30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

Parametric and Nonparametric Methods to Generate Time-Varying Surrogate Data

He Zhao, Luca Faes, Giandomenico Nollo, and Ki H. Chon

Abstract— **We present both nonparametric and parametric approaches to generating time-varying surrogate data. Nonparametric and parametric approaches are based on the use of the short-time Fourier transform and a time-varying autoregressive model, respectively. Time-varying surrogate data (TVSD) can be used to determine the statistical significance of the linear and nonlinear coherence function estimates. Two advantages of the TVSD are that it keeps one from having to make an arbitrary decision about the significance of the coherence value, and it properly takes into account statistical significance levels, which may change with time. Our simulation examples and experimental results on blood pressure and heart rate data demonstrate the efficacy and applicability of the proposed TVSD methods.**

I. Introduction

Theiler developed a time-invariant surrogate data technique for statistical evaluation of the presence of nonlinear dynamics in time series [1]. Improved surrogate data techniques have followed thereafter [2] and have found applications in many different disciplines [2, 3]. While these time-invariant surrogate data techniques were originally developed to determine the presence of nonlinearity, they have also been applied to evaluating the statistical significance of linear time-invariant coherence functions [4, 5].

With recent advances in the development of time-varying coherence techniques [6, 7], the need for a time-varying surrogate data technique is apparent. Surrogate data techniques are designed to destroy any coupling present in the signal, and because they are designed to generate multiple realizations of the non-coupled data, the statistical significance of the coherence can be evaluated. Without surrogate data, quantification of the strength of coherence is arbitrary as any values higher than 0.5 are considered to be an indication of highly coherent signals. Thus, this arbitrary demarcation is most appropriate for highly coherent signals and incorrectly ignores any coherent values that are less than 0.5. Time-invariant surrogate data techniques remove this bias toward only the highly coherent signals, and were found to be sensitive even for weakly coupled signals [1, 8]. However, they are inappropriate for time-varying data as they provide time averaged statistical significance values. Thus, in a case where coherence values wax and wane with time, and waning of the coherence values is more prevalent, the time-invariant surrogate data technique is most likely to result in the incorrect interpretation that there is no coherence for all time points. However, with the TVSD technique, the statistical significance of the time-dependent changes in the observed coherence values at a particular frequency can be evaluated. It should be noted that the TVSD method is also applicable to TV bispectrum analysis as it can be used to determine the statistical significance of the nonlinearly coherence values.

In this paper, we introduce a short-time Fourier transform (STFT) and a time-varying autoregressive model based approach to generate time varying surrogate data for nonstationary signals. The method has been tested using simulation examples and applied to experimental data. These results are provided in Results section.

II. Method

A. Short-Time Fourier Transform

Our approach to generating time-varying surrogate data is more similar in concept to the STFT for nonstationary signals than it is to the power spectral density for time-invariant systems. Our technique, similar to the STFT, is to segment the data and compute surrogate time series for each of the segmented time series. Within each segment, the signal is assumed to be stationary. The length of the segment depends on the trade-off between time and frequency resolution as well as the validity of the stationarity assumption within the chosen segment length. For example, for highly time-varying systems, small segment lengths are necessary, but the consequence is decreased frequency resolution, and vice versa.

Based on the stationarity assumption within the chosen segment length, we can then use any of the many known time-invariant surrogate data techniques. We chose the iteratively refined surrogate data technique (IRSDT) [8]. The IRSDT will destroy any nonlinearity in the signal, and has been shown to be more accurate than the amplitude adjusted Fourier transform technique [1] because it iteratively corrects for deviations in the spectrum as well as maintains the correct distribution of the signal.

Manuscript received April xx, 2008.

H. Zhao is with the Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794 USA.

L. Faes is with the Laboratorio Biosegnali, Dipartimento di Fisica, Università di Trento,and INFM, 38050 Povo, Trento Italy.

G. Nollo is with the Dipartimento di Fisica, Università di Trento, and ITC-irst, 38050 Povo, Trento, Italy.

K. H. Chon is with the Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: ki.chon@sunysb.edu).

B. Time-Varying Autoregressive (TVAR) Model

The first step involves fitting a TVAR model to the time series. While many different TVAR methods are appropriate, the recursive least squares (RLS) was used to estimated TVAR coefficients. It should be noted that AR modeling in general is appropriate for slowly time-varying systems. A properly chosen AR model should result in a low mean-square error (MSE) or a residual that is essentially white noise, which implies that the linear dynamics of the signal are well accounted for. The second step involves generation of surrogate data realizations by regressing the TVAR coefficients with white noise signals. This last procedure is equivalent to a TV moving average process, which yields a nearly identical spectrum to that of the original time series but which is uncorrelated with the original time series.

