International Journal of Earth & Environmental Sciences

Research Article **Open Access**

Effects of Time Series Undersampling on the Detection of Deterministic Structure in Volcanic Tremor: An Example from Sangay Volcano, Ecuador

Kostas I. Konstantinou

Department of Earth Sciences, National Central University, Jhongli, 320 Taiwan

Abstract

Even though theoretical considerations and several observations point to the possibility that volcanic tremor may be a low-dimensional deterministic phenomenon, there has been no effort in the past to rigorously establish the presence of nonlinear dynamics. In this study we use volcanic tremor time series recorded at Sangay volcano, Ecuador, in order to find such evidence. Initially, the surrogate data test is employed and nonlinear prediction errors are used as a suitable quantity for such a test. Results indicate that tremor time series yielded slightly higher prediction errors than their corresponding surrogates, implying that they can be modeled better by a linear stochastic process at a 95% confidence level. However, the application of Casdagli's test to the same data shows reduced prediction errors in smaller length scales that is compatible with low-dimensional chaos. The apparent conflict of the two tests can be resolved by taking a closer look at the prediction errors as a function of neighborhood size in phase space. This shows a sharp drop in the number of points available for nonlinear prediction at small length scales, probably due to the undersampling of the process that generates the tremor signal. This puts constraints on the efforts to analyze the nonlinear dynamics of volcanic tremor and calls for a reevaluation of seismic data acquisition procedures at active volcanoes.

Introduction

A central issue emerging during the analysis of time series that measure a property of a given physical system, is what kind of process is generating the observed signal. Linear stochastic processes with infinite degrees of freedom have been used in the past to explain irregular or broadband signals and their Fourier spectrum was believed to be the only tool needed in order to characterize them [1]. On the other hand, nonlinear deterministic processes offer a plausible alternative to linear models for such signals and suggest that a much richer structure may exist in the data. This structure takes the form of fractal geometrical objects, often called strange attractors, that consist of trajectories moving in an m-dimensional Euclidean space (called phase or state space) and represent the evolution of the states of the system under study [2]. Furthermore, a whole new discipline of time series analysis has been established that provides methods to detect and exploit such structure in the data [1,3].

Volcanic tremor is a seismic signal recorded before and during eruptions and is probably related to the subterranean movement of magma and/or venting of gas. On theoretical grounds, volcanic tremor has been identified as being generated by nonlinear processes related to flow of fluids inside volcanic conduits [4,5,6]. Some previous studies have attempted to apply nonlinear dynamics methods to tremor time series recorded at volcanoes around the world, in order to detect the presence of deterministic structure [7,8,9,10]. The main focus of these studies was the reconstruction of the phase space and subsequently either the calculation of the fractal dimension of the resulting attractor, or the detection of temporal changes in the phase space by checking how well the short-term future states of the system are predicted. However, none of the studies mentioned above tried to establish the presence of chaos more rigorously, through a series of tests designed to detect nonlinear dynamics. This is an important point, as it would determine which set of models (linear stochastic or nonlinear chaotic) are more appropriate for describing the tremor source.

The purpose of this study is to investigate the presence of deterministic structure in volcanic tremor time series recorded at Sangay volcano, Ecuador. The selection of the particular dataset for this kind of study is motivated by previous nonlinear time series analysis that had already determined suitable embedding parameters for this tremor data [11]. The paper starts with a summary of the dataset characteristics and of results from previously published studies. Then it proceeds to the description and application of two different tests to the tremor events, followed by the presentation of results and closing with the main conclusions.

