ATS_0 value equals the pre-specified ATS_0 value of 200×16 days. The GLR-L chart with the searched h value is shown in Figure 8. The first signal given by the GLR-L chart is at July 7, 2003 which is shown in the lower panel of Figure 7 as well. From Figure 7, we can see that the GLR-L chart can detect the mean drift in the lake area quite quickly. To further investigate the root cause of the detected mean drift, the observed satellite images of the Salton Sea area taken on April 13, 1984 and July 7, 2003 are shown in Figure 9. By comparing the two images carefully, it can be seen that the lake shape changed at certain places after 19 years. For instance, the southeast shore of the Salton Sea highlighted in the dashed ractangle in each image indeed shrank over time.

4 Concluding Remarks

Sequential monitoring of the area of the Salton Sea over time is important for protecting the local environment and minimizing the damage of the local ecosystems for animals and plants. To this end, traditional methods calculate certain indices of the lake from the satellite images and then study their trends manually. In this paper, we have developed a procedure for online monitoring of the area of the Salton Sea automatically using a computer algorithm, based on the observed

Figure 7: Upper panel: The calculated areas of the Salton Sea during April 13, 1984 and November 2, 2011. Lower panel: The decorrelated data of those in the upper panel. In each plot, the vertical solid line separates the IC data from the data for process monitoring. The dashed vertical line in the lower panel denotes the signal time of the GLR-L chart.