III. Results

A. Simulation Results

Surrogate data technique can serve a dual purpose as it can be used to detect both nonlinearity and nonstationarity in the signal. For the simulation examples to follow, time-varying open-loop [6] and causal coherence functions [6], and their time-varying surrogate data, were generated. The time-invariant causal coherence function was introduced by Porta *et al.* [4], and we have extended it to be applicable for time-varying systems [6]. The threshold values for statistical significance of the time-varying surrogate data to be provided in the proceeding simulations for both STFT and TVAR were based on the mean plus two standard deviations, which were derived from 20 realizations. Any time-varying coherence value (both causal and open-loop) greater than the statistically significant threshold value for each frequency represents 95% statistical confidence that it did not occur by some random occurrence. The choice of AR model order for the RLS was based on the use of minimum description length. Further, we assumed the chosen model order to be stationary.

The first simulation was adopted from Porta et al. [4], but with varying coupling magnitudes. A simulation model of a time-varying closed-loop system consisting of two 512 data point segments each for signals y_1 and y_2 was generated using the following expression:

$$\begin{cases} y_1(n) = -c_{21}(n)\, y_2(n-2) + e_1(n) \\ y_2(n) = c_{12}(n)\, y_1(n-1) + e_2(n) \end{cases} \quad (1)$$

where e_1 and e_2 are Gaussian white noise with unit variance, and c_{21} and e_{12} represent time-varying coupling coefficients with subscripts indicating the directionality of the coupling strengths. The model structure indicative of Eq. (1) is shown in the top panel of Fig. 1, and the time-varying coupling magnitudes we have chosen for this simulation example are shown in the bottom panel of Fig. 1. For the first half of the data, only the coupling coefficient c_{21} is present, while in the last half of the data, only the coupling coefficient c_{12} is present.

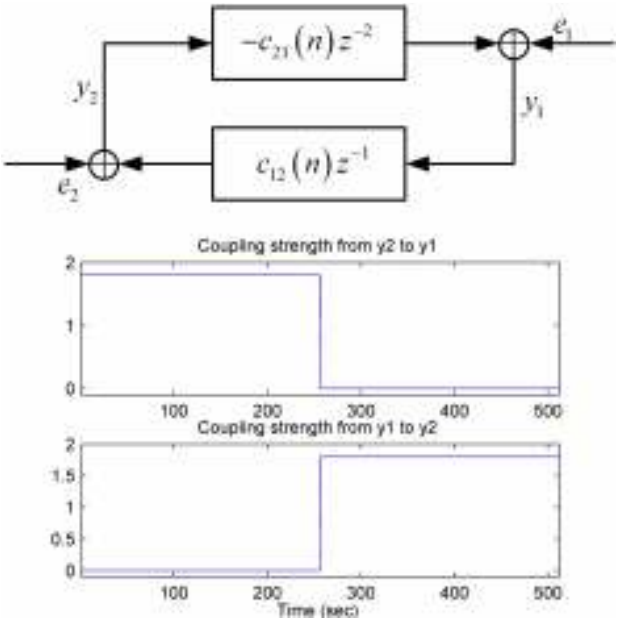

Fig. 1. Time-varying coupling strength of the first simulation example.

Time-varying causal coherence function estimates (TVCCF) are illustrated in Fig. 2a. The TVCCF estimates were based on the method we have developed, which determines only the significant terms among the initially chosen model candidate terms [6, 9]. For all simulations as well as application to the experimental data, the initial model order selected was 10 autoregressive and 10 moving average terms. The left and right panels show the TVCCF estimates from y_2 to y_1 and vice versa, respectively. With high coupling magnitudes for c_{21}, we correctly observe high coherence values (~0.8) for the first half of the data followed by low coherence function values thereafter since the coupling strengths were switched from high to low values, as shown in the bottom panel of Fig. 1. The right panel of Fig. 2a correctly shows a transition of low to high coherence values with increasing time since the coupling magnitudes were designed to have high values only in the last half of the data segment. Results based on the time-varying surrogate data generated by the STFT and the RLS are shown in Fig. 2b and Fig. 2c, respectively. To estimate the signals y_1 and y_2 using the RLS, AR model orders of 40 and 30 were chosen, respectively. Our choice of these model orders resulted in the normalized MSE values of 3.03% for the signal y_1 and 3.49% for the signal y_2. From the TVAR coefficients associated with the signals y_1 and y_2, 20 different realizations of surrogate data were generated for each of the two signals. As the left panel of Figs. 2b – 2c shows, the statistical threshold coherence values derived from the TVSD are approximately 0.5 in the time interval from 0 to approximately 256 seconds, which is lower than those values in Fig. 2a, thereby correctly indicating the statistical significance of the TVCCF values. In the second half of the data (from 257 to approximately 512 seconds), the values in the left panel of Fig. 2a are very close to 0 while the values in the left panel of Fig. 2b are approximately 0.03 for the