Data description

The data used in this study were collected during 21-26 April 1998 and recorded at a sampling interval of 125 "samples per second by two broadband three-component seismometers installed near the active summit vent of Sangay volcano, Ecuador (Figure 1). A spectral analysis of the data published by Johnson and Lees [12] revealed (a) a class of tremor events that are broadband and contain significant energy above 5 Hz, and (b) extended degassing 'chugging' events with duration 2-5 minutes that contained well-defined integer overtones (1-5 Hz) and are followed by audible sounds (Figure 2). Events that exhibited high signal-to-noise ratio were selected by Konstantinou and Lin [11] for phase space reconstruction by applying the delay embedding theorem of Takens that uses points y with co-ordinates [1,3]
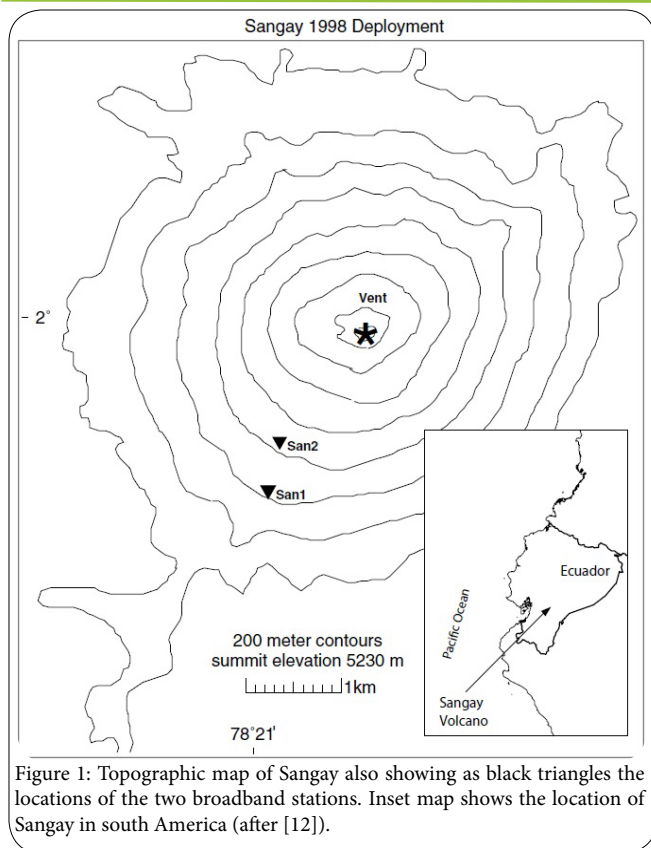
$$y = s(t), s(t+\tau), ..., s(t+(m-1)\tau) \qquad (1)$$

where $s(t)$ are the scalar measurements that compose the seismogram, τ is the delay time and m is the sufficient dimensionfor phase space reconstruction (also known as embedding dimension).

*Corresponding Author:** Kostas I. Konstantinou, Department of Earth Sciences, National Central University, Jhongli, 320 Taiwan; E-mail: kkonst@cc.ncu.edu.tw

Figure 1: Topographic map of Sangay also showing as black triangles the locations of the two broadband stations. Inset map shows the location of Sangay in south America (after [12]).

No	Day	Time	Durations(s)	τ	m	Type
1	112	04:23:41	145	15	5	Broadband
2	112	07:27:36	120	5	7	Broadband
3	112	21:33:36	92	5	5	Broadband
4	112	14:19:06	190	5	6	Broadband
5	111	21:18:56	160	18	4	Chugging
6	112	11:57:46	130	14	4	Chugging
7	112	18:15:26	150	17	4	Chugging
8	114	19:51:19	110	14	4	Chugging
9	115	13:49:00	140	15	4	Chugging
10	115	15:24:50	130	14	4	Chugging
11	116	02:00:00	250	15	5	Chugging
12	116	09:10:20	110	16	4	Chugging

Table 1: Parameters of the tremor events considered in this study after [11]. Note: Time is the start time of each event window given in GMT, τ is the delay time and m is the embedding dimension.

where n_n are Gaussian uncorrelated random increments. The time series from such a process can be regarded as linear since the only nonlinearity is contained in the measurement function s. Surrogates can be easily calculated by multiplying the Fourier transform of the data by random phases and then transforming back to the time domain. However, since the same first/second order properties and spectrum with the data cannot be achieved at the same time, an iterative procedure is used in order to minimise any such differences between the surrogates and the data [13].

