Denoising Dolphin Click Series in the Presence of Tonals, using Singular Spectrum Analysis and Higher Order Statistics

Paul James Seekings[1], Jolyn Tan[2],
John Potter[1], Matthias Hoffman-Kuhnt[2], Adam A. Pack[3], Louis M. Herman[3]
[1]Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore,
[2]Marine Mammal Research Laboratory, Tropical Marine Science Institute, National University of Singapore,
[3]Kewalo Basin Marine Mammal Laboratory

Abstract – **We examine the use of Singular Spectrum Analysis (SSA) technique as an alternative technique to using standard wavelet shrinkage schemes for the purpose of de-noising mixtures of tonals, transients and Gaussian noise. Wavelet schemes require a calculation of a threshold to determine which components are taken to be signal and noise. If the noise component is Gaussian, then threshold can be determined by using an appropriate estimator. However, in the presence of strong tonal content the Gaussian threshold estimators do not give optimal performance. One method is to iteratively shift the threshold until some performance criterion has been maximized. However this frequently leads to over de-noising this time series. Since the wavelet basis is chosen to best represent the signal of interest, over de-noising can cause artifacts to appear similar to the signal of interest. In most applications this can not be tolerated.**

SSA has advantages in that the basis of decomposition is derived from the time series itself. So-called Empirical Orthogonal Functions (EOFs) are derived from a lag matrix created from the time series. Singular Value Decomposition (SVD) is then used to decompose a time series into a number of time series components.

In the case of signal separation or de-noising the time series components can be combined by using their statistical properties. We examine the use of higher order statistics, to group components into tonals, transient, and Gaussian noise. By using the properties of the kurtosis for these three types of signal, the grouping of components can be done in a more formal manner, than the thresholding technique found in wavelet schemes.

The technique is demonstrated on test data consisting of dolphin clicks in the presence of tonal and Gaussian noise. Results are also shown for real data of a dolphin click series while echo- locating on a target. It is critical for future work that after de-noising, the shape of the dolphin clicks is preserved, and the recorded reflections from the target are adequately de-noised, without introducing artifacts which could be mistaken for reflections. We discuss the results of the SSA and evaluate its potential for de-noising applications.

I. INTRODUCTION

Wavelet packet shrinkage has been one of the main methods to de-noise signals. Threshold estimators are used to determine a threshold below which coefficients of the wavelet packet decomposition are reduced (Soft threshold) or removed (Hard threshold). The assumption is that the noise component is not well represented in the wavelet basis, and constitutes the smaller coefficients. Threshold estimators like Donoho-Johnson estimator (DJE) [1] has been shown to be the optimum threshold estimator when the noise component is Gaussian. This approach has been used efficiently in classification of humpback whale song [2]. The ASC algorithm was designed to perform the automated detection and extraction of the four different types of signals, Gaussian (spectrally smooth noise), tonals, transients and time-frequency transients [3]. Each stage of extraction relied on the DJE.

While the ASC algorithm is successful at decomposing a signal into spectrally smooth noise, tonals, transients and time-frequency transients, it is found necessary to provide a bias to threshold value calculated by the DJE where the noise is non-Gaussian. The bias is set manually by a process of trial and error.

This paper explores the technique of singular spectrum analysis as a tool for de-noising acoustic signals. We show that SSA can successfully decompose a time series into many components which can then be grouped together based on the statistic properties such as Kurtosis to form tonals, transients and spectrally smooth noise time-series.

II. SINGULAR SPECTRUM ANALYSIS

Singular Spectrum Analysis (SSA) is a Singular Value Decomposition (SVD) based procedure that decomposes a time series into a number of times series components. The procedure uses Empirical Orthogonal Functions (EOF) as the basis of representation. SSA has seen application in financial and applied physics to detect trends in data [4].

Briefly the structure of the SSA algorithm is this: A lagged trajectory matrix is formed from the times series to be decomposed. The number of lags determines the number of components the time series will be decomposed into. An SVD is taken of the trajectory matrix; the resulting eigentriples are reconstructed into time series by a process of diagonal averaging (Hankelisation).

Let f_k for $k = 0,1,\ldots,N-1$ be a time series of length N, and L be an integer corresponding to the maximum number of lags where $1 < L < N$.

