Characterization of the complexity in short oscillating time series: An application to seismic airgun detonations

R. Miralles,[a] A. Carrion, D. Looney,[b] G. Lara, and D. Mandic[b]

Instituto de Telecomunicación y Aplicaciones Multimedia (iTEAM),

Universitat Politècnica de València,

Spain.

Abstract

Extracting frequency-derived parameters allows for the identification and characterization of acoustic events, such as those obtained in passive acoustic monitoring applications. Situations where it is difficult to achieve the desired frequency resolution to distinguish between similar events occur, for example, in short time oscillating events. One feasible approach to make discrimination among such events is by measuring the complexity or the presence of non-linearities in a time series. Available techniques include the Delay Vector Variance (DVV) and Recurrence Plot (RP) analysis, which have been used independently for statistical testing, however, the similarities between these two techniques have so far been overlooked. This work suggests a method that combines the DVV method with the recurrence quantification analysis parameters of the RP graphs for the characterization of short oscillating events. In order to establish the confidence intervals, a variant of the pseudo-periodic surrogate algorithm is proposed. This allows to eliminate the fine details that may indicate the presence of non-linear dynamics, without having to add a large amount of noise, while preserving more efficiently the phase-space shape. The algorithm is verified on both synthetic and real world time series.

PACS numbers: 4360Wy, 4350Yw, 4380Nd, 4330Sf

I. INTRODUCTION

The power spectrum has been widely used in Passive Acoustic Monitoring (PAM) to distinguish among different cetacean species as well as other anthropogenic noises. Although this approach has provided very good results in a large number of applications, there are many challenging situations where we still have not managed to achieve the desired discriminating percentages. Dealing with these situations may require using larger temporal series to obtain the appropriate frequency resolution[1]. Unfortunately, either because some PAM events have a limited duration or because short time windows have to be employed to deal with non-stationary regimes, increasing the temporal length is not always a feasible alternative. A clear example of this can be found in the short non-stationary explosive PAM events such as those produced by seismic airgun surveys. The ability to accurately detect these detonation sounds and to be able to distinguish them from other sounds created by marine mammals has been a problem of great concern for biologists and for the preservation of marine life. These non-stationary detonations occupy the frequency spectrum, which can partially (or in some cases completely) mask the vocal communication of many cetacean species (humpback whales (*Megaptera novaeangliae*) communicate at 400-900 Hz, fin whales (*B. physalus*) at around 20 Hz and minke whales (*Balaenoptera acutorostrata*) from 30-450 Hz.) Some authors have strongly suggested that these anthropogenic noises can be clearly audible[2] and have both short- and long-term effects on cetacean communication and behaviour[2,3]. Additionally, these airgun survey sounds can be transmitted over 200 Km away[3]. Achieving an accurate characterization of the PAM seismic airgun detonations could therefore help in distinguishing maritime mammal calls from airgun detonations. This is important in regions where seismic airgun surveys are taking place so that this activity could be temporally stopped when maritime mammals are detected. The non-stationary behaviour of

[a]Electronic address: `rmiralle@dcom.upv.es`

[b]Communication and Signal Processing Research Group, Department of Electrical and Electronic Engineering, Imperial College, United Kingdom.

seismic airgun acoustic events, along with the presence of other noises, makes it even more difficult to distinguish between marine mammals sounds and airgun detonations when using only frequency derived parameters. Thus, further research on alternative parameters that allow knowledge extraction in PAM may become an interesting area of research. Among the many possible techniques and parameters that could be used for this purpose are signal modality related parameters (non-linearity detection, presence of chaos, etc.). This alternative perspective has provided good results in areas ranging from magnetic resonance imaging to condition monitoring of wind turbines, but it has hardly ever been applied to PAM. The complex nature of the explosions appearing in seismic surveys are reflected in the acoustic event and, in some situations, these events exhibit a non-linear chaotic behaviour[4]. The characterization of this chaotic behaviour could also be performed in the frequency domain. However, as previously mentioned, the non-stationarity and the presence of noise makes this difficult. The characterization of the possible chaotic (or non-linear) behaviour in time series can be achieved using a range of techniques such as: surrogate data bootstrapping methods[5], the Delay Vector Variance Method (DVV)[6], and Recurrence Plots (RP)[1]. Surrogate data bootstraping methods are very powerful, but are hardly ever used when the data oscillates; also, very little is known about the choice of the discriminating statistics[7]. On the other hand, the DVV technique depends only on a small number of parameters and it has been successfully applied in a large number of situations[8,9] (EEG, financial data, etc.). Nevertheless, it is not the most appropriate technique for the characterization of non-linearities in short length oscillatory signals as shown in the sequel. On the other hand, despite its great potential to characterize a large number of signals (including oscillatory signals[1]), RP depends on a crucial parameter for which a systematic study has yet to be done[10].

In this work, we are going to show that by combining two different surrogate generation techniques it is possible to introduce a test for the characterization of complexity of oscillatory signals. We also present an algorithm that combines some of the ideas from the DVV method to produce a less parameter-dependent Recurrence Quantification Analysis (RQA). The potential of the proposed technique is demonstrated on the differentiation of

quasi-periodic oscillations from chaotic ones in short time series.

This work is structured as follows. Section II, we study the particularities of the PAM signals and propose a surrogate generation algorithm to establish the confidence intervals of signals of this kind when using RQA. Section III, we establish some relationships between the DVV method and the quantification of the RP that allow us to design a robust and less parameter-dependent technique in order to characterize time series and look for chaos and quasi-periodicity. Section IV presents an application of the proposed technique to characterize airgun detonations. Finally, we present our conclusions and future lines of research in Section V.

II. A SURROGATE DATA GENERATION ALGORITHM FOR THE CHARACTERIZATION OF CHAOTIC AND QUASI-PERIODIC SIGNALS

PAM events are oscillatory signals that often contain complex patterns of amplitude and frequency modulation. Similar complex patterns, or chaotic behaviour, have also been studied in human voice[11] and have been reported to be related to voice disorders or some pathologies in some situations[12]. Because traditional PAM methods are limited, complimentary objective measures that are able to characterize and distinguish between chaotic and quasi-periodic acoustic events are desirable. Surrogate underpin many such measures, and are based on bootstrapping to generate ensembles of data used to estimate the confidence intervals of the measurements.

