


## Long-range persistence in sea surface temperature off the coast of central California

Journal of Ocean and Climate  
Volume 9: 1–13  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1759313118791113  
journals.sagepub.com/home/ocs  


Laurence C Breaker<sup>1,2</sup> 

### Abstract

We estimate long-range persistence in ocean surface temperature off the coast of central California, a region where similar observations have not been made. The database consists of 20-year records of daily sea surface temperature from three locations: Pacific Grove and Granite Canyon along the coast, and Southeast Farallon Island located 40 km off the coast and slightly further north. Long-range persistence is important for a number of reasons: on the negative side, it can have serious detrimental effects for statistical inference and on the positive side, it provides access to the ocean's memory which can lead to a greater understanding of the processes involved and thus to better prediction. Long-range persistence also provides important insights into the relationship between the scaling that is obtained and the time scales employed. The first step in the analysis was to remove the annual cycle from the data at each location because of its detrimental effect on estimating long-range persistence. Then detrended fluctuation analysis was used to calculate long-range persistence where a single scaling exponent is obtained that relates the magnitudes of the fluctuations in the data to the time scales involved. Similar scaling exponents were obtained for Granite Canyon and Pacific Grove with values of 1.04 and 1.05, respectively. At Southeast Farallon Island, a value of 1.16 was obtained. The increase in the scaling exponent at Southeast Farallon Island is consistent with observations made elsewhere and model results, which indicate that as coastal influence decreases further offshore, the scaling exponents for sea surface temperature tend to increase. Because Southeast Farallon Island is exposed to subarctic waters offshore, whereas Pacific Grove and Granite Canyon are exposed to warmer waters from the California Undercurrent along the coast, these exposures to different water masses may contribute to the observed change in scaling behavior.

### Keywords

Long-range persistence, sea surface temperature, central California coast, detrended fluctuation analysis, singular spectrum analysis, memory of the ocean

Date received: 12 March 2018; accepted: 11 June 2018

### Introduction

One of the basic properties of the observations that comprise most geophysical time series is that they are not statistically independent. As a result, any given observation is influenced by, correlated with, or related to all of the previous values in the series, to a greater or lesser extent. Thus, one of the classic problems in time series analysis arises, autocorrelation between the observations in the record. In some cases, we only need to consider the correlation between a given value and its closest neighbors that precede it in time, which we refer to as short-term correlation or short-term persistence. However, a given value may

also be correlated with distant neighbors in the time series as well. When there is significant correlation between values that are far apart relative to the sampling interval, then long-range correlation or long-range persistence (LRP) enters the picture. LRP reflects long-term memory and the

<sup>1</sup>Moss Landing Marine Laboratories, Moss Landing, CA, USA

<sup>2</sup>School of Marine Science and Policy, University of Delaware, Newark, DE, USA

#### Corresponding author:

Laurence C Breaker, Moss Landing Marine Laboratories, 8272 Moss Landing Road, Moss Landing, CA 95039, USA.

Email: lbreaker@mml.calstate.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

long-term memory of the ocean far exceeds that of most other components in the climate system.

In the study of LRP, different methods have been used to estimate this quantity. Witt and Malamud (2013) compare four analysis techniques that have been commonly used to estimate LRP. Of the four methods, they recommend that power spectrum analysis (PSA) and detrended fluctuation analysis (DFA) be used to estimate LRP. Although PSA is highly recommended, not all agree that it should be the method of choice. According to Tsonis et al. (1999), DFA is superior to PSA because the basis function that is obtained for estimating the scaling exponent is generally more representative of the entire time series than one obtained using PSA, and, furthermore, that fluctuations that typically arise in PSA result in scaling regions that are often masked. According to Talkner and Weber (2000), DFA is a more systematic procedure than PSA, which is something of an art. We have also chosen DFA because the majority of previous studies that we reference have employed this method, making our results easier to compare with past work.

In DFA, a single scaling exponent is derived that relates the magnitudes of the fluctuations in the data to the scales involved, be they temporal or spatial. The calculation, although similar to that of the standard deviation, differs in one important way. The reference in this case is not the mean value but is the residual that is produced when the various sources of non-stationarity have been removed from the original data. When this relationship, expressed in log-log coordinates, yields a straight line (or at least a reasonably close approximation), the data are said to be self-similar or scale invariant (e.g. Malamud and Turcotte, 1999). For self-similarity to apply, the points on the log-log plot must be sufficiently collinear across a relatively wide range of scales. This scaling exponent represents the slope of the straight line so obtained. If the scaling exponent,  $C_s$ , lies between 0.0 and 0.5, the data are classified as anti-correlated or anti-persistent. If  $C_s$  is close to 0.5, then the data are randomly distributed and thus resemble white noise. If  $C_s$  is greater than 0.5, then the data are positively correlated and they exhibit persistence. When  $C_s$  exceeds 1.0, the data are considered to be strongly persistent and/or non-stationary (Malamud and Turcotte, 1999). Finally, although  $C_s$  is often referred to as the Hurst exponent (Hurst, 1951), we simply refer to  $C_s$  as the DFA scaling exponent.

Why is LRP important? On the negative side, LRP can have disastrous effects on statistical inference (e.g. Beran, 1994). As an example, consider the variance of the sample mean given by  $\sigma^2/N$ , where  $\sigma$  is the standard deviation and  $N$  is the sample size. When serial correlation is present whether it is due to short-range or long-range correlation, or both, the effective degrees of freedom can be significantly reduced. The result is that higher estimates of the variance in the sample mean are obtained. Thus, this problem arises whenever we try to estimate confidence limits and confidence intervals using correlated data. Finally, it is

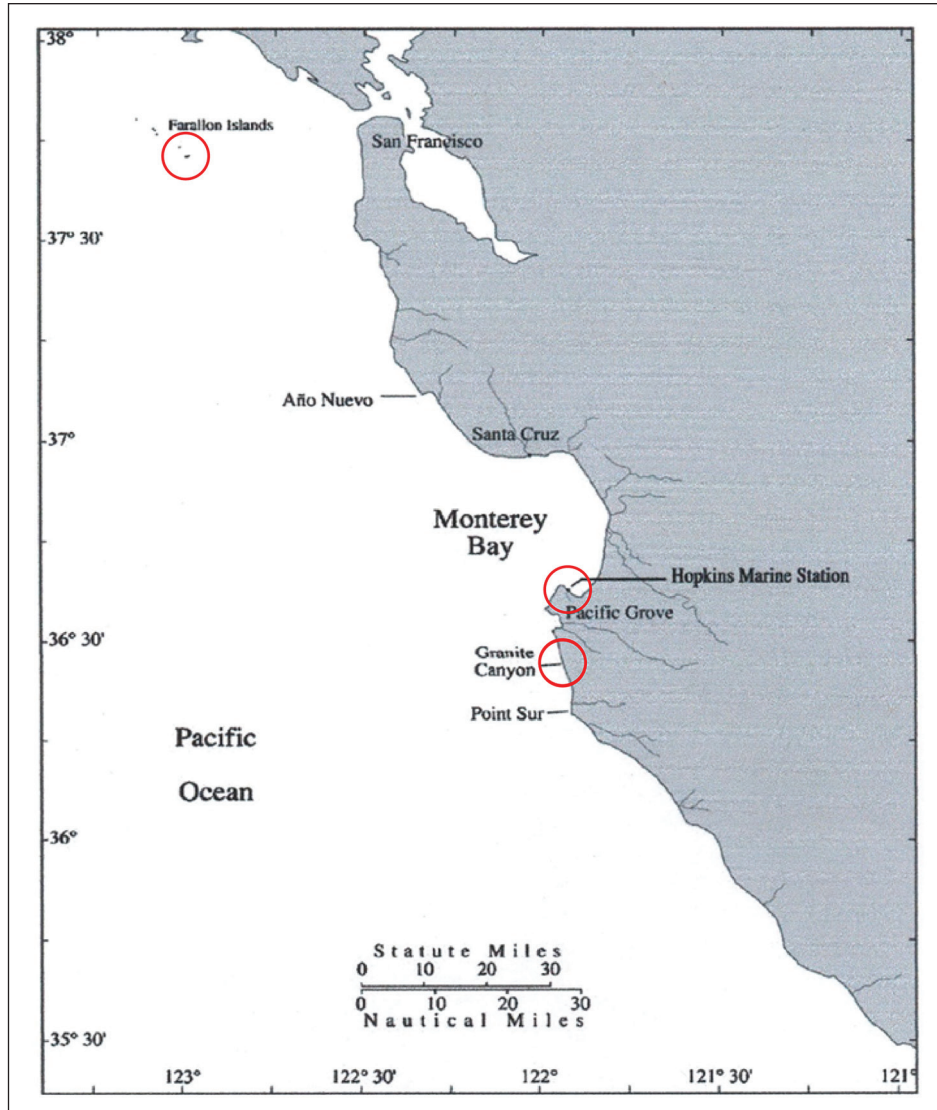
often difficult to distinguish between the effects of LRP and non-stationarity. Without making this distinction, non-stationarity could contribute to LRP leading to values that are unrealistically high.