Figure 2: Vertical velocity waveforms recorded at stations near the active vent of Sangay of (a) broadband tremor event and (b) chugging tremor event.
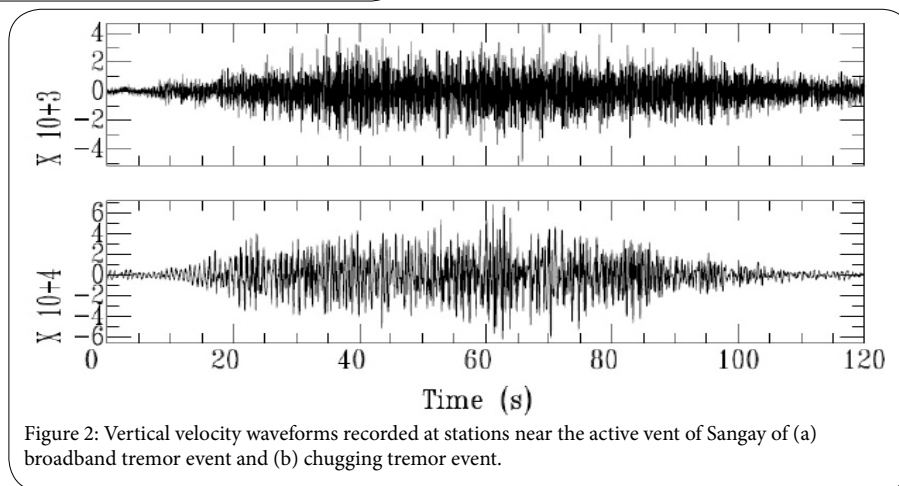
The delay time was estimated after comparing the first minimum of the mutual information and the first zero crossing of the autocorrelation function, while the embedding dimension was determined using the false nearest neighbours algorithm [1,3]. Table 1 gives a summary of these events along with their embedding parameters.

Surrogate data test

The term 'surrogate' refers to time series having the same linear properties (mean/standard deviation, Fourier spectrum) as the original data, but are otherwise resembling some stochastic process, distorted by the action of a nonlinear static measurement function s_n (similar to linearly filtered noise) given by

$$s_n = s(x_n), \; x_n = \sum_{i=1}^{M} a_i s_{n-i} + \sum_{i=0}^{N} b_i n_{n-i} \quad (2)$$

Once the surrogates are calculated a comparison of a quantity that can be easily estimated from the data against the same quantity estimated from the surrogates is used as a test for detecting deterministic structure. The question of which quantity is more suitable for such a testing has been addressed by Schreiber and Schmitz [14] after performing several tests with time series generated by known chaotic systems (like the tent map or the H´enon equations) and their corresponding surrogates. The authors showed that nonlinear prediction errors performed better in detecting deterministic structure than other tested quantities. Nonlinear prediction errors can be calculated in the following way: assuming that the signal is deterministic, then after reconstructing the phase space the dynamics of the trajectories may be approximated by the difference equation

$$y_{n+1} = F(y_n) \quad (3)$$

where $\mathbf{y}_{n+1}$ and $\mathbf{y}_n$ represent points in the m-dimensional Euclidean space that the phase space lies. Knowing the present state of the system $\mathbf{y}_n$ it is possible to predict one step into the future $\mathbf{y}_{n'+1}$; since the function $\mathbf{F}$ describing the evolution of the dynamics is continuous, a past state $\mathbf{y}_{n'}$ very close to the present one is taken and its image $\mathbf{y}_{n+1}$ is used as the predicted value. Because all interpoint distances are contaminated with an uncertainty due to the finite resolution of the data, all points in the phase space closer than a distance r are equally good for predicting $\mathbf{y}_{n+1}$. By forming a neighborhood $\boldsymbol{u}(\dagger)$ of radius r around $\mathbf{y}_n$, the final predicted value will be the arithmetic mean of all the images $\mathbf{y}_{n'+1}$. The nonlinear prediction error can then be defined as [13,14]

$$\gamma = \sqrt{\sum \left[y_{n+1} - \left(\frac{1}{|\boldsymbol{u}_{(\dagger)}|} \sum y_{n'+1} \right) \right]^2} \qquad (4)$$

where $|\boldsymbol{u}(\dagger)|$ is the number of points inside the neighborhood $\boldsymbol{u}(\dagger)$. Since the primary assumption of the method described above is that the signal is deterministic, its application to a time series generated by a linear process with no structure in the phase space (like the surrogates), is expected to yield a larger value for γ.

