Hurst exponent for fractal characterization of LANDSAT images

Juan C. Valdiviezo-N.,[a*] Raul Castro,[a] Gabriel Cristóbal,[b] Anna Carbone,[c]

[a]Engineering Division, Polytechnic University of Tulancingo, Tulancingo, Hgo. 43629, Mexico.

[b]Instituto de Óptica, Consejo Superior de Investigaciones Científicas, Serrano 121, Madrid 28006, Spain.

[c]Applied Science and Technology Department, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy.

ABSTRACT

In this research the Hurst exponent H is used for quantifying the fractal features of LANDSAT images. The Hurst exponent is estimated by means of the Detrending Moving Average (DMA), an algorithm based on a generalized high-dimensional variance around a moving average low-pass filter. Hence, for a *two*-dimensional signal, the algorithm first generates an average response for different subarrays by varying the size of the moving low-pass filter. For each subarray the corresponding variance value is calculated by the difference between the original and the averaged signals. The value of the variance obtained at each subarray is then plotted on log-log axes, with the slope of the regression line corresponding to the Hurst exponent. The application of the algorithm to a set of LANDSAT imagery has allowed us to estimate the Hurst exponent of specific areas on Earth surface at subsequent time instances. According to the presented results, the value of the Hurst exponent is directly related to the changes in land use, showing a decreasing value when the area under study has been modified by natural processes or human intervention. Interestingly, natural areas presenting a gradual growth of man made activities or an increasing degree of pollution have a considerable reduction in their corresponding Hurst exponent.

Keywords: Fractals, Hurst exponent, LANDSAT, satellite images

1. INTRODUCTION

Complex systems are probed by observing a relevant variable over a certain temporal or spatial range, which can be represented as long-range correlated sequences or arrays. The Hurst exponent H, whose values vary in the range $0 \leq H \leq 1$, quantifies the tendency of such correlated variables to regress to the mean or to cluster in a specific direction. In particular, the Hurst exponent indicates positive correlation, uncorrelation, and negative correlation for values $H > 0.5, H = 0.5$ and $H < 0.5$, respectively. The estimation of H has been exploited in a variety of fields, including but not limited to climate, genomic, and financial markets. Such correlated variables share the remarkable feature of displaying ordered/disordered patterns, which occur and emerge from the seemingly random structure.[1,2] The interplay of order to disorder is intrinsically linked to the information embedded in the patterns, whose extraction and quantification might add clues to the complex phenomena ruling the evolution dynamic of the underlying system. The most relevant features of complex adaptive systems, as for example biological organisms, social organizations, natural and man-made technological systems obey self-similar scaling laws and fractal-like behavior. Interestingly the complex, spatially and temporally varying patterns observed in ecological systems are largely consequences of the intrinsic dynamics underlying fundamental biological processes: energy and chemical elements feed the main vital activities and produce new organic structures, hence developing a unique biological, chemical, and physical environment ruling the ecosystem development. Therefore, the interest in the application of methods and tools of statistical physics and complex systems science in the analysis

*Corresponding author information:
E-mail: carlos.valdiviezo@upt.edu.mx, Phone: +52 775 755 8202, ext. 1509

Remote Sensing and Modeling of Ecosystems for Sustainability XI, edited by Wei Gao, Ni-Bin Chang, Jinnian Wang, Proc. of SPIE Vol. 9221, 922103
 · doi: 10.1117/12.2060281

of data and modeling of environmental systems is steadily growing. Statistical physics concepts and tools are used to model pollutant flow and transport, seismology and climate, including physical, chemical and biological processes that globally affect the Earth ecosystems. Prominent contributions to a systematic understanding and modelling of our Planet have been provided by the analysis of satellite observations.[3–8]

Motivated by such general observation and the underlying theoretical concepts, the present investigation is thus addressed to identify characteristic patterns and quantify the structure and dynamics of image sets acquired by the LANDSAT Thematic Mapper (TM) satellite in different locations worldwide with the general purpose to obtain meaningful indicators for assessing ecosystem organization, dynamics, stability, and functions of key relevance to the sustainable management of natural resources, social organizations and technological infrastructures.