There are major benefits to LRP as well. First, by examining LRP, we gain insight into the memory of the ocean including not only the information regarding longevity but also with regard to its contents. Because LRP is often associated with processes that are deterministic in nature, they may leave an imprint in the data that provides the basis for a greater understanding of the processes involved. Along similar lines, the information obtained from LRP for variables such as sea surface temperature (SST) should lead to better forecasts from the appropriate models, and, according to Chatfield (1995), this premise can be theoretically verified. As a case in point, the results of Zhu et al. (2010) using SST demonstrate that in high latitude regions where the values of LRP tend to be relatively high ( $\sim 0.9$  or greater), a simple autoregressive model produced forecast skills in many cases that exceeded those expected from red noise correlations alone.

At a deeper level, LRP provides information on the relationship between the scales employed in the scaling analysis. This relationship is often assumed to be linear (i.e. scale invariant), and when the relationship is scale invariant, it implies that the dynamics at smaller scales are connected to the dynamics at larger scales by a simple power law, and, as a result, the memory of the system is not confined only to the largest scales but extends to smaller scales as well (Tsonis et al., 1999). When scale invariance does not hold such as is the case when changes in the scaling or cross-over points occur, then other processes (and/or non-stationarities) are at work which interfere with this relationship.

Historically, there has been a significant effort to demonstrate the universal nature of the scaling exponents that have been obtained for air temperature through power law scaling (e.g. Fraedrich and Blender, 2003; Koscielny-Bunde et al., 1998; Pelletier, 1997; Peng et al., 1994). On a global basis, it has been shown that scaling exponents for the inner continents tend to cluster around 0.5, over the oceans they often approach 1.0, and in the transition zones that separate them, values in the neighborhood of 0.65 are frequently reported (e.g. Fraedrich and Blender, 2003). Within the past decade or so, there has also been interest in the variability of scaling behavior both for the atmosphere (e.g. Tsonis et al., 2000) and the ocean (e.g. Luo et al., 2015). This interest belies the possibility that changes in scaling behavior, both temporal and spatial, may provide additional insight into the physical and dynamical processes that govern these changes. The present study falls into this category.

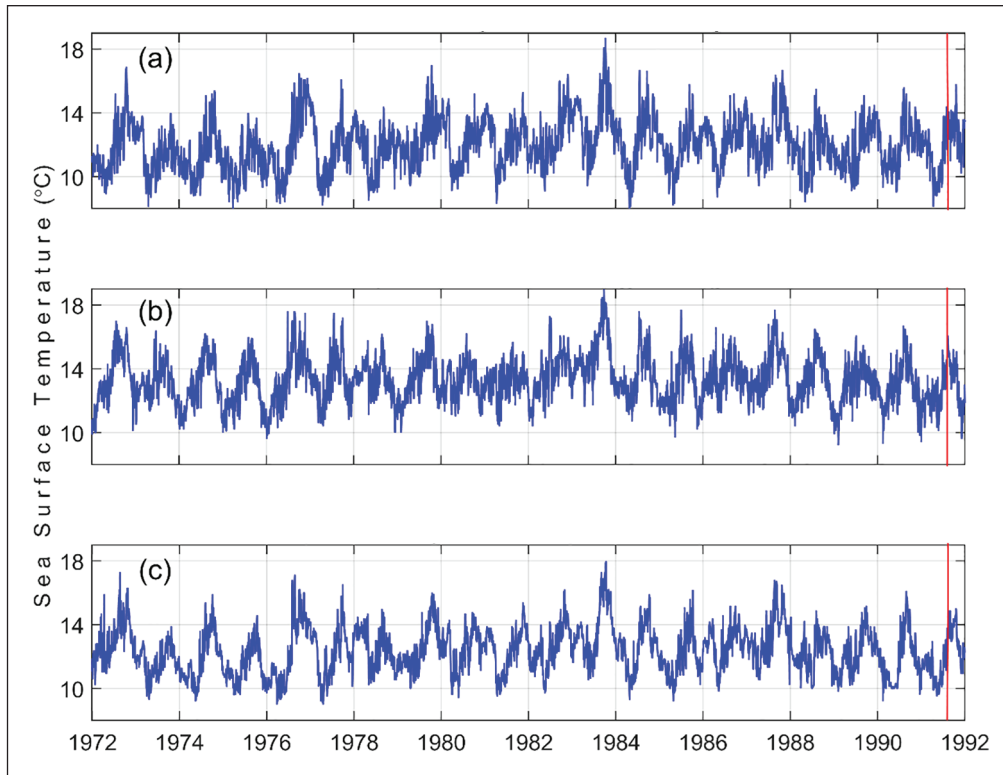
In the present study, the data we employ are daily observations of SST acquired at three locations off the central California coast. These locations provide slightly different oceanographic settings with somewhat different exposures to the coastal ocean and yet they have much in common.



**Figure 1.** This map shows the coast of central California. Daily observations of sea surface temperature have been acquired at Granite Canyon located approximately half-way between the Monterey Peninsula and Pt. Sur, at the Hopkins Marine Station in Pacific Grove at the southern end of Monterey Bay, and at Southeast Farallon Island at the southern extremity of the Farallon Island chain. Each location is highlighted by a small red circle.

These sites are Granite Canyon, Pacific Grove, and Southeast Farallon Island (SEFI), and are shown in Figure 1. Daily observations of SST have been acquired at these three locations for many years, and we have chosen a base period of 20 years for this study that extends from 1 January 1972 through 31 December 1991. We say “base period” because although we employ the full 20 years of data at each site in the early stages of the calculations, when estimating the scaling exponents themselves, we use only the first 19.64 years. This was done strictly for convenience in setting up the range and size of the scales we use in estimating these exponents. In any case, the difference between 19.64 and 20 years is small and so the effect on our results should be negligible.