The surrogate statistical test is posed as the formulation of a null hypothesis that states that the data are generated by the process described by equation 1. This hypothesis can be rejected only in the case when the data have a smaller nonlinear prediction error than their corresponding surrogates. The level of significance of this rejection is given by $(1 - \alpha) \times 100\%$, where α is the probability of a false rejection [13]. Then the number of surrogates that have to be generated as a function of α is $1/\alpha - 1$. If we aim for a significance level of 95% ($\alpha = 0.05$), then 19 surrogates have to be generated for each tremor event tested. This test is onl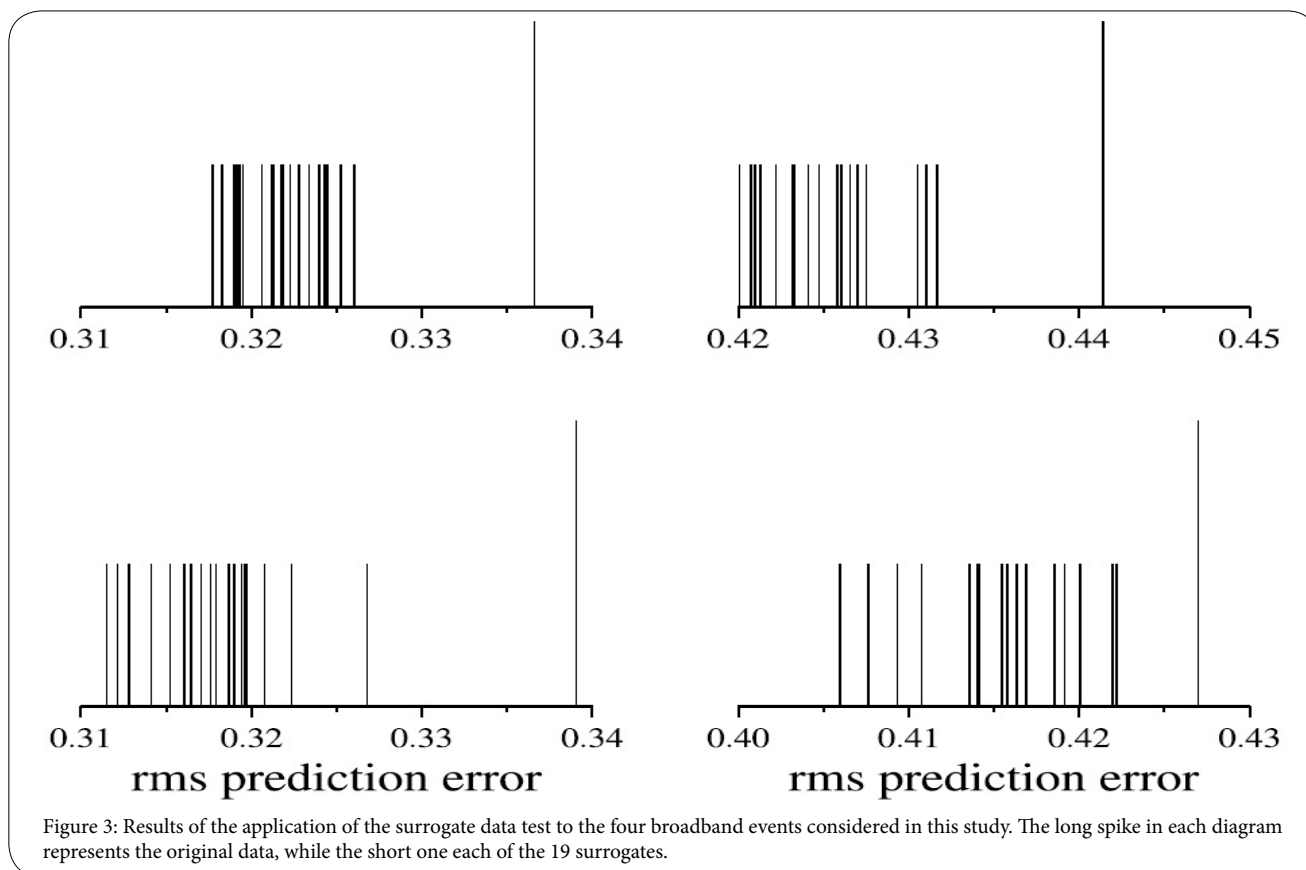y suitable for broadband tremor, as the null hypothesis assumes that the time series is linearly filtered noise; on the contrary, chugging events have a strong periodic component which is incompatible with this null hypothesis.

