

WAVELETS IN MULTI-STEP-AHEAD FORECASTING

M.A. Kaboudan

*School of Business, University of Redlands,
1200 E. Colton Avenue, Redlands, CA 92373, USA*

Abstract: This paper investigates the possibility of obtaining long-into-the-future reliable forecasts of observed nonlinear cyclical phenomena. Unsmoothed monthly sunspot numbers that are characteristically cyclical with nonlinear dynamics as well as their wavelet-transformed and wavelet-denoised series are forecasted through October 2008. The objective is to determine whether modelling wavelet-conversions of a series provides reasonable forecasts. Two computational techniques – neural networks and genetic programming – are used to model the dynamics of the series. Statistical comparison of their *ex post* forecasts is then used to identify the data set and computational technique to use under the circumstances. *Copyright © 2005 IFAC*

Keywords: Artificial intelligence, Genetic algorithms, Nonlinear systems.

1. INTRODUCTION

The purpose of this paper is to investigate whether it is possible to produce reasonable forecasts of cyclical phenomena for many periods into the future when their wavelet-converted data is fitted using two computational techniques – artificial neural networks (ANN) and genetic programming (GP). Modelling and forecasting real-world nonlinear cycles are particularly challenging because their dynamics tend to be irregular as well as noisy. *Nonlinear* models are known for their sensitivity to initial conditions. Predicted values of input variables thus can only deliver poor extended forecasts. Having irregular noisy cycles only aggravates the quality of such forecasts. It is therefore logical to obtain forecasts using “actual” rather than “predicted” values as inputs. This is possible when models are specified and estimated with lagged explanatory variables having minimum lag lengths = the number of periods or steps ahead to forecast. It is impossible to estimate such models using standard statistical techniques because of the large number of degrees of freedom lost when forecasting a large number of periods into the future. ANN and GP are used instead. They deliver forecasts without estimating equation coefficients and therefore pose fewer statistical problems.

Sunspot numbers have cyclical behavior and their annual averages rank among the most statistically analyzed time series. See for example Tong (1990),

and Lin and Pourahmadi (1998). However, forecasts of annual averages may be too aggregated over time to really help in decision making, and forecasts of monthly averages may be more useful. Examples of studies that forecast monthly averages are few. They include those of Mundt *et al.* (1991) and Hathaway *et al.* (1999). Unsmoothed monthly sunspot numbers is the variable selected to experiment with in this study. Forecasting the numbers is of interest when making decisions about satellite orbits and space missions. In addition, accurate predictions of them have significant economic implications for technologies (such as high-frequency radio communications and radars) and help in weather forecasting..

Sunspots are huge dark areas (sometimes exceeding Earth’s size) that appear on the Sun’s visible surface then disappear in a few hours, days, or even months. A sunspot number is a daily measure $r = A (10 G + I)$, where I is actual number of visible spots on a given day observed from twelve centers at different locations worldwide, A is an adjustment factor that accounts for differences between observatories and observers collecting data at the different centers, and G is a count of the number of groups of observed sunspots on that day.

The initiative to integrate wavelet-converted data with computational techniques to obtain forecasts is not entirely new. Given noisy data Y_t where X_t is an unknown signal, Donoho and Johnstone (1995) proposed a thresholding method to reconstruct the

unknown signal. Advances in thresholding by Donoho (1995), Donoho *et al.* (1998), and Abramovich *et al.* (1999) followed. The term “thresholding” is typically used to describe ways to filter signal from noise and obtain smoother dynamics that may be easier to estimate. Applications using wavelets in estimation were reviewed in Lee (1998). More recent applications are in Nason and Theofanis (2000), and Cherkassky and Shao (2001). Contributions on using wavelets in forecasting (rather than only in estimating a function) are less ubiquitous and they focus only on one-step-ahead forecasts. Aussem and Murtagh (1997) apply neural networks to ‘à-trous’ wavelet-transformed annual sunspot numbers to obtain one-step-ahead forecasts for 59 sunspot values (1921 to 1979). Pan and Wang (1998) introduced a new estimator that combines a state-space model with wavelet transforms to forecast S&P 500 as a function of the S&P dividend yield. Renaud *et al.* (2002) experimented with AR(4) noisy data to provide one-step-ahead forecast.

