Testing for intra-cycle determinism in pseudo-periodic time series

Mara C. S. Coelho*

Programa de Pós-Graduação em Engenharia Elétrica,
Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627,
31.270-901 Belo Horizonte, Minas Gerais, Brazil

Eduardo M. A. M. Mendes† and Luis A. Aguirre‡

Departamento de Engenharia Eletrônica,
Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627,
31.270-901 Belo Horizonte, Minas Gerais, Brazil

Abstract

A determinism test is proposed based on the well-known method of the surrogate data. Assuming predictability to be a signature of determinism, the proposed method checks for intra-cycle (e.g. short-term) determinism in pseudo-periodic time series for which standard methods of surrogate analysis do not apply. The approach presented is composed of two steps. First, the data are pre-processed to reduce the effects of seasonal and trend components. Second, standard tests of surrogate analysis can then be used. The determinism test is applied to simulated and experimental pseudo-periodic time series and the results show the applicability of the proposed test.

Keywords: Surrogate Method, Periodicity, Nonlinear Time Series Analysis
In the analysis of a time series, one of the first tests to be performed is to determine whether there is any signature of determinism underlying the data. The presence of determinism will suggest the construction of deterministic models or other ways of analysis, whereas the lack of determinism demands a statistical approach to modeling and analysis. In cases where the original time series exhibits a pseudo-periodic component, short-time intra-cycle determinism cannot generally be determined using standard surrogate methods. In this work a test is proposed to check for possible intra-cycle determinism in pseudo-periodic time series.

I. INTRODUCTION

The interest in detecting determinism in time series arose out of the need to discriminate a chaotic time series from a random time series in cases where the chaotic behavior was noise-like. In the eighties, the distinction between chaotic determinism and randomness was made by calculating well-known dynamic invariants such as correlation dimension and Lyapunov exponents. Later, several other methods were proposed for the analysis of determinism in time series. Among such methods, those based on surrogate data have become quite popular.

The surrogate time series is a set of data artificially generated from the original time series under investigation. Once the property under investigation (for instance nonlinearity, determinism, and so on) is defined, a method is sought to generate the surrogate data that preserves a specific characteristic of the original time series. A discriminant factor is then used to differ the original time series from the artificially set of time series generated by the chosen surrogate method.

Several algorithms have been proposed to generate the surrogate data in the literature. The most commonly used methods are: the RS (Random Shuffle) algorithm, the FT (Fourier Transform) algorithm and the AAFT (Amplitude Adjusted Fourier Transform) algorithm. These methods were designed to detect determinism in nonperiodic and stationary time series.

Economic, physics and biological real time series often exhibit periodic or pseudo-periodic characteristics that hide any possible underlying dynamics. See for an interesting discus-
sion of the importance of determining such dynamics in the context of load forecasting. At first, a visual inspection or a linear technique of analysis of time series, such as the autocorrelation function (ACF) and the power spectrum, could be attempted. However it is not always possible to check for the existence of deterministic dynamics in addition to the pseudo-periodic behavior. The algorithm SSS (Small Shuffled Surrogate) proposed by Nakamura and Small\textsuperscript{11} is perhaps the sole method available for such a purpose. The goal of this paper is then to propose an alternative approach to detect intra-cycle determinism in pseudo-periodic time series.

In this work, the definition of intra-cycle determinism is based on whether a time series can be predicted or not\textsuperscript{6}. This definition was also used in the work by Gomes et al\textsuperscript{12}. The background material, some definitions and the techniques borrowed from the statistical analysis of time series are given in Section II. In a statistical framework, the null hypothesis, $H_0$, to be tested is that the pseudo-periodic time series is random apart from its pseudo-periodic characteristic. $H_0$ will be rejected if the time series exhibits a deterministic characteristic in the sense just defined. This will be made more precise in Section III. Experiment results are given in Section IV and the conclusions in Section V.

II. BACKGROUND MATERIAL

In this section, a simple example is given to illustrate the problem of distinguishing the underlying dynamics in a given pseudo-periodic time series. A brief review of the surrogate method is also provided. In order to pre-process the data for further analysis, methods to remove seasonal and trend components are listed.