In this research paper, the approach to estimate the Hurst exponent proposed in[9,10] for long-range correlated sequences and in[11,12] for higher-dimensional arrays is employed for the characterization of real images or, in general, data obtained from two-dimensional disordered random structures. The algorithm to estimate the Hurst exponent is used for the analysis of two image sets taken at different times. Among the several image sets acquired by the different LANDSAT missions, available through the U.S. Geological Survey site,[13] we have selected some scenes of particular interest showing changes over time in land use and water management (such as the Mangystan Province in Kazakhstan and the Elephant Butte Reservoir located in New Mexico USA). The organization of the document is as follows. Section 2 presents the fundamental steps of the Detrending Moving Average algorithm to estimate the Hurst exponent. Section 3 describes the preprocessing steps required to extract the same area on Earth surface from the LANDSAT satellite images. Section 4 presents the experimental results obtained by the application of the algorithm for estimating the Hurst exponent of a pair of image sets. Finally, conclusions and directions for future work are given in Section 5.

2. DETRENDING MOVING AVERAGE VARIANCE ALGORITHM

An efficient method to estimate the Hurst exponent of high-dimensional fractals using the Detrending Moving Average (DMA) has been recently introduced by Carbone *et al.* in Refs.[9, 11]. The method is based on the generalized variance σ^2_{DMA} computed on d-dimensional *fractional Brownian* functions. In the *one*-dimensional case, i.e. $d = 1$, the generalized variance computed on a series $f(i)$ of length N with respect to the moving average $\tilde{f}_n(i)$ is defined by:

$$\sigma^2_{DMA} = \frac{1}{N - n_{max}} \sum_{i=n-u}^{N-u} [f(i) - \tilde{f}_n(i)]^2, \tag{1}$$

where the moving average is defined by:

$$\tilde{f}_n(i) = \frac{1}{n} \sum_{k=-u}^{n-1-u} f(i-k). \tag{2}$$

In these expressions n is the sliding window and $n_{max} = \max(n)$ refers to the maximum size of the window, which must satisfy $n_{max} << N$. The quantity $u = \mathrm{int}(n \cdot \theta)$ is the integer part of the product $n \cdot \theta$, and $\theta \in [0, 1]$. The moving average $\tilde{f}_n(i)$ is calculated for different values of the window n, ranging from 2 to the maximum value n_{max}. By changing the values of θ, the indexes i and k of the above equations are accordingly set within the subarray. For example if $\theta = 1/2$, then i coincides with the center of the subarray. Therefore, once the variance σ^2_{DMA} is calculated for each n, it is plotted as a function of n on log-log axes. The plot is a straight line, as expected for a power law dependence of σ^2_{DMA} on n, such that $\sigma^2_{DMA} \sim n^{2H}$.

On the other hand, consider a *two*-dimensional surface of size $N \times M$, for $d = 2$ the generalized variance is defined by

$$\sigma^2_{DMA} = \frac{1}{(N - n_{max})(M - m_{max})} \sum_{i=n-u}^{N-u} \sum_{j=m-v}^{M-v} [f(i,j) - \tilde{f}_{n,m}(i,j)]^2, \tag{3}$$

with $\tilde{f}_{n,m}$ given by,

$$\tilde{f}_{n,m}(i,j) = \frac{1}{nm} \sum_{k=-u}^{n-1-u} \sum_{l=-v}^{m-1-v} f(i-k, j-l). \tag{4}$$

Notice from the above expressions that the moving average is calculated over subarrays of size $n \times m$; N and M correspond to the size of the sequence, $n_{max} = \max(n) << N$, and $m_{max} = \max(m) << M$. The quantities u and v are defined, respectively, as $u = \text{int}(n \cdot \theta_1)$ and $v = \text{int}(m \cdot \theta_2)$, where $\theta_1, \theta_2 \in \{0, 1\}$. Note that the choice $\theta_1 = \theta_2 = 1/2$ implies that the moving average $\tilde{f}_{n,m}$ is calculated over squares $n \times m$ whose center coincides with (i, j). In addition, if $\theta_1 = \theta_2 = 1$, or $\theta_1 = \theta_2 = 0$, or $\theta_1 = 1, \theta_2 = 0$ or $\theta_1 = 0, \theta_2 = 1$, then one of the vertices of $\tilde{f}_{n,m}$ coincides with (i, j).