In the literature, there are two methods that preserve the higher order moments for the surrogate generation of oscillatory signals[13,14]. The first method, called pseudoperiodic surrogates (PPS), was proposed in[13] by Small et al. (2001). It mimics the phase space of the original data, by adding some dynamic noise in such a way that any existing fine dynamics (which are closely related to chaos) are destroyed. This method was first proposed to test the null hypothesis that an observed time series is consistent with an (uncorrelated) noise-driven periodic orbit. However, this method does not provide good results when applied

to the problem of characterizing/ differentiating complex patterns of amplitude modulation (chaos) from quasi-periodic signals obtained as the sum of sinusoids with incommensurate frequencies. The amount of dynamic noise that needs to be added to remove the fine patterns in an oscillatory signal strongly depends on the complexity of the signal (more complex signals need more dynamic noise to remove details). Thus, the method cannot be applied to compare two low-noise oscillatory orbits that have a different degree of complexity. The second method, called twin surrogates (TS), was proposed in[14] by Thiel et al. (2006). It reproduces the trajectory of the underlying system by visiting the attractor in a different way. These surrogates were devised to test for complex synchronization problems. Importantly, this algorithm can also be used to distinguish chaos from quasi-periodicity using RQA. To this end we have to deal with some practical implementation problems when generating surrogates with the TS algorithm for short time series (such us the limited number of twins[15]).

The proposed algorithm combines ideas of the PPS and the TS methods to generate surrogates that preserve the shape of the phase space orbit while at the same time removing the fine dynamics of the original signal in order to overcome the problem of surrogate generation in short time series. The algorithm, which has been named PPTS (PseudoPeriodic Twin Surrogate) allows sufficient detail to be removed independently of the signal complexity. To achieve this, the proposed algorithm embarks upon the idea of jumping among twin points as proposed in the TS algorithm. The definition of twin points is based on the recurrence plot (RP) of the signal (Eckmann et al.[16]), which allows the recurrence of states in phase space to be to visualized in a simple way. For an N-point time series the RP is calculated as shown in equations (1) and (2),

$$D_{i,j} = ||\boldsymbol{x}(i) - \boldsymbol{x}(j)||, \qquad i,j = m \cdot \tau, \ldots, N-1 \tag{1}$$

$$R_{i,j} = \Theta(\varepsilon - D_{i,j}) \tag{2}$$

where $|| \cdot ||$ is the Euclidean norm, N is the length of the temporal signal, and $\boldsymbol{x}(i) = [x_{i-m \cdot \tau}, x_{i-(m-1) \cdot \tau}, \ldots, x_{i-\tau}]$ are the Delay Vectors (DVs) obtained for a given embedding dimension (m) and time lag (τ). Every DV $\boldsymbol{x}(i)$ has a corresponding target, namely the following sample $x(i) = x_i$. Equation (1) defines the so called Distance Plot (or Unthresholded Recurrence Plot), whereas Equation (2) defines the RP. The parameter ε is a threshold distance (or recurrence threshold), and $\Theta(\cdot)$ is the Heaviside step function.

Twin points are points that are not only neighbours $||\boldsymbol{x}(i) - \boldsymbol{x}(j)|| < \varepsilon$; they also share the same neighbourhood $R_{i,l} = R_{j,l} \; \forall l$. Twin points are indistinguishable regarding their neighbourhoods, but, in general, they have different pasts and, more importantly, different futures. We can generate surrogates by changing the structures in the RP consistently with those produced by the underlying dynamical system. This can be performed by looking in the RP for twin points and then jumping randomly among them. It is easy to see that this is equivalent to a block-wise random shuffling of the columns of the RP matrix in order to break the RP in the position where twin points appear $\{(i,j) \mid R_{i,l} = R_{j,l}, \forall l\}$. As a result of this shuffle, we obtain very similar RPs representations of the surrogates (and thus very similar surrogates) if the time signal is periodic or quasi-periodic (see Figure 1). In contrast, RP representations of the surrogates for a chaotic signal are quite different.

The second idea for generating the surrogates is to move from point to point in the phase space in accordance with a probability that is inversely proportional to the distance between the two points (following the equation in the PPS algorithm[13]). As a result, the proposed PPTS algorithm is summarized as follows:

i) Compute the RP of the original signal using Equation (2), with an appropriate choice of ε, denoted by ε_{TP}, and identify the twin points $(R_{i,l} = R_{j,l} \; \forall l)$. The choice of ε_{TP} is not cruccial for the PPTS; it has been shown in[14] that a choice of ε_{TP} corresponding to 5% -10% of black points in the RP is appropriate.

ii) Randomly choose an initial condition i_0 and make $i = i_0$. Initialize $n = 1$.