The challenge undertaken in this paper is to obtain forecasts of sunspot numbers for a relatively large number of periods ahead. Compared with one-step-ahead forecasts, multi-step-ahead forecasts are more useful in planning and decision making. It is assumed here that a model that produces accurate forecasts for a small number of periods into the future, especially if it succeeded in reproducing historical values, may be more reliable in producing reasonable forecasts further into the future. Alternatively, this assumption is the conditional probability

$$P(\text{Reliable } F_b / \text{acceptable } F_a \text{ and low historic MSE}) \geq P(\text{Reliable } F_b / \text{min. historic MSE}), \quad (1)$$

where F = forecast, $a = T+1, \dots, T+A$, T is the number of observations used to obtain a model, A is the number of forecast periods for which outcomes are known *a priori*, and $b = T+A+1, \dots, T+A+B$, where B is the number of forecast periods for which outcomes are unknown. Accordingly, $F_b = \text{ex ante}$ forecast or forecasts of unknown future values, and $F_a = \text{ex post}$ forecast or forecasts of values whose outcomes are already known. Alternatively, the LHS of (1) is the conditional probability of obtaining a reliable *ex ante* forecast if a model delivers acceptable *ex post* forecast and relatively low (but not necessarily minimum) residuals’ MSE, and its RHS is the conditional probability of obtaining a reliable *ex ante* forecast if only residuals of the historic fit from a model deliver the lowest MSE. Thus, the objective function:

$$F_a \text{ error} \quad \text{subject to: low } \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 \quad (2)$$

is minimized to obtain a forecasting model.

Forecasts using ANN and GP are obtained and compared below. The results are different from any reported before on forecasting sunspot numbers since the lag structures used deliver forecasts for a very large number of periods without having to use predicted values (or model solutions) as input.

Before presenting the results, the difference between *wavelet-transformed* and *wavelet-denoised* data conversions is explained in the next Section. GP is briefly reviewed in Section 3. Only a short introduction of ANN is in Section 4 that mainly presents fitting outcomes along with forecast results. A few concluding remarks are in Section 5.

2. WAVELETS

In and by itself, wavelet analysis is not a forecasting technique. In wavelet analysis, data can be converted into forms to which it may be easier to fit models. Data conversions can be one of two forms: *wavelet transforms* and *denoised* series. Conversions can be carried out using any of many available wavelets (or newly created ones). The Haar and Daubechies wavelets are perhaps the most popular and are explained by many (e.g. Daubechies, 1992 and Gençay *et al.*, 2002).

2.1 Wavelet transforms

A wavelet transform is a scaling function used to transform a signal into father and mother wavelets. Father wavelets are representations of a signal’s smooth or low-frequency component. Mother wavelets are representations of the details or high-frequency component in a signal. The Haar wavelet is the simplest to use. It is a decimated process where at each level of scaling half the number of observations disappears. When used, it transforms a series Y_t to obtain mid-point averages (s_1) and mid-point differences (d_1) of consecutive pairs of observations first. Averages preserve the main signal while differences capture the series’ detailed fluctuations. In turn, it transforms mid-point averages (s_1) to obtain their mid-point averages (s_2) and their mid-point differences (d_2), and so on. For discrete time series, the values of the mid-point averages and differences are known as “coefficients” in the literature on wavelets. Obtaining these values is known as a *discrete wavelet transform* process (DWT). Alternatively, DWT maps a vector of Y_t values to a vector of wavelet coefficients w , or

$$w = \begin{pmatrix} s_J \\ d_J \\ d_{J-1} \\ \dots \\ d_1 \end{pmatrix} \quad (3)$$

where J is the number of scales or multiresolution components, and

$$\begin{aligned} s_J &= (s_{J,1}, s_{J,2}, \dots, s_{J,T/2})^J \\ d_J &= (d_{J,1}, d_{J,2}, \dots, d_{J,T/2})^J \\ d_{J-1} &= (d_{J-1,1}, d_{J-1,2}, \dots, d_{J-1,T/2})^J \\ &\dots \\ d_1 &= (d_{1,1}, d_{1,2}, \dots, d_{1,T/2})^J \end{aligned} \quad (4)$$