A. Statement of the problem

The sole use of the autocorrelation function (ACF) or the power spectrum to detect if there is deterministic dynamics underlying the pseudo-periodic behavior of a time series is not always successful. For instance, compare time series $y_1(k)$ and $y_2(k)$ in Figures 1(a) and 2(a). At first sight, there is no obvious qualitative difference between them. Could the autocorrelation function or the power spectrum be used to distinguish the underlying dynamics of such time series? To answer this question, consider the pseudo-periodic component $s(k)$
with seasonal periods $d_1 = 13$ and $d_2 \approx 314.16$ defined as follows

$$s(k) = 5\sin(\omega_1 k) - \sin(\omega_2 k),$$

where $\omega_1 = 2\pi/d_1$ and $\omega_2 = 2\pi/d_2$. The deterministic component of one of the time series shown in the Figure 1(a) is given by the following auto-regressive, AR(3), model:

$$x_1(k) = 0.9x_1(k - 1) + 0.02x_1(k - 2) + 0.03x_1(k - 3) + 0.2e(k),$$

where $e(k)$ is zero-mean, Gaussian white noise with variance equal to 0.04, $\mathcal{N}(0.0, \sqrt{0.04})$. The resulting pseudo-periodic time series with short-term deterministic signature can be obtained by adding the pseudo-periodic and deterministic components as follows

$$y_1(k) = s(k) + x_1(k). \quad (1)$$

The other pseudo-periodic time series, $y_2(k)$, with underlying random behavior can be generated by adding a random time series $x_2(k)$, with distribution $\mathcal{N}(0.0, \sqrt{0.2})$, to the same pseudo-periodic component as shown below

$$y_2(k) = s(k) + x_2(k). \quad (2)$$

The auto-correlation function and the power spectrum of time series $y_1(k)$ and $y_2(k)$ are shown in Figures 1(c)-(d) and Figures 2(c)-(d), respectively. Note that nothing can be said about the underlying dynamic behavior of the series. The linear techniques used in this example were not adequate to distinguish the underlying dynamic behavior.

The aim of this work is to detect intra-cycle deterministic dynamics. In other words, the proposed method should distinguish between $y_1(k)$ and $y_2(k)$. A review of the surrogate method on which the method is based is given next.

B. Hypothesis test based on surrogate data

The hypothesis test based on surrogate data is widely used to detect nonlinearity and determinism in time series\textsuperscript{13,14}. The hypothesis test consists of generating a set of artificial time series that preserves some statistical characteristics of the time series and destroy other statistical characteristics\textsuperscript{14}. The surrogate data set should be consistent with the null
FIG. 1: (a) Pseudo-periodic time series with short-term deterministic dynamics, \( y_1(k) \) in (1). (b) Power spectrum of \( y_1(k) \) with periods \( d_1 = 13 \) and \( d_2 = 100\pi \ldots \). (c) ACF of \( y_1(k) \).

hypothesis \( H_0 \) of interest.\(^1\) To carry on the hypothesis test, some statistical discriminant must be applied to the time series and to the surrogate data set. If the discriminant factor of the time series is significantly different from that one obtained from the set of surrogate data, the null hypothesis \( H_0 \) can be rejected as the generating mechanism of the original time series. If the opposite happens, then nothing can be said about the generating mechanism of the time series.

Few methods were developed to detect determinism in pseudo-periodic time series. The

\(^1\) \( H_0 \) attempts to define the nature of the underlying process that generated the original time series under analysis. The statistical framework under \( H_0 \) establishes whether the process can or cannot be appropriately explained by the available data.
algorithm CSS (Cycle Shuffled Surrogate) proposed by Theiler et al.\textsuperscript{8} searches for \textit{inter-cycle} determinism by shuffling the seasonal cycle within a time series. The algorithm PPS (Pseudo-Periodic Surrogate) proposed by Small et. al\textsuperscript{15} also searches for \textit{inter-cycle} determinism, but the authors use different embeddings of the time series to generate the surrogate data. The null hypothesis, $H_0$, in those methods is different from $H_0$ used in this work because they search for \textit{inter-cycle} determinism rather than \textit{intra-cycle} determinism.