We set $K = N - L + 1$ and define the L-lagged trajectory (Hankel) matrix to be

$$X = \begin{pmatrix} f_0 & f_1 & f_2 & \cdots & f_{K-1} \\ f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \cdots & f_{N-1} \end{pmatrix} \qquad (1)$$

The singular value decomposition (SVD) is taken of the trajectory matrix; it is a decomposition of X in the form

$$X = \sum_{i=1}^{L} \sqrt{\lambda_i} U_i V_i^T \qquad (2)$$

where U_i and V_i are the left and right eigenvectors respectively, and $\sqrt{\lambda_i}$ are singular values arranged in decreasing order of magnitudes.

The left eigenvector are the orthogonal basis functions and are generally referred to as the Empirical Orthogonal Functions (EOF). They are referred to as empirical since they are derived from the data itself, as opposed to Fourier basis or wavelet basis.

Let $X_i = \sqrt{\lambda_i} U_i V_i^T \qquad (3)$

such that from (2)

$$X = \sum_{i=1}^{L} X_i \qquad (4)$$

Then each X_i forms an 'eigentriple' which is converted into a new time series of length N by a process of diagonal averaging (Hankelisation). Let $\mathrm{H}(X_i)$ be the Hankelisation operator which by a process of diagonal averaging inverts each eigentriple back into a time series $\tilde{f}_k$.

The Hankelisation operation $\mathrm{H}(Y)$ on matrix Y size L by K, is defined by

$$\tilde{f}_k = \begin{cases} \dfrac{1}{k+1} \displaystyle\sum_{m=1}^{k+1} Y_{m,k-m+2} & \text{for } 0 \le k < L-1, \\ \dfrac{1}{L} \displaystyle\sum_{m=1}^{L} Y_{m,k-m+2} & \text{for } L-1 \le k < K, \\ \dfrac{1}{N-k} \displaystyle\sum_{m=k-K+2}^{N-K+1} Y_{m,k-m+2} & \text{for } K \le k < N. \end{cases} \qquad (5)$$

Thus the Hankelisation operator is applied to each eigentriple, X_i, resulting in

$$\tilde{f}_k^i = \mathrm{H}(X_i) \text{ for } i = 1,2,\ldots L \qquad (6)$$

From (4)

$$f_k = \mathrm{H}(X) = \sum_{i=1}^{L} \mathrm{H}(X_i) = \sum_{i=1}^{L} \tilde{f}_i^k \qquad (7)$$

Thus $\tilde{f}_k^i$ represents L time series components whose sum is equal to the original time series.

Normally a process called grouping is carried out in the eigentriple domain prior to Hankelisation. Grouping is normally done by considering the singular values $\left(\sqrt{\lambda_i}\right)$. In de-noising applications components with large singular values are taken to be signal, and rest noise. This approach presents two problems:

1) The need to define a threshold to decide what is taken to be large singular values.
2) If there is more than one signal present, the signal of interest may not have the largest singular value.

Hence we consider grouping by using the kurtosis of each $\tilde{f}_k^i$. The kurtosis of a random variable is given by

$$kurtosis(x) = \frac{E\left[(x-\mu)^4\right]}{\sigma^4} \qquad (8)$$

where μ and σ is the mean and standard deviation of x respectively and $E[\ ]$ is the expectation operator.

A Gaussian distribution has a kurtosis of 3. Heavy-tailed distribution has kurtosis of more than 3 and is termed leptokurtic. Platykurtic describes a distribution with thinner tails and kurtosis less than 3. Tonals are defined as a signal localized in frequency. It can be shown that tonals such as the sine wave of constant amplitude and frequency has a kurtosis of 1.5, thus have platykurtic distribution. By definition Gaussian noise will have kurtosis equal to 3. Transients are broadband time-localized signals that will have a leptokurtic distribution.