A DWT process of the Haar wavelet has a desirable property that may help in forecasting. Given a series’ wavelet transformed coefficients s_j and d_j, \dots, d_1 in (4) above, original values of that series can be reconstructed from the transformed data (Mallat, 1989). If a series has $T = 512$, DWT with $J = 4$ (for

example) delivers five series: s_4 , d_4 , d_3 , d_2 , and d_1 with 32, 32, 64, 128, and 256 coefficients in each, respectively. For each of the five data sets (s_4 , d_4 , d_3 , d_2 , and d_1), different models can then be obtained. The idea of using these five series basically amounts to adopting a “divide and conquer strategy”. The DWT of the unsmoothed monthly sunspot numbers with scaling level $J = 4$ as well as their IDWT are shown in Figure 1.

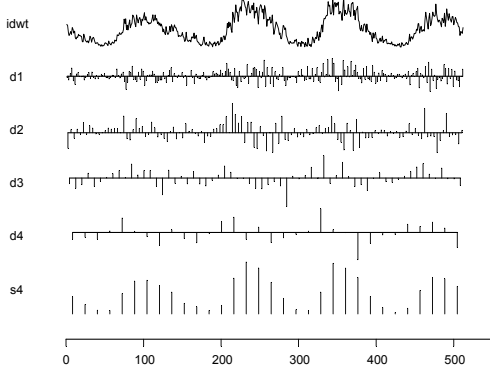


Figure 1. DWT and IDWT of sunspot numbers using the Haar wavelet transformation.

The five DWT series are independent inputs to model. The model specification assumed for s_4 is:

$$s_{4,t} = f(s_{4,t-4}, s_{4,t-5}, s_{4,t-6}, s_{4,t-7}). \quad (5)$$

According to (5), s_4 is assumed to be a function of four distant but consecutive lagged values. Similarly,

$$d_{4,t} = f(d_{4,t-4}, \dots, d_{4,t-7}); \quad (6)$$

$$d_{3,t} = f(d_{3,t-8}, \dots, d_{3,t-15}); \quad (7)$$

$$d_{2,t} = f(d_{2,t-16}, \dots, d_{2,t-27}); \quad (8)$$

$$d_{1,t} = f(d_{1,t-32}, \dots, d_{1,t-55}). \quad (9)$$

Once the five models are obtained, they are used to compute fitted values and forecasts of the five series.

The reconstructed \hat{Y}_t has the same number of observations as Y_t , except that t in \hat{Y}_t is shifted forward relative to the t in Y_t to include forecast periods. To obtain fitted and forecasted values, the decimation process to use with DWT and its inverse must satisfy the condition of perfect reconstruction. This means that the number of observations in each DWT series must remain the same to produce IDWT - the reconstructed series. Accordingly, to reconstruct the series, the number of observations T are shifted forward by $(n+a+b)$ where $b =$ the desired number of observations to forecast *ex ante* which is set = L . More specifically, s_4 and d_4 that have 32 observations will be reduced by 4 observations to leave 28 for training or fitting models. The resulting models will then produce the 28 fitted and 4 forecast values. The number of observations used to fit and forecast are doubled for d_3 , then doubled again for d_2 , and so on. Using this method, equations (5) and (6) provide four-step-ahead forecast, equation (7) provides eight-step-ahead forecast, equation (8) provides 16-step-ahead forecast, and (9) provides 32-step-ahead forecast. The final forecast delivered by the inverse DWT is 64 steps ahead.