The steps we take in conducting the subsequent analyses are as follows. First, we employ a well-known method of decomposing the data into independent modes called singular spectrum analysis (SSA) to remove the influence of the annual cycle (Elsner and Tsonis, 1996; Golyandina et al., 2001). Although DFA contains a detrending component, we felt that because of the relative magnitude of the annual cycle we would remove it separately before conducting DFA, consistent with the results of Hu et al. (2001). Second, we calculate the DFA scaling exponent ( $\mathcal{C}_s$ ) from the data in each case in order to estimate the magnitude of LRP and to examine the relationship between the scales. The detrending component in DFA is based on polynomial regression and most reports on its performance have been



**Figure 2.** (a) Shows the daily observations of SST from Granite Canyon for the period from 1 January 1972 through 31 December 1991. (b) Shows the daily observations of SST from Pacific Grove for the same period. (c) Shows the daily observations of SST from the Southeast Farallon Island for the same period. The vertical red line near the end of each record shows the 19.64 years of data that are included in calculating the scaling exponents.

favorable, but its effectiveness has been challenged on several occasions (e.g. Bardet and Kammoun, 2008; Bryce and Sprague, 2012).

Our goals in this study are to acquire information on LRP in a coastal region, where similar observations, to the best of our knowledge, have not been made. We are also interested in determining how well these results agree with observations of LRP made elsewhere in the North Pacific basin and in other coastal regions. We are particularly interested to see how well our estimates of LRP along the coast agree with the data we analyze at the SEFI since previous results suggest that our scaling estimate obtained at SEFI may significantly exceed the values we obtain along the coast. Furthermore, we briefly address the question of whether or not our results support the assumption that the coastal data we employ are representative of waters further offshore. In addition, we ask if our results suggest the importance of one or more ocean processes at work in the region. Finally, we provide an illustrated introduction to DFA that might benefit those who are unfamiliar with the methodology.

## The observations

Daily observations of SST have been acquired at Granite Canyon on the central California coast by the Marine

Culture Laboratory of the California Fish and Game Commission since March 1971. Granite Canyon is located about 15 km south of Monterey Bay and almost the same distance north of Pt. Sur (Figure 1). The nominal period of the data we employ from Granite Canyon is from 1 January 1972 to 31 December 1991 yielding a record length of 20 years (but only the first 19.64 years are used to calculate  $\epsilon_s$ ). The data are collected at approximately at 08:00 am local time. Granite Canyon, because of its location, has direct exposure to the coastal ocean. Waters at this site reflect the strong influence of coastal upwelling during the spring and summer. The data from Granite Canyon are shown in Figure 2(a). The vertical red line located near the end of the record delineates the data that are actually used to calculate the scaling exponents.

At the Hopkins Marine Station in Pacific Grove located at the southern end of Monterey Bay (Figure 1), daily observations of SST have been acquired since January 1919. (In the text, we refer to both Hopkins Marine Station and Pacific Grove—they are one in the same.) The waters that circulate inside Monterey Bay originate almost entirely from outside the bay. Due to local heating inside the bay, however, SSTs tend to be slightly higher than they are further offshore except during the winter when SSTs inside and outside the bay are about the same. The data are collected at 08:00 am

local time each day. For this study, we employ the same nominal 20-year period between 1 January 1972 and 31 December 1991, again, using only the first 19.64 years of data to calculate the scaling exponents. The data from Pacific Grove are shown in Figure 2(b).

Finally, daily SSTs have been acquired at the SEFI since about 1925. Members of the Point Reyes Bird Observatory have been instrumental in maintaining this data collection program for many years. Data collection is difficult during the winter, when storms frequent the area, and so usually fewer observations are acquired during the winter months. It has also made it more difficult to maintain a fixed time-of-day for collecting the data. Because the Farallon Islands are located approximately 40 km offshore they experience greater influence from the California Current per se than either of the other two locations. Thus, SEFI has greater exposure to subarctic waters from the north that are lower in temperature and salinity than waters along the central coast that often originate in the California Undercurrent which tend to be warmer and of higher salinity.

At SEFI, 7.9% of the observations were missing, with most of the missing values occurring between November and February. The majority of gaps were single days with some gaps that lasted for several days and one gap that lasted for 9 days. A shape-preserving piecewise cubic spline interpolation procedure was used to fill the gaps. Although any interpolation procedure tends to reduce natural variability in the data, most of the missing values occurred during the winter when the gradients in SST across the region are at a minimum. As in the two previous cases, the period we have employed is from 1 January 1972 to 31 December 1991, and the scaling exponents were calculated for only the first 19.64 years. The daily observations for this location with the gaps filled are shown in Figure 2(c).

Before moving on, we point out that the data often used in estimating scaling exponents come from climate models and raises the question of whether such models can faithfully reproduce empirical findings (e.g. Govindan et al., 2002). We also note more specifically that most of the SST data used in estimating LRP to date have been acquired from global analyses where spatial interpolation of the gridded data is performed. Interpolation, by definition, implies some degree of smoothing, which reduces natural variability in the data. Thus, a bias, however small, is introduced into calculations that attempt to estimate the magnitude of variability in the data. In the present work, it should be clear that since we use point observations of SST that no such problem exists.

## Background

The formal background that justifies the study of LRP is well documented in the literature on the subject of long-range power law correlation (e.g. Beran, 1994).

Accordingly, we begin with the autocorrelation function, which provides a quantitative measure of the correlation between observations that are separated in time by a lag that can vary from zero to a significant fraction of the record length. This relationship can be expressed as

$$C_s = 1/n \sum_{i=1}^{n-s} (x_i - \langle x_t \rangle)(x_{i+s} - \langle x_t \rangle) \quad (1)$$

where  $x_t$  represents the time series (i.e. daily observations of SST),  $x_i$  is the  $i$ th value of the time series,  $x_{i+s}$  is a lagged version of  $x_i$  at lag  $s$ ,  $\langle x_t \rangle$  is the mean value of the record, and  $n$  is the record length.

When long-range correlation is present and  $s$  exceeds the time scale we usually associate with short range correlation, these correlations take on the following form

$$C_s \propto s^{-\alpha} \quad (2)$$

if the data follow a power law where  $\alpha$  ranges from 0 to 1. To obtain  $\alpha$  we simply regress  $\log C_s$  with  $\log s$  over a selected range of time scales or values for  $s$ . Although, in principle, the autocorrelation function itself could be used to estimate  $\alpha$ , in practice, noise in the data and trend-like variations often make the interpretation of  $C_s$  difficult. However, there are several other methods that can be used to estimate  $\alpha$  and we have chosen DFA for this purpose.

## Methods

### SSA

In conducting DFA, it is important to first remove all major sources of non-stationarity in the data. These include, but are not necessarily limited to, long-term trends and the annual cycle. Regarding the annual cycle, the concept of LRP refers only to non-periodic processes, according to Maraun et al. (2004). In this regard, Markovic and Koch (2005) removed the annual cycle from their precipitation data in order to estimate the Hurst exponent. They state that by removing the annual cycle, their estimate of the Hurst exponent better reflected long-term correlation in the data.

In order to remove the annual cycle using the detrending capability in DFA a relatively high-order polynomial would be required. This in turn could lead to overfitting of other non-stationarities in the data, such as the long-term trend. Thus, we remove this component separately before conducting DFA. As stated above, to remove the annual cycle, we have applied SSA (Elsner and Tsonis, 1996; Golyandina et al., 2001). In our experience, we have found that SSA performs well in removing the influence of the annual cycle whether or not it departs from a pure sinusoid.