The algorithm SSS (Small Shuffled Surrogate) proposed by Nakamura and Small\textsuperscript{11} aims to detect \textit{intra-cycle} determinism by shuffling of the pseudo-periodic time series locally; the same goal of the method proposed in this work. As shown in paper\textsuperscript{11} the algorithm SSS has been successfully applied to several cases such as the cobalt data and daily sunspots numbers. However, SSS can fail to detect determinism in pseudo-periodic time series with
different underlying behavior. For instance, the algorithm SSS is applied to the time series \( y_1(k) \) and \( y_2(k) \) shown in Figures 1(a) and 2(a), respectively. For the time series \( y_1(k) \) with 4000 observations, a set of 400 surrogate data was generated using the algorithm SSS\(^{11}\). Figure 3 (a) shows one example of such surrogate series. One of the discriminant factors used in\(^ {11}\) is the ACF. In the same paper, the authors show that if ACF of the original series fall within the distributions of the surrogate data they consider that the original and the surrogate data may come from the same population and then the surrogate null hypothesis may not be rejected. This procedure was also followed in this work.

The ACF of \( y_1(k) \) and the ACF of the 400 surrogate time series are shown in Figure 3(b). Note that the ACF of the time series \( y_1(k) \) is different from the ACF of the surrogate data sets. Therefore, it can be concluded that the pseudo-periodic time series \( y_1(k) \) exhibits short-term determinism.

For the pseudo-periodic time series with underlying random behavior \( y_2(k) \) shown in Figure 2(a), a set of 400 surrogate data was generated. The ACF of both \( y_2(k) \) and the surrogate data are shown in Figure 4(b). It can be noticed that the ACF of \( y_2(k) \) is also different from the ACF of surrogate data sets which implies that the underlying behavior of \( y_2(k) \) is different from the surrogate data. In other words, it can be inferred that \( y_2(k) \) has short-term determinism in addition to the pseudo-periodic behavior. This result is not in agreement with what is known about \( y_2(k) \).

It is worth pointing out that, in order to detect determinism in pseudo-periodic time series, the three aforementioned methods - CSS, PPS and SSS - generate surrogate data in a different fashion when compared to the work of Theiler\(^ {8}\). However these methods were developed to be directly applied to the pseudo-periodic time series avoiding the necessity of pre-processing the time series under investigation. In this work, the test of determinism combines some available and well established methods in the literature to generate the surrogate data\(^ {8}\). In order to use the proposed algorithm, the pseudo-periodic time series must be pre-processed, since the aforementioned methods cannot be applied to periodic time series. In particular, it is necessary to extract pseudo-periodic and trend components of the pseudo-periodic time series.

In the literature of time series, traditional approaches for modeling seasonal time series are to remove the seasonal components using seasonal adjustment methods and then the models are scaled back using the estimated seasonal components for forecasting purposes.
FIG. 3: (a) Pseudo-periodic time series with short-term deterministic dynamics, \( y_1(k) \) in (1) in (-x-), one of surrogate data (-o-). (b) ACF of \( y_1(k) \) and of the surrogate data sets.

FIG. 4: (a) The pseudo-periodic time series with underlying random behavior \( y_2(k) \) in (2) in (-x-) and, one of the surrogate data (-o-). (b) ACF of \( y_2(k) \) and of the surrogate data sets.

To the best of our knowledge there is no mathematical proof that preprocessing does not affect the underlying dynamics. For a recent debate on whether these adjustment methods affect directly the deterministic components or not, please see\(^{16}\).

One of the most used techniques in financial time series to extract seasonal and trend components is known as X-11-ARIMA\(^{17}\) and its extended version X-12-ARIMA\(^{18}\). Ano-
ther well-known method used to reduce seasonal components is the moving average seasonal filter\textsuperscript{19}. This filter will be used in this work and will be reviewed in Section II C.

C. Reduction of seasonal and trend components

A time series may be represented as the combination of four distinct components: trend, cyclical, seasonal and residual\textsuperscript{19}. In order to isolate these components, several procedures to extract them from a time series have been developed.

This section aims at briefly describing a method to estimate and reduce trend and seasonal components. As a result a residual time series to which the determinism test will be applied is obtained. Quite possibly (the degree of uncertainty is imposed by the user) the original pseudo-periodic time series has intra-cycle determinism behavior if determinism is found in the residual time series. In the statistical framework, the null hypothesis can be safely rejected.