2.2 Thresholding

In thresholding, insignificant values in the DWT transformed signal are set equal to zero; then the altered series goes through the inverse DWT to produce an approximation of the original signal. The input series to model and forecast is the IDWT. A threshold value is used to distinguish between what is significant and what is not. Donoho and Johnstone (1995) describe the process and show the optimality properties of such wavelet estimator. The steps leading to delivering a denoised data set to model are in Figure 5, where

$$\tilde{s}_4 = \delta_{\lambda,4} \sigma_4 (s_4) \quad (10)$$

$$\tilde{d}_4 = \delta_{\lambda,4} \sigma_4 (d_4) \quad (11)$$

$$\tilde{d}_3 = \delta_{\lambda,3} \sigma_3 (d_3) \quad (12)$$

$$\tilde{d}_2 = \delta_{\lambda,2} \sigma_2 (d_2) \quad (13)$$

$$\tilde{d}_1 = \delta_{\lambda,1} \sigma_1 (d_1) \quad (14)$$

$\delta_{\lambda,j}$ is a function that shrinks the detailed coefficients, λ is a shrinking threshold below which values are set equal to zero, and $\sigma =$ estimate of the scale of noise. The shrinkage function is:

$$\delta_{\lambda}(Y) = \begin{cases} 0 & \text{if } |Y| \leq \lambda \\ \text{sign}(Y)(|Y| - \lambda) & \text{if } |Y| > \lambda \end{cases} \quad (15)$$

To complete the thresholding process for the monthly data of sunspot numbers, following Bruce and Gao (1996, pp. 90-101), λ_j and σ_j were used, where $\lambda_j = \sqrt{2 \log(T)}$ and σ_j is a single value to estimate the scale of noise obtained from the fine scale coefficient d_1 . The hypothetical model to train or fit is:

$$\tilde{Y}_t = (\tilde{Y}_{t-64}, \dots, \tilde{Y}_{t-101}) \quad (16)$$

where \tilde{Y}_t is the inverse DWT of the denoised coefficients. The number of observations in this series is equal to that of the original data ($T=512$). To obtain a forecast, the data is shifted in the manner described earlier to deliver 64-step-ahead forecasts.

3. GENETIC PROGRAMMING

GP is a computerized optimisation technique employed to solve diverse problems in different disciplines. Foundations of GP are in Koza (1992). Description of how GP is used in forecasting and its statistical properties are in Kaboudan (2001). The GP software used in this study is TSGP (Kaboudan, 2003) written in C++ for Windows environment. It takes two types of input files: data files and a configuration file. Data input files contain values of the dependent and each of the independent variables. The configuration file contains execution information such as: name of the dependent variable, number of observations to fit, number of observations to forecast, number of equation specifications to evolve, and other GP-specific parameters. Basically, TSGP randomly assembles an initial population of individual specifications (say 1000), computes their fitness, and then breeds new equations as members of a new generation with the same population size. An individual specification is represented by a tree

consisting of nodes connected with arcs. The inner nodes contain mathematical operators (such as +, -, *, /, sin, cos, etc.). A tree continues to grow until end nodes contain explanatory variables or a constant term. A node containing a variable or constant is thus known as a 'terminal'. A new population is bred using mutation, crossover, and self reproduction. Fitter equations in a population get a higher chance to participate in breeding. In mutation, TSGP is designed to randomly assemble a sub-tree that replaces a randomly selected existing part of a tree. In crossover, randomly selected parts of two existing trees are swapped. In self reproduction, a top percentage of the fittest individuals in an existing population are passed on to the next generation. The idea is to continue generating new populations while preserving good genes. After a specified number of generations, the program terminates and saves to an output file the specification that captures the dynamics of a series best. That best equation is then used to forecast the series' future values.

GP is employed to obtain a model for each of the series representing sunspot numbers. These include the observed data as well as their wavelet-transformed and wavelet-denoised data. For each of those equations are evolved assuming the delayed autoregressive specification discussed earlier, or $Y_t = f(Y_{t-L}, Y_{t-(L+1)}, \dots, Y_{t-(L+c)})$ where the number of periods to forecast $L > 1$ and can exceed the number of explanatory variables, and the number of explanatory variables $c > 1$. Such specification is possible when using GP because coefficients in the evolved equations are not computed. They are random numbers (between -128 and 127) TSGP is programmed to generate using a random number generator. Given that there are no coefficients to compute, there is no restriction on the number of lagged dependent variables to include since there are no lost degrees of freedom. Such advantage makes it possible to set 'the number of *ex post* forecast periods to evaluate' = 'the number of *ex ante* forecast periods desired'. Using the long lag structures ($L = 64$) means that MSN_t dynamics over the past two to six years are assumed to help explain its dynamics over the next 64 months. This assumption is similar to that made in the precursor model and solar dynamo theory of sunspot cycles in Schatten and Pesnell (1993).