STFT. For the RLS in Fig. 2c, the value is approximately 0.15. This indicates that the coherence from y_2 to y_1 is non-significant from 257 to about 512 seconds. Similar correct observations are found for the right panels of Figs. 2a – c.

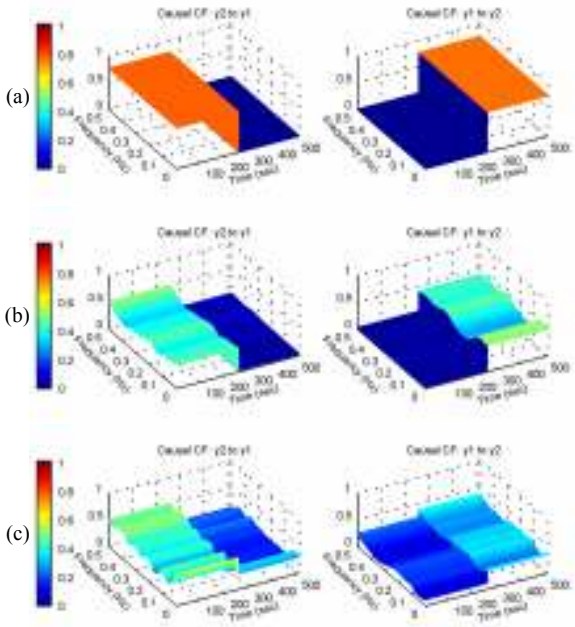

Fig. 2 Estimated causal coherence functions: (a) estimated TVCCFs of the simulation data, (b) estimated statistical threshold coherence values based on the STFT surrogate data, and (c) estimated statistical threshold coherence values based on the RLS surrogate data.

As a second simulation example, we consider the open-loop system created by setting the coefficient c_{12} equal to zero in Eq. (1). Further, we set the coupling strength to monotonically decrease from 3 to 0, as depicted in Fig. 3a. The estimated open-loop coherence function values, using the TV method we have previously developed [6], are shown in Fig. 3b. We correctly observe high values (~ 0.9), followed by a gradual decrease to near zero values as time increases. We also note the surrogate data generated threshold values (STFT: Figs. 3c and RLS: Fig. 3d) exhibiting a similar gradually decreasing pattern. To estimate the signals y_1, an AR model order of 35 was chosen, which resulted in the normalized MSE value of 3.4%. 20 surrogate data realizations were generated from the TVAR coefficients pertaining to the signal y_1. In Figs. 3b-d, we note that the coherence function values are greater than the surrogate threshold values in the time interval 0 to 440 seconds. Thereafter, the coherence function values are lower than the surrogate threshold values, suggesting that when the coupling strength (c_{21}) decreases to a value around 0.4, the coherence between signals y_1 and y_2 becomes insignificant.

B. Application to blood pressure and heart rate data

We present an application of the TVSD to previously collected blood pressure (BP) and heart rate (HR) data [10]. Details regarding data collection and data preprocessing procedures are described in our previous study [10]. The purpose of this section is to illustrate the importance of using