D. The seasonal moving average filter method (SMAF)

The pseudo-periodic time series can be represented by the following classic additive decomposition model\textsuperscript{19}:

\[ y(k) = m(k) + s(k) + x(k), \quad k = 1, \ldots, n, \quad (3) \]

where \( n \) is the number of observations, \( m(k) \) is the trend component, \( s(k) \) is the seasonal component with seasonal period \( d \), and \( x(k) \) is the residual component.

The components \( m(k), s(k) \) and, \( x(k) \) in (3) will be estimated by following a sequence of steps, detailed below.

Considering the original time series \( y(k) \) with \( n \) observations, the trend component \textit{over the sliding window}, \( \hat{m}_0(k) \), can be estimated by applying a moving average filter as follows (\( \hat{m}(k) \) will be estimated later)

\[
\begin{cases}
\hat{m}_0(k) = \frac{0.5[y(k-q)+y(k-q+1)+\ldots+y(k+q-1)+0.5y(k+q)]}{d}, & \text{for } d \text{ even, then } q = d/2, \\
\hat{m}_0(k) = \frac{1}{2q+1} \sum_{i=-q}^{q} y(k-i), & \text{for } d \text{ odd, then } q = (d-1)/2, 
\end{cases}
\quad (4)
\]

where \( q < k \leq (n - q) \).
The next step consists of removing from the data the trend component, that is,

\[ \alpha(k) = y(k) - \hat{m}_0(k), \quad q < k \leq (n - q). \]

Because of edge effects due to the moving average filter (4), \( \hat{m}_0(k) \) and \( \alpha(k) \) are pruned at each end, notice that such series exist for \( q < k \leq (n - q) \).

Subsequently, it is desired to find the “average cycle” among all the complete cycles within \( \alpha(k) \). This average is a sample average, that is, we take the average of all the first values of each cycle in \( \alpha(k) \), then the second values are considered, and this is repeated until the \( d \)th values (the last values) of each complete cycle in \( \alpha(k) \). Mathematically, this can be expressed as

\[
\hat{w}(k_0) = \frac{1}{j_c} \sum_{j=1}^{j_c} \alpha((j-1)d + k_0), \quad k_0 = 1 \ldots d,
\]

where \( j_c \) is the number of complete cycles in \( \alpha(k) \). Because \( \hat{w}(k_0) \) can have an offset, \( \overline{w} \), this is removed thus

\[
\beta(k_0) = \hat{w}(k_0) - \overline{w}, \quad k_0 = 1, \ldots, d.
\]

Now a periodic time series, composed of the average zero-mean cycle \( \beta(k_0) \), has to be produced. This is done as

\[
\hat{s}(k_0) = \beta(k_0), \quad k_0 = 1, \ldots, d \\
\hat{s}(d + k_0) = \beta(k_0), \quad k_0 = 1, \ldots, d \\
\hat{s}(2d + k_0) = \beta(k_0), \quad k_0 = 1, \ldots, d \\
\hat{s}((j-1)d + k_0) = \beta(k_0), \quad k_0 = 1, \ldots, d \text{ and } j = 1, \ldots, j_c.
\]

The time series without the seasonal component is obtained as

\[
\hat{D} = y - \hat{s},
\]

where it is important that the data \( y \) and \( \hat{s} \) be in phase.

Now we are able to estimate \( \hat{m}(k) \). Notice that \( \hat{m}_0(k) \) is the trend component along the sliding window and not of the time series as a whole. \( m(k) \) can be estimated from \( \hat{D}(k) \) as

\[
\hat{m}(k) = a_0 + a_1 k + a_2 k^2 + \ldots + a_j k^j.
\]
where $j$ is the degree of the polynomial chosen by visual inspection. Finally, the residual time series can be obtained as:

$$\hat{x} = y - \hat{m} - \hat{s}. \quad (10)$$

In equations (8) and (10) the time index $k$ was omitted to call attention to two facts. First, $y$ is longer than $\hat{s}$. Although $\hat{m}(k)$ can be, in principle, evaluated at any value of $k$, it is strongly recommended that $k$ should range only within the limits of $\hat{D}$, which is shorter than $y$. Second, $y$ and $\hat{s}$ must be in phase. In practice, these remarks become obvious and are easily implemented.