This method decomposes a time series into a set of independent modes. These modes are reconstructed

from the eigenvectors and principal components that result directly from the decomposition. Thus, it is similar to principle component analysis in its formulation. In SSA, the number of modes is determined by the user and provides the needed flexibility to adjust the filtering properties of the method to remove the annual cycle. A window length,  $L$ , must be specified that determines the spectral resolution of the technique and thus the number of modes that are produced. Typical values of  $L$  range between 15% and 25% of the record length. Finally, distortion in the annual cycle, which is common in records of air temperature (e.g. Capparelli et al., 2011) and SST (e.g. Breaker, 2006) often produces higher harmonics that are easily identified in the SSA decompositions. These harmonics are then removed together with the fundamental oscillation to effectively remove the annual cycle.

## DFA

DFA is a well-established method for estimating the scaling behavior of noisy data in the presence of trends without knowing their or shape (e.g. Kantelhardt et al., 2001, 2002). DFA is composed of five steps, following Kantelhardt et al. (2001). The first step can be expressed as

$$y_n = \sum_{i=1}^n (x_i - \langle x_i \rangle) \quad (3)$$

where the notation is as before and  $y_n$  is the cumulative sum. This step is often referred to as creating the profile. According to Ihlen (2012), this step is necessary if the time series possesses significant noise-like structure, which is true for most geophysical time series.

In the second step,  $y_n$  is divided into  $n_s$  non-overlapping segments ( $n/s$ ) of equal length,  $s$ . If  $n$  is not a multiple of  $s$ , there will be a small portion of the record that will not be included in the calculation. In our case,  $n=7300$  (i.e. 20 years) and  $n_s \cdot s=7168$ , where 7168 days corresponds to 19.64 years, consistent with our comments earlier concerning the nominal record length.

In the third step, the local trend,  $v$ , is calculated for each segment based on a least-squares fit to the data. The basis function that is chosen to provide the fit can be linear, quadratic, and cubic, a higher polynomial, or an independent basis such as could be obtained from Empirical Mode Decomposition (e.g. Qian et al., 2009). To simplify the notation, Kantelhardt et al. (2001), and others, refer to linear detrending as DFA1, quadratic detrending, DFA2, and so forth. We follow this notation from hereon. In any case, selection of the detrending polynomial may affect the results and thus should be considered a variable in the subsequent analyses.

In the fourth step, the variance,  $F^2(s)$ , for each of the  $n_s$  segments is calculated as

$$F_s^2(v) = 1/s \sum_{j=1}^s \left\{ y[(v-1)s + j] - p_v(j) \right\}^2 \quad (4)$$

for each segment  $v$ , where  $v=1, \dots, n_s$ , and  $p_v(j)$  is the fitting polynomial for segment  $v$ .

The fifth step is to average over all of the segments and then take the square root of the mean squared values to obtain the overall fluctuation function, as it is called, for each value of  $s$ , according to

$$F(s) = \left[ \frac{1}{n_s} \sum_{v=1}^{n_s} F_s^2(v) \right]^{1/2} \quad (5)$$

If the data follow a power law resulting from long-range correlation, the fluctuation function,  $F(s)$ , increases according to

$$F(s) \propto s^{E_s} \quad (6)$$

To obtain the final scaling exponent,  $E_s$ , we plot  $F(s)$  versus  $s$  in log-log coordinates and then determine the slope of the line from a least-squares linear fit to the data. (Non-linear regression, of course, could also be used to obtain the scaling exponent.)

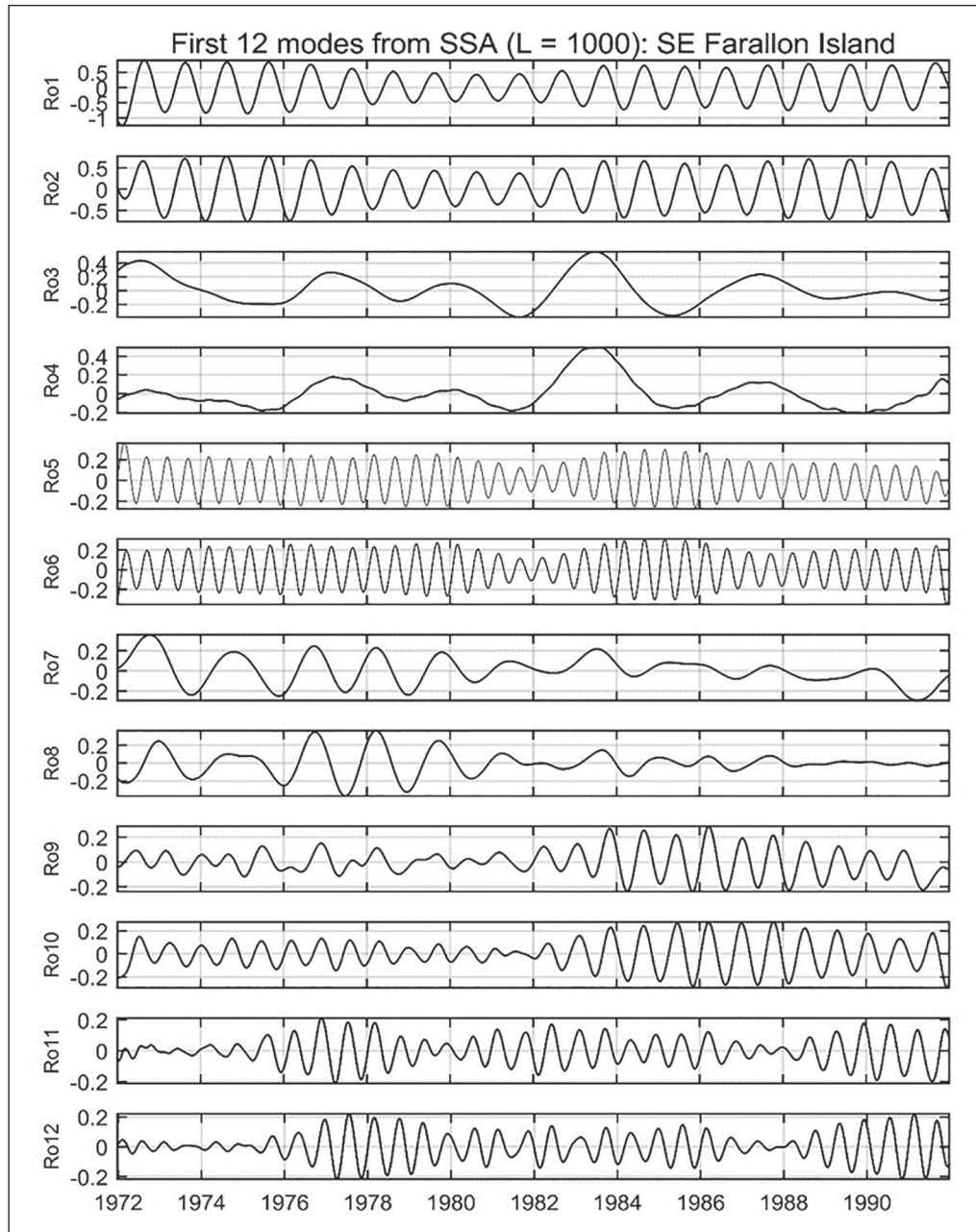
To help evaluate the results, it is recommended that white (Gaussian) noise be included in these calculations as we have done. If the value of  $E_s$  obtained for white noise is not close to 0.5 in all cases, then the calculations should be examined.