The delay times and embedding dimensions listed in Table 1 were used and the initial neighbourhood size for making predictions was set to the interval of the samples in the time series divided by 1000. A minimum number of 30 neighbours was required in order to predict one step into the future. If the neighbours found were less than the minimum number, the neighbourhood size increased by a factor of 1.2 until enough neighbours could be found. For all of the broadband events tested we found that the null hypothesis could not be rejected at a 95% significance level, since they exhibited a slightly higher prediction error than the surrogates (Figure 3). According to Schreiber and Schmitz [13] such a negative result may have two possible interpretations: (a) the process generating the data is a linear stochastic one and the null hypothesis is true, or (b) the prediction error quantity may not be able to detect any structure, possibly because the tremor time series cover at this sampling rate only a fraction of the nonlinear dynamics.

Casdagli's test

Another method to perform short-term nonlinear predictions is to try to approximate the function $\mathbf{F}$ that describes the evolution of the trajectories in the phase space (equation 2) by fitting a simple linear function to a part of the trajectory inside a neighbourhood $\boldsymbol{u}(\dagger)$ of radius r. In this way a predicted value (one step into the future) $\hat{\mathbf{y}}_{n+1}$ will be given by [3]

$$\hat{y}_{n+1} = a_n y_n + b_n \qquad (5)$$

Figure 3: Results of the application of the surrogate data test to the four broadband events considered in this study. The long spike in each diagram represents the original data, while the short one each of the 19 surrogates.

the coefficients of the model a_n and b_n can then be easily calculated by minimizing in a least-squares sense the difference between actual and predicted values:

$$\sigma^2 = \sum (y_{n+1} - a_n y_n - b_n)^2 \qquad (6)$$

Casdagli [15] suggested that locally linear models can be used for testing data for the presence of deterministic structure. If the prediction error is lower at large neighbourhood sizes, the data are in this embedding dimension best described by a linear stochastic process. On the other hand, low prediction errors at small neighbourhood sizes support the idea of the existence of a nonlinear, almost deterministic equation of motion. Casdagli's test is much more general in its assumptions, therefore it can be applied to broadband as well as chugging tremor events.

The test is first applied to the broadband tremor events and their surrogates using the same embedding parameters (τ, m) as in Table 1 and predicting one step into the future. For consistency with the surrogate data test, we chose the neighbourhood radius r to range from a minimum value equal to the interval of the data divided by 1000, to a maximum value equal to the interval of the data increasing it each time by a factor of 1.2. Figure 4 shows plots of the prediction error curves for the tremor events and their surrogates while also plotted is the distribution of the number of neighbours found for different r values in each case. It can be seen that the curves corresponding to tremor events show a trend of reduced prediction errors as the neighbourhood size is decreasing, which is absent in the surrogate data curves. The application of this test to the chugging tremor events shows very similar results to the broadband ones, with reduced errors