E. Application of the SMAF method to reduce seasonal and trend components

The SMAF method described in the section II C is applied to the two simulated pseudo-periodic time series, $y_1(k)$ and $y_2(k)$, shown in Figures 1(a) and 2(a) respectively, and to four experimental pseudo-periodic time series, namely:

- $y_C(k)$ is the reported monthly cases of mumps during 1928-1972 in New York City, with seasonal period $d = 12$. The corresponding residual time series is represented by $\hat{x}_C(k)$, shown in Figure 6(a) and (b), respectively.

- $y_L(k)$ is the monthly observations of milk production during 1962-1975. This time series has the seasonal period $d = 12$. The corresponding residual time series is represented by $\hat{x}_L(k)$, shown in Figures 7(a) and (b), respectively.

- $y_Q(k)$ is the hourly consumption of electric power (electric load) with seasonal period $d = 24$. The corresponding residual time series is represented by $\hat{x}_Q(k)$, shown in Figures 8(a) and (b), respectively.

- $y_E(k)$ is the monthly consumption of electric power (electric load) with seasonal period $d = 12$. The corresponding residual time series is represented by $\hat{x}_E(k)$, shown in Figures 9(a) and (b), respectively.

For the simulated pseudo-periodic time series, $y_1(k)$ and $y_2(k)$, depicted in Figure 1 and 2 respectively, the corresponding residual time series, $\hat{x}_1(k)$ and $\hat{x}_2(k)$ are shown in Figure 5(a) and (d), respectively.
From Figures 5(a) and (d) it can be noticed that the SMAF method has successfully removed the seasonal component. Therefore, the resulting residual time series, \( \hat{x}_1(k) \) and \( \hat{x}_2(k) \), do not show seasonal behavior with respect to \( d_1 = 13 \) and \( d_2 \approx 314.15 \), as can be seen in the power spectrum and the ACF of residual time series (Figures 5(b)–(c) and Figures 5(e)–(f), respectively). Therefore these results show that the procedure of removing the cycles did not affect the deterministic component.

The SMAF method for reduction of pseudo-periodic component has also a good performance when used to reduce seasonal and trend components of the experimental pseudo-periodic time series \( y_C(k), y_L(k), y_Q(k) \) and \( y_E(k) \). From the analysis of the ACF and power spectrum shown in Figure 6 (c) and (d), it can be verified that the seasonal component with \( d_1 = 12 \) and \( d_2 = 31 \) of the mumps time series, \( y_C(k) \), were removed. The same occurs with the time series of milk, \( y_L(k) \), with with \( d = 12 \), as shown in Figure 7 (c) and (d). Likewise, in the case of the load time series \( y_Q(k) \), the seasonal component with \( d_1 = 168 \) was strongly reduced. However a lesser reduction of the seasonal component with \( d_2 = 24 \) can be observed (see Fig. 8 (c) and (d)). In principle, unremoved portions of the seasonal component can affect the performance of the proposed method, although this seems not to have been the case in the examples given here, as argued in Section IV.

The reduction of the pseudo-periodic component in the time series \( y_E(k) \) also had a good performance as it can be seen from the analysis of the ACF and power spectrum shown in Figure 9 (c) and (d), it can be verified that the seasonal component with \( d = 12 \) of the electric load time series, \( y_E(k) \), were removed.

III. METHODOLOGY

In order to evaluate whether there is a deterministic feature in the original time series a discriminating statistic must be chosen. Since the objective of this work is to verify the existence of intra-cycle determinism in pseudo-periodic time series based on the existence of some predictability, the original time series and the surrogate data set are submitted to NARMA (Nonlinear Autoregressive and Moving Average model) polynomial modeling\(^{20-23}\) and subsequent prediction.

To quantify the forecasting performance obtained, each time series used in the test (ori-
FIG. 5: Simulated pseudo-periodic time series. (a) Residual time series $\hat{x}_1(k)$. (b) Power spectrum of the residual time series $\hat{x}_1(k)$ and, in (c) ACF of $\hat{x}_1(k)$. (d) Residual time series $\hat{x}_2(k)$. (e) Power spectrum of the residual time series $\hat{x}_2(k)$ and, (f) ACF of $\hat{x}_2(k)$. The interval indicated by dashed lines corresponds to the 95% confidence level.