In order to calculate the Hurst exponent, the algorithm is implemented as follows. A moving average $\tilde{f}_{n,m}(i,j)$ is calculated for different subarrays by varying n, m from 2 to the maximum values n_{max}, m_{max}; these values depend on the maximum size of the fractal domain. For each subarray, the corresponding value of σ^2_{DMA} is computed by the difference between the original and the averaged signals, as shown in Eq.(3). The variance obtained at each subarray is then plotted as a function of $s = |n \times m|$ on log-log axes; the slope of the regression line corresponds to the Hurst exponent.

3. PREPROCESSING OF SATELLITE IMAGES

The algorithm to estimate the Hurst exponent, based on the generalized variance, was used for the analysis of image sets acquired by the LANDSAT Thematic Mapper (TM) satellite. Several image sets acquired by the different LANDSAT missions are available through the U.S. Geological Survey site.[13] For the research work here presented, we have selected some scenes of particular interest showing changes over time in water management and land use, such as the Mangystan Province (Kazakhstan) and Elephant Butte Reservoir located in New Mexico, USA. These sites have been monthly imaged by the LANDSAT-TM satellite, carried on board LANDSATS 4 and 5, from July 1982 to May 2012, with a 16 day repeat cycle. The data acquired by the sensor consists of seven spectral bands with a spatial resolution of 30 m, in the wavelength intervals from 0.45 to 0.9 nm (bands 1 to 4), 1.55 to 2.35 nm (bands 5 and 7), and 10.4-12.5 nm (band 6). A single scene covers approximately 170 km in height and 183 km in width. Figure 1 displays a natural color image (bands 1,2,3) and a thermal image (band 6) registered by the LANDSAT-TM satellite from the Mangystan province.

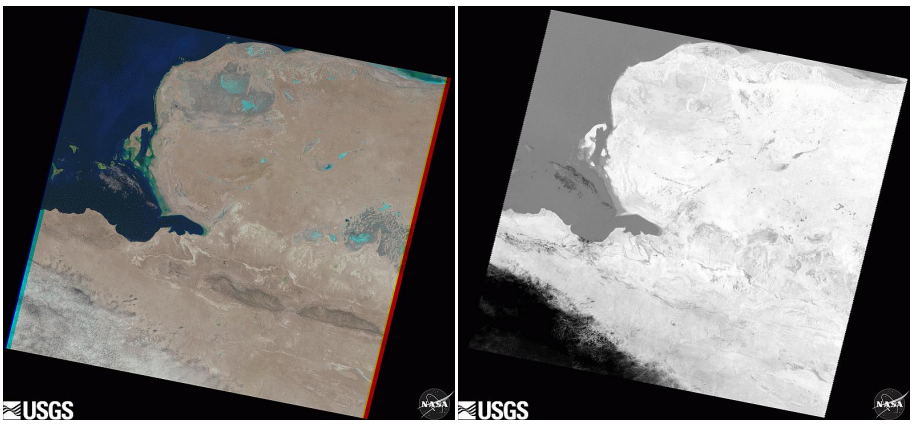

Figure 1. Images registered by LANDSAT-TM from the Mangystan province. Left: color composite RGB image. Right: thermal image. Images taken from the U.S. Geological Survey site[13]

For each area under study, the H value was measured from images registered at different dates between 1985 and 2011, showing a cloud cover less than 10 %. Since each image in the sequence is shifted and sloping due to the repeated cycle of the satellite around the Earth surface. For this reason, a preprocessing procedure to select exactly the same area from the images before performing the simulations was necessary. In order to locate the

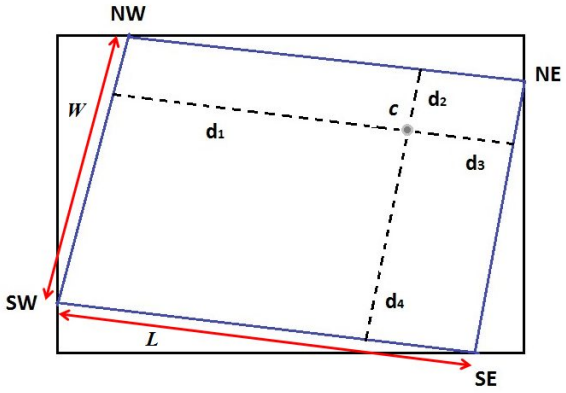

Figure 2. Distances used to locate a center c from a LANDSAT image

center of the subimage, we made use of the latitude and longitude coordinates associated to the image corners, whose information is available in the USGS site. These coordinates served to transform the image pixels into a coordinate space, from which a pixel at a given latitude and longitude can be located and extracted. The preprocessing procedure is described in the next subsection.