In practice, it is necessary to state a confidence interval around 3, within which components are regarded to be spectrally smooth noise, not strictly Gaussian. Components with kurtosis below the lower bound of this interval are regarded to be tonals. Component with kurtosis above the upper bound of this interval are regarded to be transients. Standard error for kurtosis is defined as [5]

$$s.e = \sqrt{\frac{24}{N}} \qquad (9)$$

Thus a 95% confidence interval can be defined with lower bound given by

$$L_b = 3 - 1.96\sqrt{\frac{24}{N}} \tag{10}$$

and upper bound given by

$$H_b = 3 + 1.96\sqrt{\frac{24}{N}} \tag{11}$$

The kurtosis of the each component is measured

$$K_i = kurtosis\left(\tilde{f}_k^i\right) \tag{12}$$

Components are then grouped into

Tonals:

$$T_k = \sum \tilde{f}_k^T \text{ for } K_T < L_b \tag{13}$$

Spectrally smooth noise:

$$G_k = \sum \tilde{f}_k^T \text{ for } L_b \le K_T \le H_b \tag{14}$$

and Transients

$$Tr_k = \sum \tilde{f}_k^T \text{ for } H_b < K_T \tag{15}$$

The clear advantage is that no threshold as such is needed and signals with different statistical properties can be separated. However the number of components from the decomposition has to be sufficient such that they can be grouped together adequately. Increasing the number of components will increase the efficiency at which the three components will be separated. Since this is a SVD based technique the computation requirements are high. As the length of the time-series (N) and number of components (L) increase the computation time and memory requirements increase. On a typical personal computer N = 4096 and L =80 are reasonable.

III. METHOD OF TESTING

Test data was extracted from data collected of an echolocating dolphin. The data was sampled at 500 kHz. The objective of the recordings was to capture the echolocation clicks that a dolphin sent while interrogating an object placed behind a translucent screen. Each recording consisted of a 10-second long data, which also included noise from the equipment, radio and various dolphin sounds besides the echolocation clicks. The objective is to de-noise the data with the minimum of distortion to the clicks. Noise samples were taken from the beginning of the file and where there was no echolocating activity. An echolocation click can be described as a broadband transient with frequency range between 0.25 and 220 kHz, and sound pressure levels range from 150 to 230 dB re: 1 µPa peak to peak [6].

As an example a time series is synthesized containing the summation of a section of noise, and an echolocation click. In order to quantify the signal to noise ratio it is expressed in standard terms of the energy of the signal divided by the energy of the noise. However since the signal is of a transient nature the energy of the transient it is dependent on the length of the times series used to measure the energy. Thus it must be noted that signal to noise ratio is stated for the time series length of 2048 samples. The other measure of signal energy to noise energy adopted in this paper is the signal peak to noise ratio. The ratio of the maximum absolute value the transient divided by the energy and noise.

For the example time series shown in Fig. 1, the signal to noise ratio was -19.67 dB and signal peak to noise ratio of 9dB. The power spectrum of the click and the click plus noise is shown in Fig. 2.

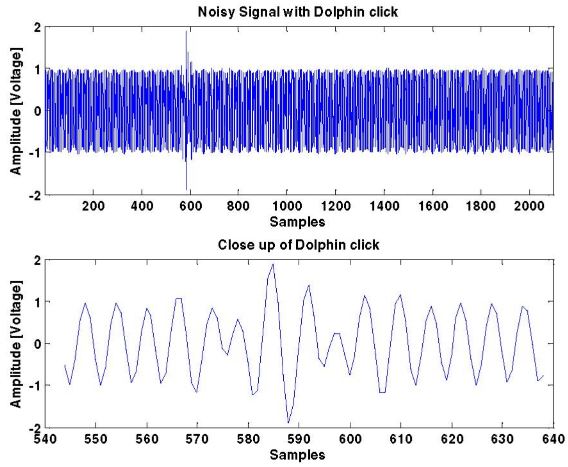

Fig. 1. Noisy time series

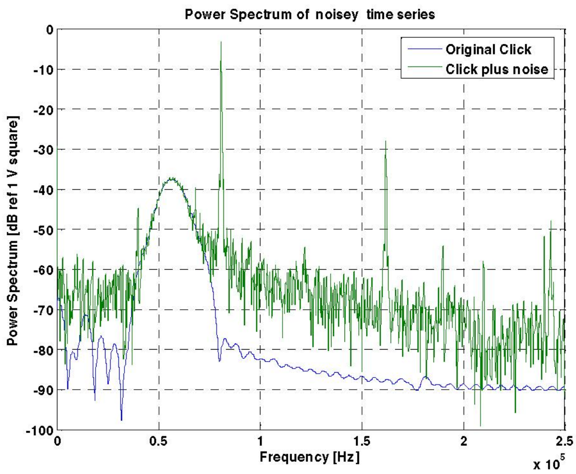

Fig. 2. Power spectrum of noisy time series

The data was de-noised using the algorithm shown above with N = 2048 and L =80. The separated signals are shown in Fig. 3. The kurtosis of each component is shown; they correctly fall within expected bounds.