iii) If there is a twin point for $\boldsymbol{x}(i)$, make the next point of the surrogate $\boldsymbol{x_s}(n) = \boldsymbol{x}(j)$,
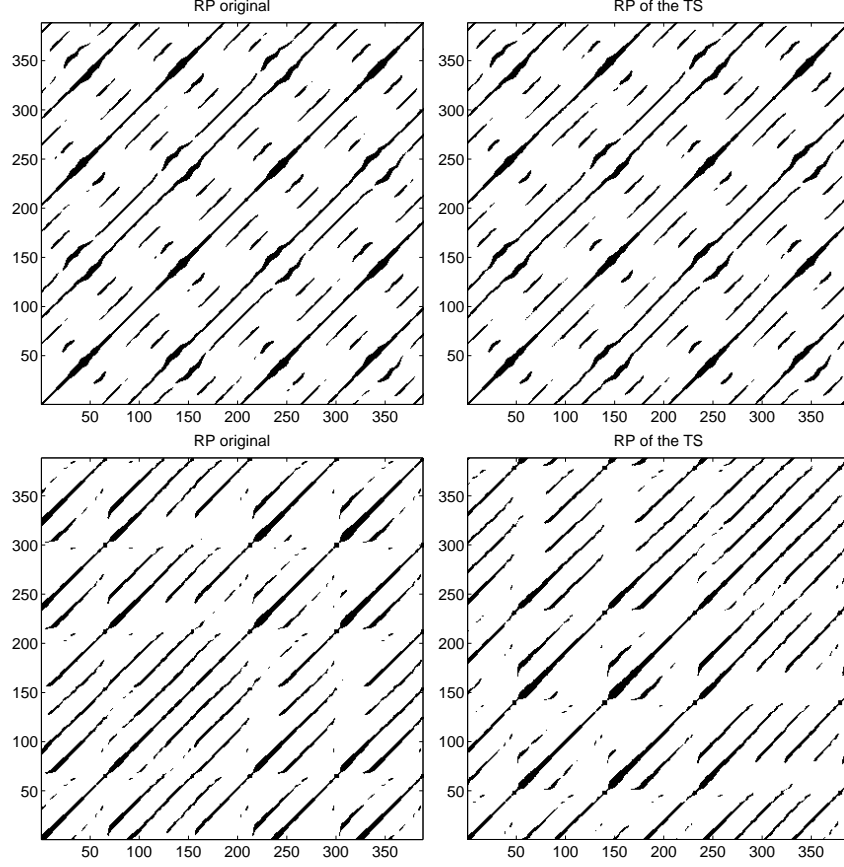
7

FIG. 1. (Top left panel): RP of a quasi-periodic time series. (Bottom left panel): RP of a Rössler time series. (Top right panel): RP example of a surrogate computed using the TS from the quasi-periodic time series. (Bottom right panel): RP example of a surrogate computed using the TS from the Rössler time series. The twin points in the TS algorithm were obtained using a recurrence threshold ε_{TP} corresponding to 10% of black points in the RP. For details regarding the generation of the time series, we refer to Section II.A.

where j is randomly chosen among the twin points with the probability $1/T$ (T is the number of twin points for the state $\boldsymbol{x}(i)$). Let $i = j$ and $n = n + 1$.

iv) For $\boldsymbol{x}(i)$, choose a neighbour $\boldsymbol{x}(j)$ from all of the elements of the phase space represen-
tation ($j = m \cdot \tau, \ldots, N - 1$) with probability

$$Prob(\boldsymbol{x}(j)) \propto exp\frac{-||\boldsymbol{x}(i) - \boldsymbol{x}(j)||}{\rho} \tag{3}$$

, where ρ is the noise radius studied in[13]. Make the next point of the surrogate $\boldsymbol{x_s}(n) =$ $\boldsymbol{x}(j)$. Let $i = j$ and $n = n + 1$.

v) Repeat from Step (iii) until $n = N$.

The surrogate is formed from the first scalar component of $\boldsymbol{x_s}(n)$.

The proposed PPTS algorithm generates surrogates that are very similar to the original signal as long as the original signal is periodic or quasi-periodic. When the original signal deviates from a periodic or quasi-periodic oscillation the PPTS generates surrogates that have quite a different RP matrix while still preserving the approximate phase space shape of the original signal. Thus, these surrogates are appropriate for testing the null hypothesis that the observed time series is consistent with a quasi-periodic orbit. The following examples illustrate this idea (all the PPTS examples in this work were computed using a ε_{TP} corresponding to 10% of black points in the RP).

A. Examples

Consider the Rössler equations of a chaotic time series, given by:

$$
\begin{cases}
\dot{x} &= -y - z \\
\dot{y} &= x + ay \\
\dot{z} &= b + z(x - c)
\end{cases}
\tag{4}
$$

with the initial conditions $x(0) = y(0) = z(0) = 0.1$, $a = 0.2$, $b = 0.2$, $c = 5.7$ (chaotic state) and a sampling time of $\Delta t_s = 0.2$. The system was integrated 1400 times using the Matlab ODE solver ODE45 and the time series $x(n)$ was obtained from the x component after discarding the first 1000 data points to avoid transient states (only 400 data points were employed).

Figure 2 illustrates the phase space reconstruction (with $m = 3$ and $\tau = 6$) of the following: (a) original time series, (b) and (c) the PPS , and (d) the PPTS. When comparing the (b) and (d) panels, it can be observed that for the same noise radius ($\rho = 0.25$), the

9

proposed PPTS algorithm achieves a phase-space representation that removes most of the details which evidence chaos (the large number of trajectories that run arbitrarily close together) while at the same time preserving the shape of the original time series. The PPS algorithm would need a higher noise radius ($\rho = 0.55$) to produce similar detail removal (panel (c)). However, using such a high noise radius will produce noisy surrogates, and this lowering of the signal to noise ratio is always inherent in the null test hypothesis of the PPS.
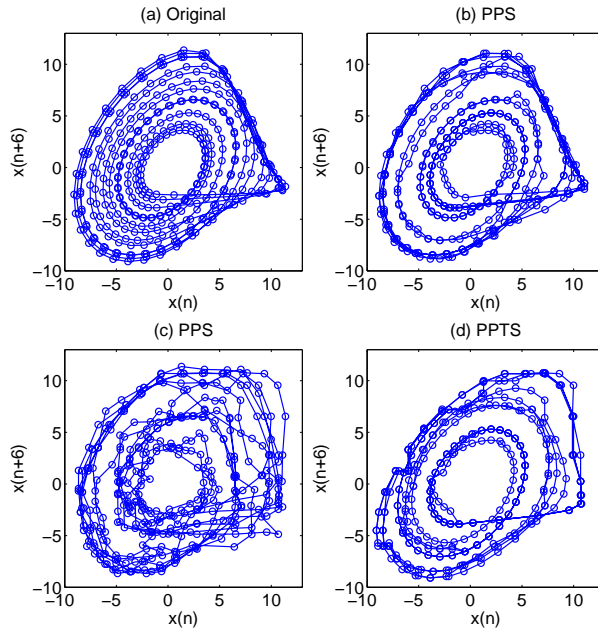
FIG. 2. (Color online) Phase space reconstruction of Rössler attractor. (a) Rössler chaotic time series. (b) Surrogates obtained with the PPS algorithm ($\rho = 0.25$). (c) Surrogates obtained with the PPS algorithm ($\rho = 0.55$). (d) Surrogates obtained with the proposed PPTS algorithm ($\rho = 0.25$).