3.1 Subimage extraction

The selection of a specific subimage covering the same geographical area in the image sequence should be realized taking into account the latitude and longitude coordinates of the image corners and its corresponding pixel positions. Hence, two steps are required to accomplish this task: a) locate the subimage center and b) define the subimage dimensions from this center position (location reference).

The diagram presented in Figure 2 shows the common frame of LANDSAT images (blue square), which are merged into a dark background. The coordinates of the North-East (NE), South-East (SE), South-West (SW), and North-West (NW) corners denoted, respectively as, NE= (x_A, y_A), SE= (x_E, y_E), SW= (x_D, y_D), and NW= (x_B, y_B) are available as information concerning the image acquisition process (metadata file). Let c denote the subimage center at corresponding latitude and longitude coordinates (x_C, y_C) and suppose we want to locate the pixel positions of this center at each image of the set. For this purpose we can specify the distances d_1, d_2, d_3, d_4, shown in Fig. 2, which are given respectively by:

$$d_1 = x_C - x_B = x_C - x_D \quad ; \quad d_2 = y_C - y_B = y_C - y_A \tag{5}$$

$$d_3 = x_C - x_A = x_C - x_E \quad ; \quad d_4 = y_C - y_E = y_C - y_D. \tag{6}$$

If L and W denote, respectively, the pixel number of columns and rows of the LANDSAT image, the above distances can be expressed in terms of pixels according to the expressions:

$$d_{1p} = L\frac{d_1}{d_1 + d_3} \quad ; \quad d_{2p} = W\frac{d_2}{d_2 + d_4} \tag{7}$$

$$d_{3p} = L\frac{d_3}{d_1 + d_3} \quad ; \quad d_{4p} = W\frac{d_4}{d_2 + d_4}, \tag{8}$$

where d_{1p}, d_{2p}, d_{3p}, d_{4p} represent the distances d_1, d_2, d_3, d_4 in pixels. Once we have located the distances to the center, it is possible to define the subimage in both vertical and horizonal directions. Having the pixel positions of the center, we should now define the height $(2h)$ and width $(2g)$ of the subimages in the sequence (see Figure 3) so that they cover the same area. Although the LANDSAT images used in our simulations belong to the same sensor, we found several images with different sizes, being necessary to realize a normalization step. Consider for example an image sequence with five images from which we want to extract the same subarea, with corresponding dimensions $N_1 \times M_1, N_2 \times M_2, \ldots, N_5 \times M_5$; suppose that $N_{max} = N_1$ and $M_{max} = M_1$, the

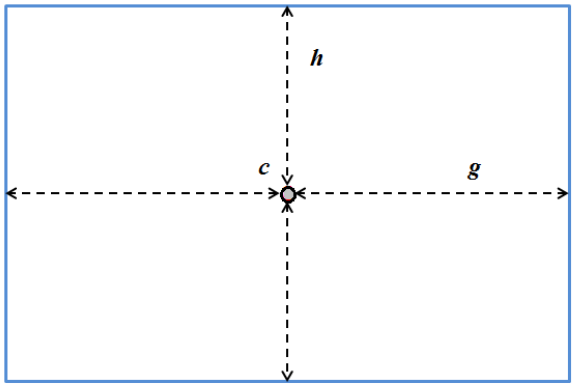

Figure 3. Determination of the subimage height ($2h$) and width ($2g$) given the location c (center of the subimage).

following relations allow to define a normalized height ($2h^*$) and width ($2g^*$) from the center c, no matter the dimensionality of the images in the set:

$$h^* = h\frac{M_{max}}{M_i} \quad , \quad g^* = g\frac{N_{max}}{N_i} \qquad \text{for } i = 1, \ldots, 5. \tag{9}$$

At the end of the procedure, the subimages extracted from the image sequence correspond exactly to the same area under study.