in smaller length scales (Figure 5). These observations seem therefore to support the suggestion of the presence of determinism in the Sangay tremor data. In the next section an explanation is given for this apparent conflict of results between the surrogate and Casdagli tests.
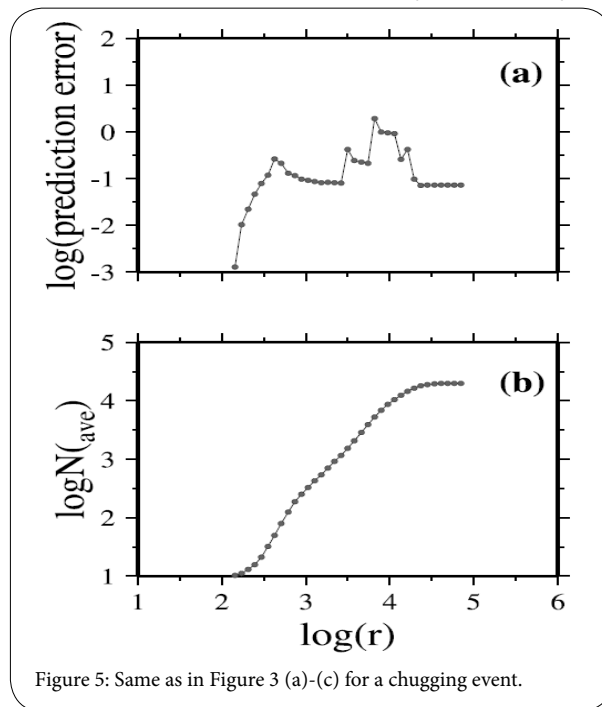

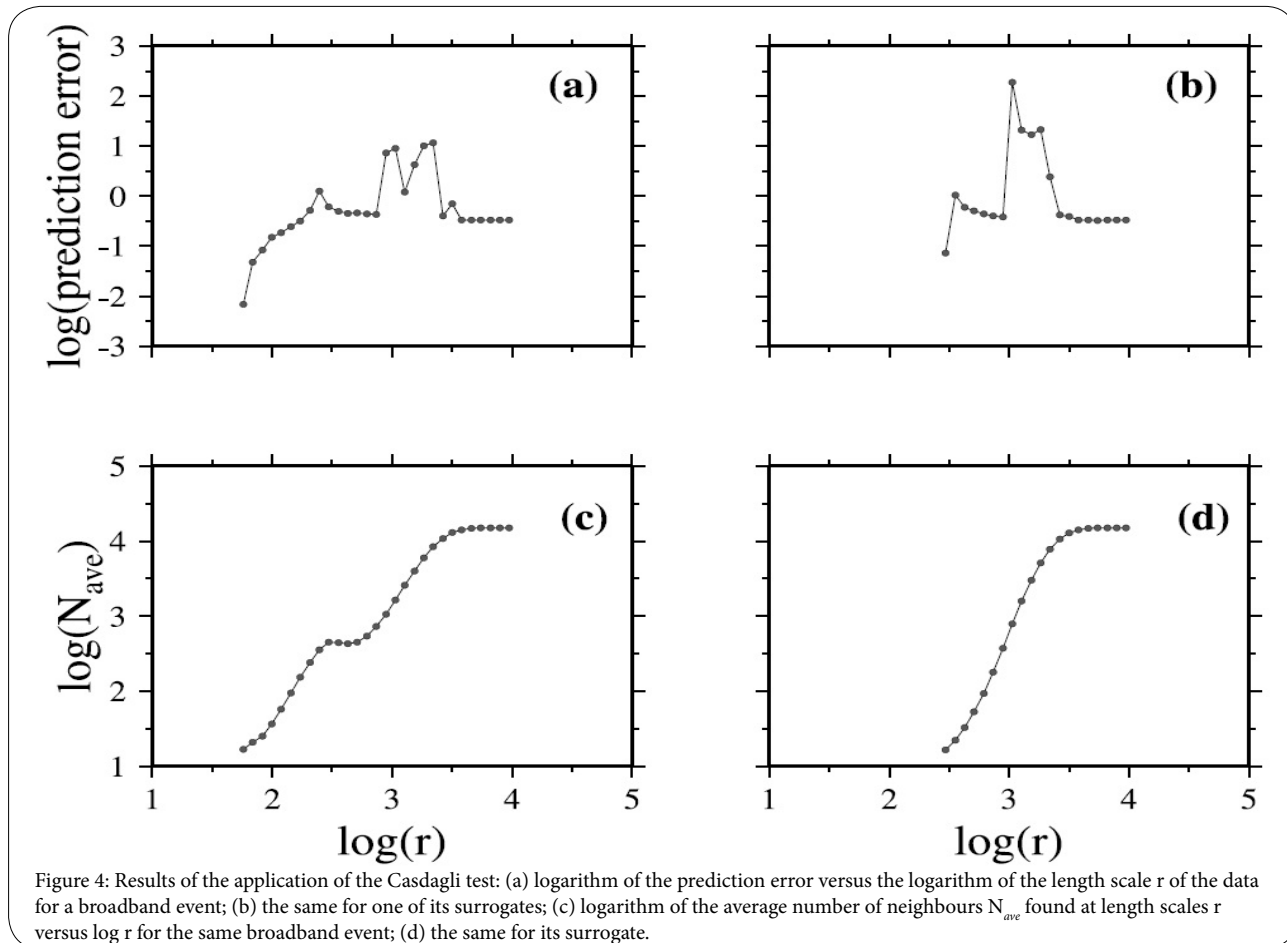Figure 5: Same as in Figure 3 (a)-(c) for a chugging event.