Original times series and the surrogates) is divided into $N_w$ intervals (windows) of equal length. The forecast error index mRSE (Mean Root Square Error) is calculated over each window as follows

$$mRSE(i) = \frac{1}{N_w} \sum_{w=1}^{N_w} \frac{\sqrt{(y_w(i) - \bar{y}_w(i))^2}}{\sqrt{(y_w(i) - \bar{y}_w)^2}}$$

(11)

where $i$ is the forecasting step and can range from 1 to $h$, where $h$ is the maximum forecasting horizon. $w = 1 \ldots N_w$ is the window used for forecasting, where $N_w$ is the maximum number of windows for which the average is calculated. In this work, $N_w = 10$. $y_w(i)$ refers to time series in the interval (window) $w$ in the instant $i$, $\bar{y}_w$ is the predicted value of time series in the interval $w$ obtained from the model and $\bar{y}_w$ is the mean value over the window $w$. 

13
FIG. 6: Experimental pseudo-periodic: (a) time series of the monthly cases of mumps with $d = 12 \, y_C(k)$, and (b) corresponding residual series $\hat{x}_C(k)$. (c) ACF and (d) power spectrum of the time series $y_C(k)$ (---) and $\hat{x}_C(k)$ (—).

A. The proposed test for detecting determinism

The proposed method for testing determinism in time series is based on the idea that if the time series is predictable to some extent, it exhibits some determinism\textsuperscript{12}. In other words, predictability (even limited) is a signature of determinism. In paper\textsuperscript{12} the seasonal component of the time series was not extracted; The FT and AAFT algorithms\textsuperscript{8} were used to generate the surrogate data set and; the index RMSE was used and it is calculated for every window. The proposed algorithm for testing intra-cycle determinism can be summarized as:

1 - from $y(k)$, estimate and remove the seasonal and trend components, yielding $\hat{x}(k)$;

2 - from $\hat{x}(k)$, generate the surrogate time series;
FIG. 7: (a) Experimental pseudo-periodic time series of the monthly milk production with \( d = 12 \) \( y_L(k) \), and (b) corresponding residual series \( \hat{x}_L(k) \). In (c) ACF and (d) power spectrum of the time series \( y_L(k) \) (--) and \( \hat{x}_L(k) \) (—).

3 - build mathematical models from \( \hat{x}(k) \) and the surrogate time series;

4 - using the models just built, forecast from 1-to-\( h \) steps ahead and calculate the forecasting error index mRSE given by equation (11);

5 - test of the null hypothesis \( H_0 \).

In this paper, step 1 is performed using the seasonal moving average filter (SMAF) method described in Section II.C.
FIG. 8: (a) Experimental pseudo-periodic time series of electric load time series with \( d = 24 \) \( y_Q(k) \), and (b) corresponding residual series \( \hat{x}_Q(k) \). (c) ACF and (d) power spectrum of the time series \( y_Q(k) \) (- -) and \( \hat{x}_Q(k) \) (---).

The second step is performed using the Shuffled Algorithm\(^2\) to generate the surrogate time series described in paper\(^8\).

In the third step, \( \hat{x}(k) \) and the surrogate time series are divided into two windows: one for identification, another for validation. This step is performed as described in\(^12\). It should be noticed that other modelling techniques can be used here.

After building the models from both the original time series and the surrogate series, 1-to-\( h \) step-ahead predictions are determined over the validation data. The mRSE in (11) is calculated for both the original time series and the surrogate data set with respect to each

\(^2\) It is important to emphasize that the Random Shuffled Algorithm proposed in\(^8\) was not developed for time series with pseudo-periodic behavior. Therefore, to use this algorithm in pseudo-periodic time series the seasonal component of the time series must be reduced as much as possible.
prediction step. The mRSE and the confidence interval obtained from the predictions for surrogate data sets are plotted and compared with the performance of the predictions over the validation window of \( \hat{x}(k) \).

In the determinism framework established in this paper, and assuming that pseudo-periodic component was removed from the time series under investigation, \( y(k) \), the null and alternative hypotheses can be defined with respect to the index mRSE as follows:

- \( H_0 : \hat{x}(k) \) is random. \( H_0 \) cannot be rejected if mRSE \( \approx 1 \).
- \( H_1 : \hat{x}(k) \) displays short-term determinism. \( H_0 \) can be rejected and \( H_1 \) accepted if mRSE < 1.
- \( H_2 : \hat{x}(k) \) is nonstationary or cannot be represented by the chosen model class. \( H_0 \)
can be rejected and $H_2$ accepted if mRSE > 1.