4. RESULTS

The images used in our experiments were registered by the LANDSAT-TM satellite from the Mangystan Province in Kazakhstan and Elephand Butte Reservoir, in USA. LANDSAT images are available in the USGS database in two modalities: natural color and thermal infrared images; in particular, bands 1, 2, and 3 are merged into an RGB natural color image, while band 6 corresponds to the thermal image (see Fig. 1). Hence, the preprocessing algorithm discussed in the previous section was applied to images acquired in the period from 1985 to 2011.

The algorithm to estimate the Hurst exponent, presented in Section 2, was applied to the set of color composite images. A gray-scale transformation of the RGB color image is the simplest approach to estimate the Hurst exponent of a given scene. For this purpose, a gray-scale image can be obtained from the color image by a simple average of its corresponding components using the equation,

$$T_G = \frac{R + G + B}{3}, \tag{10}$$

where R, G, and B denote, respectively, the red, green, and blue components of the image and T_G is the gray-scale transformation. In the following lines we present the results obtained for both areas under study.

4.1 Mangystan Province

Characterized by significant oil and gas deposits that have changed the landscape of the area, the site is located at latitude 45.1665 ° and longitude 51.7868 °. Since 1990 many derricks and other production facilities have been constructed in the desert, which have introduced concerns of the quality and availability of fresh water. Unfortunately, the Caspian Sea is the repository of all that is transported during the oil production, including pollutants.[13]

For each image in the sequence, the DMA variance was computed by means of a square moving average window of size $n \times n$, with $n \in \{7, 11, \ldots, 256\}$. The moving average window can be thought as low-pass filter of varying bandwidth applied to the image. The dimensions of the images $N \times M$ are, respectively, 1121×1253 (for images registered in 1985, 1986, 2007 and 2010) and 1122×1256 (for image in 1992). Figure 5 presents a log-log plot of σ^2_{DMA} respect to $|n \times n|$ for the previous mentioned images. By computing the slop of the regression line in each case, we estimated the Hurst exponent, whose values are displayed in Table 1.

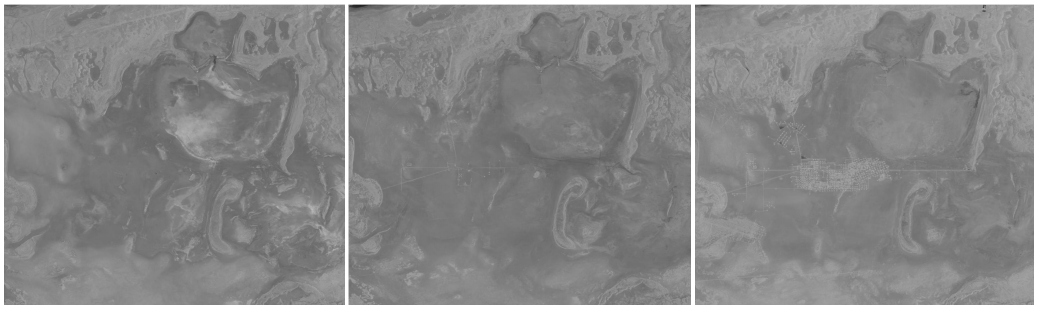

Figure 4. Subimages extracted at different years from the same area in Mangystan. From left to right: subimages in 1986, 2000, and 2011, respectively.

Table 1. Hurst exponent values estimated from a time sequence imagery, registered at the Mangystan site in the period between 1985 to 2011.

IMAGE	H VALUE
1985	0.3005
1988	0.2956
1998	0.2344
2007	0.2259
2010	0.2193
2011	0.2076

4.2 Elephant Butte Reservoir

Elephant Butte Reservoir is a water body, located along the Rio Grande River, north of Truth or Consequences, New Mexico, USA, at latitude 33.2617 ° and longitude -107.1662 °. The site is characterized by declining water levels over the past 20 years, as is shown in Figure 6. Notice the significant reduction in water levels for image in 2012 compared to that registered in 1988.