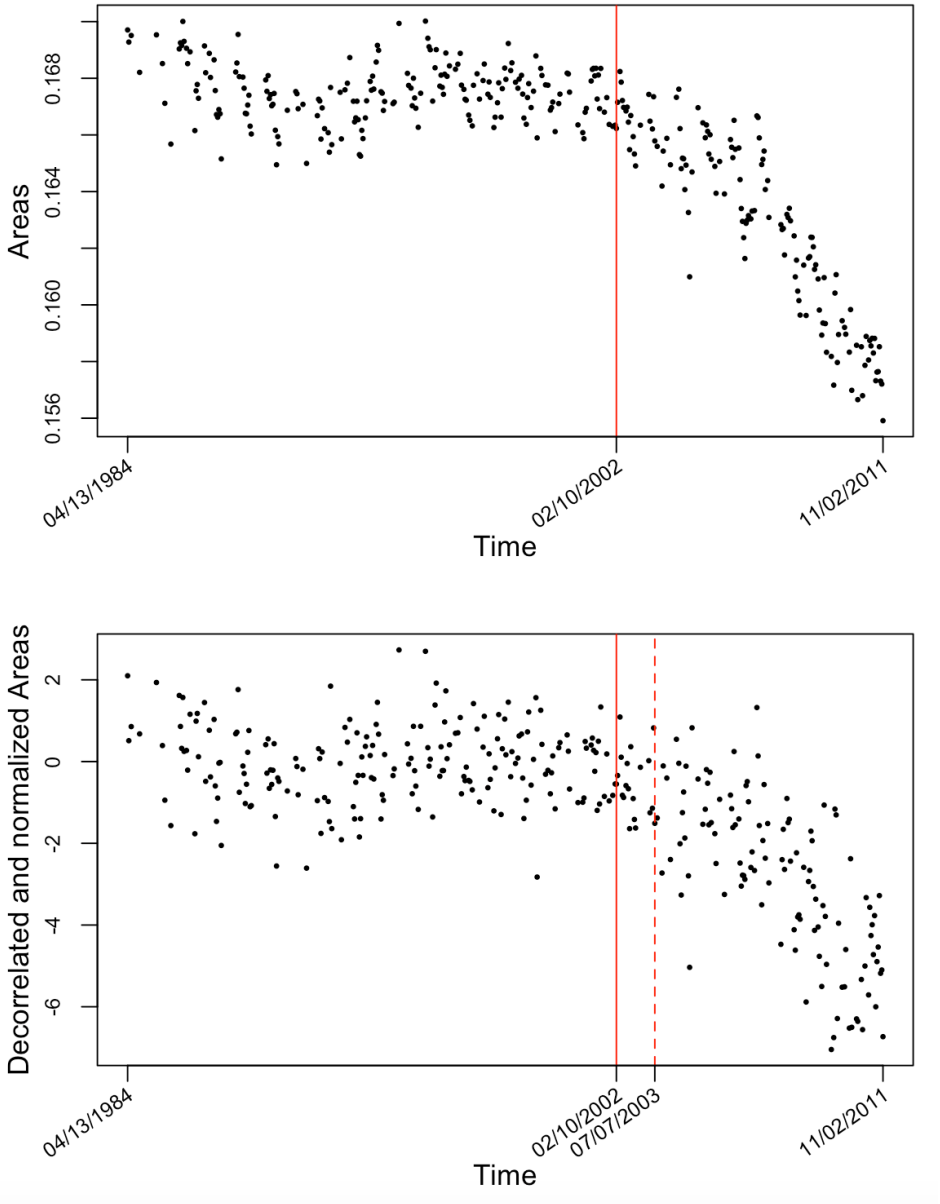
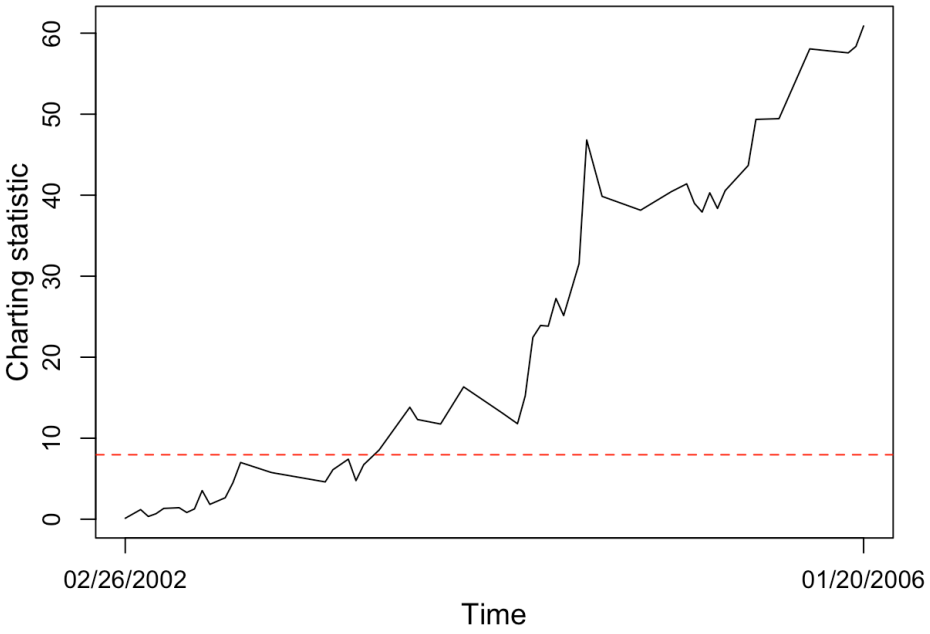
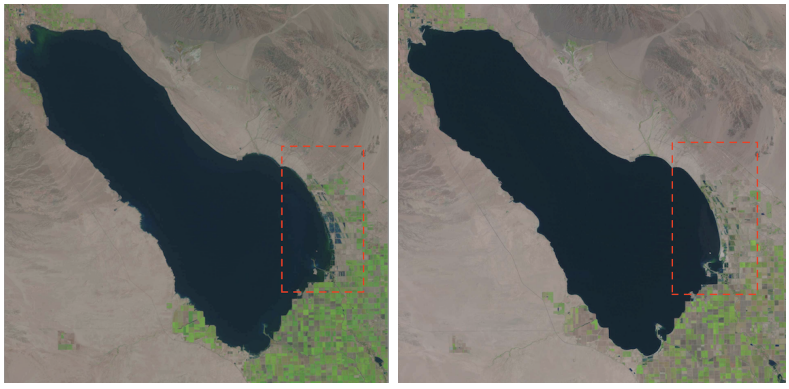


Figure 8: The GLR-L chart for monitoring the estimated areas of the Salton Sea during a time period between February 26, 2002 and January 20, 2006.



Landsat image sequence of the Salton Sea area. Our method consists of several steps, including the lake boundary curve detection, lake area estimation, and sequential monitoring of the estimated lake areas over time. It also has been shown in the paper that the proposed sequential monitoring procedure is effective for detecting a mean drift in the estimated lake areas. One limitation of our method is the potential assumption that the boundary curve of the lake has the expression (1). As pointed out in Section 2.1, in some other applications, this assumption could be violated when the boundary curve has a complex shape. One possible strategy to handle such a more general case is to divide the related image into several sub-images so that the part of the boundary curve in each

Figure 9: The satellite images of the Salton Sea observed on April 13, 1984 (left) and July 7, 2003 (right).



sub-image can be estimated by the proposed method here. Then, estimates of different parts of the boundary curve obtained from different sub-images can be combined as an estimate of the entire boundary curve. This more general approach will be studied systematically in our future research.