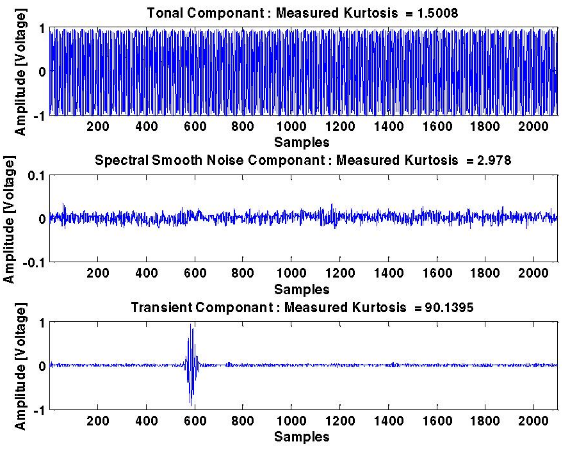

Fig. 3. Separated components

Fig. 4 shows a comparison between the original and the de-noised click. Correlation coefficient between them is calculated as 0.987.

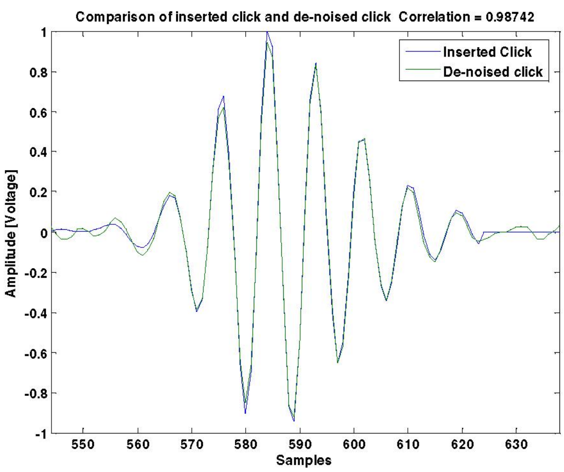

Fig. 4. Comparison of original and de-noised click

Prior to separation the signal to noise ratio between inserted click and the noise was -19.68 dB and peak noise ratio was 9dB. After de-noising the signal to noise ratio between the de-noised click (transient component) and the noise (sum of the tonal and spectrally smooth noise component) was found to be -19.67 dB and peak noise ratio of 9.1dB. The overall signal to noise ratio gain was calculated to be 32dB.

In order to further quantify the performance of the SSA algorithm, three quantities are measured

1) The mean square error between the input time series and the summation of the output components, which of course should be as small as possible i.e. equation (7) holds.

$$mse = \frac{1}{N}\sum_{k=1}^{N}\left(f_k - \left(T_k + G_k + Tr_k\right)\right)^2 \qquad (16)$$

2) Related to mean square error, the ratio of the input to output energy, which should be unity i.e. the algorithm converses energy.

$$E_i = \frac{\sum_{k=1}^{N}\left(f_k\right)^2}{\sum_{k=1}^{N}\left(T_k + G_k + Tr_k\right)^2} \qquad (17)$$

3) A separability index. The following ratio is defined as measure of separabilty,

$$S_i = \frac{\sum_{k=1}^{N}\left(T_k^2 + G_k^2 + Tr_k^2\right)}{\sum_{k=1}^{N}\left(T_k + G_k + Tr_k\right)^2} \qquad (18)$$

The numerator is equal to the denominator if the components T_k, G_k, Tr_k are uncorrelated, i.e. perfectly separated, then $S_i = 1$. If there are some correlation between components then $S_i < 1$.

For the example shown, $mse = 10^{-31}, E_i = 1$, and $S_i = 0.99$. This shows that energy is conserved, and the separation is very good.