The operation of the proposed method is next illustrated on a quasi-periodic time series composed of the sum of two sinusoids with incommensurate frequencies[17].

$$x(t) = 8 \cdot sin(2 \cdot \pi \cdot f_1 \cdot t) + 4 \cdot sin(2 \cdot \pi \cdot f_2 \cdot t) \tag{5}$$

where $f_1 = \sqrt{3}$ Hz, $f_2 = \sqrt{5}$ Hz. The discrete time series version is obtained using a sample period of $\Delta t = 0.02$ s. We also generate 400 samples of this signal and employ the PPS

and the proposed PPTS to obtain the surrogates. The phase space representation is shown in Figure 3, showing that for quasi-periodic signals, both of the algorithms achieved similar phase space reconstruction.
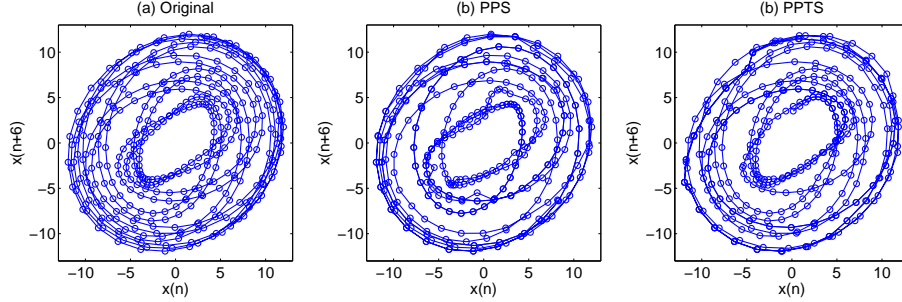

FIG. 3. (Color online) Phase space reconstruction of a sum of sine waves. (a) Quasi-periodic time series. (b) Surrogates obtained with the PPS algorithm ($\rho = 0.25$). (c) Surrogates obtained with the proposed PPTS algorithm ($\rho = 0.25$).

These two examples also show one of the weak points of the PPS algorithm when it is used to compare the complexity between orbits: the selection of the noise radius (ρ). The selection of the noise radius in the PPS is strongly related to the amount of detail that has to be removed. Thus, signals that have a chaotic structure require higher values of ρ than signals that do not have that structure. This makes the PPS algorithm unsuitable for distinguishing between chaotic and quasi-periodic time series with comparable noise levels. The proposed PPTS algorithm overcomes this weakness and achieves enough detail removal regardless of whether the signal is chaotic or quasi-periodic; therefore, it can be used as a feasible alternative to establish the confidence intervals when testing for quasi-periodic deviation.

III. FROM DVV TO DELAY VECTOR RECURRENCE QUANTIFICATION ANALYSIS (DVRQA)

RQA i s a powerful non-linear data analysis tool that can be used to measure complexity and chaos (among other characteristics) in any temporal series. This process involves com-

puting the RP with a given recurrence threshold, Equation (2), and measuring any of the RQA metrics. Table I presents some of the typical measures for the quantification of RP. Unfortunately, choosing the appropriate recurrence threshold is not an easy task and to date there is no systematic study that could help in choosing the right threshold[10]. Only a few "rules of the thumb" are given: depending on whether or not the signal is a quasi-periodic process, depending on the amount of noise, etc. The problem of choosing the most appropriate threshold in an unknown situation can be overcome using an alternative approach, similar to that employed by Gautama and Mandic in the Delay Vector Variance method (DVV)[18]. Briefly summarized, the evolution of the normalized target variance compared to that of an artificially generated set of surrogates allows the DVV plots to be obtained. The evolution curves are computed for a sufficiently representative random set of delay vectors (DVs) as the standardized distance among them increases (check[6] for a detailed description of the method).

The RQA metrics can also be computed for a range of standardized recurrence thresholds and plotted as a function of these thresholds. The thresholds play the same role in the RQA as the distances in the DVV. In fact, it can be demonstrated that the DVV can be computed from the RP using the variance of the targets instead of an RQA measurement. We next elaborate this idea in detail, as it is a key concept in the proposed algorithm.

Assume that we have a discrete temporal signal $x(i)$ and its corresponding set of delay vectors $\boldsymbol{x}(i)$, obtained with the appropriate delay embedding (m) and time lag (τ). The process of obtaining the DVV involves choosing a random DV and finding those that are within a given standardized distance (r_d)[18]. This information can be easily extracted from the RP, which is computed using Equation (2). It is important to remember that the RP is a 2D plot showing (for a given moment in time) the times at which a phase space trajectory visits roughly the same area in the phase space. Thus, looking for the "black dots" in the columns of the RP matrix that is computed with a recurrence threshold (ε) is equivalent to looking for the DVs that lie within a given distance r_d in the DVV. The metric used by the DVV to test for non-linearity is the variance of the targets of those DVs which lie within a

given value r_d (or ε in the RP) to the original. Again, this information can be computed from the RP by indexing the temporal signal in the positions where the black dots of the RP appear for that particular column. Figure 4 (panel (a)) illustrates this idea and shows how we can obtain the corresponding targets for two DVs (red lines and green lines). For a given DV $\boldsymbol{x}(i_k)$ and a threshold ε, we obtain $l_\varepsilon \subseteq \{j\}$ s.t. $R_{i_k,j} = 1$, and $\sigma_{i_k}^2(\varepsilon) = VAR[x(l_\varepsilon)]$, where $VAR[\cdot]$ is a variance estimator and $x(l_\varepsilon)$ are the corresponding targets. In order to obtain the DVV, this process has to be repeated for several DVs and standardized distances ranging from $r_d = [max\{0, \mu_d - n_d \cdot \sigma_d\}; \mu_d + n_d \cdot \sigma_d]$, with a typical $n_d = 2$ (as described in[18]). This can be achieved by averaging the estimated variance for a set of L different DVs $\boldsymbol{x}(i_k)$ (different columns of the RP) and repeating the process of computing the RP for different recurrence thresholds ($\varepsilon = r_d$) ranging in the same interval of standarized distances. Figure 4 (panel (b)) shows the appearance of one of the RPs that was computed for a larger ε. The normalized variance can be calculated by dividing by the variance of the time series (σ_x^2) as shown below in Equation (6).

$$\sigma^{*2}(\varepsilon) = \frac{\frac{1}{L}\sum_{k=1}^{L}\sigma_{i_k}^2(\varepsilon)}{\sigma_x^2} \tag{6}$$

Note that the plot of this normalized variance $\sigma^{*2}(\varepsilon)$ as a function of ε is the DVV plot.