GP delivers equations that minimize MSE. Because assembling the equations is random, the program occasionally gets trapped in a local minimum MSE rather than a global one within the search space. It is therefore necessary to generate a large number of equations and then select the best one(s) to use in forecasting. Outputs delivered by TSGP are first sorted according to lowest historical MSE. Equations with the 20 lowest MSE (where the number 20 is arbitrarily set) are then sorted in an ascending order according to *ex post* forecast mean absolute percent error (MAPE). That equation among the 20 with the lowest MAPE is then selected.

The variable to model and forecast in this paper is MSN_t = monthly averages of r at time period t . The sample selected for experimentation starts November 1960 and ends June 2003 with a total of 512 observations. Using ANN and GP, fitted and forecasted values of monthly sunspot numbers were obtained for original unsmoothed monthly sunspots numbers data, MSN , the inverse of the fitted values of the Haar DWT five series ($s_4, d_4, d_3, d_2,$ and d_1), $IDWT_H$, and the denoised numbers using the Daubechies s_8 wavelet, $Daub_T$. Of the 512 observations available, the first 63 were lost to provide the lengthy lag. An additional 48 more degrees of freedom were lost as lagged-dependent variables. The last 64 observations (March 1998 through June 2003) were reserved to evaluate the *ex post* forecasts. This left 337 observations ($= 512 - 63 - 48 - 64$) starting February 1970 through February 1998 for model training and fitting. It was then possible to forecast 64 more periods *ex ante* starting July 2003 through October 2008.

4.1 ANN & its Results

ANN is an information-processing paradigm based on the way densely interconnected parallel structure of human brains process information. The technique can detect structure in time-series. A network is a collection of mathematical models that emulates the nervous systems and draws on the analogies of adaptive learning. Input data is presented to the network that learns to predict future outcomes. Principe *et al.* (2000) among many others provide a complete description of types of networks to choose from and how ANN can be used in forecasting. NeuroSolutions is the software used here in training the networks and forecasting sunspot numbers. After experimenting with different network configurations, two types – multilayer perceptions (MLP) and generalized feedforward (GFF) – were selected to model and forecast sunspot numbers.

Statistics on ANN's training results are summarized in Table 1. These statistics are reported to help determine normality of training residuals, and to evaluate quality of fitting the historic data. In the first row, residuals' means are reported followed by their standard errors and t-statistic. These are followed by measures of kurtosis and skewness. Training residuals are assumed normal if (i) the t-test suggests failing to reject the null that a mean error is statistically equal to zero, (ii) the coefficient of kurtosis is approximately equal to zero, and (iii) the coefficient of skewness is approximately equal to zero. The results suggest that although the mean MSN residuals is not significantly different from zero and the coefficient of skewness suggests slight right skewness, the residuals are leptokurtic. Residuals from the $IDWT_H$ have a mean that is not statistically different from zero but are also leptokurtic. Residuals from the S_4 data may not have a mean of zero but are slightly platykurtic. Residuals from the $Daub_T$ data may be statistically equal to

zero but are also leptokurtic. The second set of statistics in Table 1 provides a comparison of the statistics measuring fitness of each network's simulated data. Low MSE and high R^2 are indicators of reasonable fitness. The lowest MSE = 53.55 and highest $R^2 = 0.98$. Both belong to the *Daub_T* data.

Table 1. ANN training results

	<i>MSN</i>	<i>Haar</i>	<i>Daub_T</i>
Mean Error	0.19	-4.54	0.45
Standard Error	0.72	1.24	0.39
t-test (Ho: mean = 0)	0.26	-3.66	1.14
Kurtosis	1.27	1.75	0.97
Skewness	0.29	-0.67	-0.25
MSE	174.00	537.71	52.55
R^2	0.94	0.82	0.98

4.2 GP Results

The parameters used in TSGP to evolve GP equations were: population size = 1200, number of generations = 120, mutation rate = 60%, crossover rate = 20%, cross self rate = 10%, and number of best-fit equations to evolve = 100. The statistics on fitted values of the GP-evolved models equations are in Table 2. They suggest that the *MSN* residuals have a mean that is significantly different from zero, the coefficient of kurtosis shows slight peakedness, and the coefficient of skewness suggests slight right skewness. These residuals are close to being normally distributed. Residuals from the *IDWT_H* have a mean that is statistically different from zero although they have no skewness or kurtosity. The *S4* residuals are close to being normally distributed. The *Daub_T* residuals are also close to being normally distributed. The lowest MSE = 225 and highest $R^2 = 0.92$ belong to the denoised *Haar* data.