B. Confidence Intervals

In order to define the degree of confidence of the test, confidence intervals are calculated. To this end, it is necessary to know the distribution of the mRSE index. Several examples show that the mRSE index does not have a normal distribution for short prediction horizons. See for instance Figure 10. In this particular case, a set of 1000 surrogate sequences was generated and a model was fitted to each one of the surrogate time series. Each model was used to predict from $i = 1, 2, 3, 4$ steps ahead over windows of length $h$. All our examples use $h = 10$. At each forecasting step the prediction error of each model was computed and the mRSE index was calculated.

For $i = 1$ and $i = 2$ the distribution of the index mRSE$(i)$ is closer to a chi-squared distribution than to a normal distribution. See Figures 10 (a) and (b), respectively. As the prediction horizon increases, the distribution of the index mRSE$(i)$ approaches a Normal distribution as shown in Figures 10 (c) and (d). This is an interesting result. The change in the distribution of mRSE$(i)$ as $i$ increases seems to indicate the loss of predictability when the distribution gets closer to a Normal distribution.

Since the distribution of the mRSE$(i)$ changes for different prediction horizons, the confidence interval in this work is calculated using the interval among percentis and ignoring the underlying distribution. The maximum confidence level is established gradually by increasing the confidence level of the determinism test starting from the initial value of 50% up to the point for which $H_0$ cannot be rejected anymore. It should be noticed that to reject $H_0$ at 50% confidence is equivalent to decide between the null and the alternative hypotheses based on a coin toss. So, it is expected that $H_0$ should be rejected at a much higher confidence level than 50%.

C. Performance of intra-cycle determinism test

The proposed determinism test was applied to the two simulated pseudo-periodic time series, that is, $y_1(k)$ given by equation (1) and $y_2(k)$ given by equation (2), shown in Figures 1(a) and Figures 2(a), respectively.
The results of intra-cycle determinism in $y_1(k)$ and $y_2(k)$ are shown in Figure 11 (a) and (b), respectively. As stated before, the proposed method applies a random shuffle surrogate test to the residual time series $\hat{x}_1(k)$ and $\hat{x}_2(k)$. From Figure 11 (a) it can be noticed that the index mRSE from $\hat{x}_1(k)$ is smaller than the mean mRSE obtained from surrogate data set. The test suggests that $H_0$ can be rejected and $H_1$ accepted at the confidence level of 96%. It should be noticed that in Fig. 11(a) the dotted lines show the limiting case, that is, if the confidence interval is increased any further, $H_0$ cannot be rejected anymore. In Fig. 11(a) the limiting value is 96%. This means that $\hat{x}_1(k)$ is likely to have short-term determinism and consequently $y_1(k)$ has an intra-cycle deterministic characteristic.

The determinism test for the residual time series $\hat{x}_2(k)$ shows that the mRSE of $\hat{x}_2(k)$ is closer to the mRSE of the set of surrogate time series and $H_0$ cannot be rejected at the standard level of confidence (95%). Therefore it is very unlikely that $y_2(k)$ has any intra-cycle deterministic signature.

These results show that the proposed procedure works well in this simulated example. Next section the procedure will be tested in real data sets.

IV. EXPERIMENTAL RESULTS WITH PSEUDO-PERIODIC TIME SERIES

In this section, four experimental pseudo-periodic time series are used: the time series of production of milk (N=168 observations)\(^25\); the time series of indexes of mumps occurrence (N=528 observations)\(^26\); the *hourly* time series of electric power consumption (electric load,
FIG. 11: Result of the determinism test applied the residual series: in (a) $\hat{x}_1(k)$, where the test suggests that $H_0$ can be rejected (and $H_1$ accepted) at a confidence level of 96%. The dotted lines show the 96% confidence bands. In (b) $\hat{x}_2(k)$, $H_0$ cannot be rejected even at the standard confidence level of 95%. The upper plot shows the residual time series, the middle plot shows one of the surrogate data, and the under plot shows the mRSE($h$): where mRSE($h$) for residual time (—), mean mRSE($h$) for surrogate data (- -), and confidence interval of the mean mRSE for the surrogate data (···).

N=5616 observations)\(^3\), and the monthly time series of electric load (N=116 observations) (COLCOAR A REFERENCIA). The application of the SMAF filter to these data was described in Section II C.