Studying the DVV from this perspective helps to understand why the DVV does not give the expected results when applied to oscillatory short time series. The computation of the DVV only takes into account the distance between random DVs (vertical black dots) and misses important information about the possible structures that can be used to detect chaos-order[19] and chaos-chaos[20] transitions: diagonal black lines and vertical white lines[1] respectively. Examples of parameters derived from these structures are how much time the trajectories run parallel paths in the phase space representation (the DET parameter in Table I), complexity measured as the entropy of the black diagonals or the entropy of the vertical white lines (the ENTR or the ENTRWL parameters in Table I), etc. These structures are a key factor in differentiating chaos from quasi-periodicity; this is indicated by the very characteristic diagonals distribution of the oscillatory signals presented in Figure
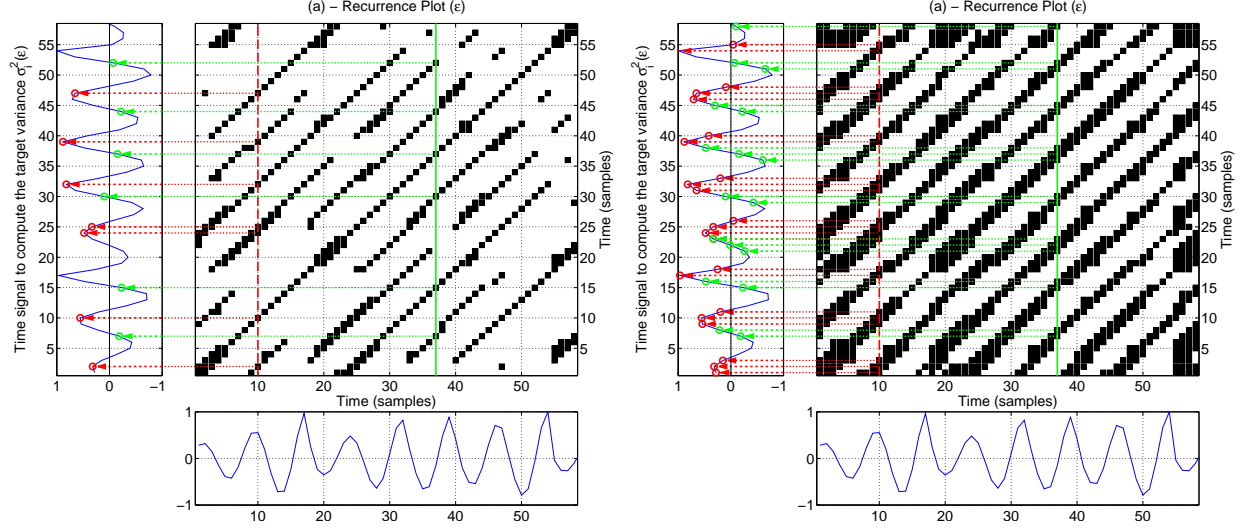
13

FIG. 4. The RP can be used to obtain the nearest DVs and to compute the normalized variance of its targets for a given recurrence threshold ε (or r_d in the DVV). The red and green vertical lines are two randomly chosen DVs ($\boldsymbol{x}(i_k)$). The red and green horizontal lines are their corresponding neighbours lying within the distance ε. The red and green circles on the left axis are their corresponding targets. Panel (a): $\varepsilon = 0.45 \cdot P_x$, panel (b): $\varepsilon = 0.85 \cdot P_x$

1. This new perspective of the RP-DVV technique opens up a large number of possibilities for signal modality analysis. The most straightforward idea is to replace the normalized variance metric of the targets from the DVV plots with any of the multiple RQA metrics that are typically employed in the RP (mean diagonal line, laminarity, entropy, etc.). With the appropriate surrogate generation method, this allows the confidence intervals to be establish so that signal modality tests can be devised for a large number of applications (in our case, complexity in oscillatory signals). The proposed Delay Vector Recurrence Quantification Analysis (DVRQA) can be summarized as follows:

i) Compute the unthresholded RP ($D_{i,j}$) using Equation (1), and obtain the mean (μ_d) and the s.d (σ_d) from the upper triangular matrix (avoiding the main diagonal).

ii) Compute the RP from $D_{i,j}$, Equation (2), for a given threshold (ε).

14

iii) Compute any of the RQA metrics given in Table (I). Repeat steps (ii) and (iii) for $\{\varepsilon \in \varepsilon_s \mid \varepsilon_s = [max\{0, \mu_d - n_d \cdot \sigma_d\}; \mu_d + n_d \cdot \sigma_d]\}$. The parameter n_d is a span factor that works the same way as it does in the DVV (n_d=2 is used in this work).

iv) The representation of the RQA parameter with respect to the standardized recurrence threshold (ε_s) is the DVRQA plot.

v) Repeat for a given number of surrogates N_s to establish the confidence interval, and plot both plots on the same graph.