Table 2. GP fitting results

	<i>MSN</i>	<i>Haar</i>	<i>Daub_T</i>
Mean	-0.27	2.52	1.14
Standard Error	1.71	0.81	1.01
t-test	-0.16	3.12	1.13
Standard Deviation	31.41	14.81	18.54
Kurtosis	0.30	0.02	0.05
Skewness	0.43	0.15	0.22
MSE	983.92	224.97	343.83
R^2	0.66	0.92	0.87

4.3 The Forecasts

Although ANN delivered the better fit when reproducing data used in training, GP delivers the better forecasts. Table 3 contains a comparison of two forecast statistics: Theil's U-statistic and the forecast MSE. Theil's U-statistic (a measure of forecast performance) is

$$U = \frac{\sqrt{MSE}}{\sqrt{A^{-1} \sum_{a=1}^A Y_a^2 + \sqrt{A^{-1} \sum_{a=1}^A \hat{Y}_a^2}}} \quad (17)$$

where $a = 1, 2, \dots, A$ (with $A =$ number of observations forecasted *ex post*), and \hat{Y}_a are forecast values of Y_a . This statistic will always fall between zero and one where zero indicates a perfect fit (Pindyck and Rubinfeld, 1998, p. 387). Among all ANN and GP results, GP delivered the lowest U-statistic and *ex post* MSE of forecasts using the *Haar* transformations.

Table 3. ANN & GP forecast results

	<i>MSN</i>	<i>Haar</i>	<i>Daub_T</i>
ANN:			
Theil's U	0.26	0.30	0.23
MSE	1823.42	2530.85	1329.82
GP:			
Theil's U	0.22	0.16	0.24
MSE	1317.25	584.12	1686.73

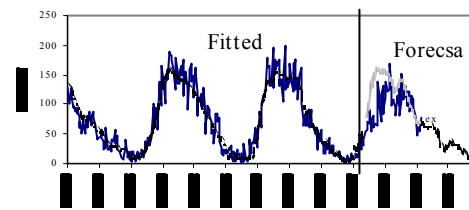


Figure 2a. Actual vs. ANN-fitted and forecasted *Daub_T* data.

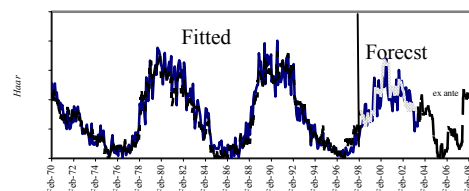


Figure 2b. Actual vs. GP-fitted and forecasted *Haar*.

Figure 2a-b through 10a-b furnish comparisons between the values of the different series (i.e., the observed unsmoothed monthly sunspot numbers and their wavelet-converted values) and each of the four fits and forecasts. Figures 7a, 8a, 9a, and 10a are outputs delivered by ANN. Figures 7b, 8b, 9b, and 10b are outputs delivered by GP. The figures suggest conclusions consistent with those Tables 3-5 suggested. All fits and forecasts seem to have captured the cycles of sunspot numbers. GP forecasts captured the beginning of a new cycle sometime around 2006 or 2007, which seems more logical. Although neither forecast is perfect, they actually exceeded expectations formulated before conducting this experiment.