## Results

### The results from SSA

First, SSA was applied to the data for the purpose of removing the annual cycle. Window lengths of 1000, 1100 and 1000 days were used for Granite Canyon, Pacific Grove, and SEFI, respectively. The results from SSA for the first 12 reconstructed modes for SEFI are shown in Figure 3. The first two modes (Ro1 and Ro2) correspond to the fundamental oscillation associated with the annual cycle. Modes 5 and 6 (Ro5 and Ro6) correspond to the first harmonic of the annual cycle. The presence of modes 5 and 6 indicate that the annual cycle is not a pure sinusoid but is to some degree, distorted. Modes 1, 2, 5, and 6 were then subtracted from the detrended data to remove the influence of the annual cycle. Similar steps were taken to remove the annual cycles from data for Granite Canyon and Pacific Grove.

Figure 4 shows power spectra of the data from Hopkins Marine Station using the Maximum Taper Method of spectral estimation (Thompson, 1982) with a time-bandwidth product of 2, for (1) the original data (black), (2) the



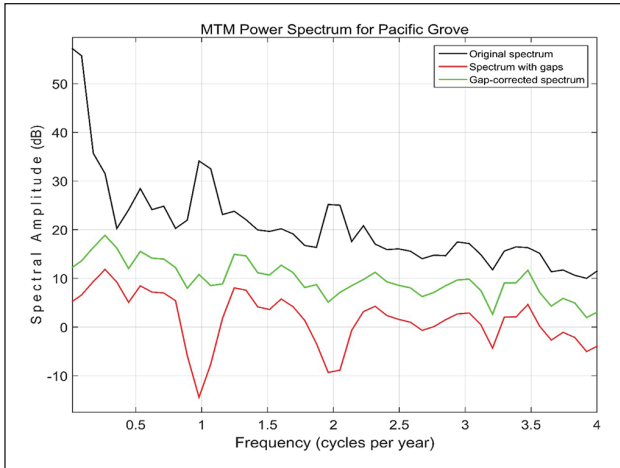
**Figure 3.** Singular spectrum analysis of SST from SE Farallon Island. A window length,  $L$ , of 1000 modes was employed. The first 12 modes are shown above. Of particular interest are modes 1 and 2, and modes 5 and 6. Together, these modes comprise the annual cycle. Modes 1 and 2 correspond to the fundamental oscillation and modes 5 and 6, the first harmonic of the fundamental. By removing all 4 modes, the annual cycle has been removed.

detrended data after the annual cycle was removed that created spectral gaps at 1 and 2 cycles per year (cpy; red), and finally (3) the detrended data with the gaps filled using a simple interpolation procedure (green). The interpolation was performed in the frequency domain from spectral estimates obtained between the gaps at 1 and 2 cpy using the discrete cosine transform (DCT) (Rao and Yip, 1990). The inverse DCT was then applied to obtain the desired gapped-corrected time series for each location. It is these data with the annual cycles removed and the

resulting gaps filled using an objective procedure that are used in the following analyses.

### DFA

In the previous section under “DFA,” we indicated that calculating the cumulative sum is the first step in conducting DFA. Thus, we have calculated the cumulative sum (CUSUM) or profile as the first step in the DFA analysis at each location. As is common practice, the mean value of



**Figure 4.** Power spectra for the data from the Hopkins Marine Station are shown. The spectrum for the original data is shown in black. Note the peaks in the spectrum at 1 and 2 cycles per year. They correspond to the annual cycle. The spectrum of the detrended data after the annual cycle has been removed is shown in red. The spectrum of the detrended data after the gaps created by removing the annual cycle have been filled using a simple interpolation procedure is shown in green. The same procedures were used for the data from Granite Canyon and SEFI. See text for further details.

each record was removed prior to calculating the CUSUMs. From this point on, we use only the first 19.64 years of data from each location.

The next step in DFA is to divide the record, which is now a CUSUM, into  $n_s$  non-overlapping segments ( $n/s$ ) of equal length,  $s$ . We have divided the record into non-overlapping segments for seven different scales: 16, 32, 64, 128, 256, 512, and 1024 days where the scale length increases by a power of 2 over the range of scales employed. This scaling arrangement is often used in DFA, although the range of scales varies depending on the record length that is available.

The step that follows is to calculate the local trend for each segment based on a least-squares fit to the data. We have fitted linear, quadratic, cubic, and 4th order and 5th order polynomials to the profiles in each case. This procedure is illustrated for the data from Granite Canyon in Figure 5 for linear, quadratic, and cubic trends for a scale or segment length of 1024 days. The local trends were then subtracted to produce detrended residuals for each segment and then for each scale. The local root-mean-square (RMS) fluctuations of the residuals are shown by dashed red lines about the trends for each segment. In each case, we have applied first (DFA1) through 5th-order (DFA5) polynomials to the data. As we might expect, higher order polynomials are increasingly effective in eliminating trends in the data, consistent with Kantelhardt et al. (2001). In our calculations, we found that as the order of the polynomial increased, the value of the final

scaling exponent also increased slightly up to and including 4th-order, after which the value of the scaling exponent decreased significantly. Ihlen (2012) found that overfitting of the data by higher order polynomials can be a problem, and so the question arises, how do we decide at what point to stop the fitting process. We have chosen to stop at DFA4 since the values of the scaling exponent, while consistently increasing up to DFA4, rapidly decreased at higher values in each case. Thus, the exponent we estimate is a relative value, that is, relative to the order of the detrending polynomial we employ.

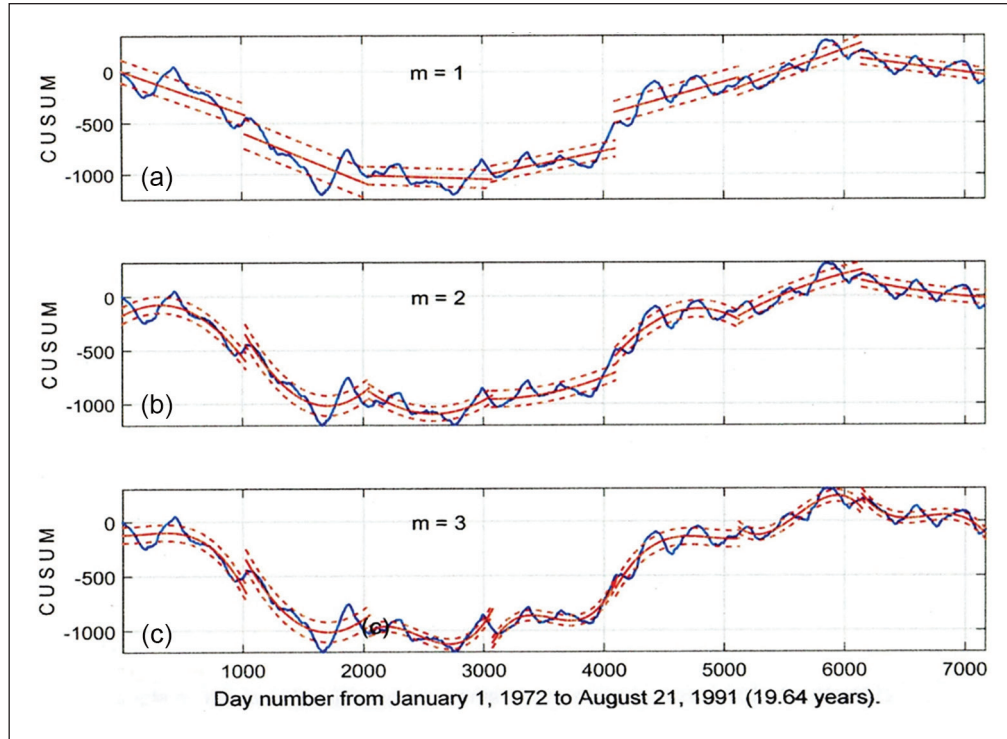
In the next step, we calculated the RMS amplitude fluctuations of the detrended residuals for each segment length (scale) and each location. In Figure 6, we show the local (dark vertical bars) and the global (red horizontal lines) RMS fluctuations for each of the seven scales for Granite Canyon. These results are based on DFA3. Note that the amplitudes of the RMS fluctuations increase rapidly as the scale or segment length increases from 16 to 1024 days, as they did for Pacific Grove and SEFI as well.