A. Monthly cases of mumps

The determinism test based on surrogates data is applied to the residual series $\hat{x}_C(k)$ shown in Figure 6(b) and the result is shown in Figure 12(a). The determinism test indicates that $H_0$ can be rejected (and $H_1$ accepted) at a confidence level of virtually 100%. This result implies that the residual time series $\hat{x}_C(k)$ is predictable up to 2 steps ahead when the seasonality has been removed from $y_C(k)$. Consequently, assuming the seasonality has been removed $y_C(k)$ is very likely to have intra-cycle deterministic features.

\(^3\) The National Electricity Market Management Company (NEMMCO).
B. Monthly milk production time series

The determinism test based on surrogates data is applied to the residual series $\hat{x}_L(k)$ shown in Figure 7(b), and the result of the determinism test shown in Figure 12(b) indicates that $H_0$ can be rejected and $H_1$ accepted at a confidence level of practically 100%. This result implies that such residual time series $\hat{x}_L(k)$ is predictable up to 5 steps ahead and consequently $y_L(k)$ is very likely to have intra-cycle deterministic behavior.

C. Hourly Electric load

The result of the determinism test applied to the residual time series $\hat{x}_Q(k)$ depicted in Figure 8(b) is shown in Figure 12(c) and suggests that $H_0$ can be rejected and that $H_1$ accepted with practically 100% confidence. This result implies that the residual time series $\hat{x}_Q(k)$ is predictable and consequently $\hat{x}_Q(k)$ is very likely to have intra-cycle deterministic features when the seasonality is removed.

D. Monthly Electric load

The result of the determinism test applied to the residual time series $\hat{x}_E(k)$ depicted in Figure 9(b) is shown in Figure 13 and suggests that $H_0$ can not be rejected with practically 100% confidence. This result implies that the residual time series $\hat{x}_E(k)$ is not predictable and consequently $\hat{x}_E(k)$ is not have intra-cycle deterministic features when the seasonality is removed.

V. CONCLUSIONS

In this work, the intra-cycle determinism test has been proposed. The method deals with the problem of finding dynamics underlying a pseudo-periodic component. To this end, the time series under investigation should have the seasonality and trend removed. When the method was applied to simulated data with known characteristics determinism was successfully detected. In real data, where short-term intra-cycle determinism seems a reasonable hypothesis but cannot be asserted, the proposed method was also able to detect determinism. In the examples given in this work the indication of determinism intra-cycle
FIG. 12: (a) Result of the determinism test applied the residual series: in (a) of the $\hat{x}_C(k)$; in (b) of the $\hat{x}_L(k)$; and in (c) of $\hat{x}_Q(k)$. For the three time series — $\hat{x}_C(k)$, $\hat{x}_L(k)$, and $\hat{x}_Q(k)$ — the dotted lines indicate the 95% confidence bands. The determinism test indicates that $H_0$ can be rejected and $H_1$ accepted with practically 100% confidence. In both the figures, the upper plot shows the residual time series $\hat{x}_L(k)$, the middle plot shows one of the surrogate data set, and the under plot shows the mRSE($h$): in (- -) mean mRSE($h$) for surrogate data set, in (···) confidence interval of 95% of the mean mRSE for the surrogate data set, and in (—) mRSE($h$) for both residual time series $\hat{x}_C(k)$, $\hat{x}_Q(k)$ and $\hat{x}_L(k)$.

implies that, in real time series such as milk and mumps, the best predictor is not the mean cycle model (known as the trivial model) and therefore modeling techniques are worth investigating. Consequently a tool to help decide when is modeling worthwhile and when is not is of great practical importance. Finally the proposed method comes as an addition to
FIG. 13: (a) Result of the determinism test applied the residual series of $\hat{x}_E(k)$, where the dotted lines indicate the 95% confidence bands. The determinism test indicates that $H_0$ can not be rejected with practically 100% confidence. In figure, the upper plot shows the residual time series $\hat{x}_E(k)$, the middle plot shows one of the surrogate data set, and the under plot shows the mRSE($h$): mRSE($h$) for $\hat{x}_E(k)$ (—), mean mRSE($h$) for surrogate data set ( - -), and confidence interval of 95% of the mean mRSE for the surrogate data set (···).

the few methods available in the literature for dealing with pseudo-periodic time series.

---