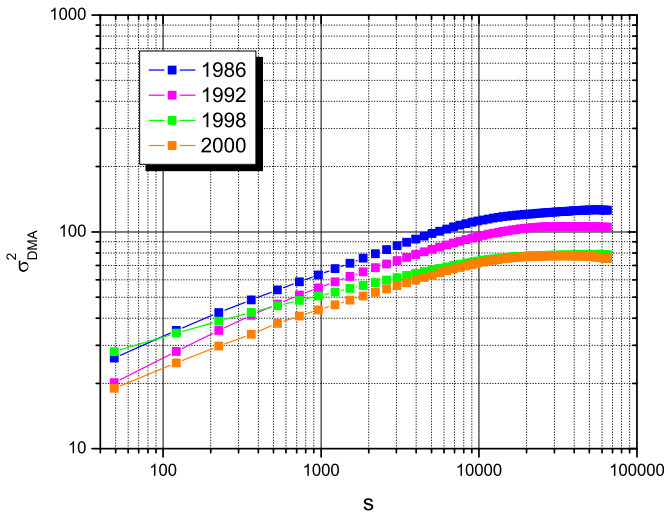

Figure 5. Log-log plot of σ^2_{DMA} respect to $s = |n \times n|$, for images between 1986 and 2000 in the Mangystan Province.

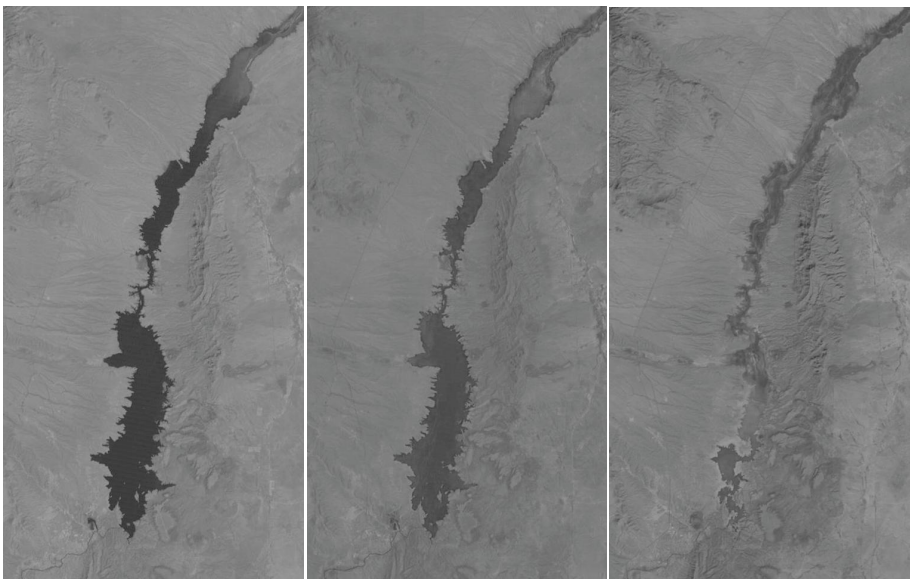

Figure 6. Subimages extracted at different years from the same area in Elephant Butte Reservoir. From left to right: subimages in 1988, 1992, and 2012, respectively.

Table 2. Hurst exponent values estimated from a time sequence imagery, registered at the Elephant Butte Reservoir in the period between 1988 to 2012.

IMAGE	H VALUE
1988	0.3019
1989	0.2732
1992	0.2323
Oct 2012	0.1262
Nov 2012	0.1110

For this experiment 5 satellite images acquired at years: 1988, 1989, 1992, October 2012, and November 2012 were used. Following the same procedure, the original images were preprocessed, as discussed in Section 3, to extract the same area of study. After preprocessing, the resulting images are of size 503×961. Hence, for each image in the set the σ^2_{DMA} was computed by means of a square moving average window of size $n \times n$, with $n \in \{7, 11, \ldots, 256\}$. A log-log plot of σ^2_{DMA} as a function of $s = |n \times n|$ for the previous mentioned images is presented in Figure 7. By computing the slop of the regression line in each case, we estimated the Hurst exponent whose values are displayed in Table 2.

According to the previous examples, it is important to note how the corresponding Hurst exponent decreases for images taken at different years. Therefore, the results obtained when the algorithm is applied to areas showing changes in land use show that the H value decreases as a function of time as polluted or human exploited areas increase. These results show a strong relation between the fractal dimension of the Earth surface and changes over time.

5. CONCLUSIONS

In this manuscript, we have estimated the Hurst exponent from two sets of satellite images registered at different years. For this purpose we used the detrending Moving Average variance algorithm. In the case of images, the algorithm first generates an average response for different subarrays by varying the size of the moving low-pass filter. For each subarray the corresponding variance value is calculated by the difference between the original and the averaged images. The variance obtained at each subarray is then plotted on log-log axes, and the slope of the regression line corresponds to the Hurst exponent.