Acknowledgments: The authors thank the editor, the associate editor and two referees for their constructive comments and suggestions, which improved the quality of the paper greatly. This paper is supported in part by an NSF grant.

Data availability

The datasets analyzed during the current study are available at https://earthexplorer.usgs.gov.

References

Adams R, and Bischof L (1994), "Seeded region growing," *IEEE Transactions on PatternAnalysis and Machine Intelligence*, 16:641-647.

Canny J (1986), "A computational approach to edge detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 8:679-698.

Cohen MJ, and Hyun KH (2006), *Hazard: The Future of the Salton Sea with No Restoration Project*, Oakland, CA: Pacific Institute.

Fahmy HM, and Elsayed EA (2006), "Detection of linear trends in process mean," *International Journal of Production Research*, 44:487-504.

Gan FF (1991), "EWMA control chart under linear drift," *Journal of Statistical Computation and Simulation*, 38:181-200.

Gan FF (1992), "CUSUM control chart under linear drift," *The Statistician*, 41:71-84.

Gonzalez RC, and Woods RE (2002), *Digital Image Processing (2nd edition)*, New York: Prentice Hall.

Guo D, Pei Y, Zheng K, Yu H, Lu Y, and Wang S (2020), "Degraded image semantic segmentation with dense-gram networks," *IEEE Transactions on Image Processing*, 29:782-795.

Han D, and Tsung F (2004) "A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes," *The Annals of Statistics*, 32:316-339.

Johnston J, Razafy M, Lugo H, Olmedo L, and Farzan SF (2019), "The disappearing Salton Sea: a critical reflection on the emerging environmental threat of disappearing saline lakes and potential impacts on children's health," *Science of the Total Environment*, 663:804-817.

Jones BA, and Fleck J (2020), "Shrinking lakes, air pollution, and human health: Evidence from California's Salton Sea," *Science of the Total Environment*, 712:136490.

Ma J, Duan H, He L, Tiffany M, Cao Z, Qi T, Shen M, Biggs T, and Xu X (2020), "Spatiotemporal pattern of gypsum blooms in the Salton Sea, California, during 2000-2018," *International Journal of Applied Earth Observation and Geoinformation*, 89:102090.

Qiu P (2005), *Image Processing and Jump Regression Analysis,* New York: John Wiley & Sons.

Qiu P (2014), *Introduction to Statistical Process Control,* Boca Raton, FL: Chapman & Hall/CRC.

Qiu P, Li W, and Li J (2020), "A new process control chart for monitoring short-range serially correlated data," *Technometrics*, 62:71-83.

Qiu P, and Sun J (2007), "Local smoothing image segmentation for spotted microarray images," *Journal of the American Statistical Association,* 18:147-164.

Qiu P, and Sun J (2009), "Using conventional edge detectors and post-smoothing for segmentation of spotted microarray images," *Journal of Computational and Graphical Statistics,* 18:147-164.

Qiu P, and Xiang, D. (2014), "Univariate dynamic screening system: An approach for identifying individuals with irregular longitudinal behavior," *Technometrics*, 56:248-260.

Shuford WD, Warnock N, Molina KC, and Sturm K (2002), "The Salton Sea as critical habitat to migratory and resident waterbirds," *Hydrobiologia*, 473:255-274

Shuford WD, Warnock N, and McKernan RL (2004), "Patterns of shorebird use of the Salton Sea and adjacent Imperial Valley, California," *Studies in Avian Biology*, 27:61-77.

Taylor M (2018), *The Salton Sea: A Status Update*, Sacramento, CA: Legislative Analyst's Office.

Tiffany MA, Ustin SL, and Hurlbert SH (2007), "Sulfide irruptions and gypsum blooms in the Salton Sea as detected by satellite imagery, 1979-2006," *Lake and Reservoir Management*, 23:637-652.

USGS (2019), *Landsat 8 (L8) Data Users Handbook,* Sioux Falls, SD: U.S. Geological Survey.

Zhou Q, Luo YZ, and Wang Z (2010), "A control chart based on likelihood ratio test for detecting patterned mean and variance shifts," *Computational Statistics and Data Analysis*, 54:1634-1645.

Zou C, Liu Y, and Wang Z (2009), "Comparisons of control schemes for monitoring the mean of processes subject to drifts," *Metrika*, 70:141-163.