From Fig. 3 and Fig. 4, it can be seen that either side of the click the time series should be zero. However in the de-noised version, there is a small amplitude signal present. Since the energy has been conserved overall, the small amplitude signal must belong in a different component, hence the separability index $S_i = 0.99$.

Further tests were conducted for different signal to noise ratios, i.e. the amplitude of the click was varied. Fig. 5 shows the signal to noise ratio of the input time series against correlation coefficient between the original click and the de-noised click. On the same figure the separability index is shown. The graph shows that as the input signal to noise ratio decreases the correlation value also decreases. The separability index remains constant at just less than one; these seem to indicate the separability for the range of input signal to noise ratio tested does not depend on the signal to noise ratio.

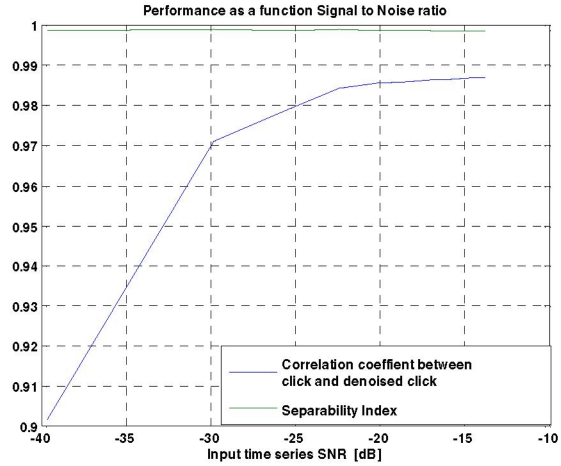

Fig. 5. Performance evaluation as a function of signal to noise ratio

IV. CLICK TRAIN DE-NOISING

In order to de-noise longer sections of data, the data was split into windows of length N = 10000, each window of data is de-noised individually. The algorithm was run on a 32 processor Beowulf cluster.

Fig. 6 shows a graph of an example of a 10-second long dolphin echolocation click train. It reveals a noisy environment where only the peaks of the largest echolocation clicks can been seen. Fig. 7. shows the components after de-noising with the SSA technique. Another example of dolphin click train before de-noising is shown in Fig. 8. This consists of a 4-second duration extracted from the original 10-second data, in which the click intensity is much higher than the rest of the time series. It represents the moment when the dolphin is intensively interrogating the object behind the translucent screen and is also the section which is useful for analytical purposes. The de-noised click data, as shown in Fig. 9, reveals many more clicks before and after the main burst.