5. CONCLUSION

The objective of this paper was to investigate whether it is possible to produce reasonable forecasts for many periods into the future when wavelet-converted data is fitted and forecasted using computational techniques. Monthly sunspot numbers were targeted to fit and forecast using three different configurations. In addition to the observed series, wavelet-converted series were used. They are less noisy and expectations were that they should produce better models and hopefully better forecasts than would the observed series. Artificial neural networks and genetic programming were used to train data or fit models then to forecast. These techniques were selected because when using them it is possible to use a relatively large number of lags and therefore obtain forecasts long into the future without relying on any forecast values as input. Comparative results suggested that the model that delivered historical best-fits was not the one that delivered the best short- or long-term forecasts. The model that delivered the best short-term forecasts also delivered the best longer-term forecasts. The series that produced the best forecast was the *Haar* transformed data using GP.

REFERENCES

- Abramovich, F., T. Sapatinas and B. Silverman (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society B* 60, 725-749.
- Aussem, A. and F. Murtagh (1997). Combining neural network forecasts on wavelet-transformed time series. *Connection Science* 9, 113-121.
- Bruce, A. and H. Gao (1996). *Applied wavelet analysis with S-Plus*. Springer, New York, NY.
- Cherkassky, V., X. Shao (2001). Signal estimation and denoising using VC-theory. *Neural Networks* 14, 37-52.
- Daubechies, I. (1992). *Ten lectures on wavelets*, Vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Donoho, D. (1995) De-noising by soft-thresholding. *IEEE Transactions on Information Theory* 41, 613-627.
- Donoho, D. and I. Johnstone (1995). Adapting to unknown smoothing via wavelet shrinkage. *Journal of the American Statistical Association* 90, 1200-1224.
- Donoho, D., M. Vetterili, R. DeVore, and I. Daubechies (1998). Data Compression and Harmonic Analysis. *IEEE Transactions on Information Theory* 44(6), 2435-2476.
- Gençay, R., F. Selçuk and B. Whitcher (2002). *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press, San Diego, CA.
- Hathaway, D., R. Wilson and E. Reichmann, (1999) A synthesis of solar cycle prediction techniques. *Journal of Geophysical Research* 104, 22375-22388.
- Kaboudan, M. (2001). Statistical properties of fitted residuals from genetically evolved models. *Journal of Economic Dynamics and Control* 25, 1719-1749.
- Kaboudan, M. (2003). TSGP: A time series genetic programming software. http://newton.uor.edu/facultyfolder/mahmoud_kaboudan/tsgp.
- Koza, J. (1992). *Genetic programming*. The MIT Press, Cambridge, MA.
- Lee, G. (1998). Wavelets and wavelet estimation: A Review. *Journal of Economic Theory and Econometrics* 4, 123-157.
- Lin, T. and M. Pourahmadi (1991). Nonparametric and non-Linear models and data mining in time series: A case-study on the Canadian lynx data. *Applied Statistics* 47 (Part 2), 187-201.
- Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7). 674-1989.
- Mundt, M., B. Maguire II and R. Chase (1991) Chaos in the sunspot cycle: Analysis and prediction. *Journal of Geophysical Research* 96, 1705-1716.
- Nason, G. and S. Theofanis (2000). Wavelet packet function modelling of nonstationary time series :<http://www.stats.bris.ac.uk/~guy/Research/papers/WPTransNonSta.pdf>.
- Pan, Z. and X. Wang (1998). A stochastic nonlinear regression estimator using wavelets. *Computational Economics* 11, 89-102.
- Principe, J., N. Euliano, and C. Lefebvre (2000) *Neural and adaptive systems: Fundamentals through simulations*. John Wiley & Sons, Inc, New York.
- Pindyck, R. and D. Rubinfeld (1998) *Econometric models and economic forecasting*. Irwin McGraw-Hill, Boston.
- Renaud, O., J. Starck and F. Murtagh (2002). Wavelet-based forecasting of short and long memory time series. Working Paper No. 2002.04, Department of Econometrics, University of Geneva. <http://www.unige.ch/ses/metri/>.
- Schatten, K. and D. Pesnell (1993). An early solar dynamo prediction: Cycle 23~Cycle 22. *Geophysical Research Letters* 20, 2275-2278.
- SIDC (2003) Solar Influences Data Analysis Center division of the Royal Observatory of Belgium. URL: <http://sidc.oma.be/DATA/monthssn.dat>.
- Tong, H. (1990) *Nonlinear time series analysis: A dynamical system approach*. University Press, Oxford