Plots of  $\text{Log}_2 F(s)$  versus  $\text{Log}_2 (s)$  for the SST data from each location are shown in Figure 7. The individual values of  $F(s)$  are plotted as large dots for Granite Canyon, small circles for Pacific Grove, and stars for SEFI. These plots include least squares fits to the data using the linear approximation. Granite Canyon is plotted with a solid blue line, Pacific Grove with a dashed blue line, and SEFI with a dotted blue line. Finally, the Gaussian white noise data, processed and plotted in similar fashion to the SST data, is shown in red.

As discussed earlier the data were initially processed by removing the annual cycles because of their relatively high amplitudes. After segmenting the profiles, DFA4 was used to locally detrend each segment for each scale. The corresponding scaling exponent for the Gaussian white noise was in most cases close to 0.5, as it should be if the calculations were performed correctly. This is an important check on the calculations and should be included in performing DFA.

As we examine Figure 7, a closer look at the plots of RMS fluctuation magnitude  $F(s)$ , versus scale  $s$ , show that departures from the straight lines that have been assumed are relatively small and do not indicate cross-over points, which, if present, would cause the slopes to change. Thus, these results appear to be generally consistent with the concept of scale invariance over the range of scales employed. One possible exception may occur for the length scale of 1024 days. Although the departures from linearity are relatively small, in each case the fitted estimate at  $s=1024$  is slightly higher than the calculated value of  $F(s)$  at that scale. This length scale corresponds to a period of about 2.8 years and could indicate that a change in slope or cross-over point is about to occur and might be resolved if a longer length scale were employed. Luo et al. (2015) found a cross-over point at 2.0 years in their SST





**Figure 5.** The figure shows the cumulative sum from the previous step (blue) together with (a) linear ( $m=1$ ), (b) quadratic ( $m=2$ ) and (c) cubic ( $m=3$ ) polynomials fitted to the data (solid red lines) from Granite Canyon for each segment with the segment length or scale equal to 1024 days. The dashed red lines show a measure of the local fluctuations about the trends. The same procedure is applied for each scale.

data from the North Pacific and attributed it to El Niño-Southern Oscillation (ENSO) modulation.

Finally, from these plots it is also apparent that although the values of 1.04 for Granite Canyon and 1.05 for Hopkins are similar the value of 1.16 at SEFI is somewhat higher. It is important to emphasize that the values we have obtained depend strongly on the framework or context upon which the calculations are based. In order to reproduce these results or compare them with those from other studies the same methodology must be employed. The significance of these differences and how our results compare with similar, results obtained elsewhere are discussed in the following section.

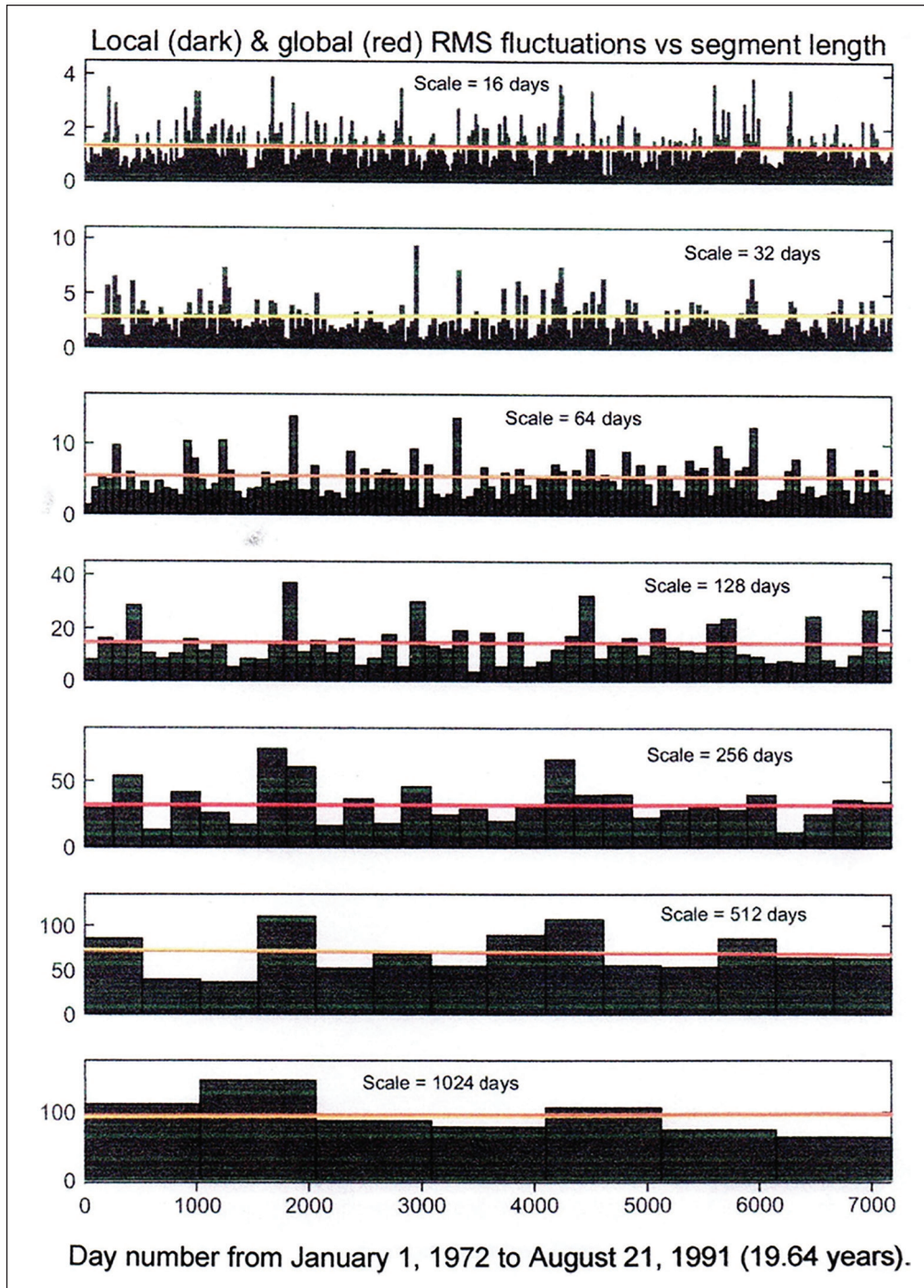
## Discussion and conclusions

Based on the foregoing analyses, we obtained values for the scaling exponent,  $\mathcal{E}_s$ , which ranged from about 1.04 at Granite Canyon, to 1.05 at Pacific Grove, and to 1.16 at SEFI. These values clearly fall into the range where long-range correlation is observed and may even reflect to some degree the effects of non-stationary behavior that has not been removed since the values in each case exceed unity. Although we cannot eliminate this possibility, between removing the annual cycles initially and then applying 4th-order detrending in the application of DFA, this issue was addressed in some detail. In removing the annual cycles

initially, we have followed the recommendations of Hu et al. (2001), who suggest removing unwanted trends first prior to conducting DFA (if there is reason to believe that DFA alone cannot accomplish the task). Finally, this information is new to the region and although it is always possible to question the absolute values because they are to some degree method-dependent, we think that the differences between the coastal values and the offshore value are significant.