A. Examples

We next illuminate how the proposed DVRQA algorithm can be used in conjunction with with the PPTS, to distinguish quasi-periodicity from chaos in short time series. To this end, we will use the signals previously presented in Section II.A; note that only 400 samples were used. The PPTS surrogates were obtained using a ε_{TP} that gives 10% of black points and a $\rho = 0.25$. The computation of the confidence intervals was performed using 50 surrogates. First, we applied the DVV to the Rössler chaotic time series (Figure 5). Only a few points lie slightly outside the confidence interval as can be observed in the magnified region of Figure 5, illustrating the difficulties of devising tests that can distinguish the complex dynamics of the Rössler in short time series using the DVV method. This behaviour exemplifies the lack of analysis of the different RP structures performed with the DVV method.

An appropriate selection of the RP structures through a suitable RQA would allow such differentiation. We have chosen two of parameters from Table I: the averaged diagonal length (Ld), and the entropy of the white vertical lines ($ENTRWL$). Jumping among twin points in the PPTS algorithm may break the diagonals that characterize chaotic signals while preserving them in quasi-periodic ones. As a result, chaotic time series should exhibit higher Ld (computed using Equation 7 in Table I) than that of its surrogates. On the other hand, the vertical white lines in the RP represent a rational approximation of the incommensurate frequencies with a tolerance on the order of ε^1. Thus, the length of the

15
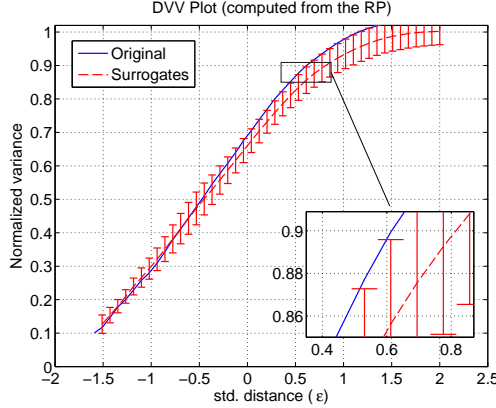
FIG. 5. (Color online) DVV of a 400-point Rössler chaotic time series (computed using RP). The results were obtained using 50 surrogates.

vertical white lines are related to the return times, and the presence of specific return times allows quasi-periodicity and chaos to be detected. The Shannon entropy of the vertical white lines, $ENTRWL$ (computed using Equation 8 in Table I), captures this relationship and gives a different value in chaotic signals compared to that of their surrogates. Figure 6 (top) shows the proposed DVRQA algorithm using the Ld metric when applied to the chaotic and the quasi-periodic time series, whereas the bottom of Figure 6 shows the $ENTRWL$ quantification parameter. The examples behave as expected: for the Rössler time series, there was higher average diagonal length and different white line length entropy; for the quasi-periodic time series, both DVRQA graphs are very similar to their surrogates.

IV. THE DVRQA METHOD FOR THE CHARACTERIZATION OF COMPLEXITY IN SEISMIC AIRGUN DETONATIONS

We next used the proposed DVRQA to investigate for the presence of chaos in sounds produced by detonations during a seismic airgun survey. The sounds used were from the NOAA database and correspond to an acoustic survey recorded by the PMEL autonomous hydrophone array deployed in central Atlantic Ocean. The nature of these acoustic explosive events strongly suggests the evidence of complex oscillations (chaos) in the recordings. We
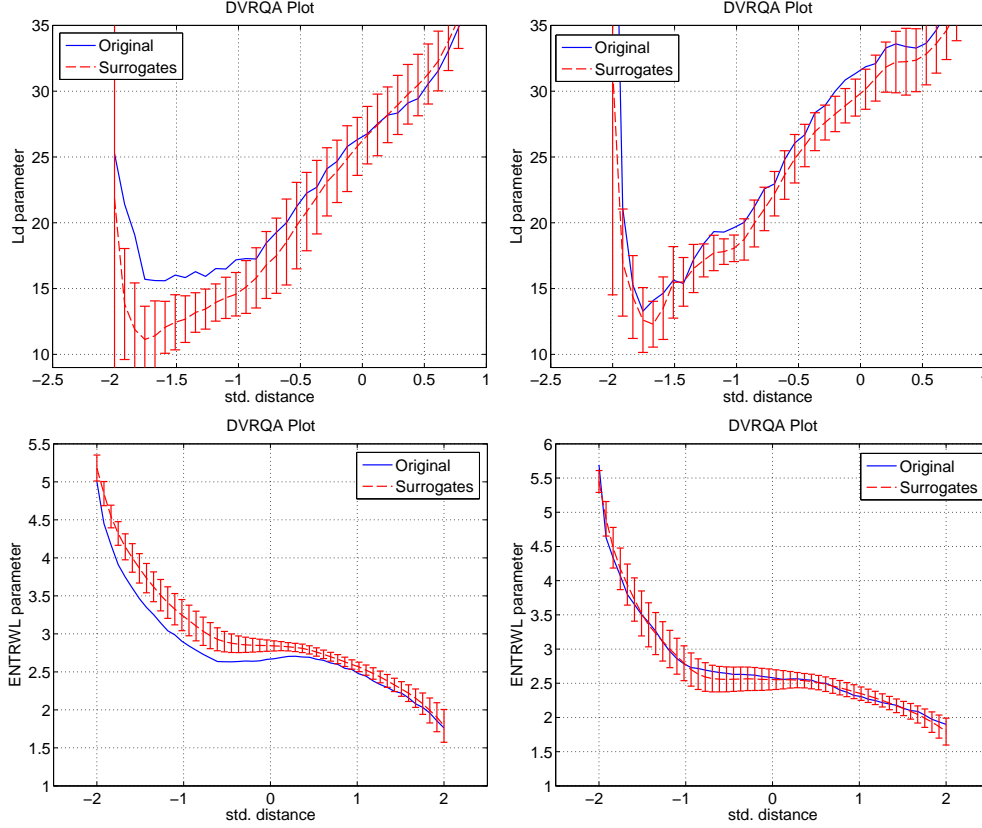
FIG. 6. (Color online) DVRQA of Rössler and quasi-periodic time series. Top left: DVRQA using Ld of a Rössler time series. Top right: DVRQA using Ld of a quasi-periodic time series. Bottom left: DVRQA using $ENTRWL$ of a Rössler time series. Bottom right: DVRQA using $ENTRWL$ of a quasi-periodic time series. The results were obtained using 50 surrogates.

analyzed a sound clip containing 46 airgun detonations recorded at $f_s = 8000Hz$. From each one of these 46 detonations, we selected a fragment of 400 samples (50 ms) which were tested for the presence of complex oscillations using the DVRQA algorithm. The 400-sample fragments were chosen around the maximum of each one of the detonation sounds and these fragments were carefully reviewed to ensure the presence of quasi-oscillatory signals. Figure 7 shows the waveforms of one of these 46 detonations.