Figure 4: Results of the application of the Casdagli test: (a) logarithm of the prediction error versus the logarithm of the length scale r of the data for a broadband event; (b) the same for one of its surrogates; (c) logarithm of the average number of neighbours N_{ave} found at length scales r versus log r for the same broadband event; (d) the same for its surrogate.

Conclusions

Despite the fact that theoretical models and many observations lend support to the idea that volcanic tremor should be a low-dimensional chaotic phenomenon, there are many difficulties involved when trying to provide concrete evidence about it. The correlation dimension [16] has been used in the past as 'indicator' of determinism, even though it is quite susceptible to the presence of small amounts of noise and heavily depends on the quantity of the available data [1,3]. Thus, any interpretation of a possible scaling region that does not extend through a range of length scales can be considered either as uncertain or as ambiguous. Surrogate data tests were designed so as to remove in theory this ambiguity, however, they also rely on the calculation of some quantity from the data that will be compared later to the value of the same quantity calculated from the surrogates.

The conflicting results of the surrogate data and the Casdagli tests for the volcanic tremor time series presented in this study, can be explained on the basis of the number of points that are available in different neighbourhood sizes. For the tremor events under study, the number of neighbours in the smallest length scale is between 13-16 (Figure 6). This means that at this smallest length scale the number of available neighbours in order to perform a prediction for the surrogate data test, is smaller than the minimum number set in section 3 (30 neighbours). The algorithm therefore will enlarge the neighbourhood size by a factor of 1.2 in order to find at least 30 neighbours. Unfortunately, this is achieved at neighbourhood sizes that yield prediction errors similar to or larger than the ones calculated from the surrogate data, leading to the acceptance of the null hypothesis.
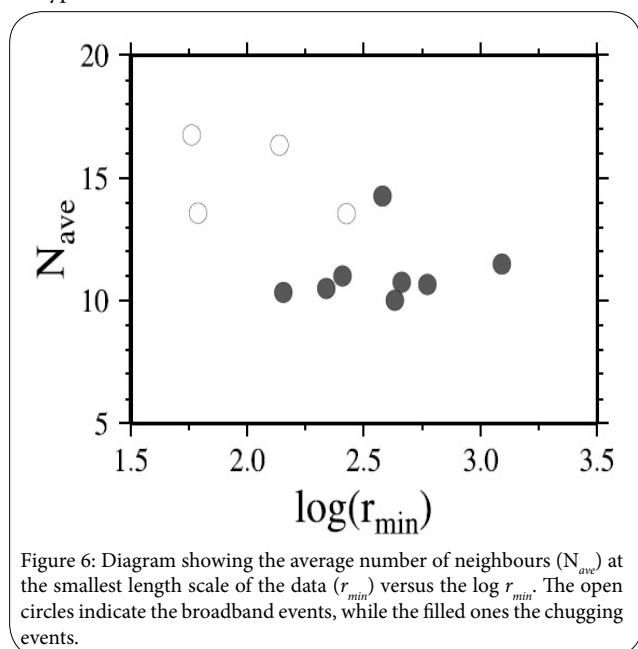

Figure 6: Diagram showing the average number of neighbours (N_{ave}) at the smallest length scale of the data (r_{min}) versus the log r_{min}. The open circles indicate the broadband events, while the filled ones the chugging events.