There are a number of factors that influence the uncertainty involved in calculating the scaling exponent,  $\mathcal{E}_s$ . These factors include the method used, the type of scaling employed, how the detrending is performed, how the scaling exponent is extracted using either a log transformation followed by linear regression or simply nonlinear regression, and finally, how, and to what extent have non-stationarities in the data been removed?

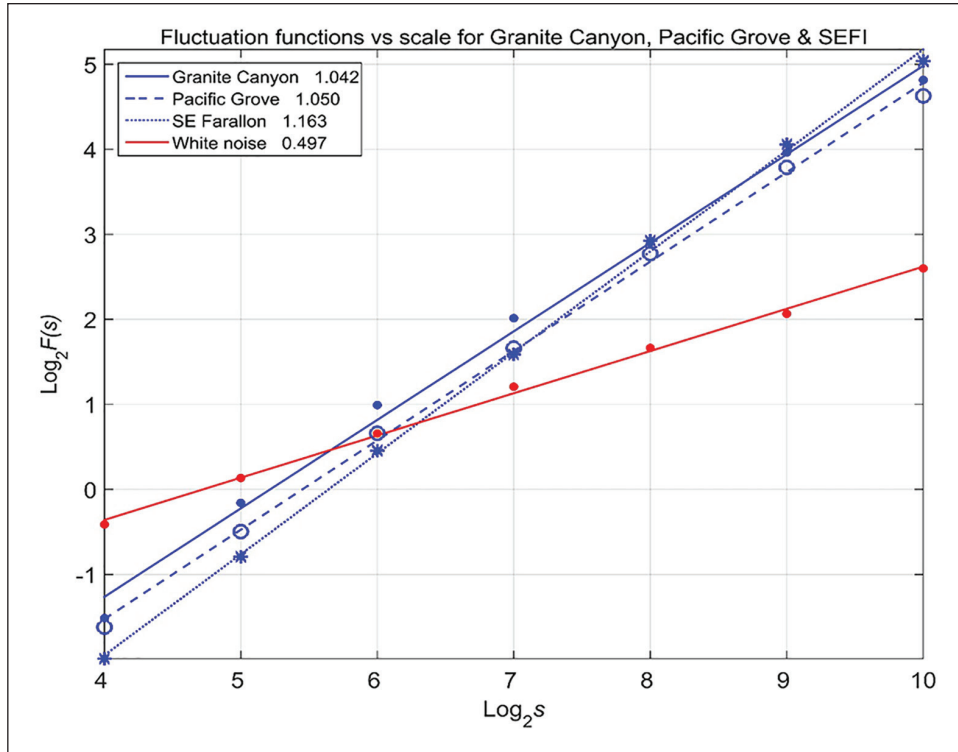
For further guidance in this matter, we turn once again to the work of Fraedrich and Blender. Fraedrich and Blender (2003) (FB) who used a 1000 year simulation based on global climate data including atmospheric and ocean temperature to estimate the variability of the scaling exponent. Using ten 100-year samples taken from their simulated data set, they obtained standard deviations of 0.025–0.05 for most regions over land and water. Blender and Fraedrich (2003) indicate that differences of 0.1 or



**Figure 6.** Local (dark bars) and global (red horizontal lines) RMS fluctuations plotted versus segment length or scale in each panel, starting with the shortest segment length (16 days) at the top and ending with the longest segment length (1024 days) at the bottom. The polynomial order ( $m$ ) was equal to 3.

greater in the scaling exponent generally exceed the uncertainty and so reflect significant changes. We find these estimates of uncertainty well within reason and consistent with our own experience in using SST. As a result, these values might serve as a useful starting point for constructing more refined confidence limits in specific situations.

How do our results compare with those of others in this field? Monetti et al. (2002) applied DFA to both weekly and monthly SST data from several locations in the Atlantic and Pacific oceans to estimate LRP. They estimated values of the scaling exponent to be in the neighborhood of 0.8 for time scales that exceeded



**Figure 7.** Overall RMS values for each segment length versus segment length are plotted in log–log coordinates. Linear least squares fits to the  $\text{Log}_2 F(s)$  versus  $\text{Log}_2 s$  plots are shown for Granite Canyon (solid—blue), Pacific Grove (dashed—blue), and for SE Farallon Island (dotted—blue). The results for Gaussian white noise are shown in red. See text for further details.

10 months. For shorter time scales, the scaling exponents were significantly higher in the range of 1.2 to 1.4, depending on location. Fraedrich and Blender (2003) used both observations of air temperature and SST together with model simulations to estimate the scaling exponents on a global basis. For the inner continents, the oceans, and the transition zones between the continents and oceans, they obtained values of  $\sim 0.5$ ,  $\sim 1.0$ , and  $\sim 0.65$ , respectively. Their results should serve as a basis for comparison in future studies of a similar nature. Gan et al. (2007) used DFA and other analysis methods to estimate the scaling exponent from weekly SST in the South China Sea where they obtained mean values for  $\mathcal{C}_s$  close to 0.95. They also found that  $\mathcal{C}_s$  depended on location with lower values observed near the coast, consistent with our results. Luo et al. (2015) estimated the scaling exponent from monthly SST anomalies for globally averaged SST and for a number of ocean basins using PSA and DFA. The scaling exponents fell into two distinct regimes separated by a cross-over point at 4.3 years for data from tropical basins and for globally averaged data and suggested that this cross-over is related to ENSO. In the North Pacific, a cross-over point was found at about 2 years, again attributed to ENSO. These crossovers were attributed to modulation from the ENSO phenomenon where this behavior was found to be more pronounced in the tropics. In addition, the scaling exponents tended to

increase with decreasing latitude, a pattern that was again linked to ENSO. Our results, as described earlier, suggest a possible cross-over point at the longest time scale, in other words, 1024 days or 2.8 years, but further work will be required to confirm this possibility. In summary, based on the available literature, there is a rather wide range of values for  $\mathcal{C}_s$  obtained from SST data. This range is consistent with a review by Bunde and Lennartz (2012) on LRP who report that long-term correlations for SST range from about 0.8 to at least 1.4, although the results for SST came from only six sources.

Returning to the scaling exponents per se, the slightly higher value for the scaling exponent at SEFI is most likely due at least in part to its location further offshore away from the coastal boundary. In this regard, Fraedrich and Blender (2003) state that there is a general tendency for scaling exponents to increase going from land to sea due to the decreasing influence of the adjacent land where the scaling exponents tend to be far smaller. Gan et al. (2007) also found slightly lower values of the scaling exponents near the coast and higher values offshore, in this case, in the South China Sea.

At SEFI, which is approximately 40 km offshore, influence from the adjacent coast should be reduced significantly. Although the explanation given by Fraedrich and Blender is consistent with our results, we find it is less than satisfying from an oceanographic perspective.