As in the previous example, 50 surrogates were employed in the statistical test, and the
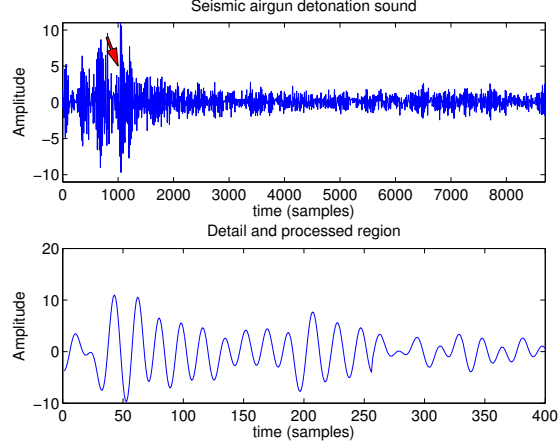
FIG. 7. (Color online) Top: Seismic airgun detonation waveform. The arrow indicates the starting point were the 400-sample fragment was extracted. Bottom: Zoom-in of one of the 400-sample segments processed with the DVRQA.

PPTS was computed using 10% of black points in the RP to obtain the twin points and $\rho = 0.25$. Figure 8 shows the results of the DVRQA for the acoustic detonation event # 27 using the Ld and the $ENTRWL$ parameter.
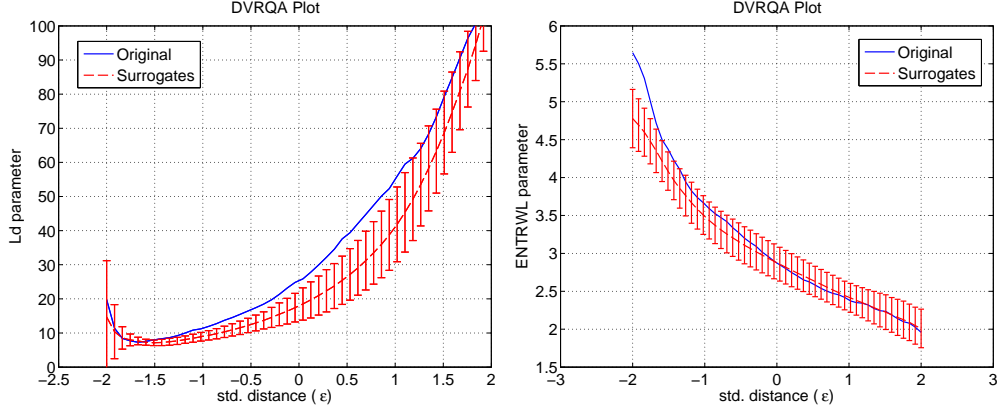
FIG. 8. (Color online) DVRQA of the detonation # 27. Left: computed using the Ld quantification parameter. Right: computed using the $ENTRWL$ quantification parameter. The results were obtained using 50 surrogates.

The analysis of the rest of the events with the DVRQA using the $ENTRWL$ parameter

indicates that 38 out of the 46 detonations lie outside the range of one standard deviation (similar results can be obtained with the DVRQA using the Ld parameter). This indicates that 82 % of the detonation acoustic events exhibit high complexity that is typically indicative of chaos or any nonlinear determinism. These results were obtained by computing the standard deviation of the white line entropy metric of the surrogate set ($\sigma_{ENTRWL}^{surro}(\varepsilon)$) and by looking for the events that lie outside this confidence interval. The bar graph in Figure 9 shows the maximum values exceeding $\max_{\varepsilon \in \varepsilon_s}(\sigma_{ENTRWL}^{surro}(\varepsilon))$ of the surrogates in the DVRQA plots. Using one standard deviation may seem to be a weak test because only 68 % of the values lie within one standard deviation of the mean. If we increase the standard deviation of the surrogates up to $2 \cdot \sigma_{ENTRWL}^{surro}(\varepsilon)$ so that nearly all (95 %) of the values lie within this interval, the number of detonation events that show a clear indication of non-linearity are only 22 % of the total number of analyzed detonation events.
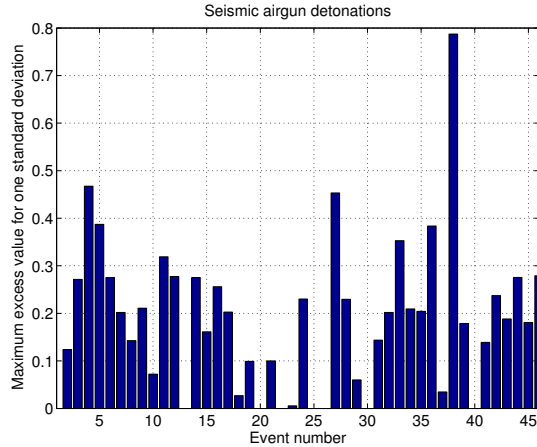
FIG. 9. (Color online) DVRQA test of the 46 seismic airgun detonation events recorded in the central Atlantic Ocean. The figure shows the $ENTRWL$ metric exceeding one standard deviation of the surrogate set.