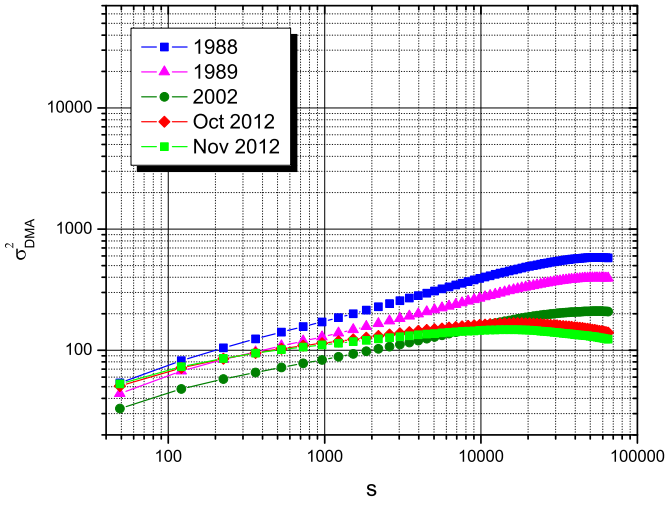

Figure 7. Log-log plot of σ^2_{DMA} respect to $s = |n \times n|$, for images between 1988 and 2012 in the Elephant Butte Reservoir.

The results obtained when our algorithm is applied to areas showing changes in land use show that the H value decreases as a function of time as polluted or human exploited areas increase. These results show a strong relation between the fractal dimension of the Earth surface and changes over the pass of time. If we think of the Earth as a complex system, one could infer that such system is loosing complexity.

ACKNOWLEDGMENTS

Juan C. Valdiviezo is grateful with the Secretary of Public Education (SEP) and the National Research System (SNI-CONACyT) for providing the financial support required to realize this project through the PADES program.

REFERENCES

1. Carbone, A., "Information measure for long-range correlated sequences", *Sci. Rep.* 3 (2013).
2. Carbone, A., Castelli G., and Stanley, H. E., "Analysis of clusters formed by the moving average of a long-range correlated time series" *Physical Review E* 69, 026105 (2004).
3. Thenkabail, P. S. and Wu, Z., "An Automated Cropland Classification Algorithm (ACCA) for Tajikistan by Combining Landsat, MODIS, and Secondary Data", *Remote Sensing*, 4, 2890–2918 (2012).
4. Potapov, P., Hansena, M. C., Stehmanb, S. V., Lovelanda, T. R., Pittmana, K., "Combining MODIS and Landsat imagery to estimate and map boreal forest cover loss", *Remote Sensing of Environment* 112, 3708–3719 (2008).
5. Hansen M. C., et al. "High-Resolution Global Maps of 21st-Century Forest Cover Change", *Science* 342, 850–853 (2013).
6. Vevatne, J. N., Rimstad, E., Hansen, A., Korsnes, R., and Hope, S. M., "Fracture Networks in Sea Ice", *Frontiers in Interdisciplinary Physics* (2014).
7. Zhou, Y., Zhang, Q. S., and Vijay, P., "Fractal-based evaluation of the effect of water reservoirs on hydrological processes: the dams in the Yangtze River as a case study," *Stochastic Environmental Research And Risk Assessment*, 28, 263–279 (2014).
8. Bapanayya, C., Raju, P. A., Das Sharma, S., et al. "Information theory-based measures of similarity for imaging shallow-mantle discontinuities" *Lithosphere*, 3, 289–303 (2011).

9. Alessio, E., Carbone, A., Castelli, G., and Frappietro, V., "Second-order moving average and scaling of stochastic time series", *The European Physical Journal B*, 27(2), 197–200 (2002).
10. Arianos, S. and Carbone, A., "Detrending Moving Average approach: a closed form appreoximation", *Physica A*, 382, 9 (2007).
11. Carbone, A., "Algorithm to estimate the Hurst exponent of high-dimensional fractals" *Physical Review E*, 76 (2007).
12. Turk, C., Carbone, A., and Chiaia, B. M., "Fractal heterogeneous media" *Physical Review E*, 81, 026706 (2010).
13. U.S. Geological Survey, www.earhexplorer.usgs.gov, (2013).