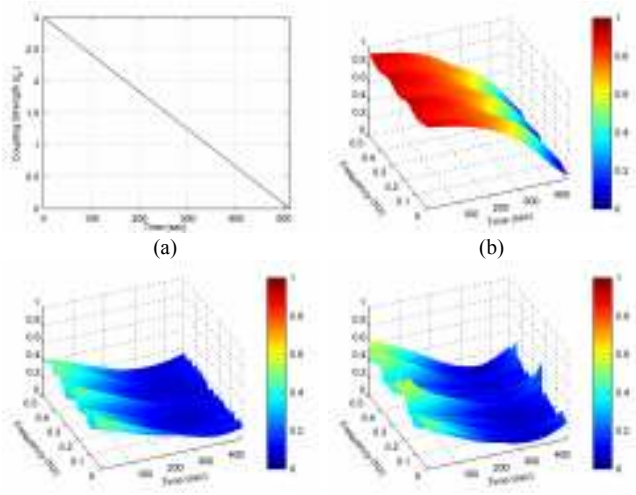

Fig. 3 An open-loop system with monotonically decreasing coupling strength. (a) Varying coupling magnitudes from y_2 to y_1, (b) Estimated open-loop TVCF of the simulated signal, (c) Estimated statistical threshold values based on 20 realizations of the STFT surrogate data, (d) Estimated statistical threshold values based on 20 realizations of RLS surrogate data

the TVSD over a time-invariant surrogate data technique. Data presented in Fig. 4 represent BP and HR data during the control state followed by the application of atropine, which blocks the parasympathetic nervous activities. The application of atropine occurs at 256 seconds in all panels of Fig. 4. The top left and right panels of Fig. 4 show time-varying closed-loop coherence functions representing BP to HR and HR to BP, respectively. [In this paper the notation "BP to HR" and similar shall mean the first variable (BP) is the input signal of the system and the second variable (HR) is the output signal.] These estimates were based on the method we have previously developed [6]. In the left panel, representing the coherence relationship from BP to HR, during the control state (time less than 256 seconds), we note statistically high coherence values especially at the low frequency (LF: 0.04 to 0.15 Hz) and high frequency (HF: 0.2 to 0.4 Hz) bands. The LF is known to contain dynamics pertaining to both sympathetic and parasympathetic nervous activities whereas the HF band is attributed to the dynamics of the parasympathetic nervous system [11]. With the application of atropine (> 256 seconds), we note insignificant coherence at all frequencies as confirmed by the STFT and TVAR generated TVSD coherence values shown in the left panels of the 2nd and 3rd rows of Fig. 4, respectively. This is the expected result, since the parasympathetic activities which reside in both LF and HF bands have been blocked with atropine. To generate TV surrogate data using the TVAR, AR model orders of 20 for HR and 15 for BP were chosen. Our choice of these model orders resulted in the normalized MSE values of 5.15% for the HR and 4.85% for the BP data. From the TVAR coefficients associated with the HR and BP, 20 different realizations of surrogate data were generated for each of the two signals. The top right panel shows the relationship between HR to BP, which is the baroreceptor activity. With application of atropine, the expected result is an increase in

HR which in turn activates the baroreceptor to decrease the HR. The baroreceptors excite parasympathetic activity to decrease HR, but since the parasympathetic nervous system is blocked, the only recourse is to decrease sympathetic nervous activity. Thus, this increased baroreceptor activity required to lower sympathetic activity should only be reflected in the LF region. This is exactly what we observe in the LF region of the top right panel; the coherence values in the LF band are greater as compared to the control state and are higher than the TVSD coherence values, as shown in the right panel of the second row of Fig. 4.