V. CONCLUSIONS

When characterizing quasi-periodicity and chaos in oscillating short time series, we have illuminated the importance of taking into account the structures appearing in the RP. We have introduced a solution for the generation of surrogates which preserves some of these structures and also overcomes some practical implementation issues in limited-length signals. The proposed surrogate generation algorithm (PPTS) has been shown to allow for sufficient detail to be removed independently of the signal complexity without adding noise. We have established the utility of this method for testing the null hypothesis that the signal deviates from a quasi-periodic orbit. We have also demonstrated that the classical DVV method can be reformulated using RP concepts. This new formulation allows RQA-based tests to be designed without the need to fine tune the crucial parameter in this kind of tests; the recurrence threshold. The proposed algorithm, referred to as the DVRQA, has been shown to characterize short duration acoustic events and to be able to distinguish quasi-periodicity from chaos. We have chosen a RQA metric that captures the complexity in the return times (a parameter that is clearly related to chaos) and a parameter related to the mean diagonal length in the RP. Finally, we have illustrated the benefits of the proposed method on a practical application of characterizing seismic airgun detonations in the central Atlantic Ocean. The results indicate a clear deviation from quasi-periodicity in most of the detonations. Future work will focus on the design of automatic detectors for passive acoustic surveillance of air gun detonations in protected areas.

20

References

[1] Y. Zou, D. Paz, M. C. Romano, M. Thiel, and J. Kurths. Distinguishing quasiperiodic dynamics from chaos in short-time series. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 76(1 Pt 2):016210, July 2007.

[2] A. O. MacGillivray, R. Racca, and Z. Li. Marine mammal audibility of selected shallow-water survey sources. *The Journal of the Acoustical Society of America*, 135(1):EL35, 2014.

[3] D. Risch, P. J. Corkeron, W. T. Ellison, and S. M. Van Parijs. Changes in humpback whale song occurrence in response to an acoustic source 200 km away. *PLoS ONE*, 7(1):e29741, January 2012.

[4] Y. Li, B. J. Yang, J. Badal, X. P. Zhao, H. B. Lin, and R. L. Li. Chaotic system detection of weak seismic signals. *International Journal of Geophysics*, 178(3):1493–1522, September 2009.

[5] X. Luo, T. Nakamura, and M. Small. Surrogate test to distinguish between chaotic and pseudoperiodic time series. *Physical Review E*, 71(2), February 2005.

[6] T. Gautama, D. P. Mandic, and M. M. Van Hulle. A novel method for determining the nature of time series. *IEEE Transactions on Bio-medical Engineering*, 51(5):728–736, May 2004.

[7] D. Kaplan. Nonlinearity and nonstationarity: The use of surrogate data in interpreting fluctuations. *Proceedings of the 3rd Annual Workshop on Computer Applications of Blood Pressure and Heart Rate Signals*, pages 1–14, 1997.

[8] M. Golz, D. Sommer, M. Chen, U. Trutschel, and D. Mandic. Feature fusion for the detection of microsleep events. *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, 49(2):329–342, October 2007.

[9] D. et al. (Eds.) Mandic. *Signal processing techniques for knowledge extraction and information fusion.* Springer Science Business Media, New York, 2008.

[10] N. Marwan. How to avoid potential pitfalls in recurrence plot based data analysis. *In-*

ternational Journal of Bifurcation and Chaos, 21(04):1003–1017, April 2011.

[11] C. Tao and J. J. Jiang. Chaotic component obscured by strong periodicity in voice production system. *Physical Review E*, 77:061922, Jun 2008.

[12] J. J. Jiang, Y. Zhang, and C. McGilligan. Chaos in voice, from modeling to measurement. *Journal of Voice*, 20(1):2–17, March 2006.

[13] M. Small, D. Yu, and R. Harrison. Surrogate test for pseudoperiodic time series data. *Physical Review Letters*, 87(18), October 2001.

[14] M. Thiel, M. C. Romano, J. Kurths, M. Rolfs, and R. Kliegl. Twin surrogates to test for complex synchronisation. *Europhysics Letters*, 75(4):535–541, August 2006.

[15] M. C. Romano, M. Thiel, J. Kurths, K. Mergenthaler, and R. Engbert. Hypothesis test for synchronization: Twin surrogates revisited. *Chaos*, 19(1):015108, 2009.

[16] J.-P Eckmann, S. Oliffson Kamphorst, and D. Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters (EPL)*, 4(9):973–977, November 1987.

[17] M. T. Rosenstein, J. J. Collins, and C. J. De Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1-2):117–134, May 1993.

[18] T Gautama, D. P. Mandic, and M. M. Van Hulle. The delay vector variance method for detecting determinism and nonlinearity in time series. *Physica D*, 190:167–176, 2004.

[19] L.L. Trulla, A. Giuliani, J.P. Zbilut, and C.L. Webber Jr. Recurrence quantification analysis of the logistic equation with transients. *Physics Letters A*, 223(4):255 – 260, 1996.

[20] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E*, 66:026702, Aug 2002.

RQA PARAMETER		EQUATION
SYMBOL	**DESCRIPTION**	
RR	Recurrence rate: density of recurrence points.	$\dfrac{1}{N^2} \sum\limits_{i,j=1}^{N} R_{i,j}$
Ld	Averaged diagonal line length. [a]	$\dfrac{\sum_{l=l_{min}}^{N} l \cdot P(l)}{\sum_{l=l_{min}}^{N} P(l)} \quad (7)$
DET	Determinism: Percentage of recurrence points that form diagonal lines. [a]	$\dfrac{\sum_{l=l_{min}}^{N} l \cdot P(l)}{\sum_{i,j=1}^{N} R_{i,j}}$
ENTR	Shannon entropy of the length of black diagonal lines. [a]	$-\sum\limits_{l=l_{min}}^{N} P(l) \cdot \ln P(l)$
ENTRWL	Shannon entropy of the length of white vertical lines. [b]	$-\sum\limits_{w=w_{min}}^{N} P(w) \cdot \ln P(w) \quad (8)$

[a] $P(l)$ is the histogram of the lengths l of the black diagonal lines and l_{min} is the minimal length of what should be considered to be a diagonal line (typically $l_{min} = 2$).

[b] $P(w)$ is the histogram of the lengths w of the white vertical lines and w_{min} is the minimal length of what should be considered to be a vertical line (typically $w_{min} = 2$).

TABLE I. Some of the posible meassures for the recurrence quantification analysis (RQA).

List of Figures