At this point it should be noted that the size of the neighbourhood or the number of neighbours needed in order to make a reliable prediction is in most cases not known a priori. In fact the value of 30 neighbors used by the prediction algorithm in the surrogate data test can be considered as a reasonable choice for many dynamical systems [18]. Another solution would be of course to lower the minimum number of neighbours (e.g. to 10), but this may result in compromising the accuracy of the prediction. It appears therefore,

that only an adequate number of neighbors at small length scales can ensure the reliability of the results of the surrogate data test. The paucity of neighbors at these scales may indicate that the process generating the tremor signals has been to some extent undersampled. There is evidence that such undersampling may result in an increase of the dimensionality in some signals [17], however, its exact effects on the nonlinear dynamics of volcanic tremor time series are not known and should be investigated in future studies. The results obtained here also highlight the need for reassessing the way seismological data are acquired. The common practice nowadays is to set the sampling interval to 100 samples per second which is even lower than the one used in Sangay. It is clear however, based on the results of Chouet and Shaw [4] that 200 samples per second is the minimum needed for obtaining higher resolution orbits in the phase space and unambiguous estimates of correlation dimension. This is likely to result in more accurate modeling of the source process of volcanic tremor.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgement

References

1. Abarbanel HDI (1996) Analysis of observed chaotic data, Springer, New York.

2. Hilborn RC (1994) Chaos and nonlinear dynamics: An introduction for scientists and engineers, Oxford University Press, Oxford.

3. Kantz H, Schreiber T (2004) Nonlinear time series analysis, Cambridge University Press, Cambridge.

4. Chouet B, Shaw HR (1991) Fractal properties of tremor and gas-piston events observed at Kilauea volcano, Hawaii. J Geophys Res 96: 10177-10189.

5. Julian BR (1994) Volcanic tremor: Nonlinear excitation by fluid flow. J Geophys Res 99: 11859-11877.

6. Rust AC, Balmforth NJ, Mandre S (2008) The feasibility of generating low-frequency volcano seismicity by flow through a deformable channel, In: Lane S.J., Gilbert J.S. (eds) Fluid motions in volcanic conduits: A source for seismic and acoustic signals, Geological Society, London, Special publications 307: 45-56.

7. Godano C, Cardaci C, Privitera E (1996) Intermittent behaviour of volcanic tremor at Mt. Etna. Pure Appl Geophys 147: 729-744.

8. Godano C, Capuano P (1999) Source characterisation of low frequency events at Stromboli and Vulcano islands (Isole Eolie Italy). J Seismol 3: 393-408.

9. Konstantinou KI (2002) Deterministic nonlinear source processes of volcanic tremor signals accompanying the 1996 Vatnaj¨okull eruption, central Iceland. Geophys J Int 148: 663-675.

10. Seidl D, Hellweg M (2003) Parameterization of multichromatic tornillo signals observed at Galeras volcano (Colombia). J Volc Geotherm Res125: 171-189.

11. Konstantinou KI, Lin CH (2004) Nonlinear time series analysis of volcanic tremor events recorded at Sangay volcano, Ecuador. Pure Appl Geophys161: 145-163.

12. Johnson JB, Lees JM (2000) Plugs and chugs - seismic and acoustic observations of degassing explosions at Karimsky, Russia and Sangay. Ecuador J Volc Geotherm Res 101: 67-82.

13. Schreiber T, Schmitz A (2000) Surrogate time series. Physica D: Nonlinear Phenomena 142: 346-382.

14. Schreiber T, Schmitz A (1997) Discrimination power of measures of nonlinearity in a time series. Phys Rev E 55: 5443.

15. Casdagli M (1991) Chaos and deterministic versus stochastic non-linear modeling. J R Statist Soc 54: 303-328.

16. Grassberger P, Procaccia I (1983) Characterisation of strangeattractors. Phys Rev Lett 50: 346-349.

17. Jing H, Takigawa M (2000) Low sampling rate induces high correlation dimension on electroencephalograms from healthy subjects. Psych Clinic Neurosci 54: 407-412.

18. Hegger R, Kantz H, Schreiber T (1999) Practical implementation of nonlinear time series methods: the TISEAN package, CHAOS, 9, 413.