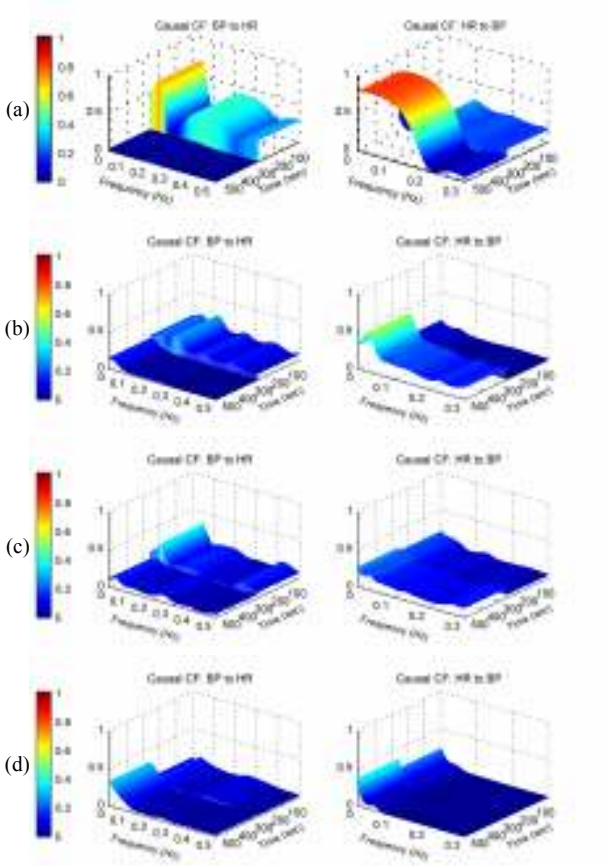

Fig. 4 Application of the proposed method to BP and HR data. (a) causal coherence functions from BP to HR (left), and from HR to BP (right); (b) statistical threshold values based on 20 realizations of the STFT surrogate data; (c) statistical threshold values based on 20 realizations of the RLS surrogate data; and (d) statistical threshold values based on 20 realizations of time-invariant surrogate data.

As expected, we do not see any changes in the coherence values in the HF region from the control state to the application of atropine in the top right panel. Insignificant coherence values in the HF band in these two states are confirmed by the STFT and RLS generated TVSD coherence values shown in the right panels of the 2nd and 3rd rows of Fig. 4, respectively. Figures in the bottom panel represent time-invariant SD results. Note that with this approach, the expected insignificant coherence in the HF region is also confirmed since time-invariant SD coherence and the TVSD coherence values are similar. It should be noted that the TVAR surrogate threshold coherence values provide the clearest demarcation of the insignificant coherence, since they are greater than the coherence values associated with the data. Thus, this example illustrates the additional insight as well as the correct physiological interpretation that can be obtained with TVSD as well as TV coherence function estimates.

IV. Conclusion

With recent new developments in time-varying open-loop and closed-loop coherence functions [6, 7], time-varying surrogate data techniques to properly take into account the statistical significance of the time-varying coherence function values are certainly needed. Towards this goal, we developed a segment-based time-varying surrogate data technique as well as the TVAR approach. One notable disadvantage of nonparametric method is the non-optimal time and frequency resolutions, as they are inversely proportional due to the use of the Fourier transform. To obviate this disadvantage, we used a time-varying autoregressive model-based spectrum to generate surrogate data, which resulted in concomitant higher time and frequency resolution than the Fourier transform surrogates provide. Both simulation and experimental results demonstrated the efficacy and applicability of the proposed methods. Our method eliminates an arbitrary decision about the significance of the coherence function, and the determination of the coherence between two signals is based on levels of statistical significance that change with time.

References

[1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," presented at Interpretation of time series from nonlinear mechanical systems, Warwick, United Kingdom, 1992.

[2] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," Phys. Rev. Lett., vol. 77, pp. 635-638, 1996.

[3] D. J. Miller, N. Stergiou, and M. J. Kurz, "An improved surrogate method for detecting the presence of chaos in gait," J Biomech, 2005.

[4] A. Porta, R. Furlan, O. Rimoldi, M. Pagani, A. Malliani, and P. v. d. Borne, "Quantifying the strength of the linear causal coupling in closed loop interacting cardiovascular variability signals," Biological Cybernetics, vol. 86, pp. 241-251, 2002.

[5] L. Faes, G. D. Pinna, A. Porta, R. Maestri, and G. Nollo, "Surrogate data analysis for assessing the significance of the coherence function," IEEE Trans Biomed Eng, vol. 51, pp. 1156-66, 2004.

[6] H. Zhao, W. A. Cupples, K. Ju, and K. H. Chon, "Time-varying casual coherence function and its application to renal blood pressure and blood flow data," IEEE Trans Biomed Eng, vol. 54(12), pp. 2142-50, 2007.

[7] H. Zhao, S. Lu, R. Zou, K. Ju, and K. H. Chon, "Estimation of time-varying coherence function using time-varying transfer functions," Ann Biomed Eng, vol. 33, pp. 1582-94, 2005.

[8] T. Schreiber and A. Schmitz, "Surrogate time series," Physica. D, vol. 142, pp. 346-382, 2000.

[9] S. Lu, K. H. Ju, and K. H. Chon, "A new algorithm for linear and nonlinear ARMA model parameter estimation using affine geometry," IEEE Trans Biomed Eng, vol. 48, pp. 1116-24, 2001.

[10] M. Elstad, K. Toska, K. H. Chon, E. A. Raeder, and R. J. Cohen, "Respiratory sinus arrhythmia: opposite effects on systolic and mean arterial pressure in supine humans," J Physiol, vol. 536, pp. 251-9, 2001.

[11] "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology," Eur Heart J, vol. 17, pp. 354-81, 1996.