Consider the following, because the Farallons are located further offshore (and ~150 km further north) they are exposed to somewhat different water masses with, most likely, different water mass histories. Within the California Current itself where SEFI is located, we expect the influence of subarctic waters with lower temperatures and lower salinities to be important (e.g. Hickey, 1979). Along the coast, and in Monterey Bay, however, we expect less influence from subarctic waters and more influence from upwelled waters that originate at lower latitudes in the California Undercurrent. Thus, exposure to water masses with different properties and evolution histories could contribute to the differences in the scaling exponents at these locations. In conclusion, we once again refer to the work of Tsonis et al. (2000) regarding changes in the scaling exponents, who state that spatial variations in the scaling properties of the variables that define our climate system are not only expected but required, based on theoretical considerations, and finally, that the spatial distribution of these scaling exponents may provide important information on the underlying dynamics that govern this system.

Although the scaling exponents obtained at Pacific Grove and Granite Canyon are of order unity and consistent with the vast number of offshore values obtained by Fraedrich and Blender in the North Pacific, they might appear to be a bit high due to their locations along the coast. However, the prevailing onshore winds and steep topography that characterize the California coast (and eastern boundaries in general) tend to favor the onshore flow of waters that originate further offshore. Thus, based on our scaling results together with the presence of an eastern boundary, we conclude that the observations from Pacific Grove and Scripps Pier are representative of the waters further offshore.

In closing, we play the devil's advocate by asking if the offshore values obtained by Fraedrich and Blender tend toward unity how do we explain obtaining a value as high as 1.16 off the same coast? There are several possibilities. First, our data correspond to point observations acquired at single locations and not obtained from global analyses, where various forms of interpolation are involved. Second, they removed the annual cycle by calculating the anomaly, a much simpler procedure than we have used. We have followed the recommendations of Capparelli et al. (2011) in this regard, who discuss the problems associated with the anomaly method of removal. Third, in applying DFA, Fraedrich and Blender used DFA1 and DFA2 in their detrending, whereas we have used DFA4 which in our experience usually produces higher values of the scaling exponent. Finally, although Fraedrich and Blender were apparently searching for uniformity in their results, we, on the other hand, have searched for non-uniformity in ours. And, perhaps it is not surprising how often we tend to find what we are looking for.

## Acknowledgements

We thank the Shore Stations Program at the Scripps Institution of Oceanography for providing the sea surface temperature data that was used in this study. We also thank the Editor-in-Chief and one anonymous reviewer for their very helpful comments regarding improvements and needed corrections to the article.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Laurence C Breaker  <https://orcid.org/0000-0002-5398-917X>

## References

- Bardet JM and Kammoun I (2008) Asymptotic properties of the Detrended Fluctuation Analysis of long-range dependent processes. *IEEE Information Theory* 54: 2041–2052.
- Beran J (1994) Statistical methods for data with long-range dependence. *Statistical Science* 7: 404–427.
- Blender R and Fraedrich K (2003) Long time memory in global warming simulations. *Geophysical Research Letters* 30: 1769–1772.
- Breaker LC (2006) Nonlinear aspects of sea surface temperature in Monterey Bay. *Progress in Oceanography* 69: 61–89.
- Bryce RM and Sprague KB (2012) Revisiting detrended fluctuation analysis. *Scientific Reports* 2: Article 315.
- Bunde A and Lennartz S (2012) Long-term correlations in earth sciences. *Acta Geographica* 60: 562–588.
- Capparelli V, Vecchio A and Carbone V (2011) Long-range persistence of temperature records induced by long-term climatic phenomena. *Physical Review E* 84: 046103.
- Chatfield C (1995) *The Analysis of Time Series: An Introduction* (Fifth Edition). Boca Raton, FL: CRC Press.
- Elsner JB and Tsonis AA (1996) *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. New York; London: Plenum Press.
- Fraedrich K and Blender R (2003) Scaling of atmospheric and ocean temperature correlations in observations and climate models. *Physical Review Letters* 90(10): 108501.
- Gan Z, Yan Y and Qi Y (2007) Scaling analysis of the sea surface temperature anomaly in the South China Sea. *Journal of Atmospheric and Oceanic Technology* 24: 681–687.
- Golyandina N, Nekrutkin V and Zhigljavsky A (2001) *Analysis of Time Series Structure: SSA and Related Techniques*. Boca Raton, FL: Chapman & Hall/CRC.
- Govindan RB, Vyushkin D, Bunde A, et al. (2002) Global climate models violate scaling of the observed atmospheric variability. *Physical Review Letters* 89: 028501.
- Hickey BM (1979) The California current system—Hypotheses and facts. *Progress in Oceanography* 8: 191–279.

- Hu K, Ivanov PC, Chen Z, et al. (2001) Effect of trends on detrended fluctuation analysis. *Physical Review E* 64: 011114.
- Hurst HE (1951) Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineering* 116: 770–808.
- Ihlen EA (2012) Introduction to multifractal detrended fluctuation analysis in Matlab. *Frontiers in Physiology* 3: 141.
- Kantelhardt JW, Koscielny-Bunde E, Rego HHA, et al. (2001) Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications* 295: 441–454.
- Kantelhardt JW, Zschiegner SA, Koscielny-Bunde E, et al. (2002) Multifractal detrended fluctuation analysis of non-stationary time series. *Physica A: Statistical Mechanics and its Applications* 316: 87–114.
- Koscielny-Bunde E, Bunde A, Havlin S, et al. (1998) Indication of a universal persistence law governing atmospheric variability. *Physical Review Letters* 81: 729–732.
- Luo M, Leung Y, Zhou Y, et al. (2015) Scaling behavior of global mean sea surface temperature. *Journal of Climate* 28: 3122–3132.
- Maraun D, Rust HW and Timmer J (2004) Tempting long—memory—On the interpretation of DFA results. *Nonlinear Processes in Geophysics* 11: 495–503.
- Markovic D and Koch M (2005) Wavelet and scaling analysis of monthly precipitation extremes in Germany in the 20th century: Interannual to interdecadal oscillations and the North Atlantic Oscillation influence. *Water Resources Research* 41: W09420.
- Monetti RA, Havlin S and Bunde A (2002) Long-term persistence in the sea surface temperature fluctuations. *Physica A: Statistical Mechanics and Its Applications* 320: 581–589.
- Pelletier J (1997) Analysis and modeling of the natural variability of climate. *Journal of Climate* 10: 1331–1342.
- Peng CK, Buldyrev SV, Havlin S, et al. (1994) Mosaic organization of DNA nucleotides. *Physical Review E* 49: 1685–1689.
- Qian XY, Zhou WX and Gu GF (2009) Modified detrended fluctuation analysis based on empirical mode decomposition. Available at: <https://arxiv.org/abs/0907.3284>
- Rao K and Yip P (1990) *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Boston, MA: Academic Press.
- Talkner R and Weber RO (2000) Power spectrum and detrended fluctuation analysis: Application to daily temperatures. *Physical Review E* 62: 150–157.
- Thompson DJ (1982) Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70: 1055–1096.
- Tsonis AA, Roebber PJ and Elsner JB (1999) Long-range correlations in the extratropical atmospheric circulation: Origins and implications. *Journal of Climate* 12: 1534–1541.
- Tsonis AA, Roebber PJ and Elsner JB (2000) On the existence of spatially uniform scaling laws in the climate system. In: Novak MM (ed.) *Paradigms of Complexity*. Singapore: World Scientific, pp. 25–28.
- Witt A and Malamud BD (2013) Quantification of long-range persistence in geophysical time series: Conventional and benchmark-based improvement techniques. *Surveys in Geophysics* 34: 541–651.
- Zhu X, Fraedrich K, Liu Z, et al. (2010) A demonstration of long-term memory and climate predictability. *Journal of Climate* 23: 5021–5029.

### Author biography

Laurence C Breaker has been a physical oceanographer since 1961 and has authored or co-authored over 50 refereed publications in his field.