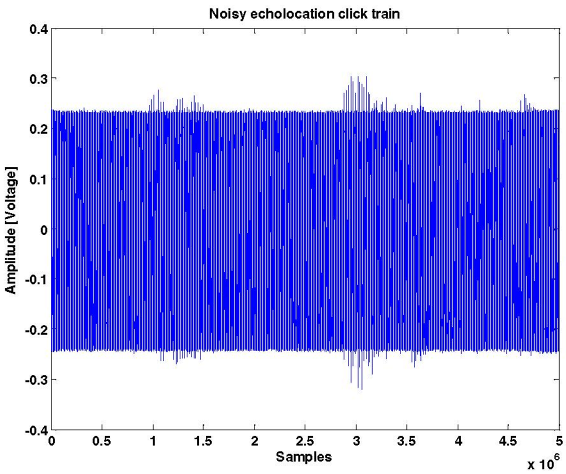

Fig. 6. Noisy echolocation click train

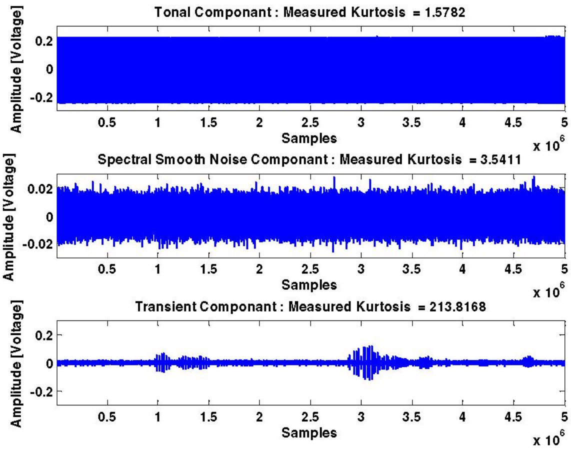

Fig. 7. Separated components after de-noising

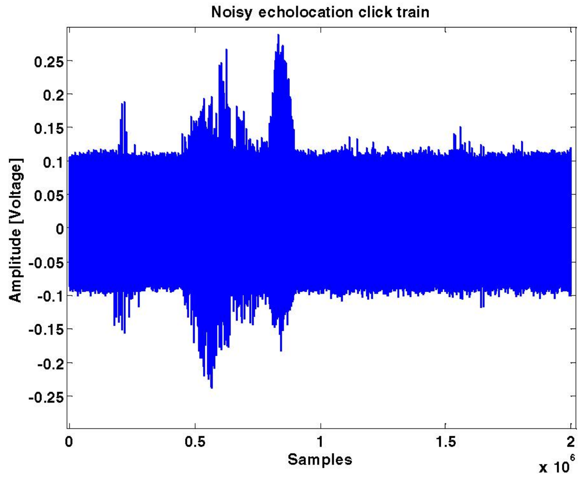

Fig. 8. Noisy echolocation click train

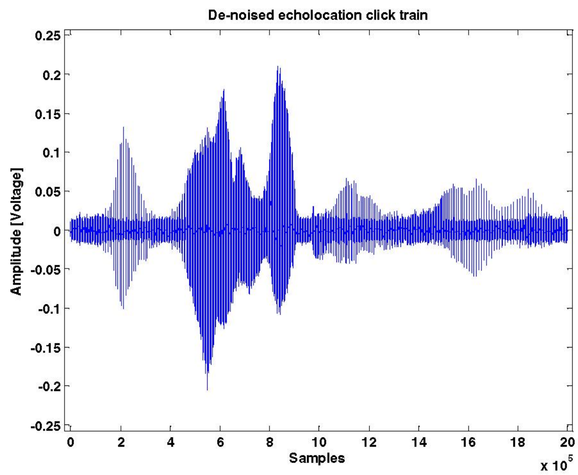

Fig. 9. De-noised echolocation click train

CONCLUSION

In this paper, higher order statistics were examined and utilized to separate the time series components generated by the SSA technique. This is explored as an alternative to grouping components by their singular values. Grouping by Kurtosis can successfully separate Tonals, Transients, and Gaussian noise.

SSA has proven to be able to de-noise dolphin clicks successfully in the presence of tonals and gaussian noise, compared to wavelet shrinkage schemes that involve threshold estimators which are optimal for Gaussian noise. Given a time series with extreme low signal to noise ratio, SSA is capable of extracting the dolphin clicks from the noise components with minimum or no energy loss. Correlation values on test data show the accuracy to which the SSA technique can de-noise a test signal.

ACKNOWLEDGMENT

This project was supported by the DSTA.

REFERENCES

[1] E.D. Kolaczyk, "Wavelet Shrinkage Estimation of Certain Poisson Intensity Signals using Corrected Thresholds", Statistica Sinica 9, 1999, pp. 119-135

[2] P. Seekings and J. Potter, "Classification of marine acoustic signals using Wavelets and Neural Networks", Proceeding of 8th Western Pacific Acoustics conference (Wespac8), Australia, April 7-9, 2003

[3] P. Seekings, F. Kerroux, and J. Potter, "Acoustic Signal Characterisation using multi-resolution transforms", Proceedings of Underwater Defense Technology Conference, Singapore, November 12-14, 2003

[4] R. Vautard and M. Ghil, "Singular-spectrum analysis in nonlinear dynamics, with applicatins to paleoclimatic time series", Physica D, 35:395-424, 1989

[5] J.D. Brown, "Skewness and Kurtosis", Shiken: JALT Testing & Evaluation SIG Newsletter, Vol. 1 No. 1 Apr 1997, pp 16-18. Available at: http://www.jalt.org/test/bro_1.htm

[6] W. W. L. Au, "Echolocation in Dolphins," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay, Springer Verlag, New York, 2000, pp. 364–408

[7] W.W.L. Au, "The sonar of dolphins", Springer Verlag, New York, 1993

[8] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, "Analysis of Time Series Structure," Chapman and Hall/CRC, 2001