10

Predicting Chaotic Time Series

One of the most common problems that financial managers have to face is the accurate prediction of time series. Various models are used for the modeling of stock returns and prices or levels of volatility. The usual approach that financial analysts follow to solve these problems is to employ statistical procedures such as the general family of ARFIMA and/or GARCH models. However, financial time series often exhibit chaotic behavior. As a result, the prediction power of linear models is limited, due to their inability to model the evolutionary dynamics of the process. The models mentioned previously do not have the ability to adapt to a change of the dynamics in a chaotic system.

Chaotic time series are dynamic systems that are extremely sensitive to initial conditions and can exhibit complex external behavior. Small differences in initial conditions can exhibit diverging results. Hence, it is difficult to find the dynamic system simply through observations of the outcome. As defined by Edward Lorenz, in chaos the present determines the future, but the approximate present does not approximately determine the future.

To improve the modeling of chaotic time series, a number of nonlinear prediction methods have been developed, such as polynomials, neural networks, genetic algorithms, dynamic programming, and swarm optimization. Recently, neural networks and local models have been employed directly for chaotic time-series prediction, and comparatively satisfactory results have been reported (Inoue et al., 2001).

Wavelet Neural Networks: With Applications in Financial Engineering, Chaos, and Classification, First Edition. Antonios K. Alexandridis and Achilleas D. Zapranis.

In this chapter, wavelet networks are used to model the dynamics of a chaotic time series. More precisely, the framework proposed in earlier chapters is evaluated in a chaotic time series where the data were generated from the Mackey–Glass equation (Mackey and Glass, 1977). Moreover, the trained wavelet network is used to predict the future evolution of the chaotic system.

MACKEY–GLASS EQUATION

The Mackey–Glass equation was proposed by Mackey and Glass (1977) as a model of white blood cell production. Subsequently, it was popularized in the neural network field due to its richness in structure (Hsu and Tenorio, 1992). The equation was first proposed to associate the onset of disease with bifurcations in the dynamics of first-order differential-delay equations that model physiologic systems (Mackey and Glass, 1977).

The Mackey–Glass equation is a time-delay differential equation given by

$$\frac{\partial x}{\partial t} = \frac{ax(t-\tau)}{1 + x^c\,(t-\tau)} - bx(t) \tag{10.1}$$

The equation displays a broad diversity of dynamic behavior, including limit-cycle oscillation, with a variety of waveforms, and apparently aperiodic or “chaotic” solutions (Mackey and Glass, 1977). The behavior of the Mackey–Glass equation depends on the choice of the time delay τ. If $\tau < 4.53$, the behavior of the Mackey–Glass equation is characterized by a stable fixed-point attractor. Similarly, if $4.53 < \tau < 13.3$, there is a stable limit-cycle attractor. For $13.3 < \tau < 16.8$, the period of the limit cycle doubles. Finally, if $\tau > 16.8$, the behavior of the Mackey–Glass equation is characterized by a chaotic attractor, which is characterized by τ.

The common value used is $\tau = 17$, and this is also the case in this case study (Hsu and Tenorio, 1992; Yingwei et al., 1997). The Mackey–Glass equation with $\tau = 17$ has chaotic behavior and an attractor with a fractal dimension of about 2.1. The usual function approximation approaches are disadvantageous when the fractal dimension is greater than 2 (Cao et al., 1995; Iyengar et al., 2002). At $\tau = 17$ the time series appear to be quasiperiodic and the power spectrum is broadband with numerous spikes, due to the quasiperiodicity (Hsu and Tenorio, 1992).

Different values of the parameters a, b, and c can be used; however, the usual values of the parameters are $a = 0.2$, $b = 0.1$, and $c = 10$. The series is initialized at x_0 0.1. For the initialization transients to decay, the first 4000 data points were discarded (Platt, 1991; Yingwei et al., 1997). As in previous studies, the series is predicted with $\upsilon = 50$ sample steps ahead using four past samples: $x_{n-\upsilon}, x_{n-\upsilon-6}, x_{n-\upsilon-12}$, and $x_{n-\upsilon-18}$.

The chaotic Mackey–Glass differential delay equation is recognized as a benchmark problem that has been used and reported by a number of researchers for comparing the learning and generalization ability of different models (Chen et al., 2006).

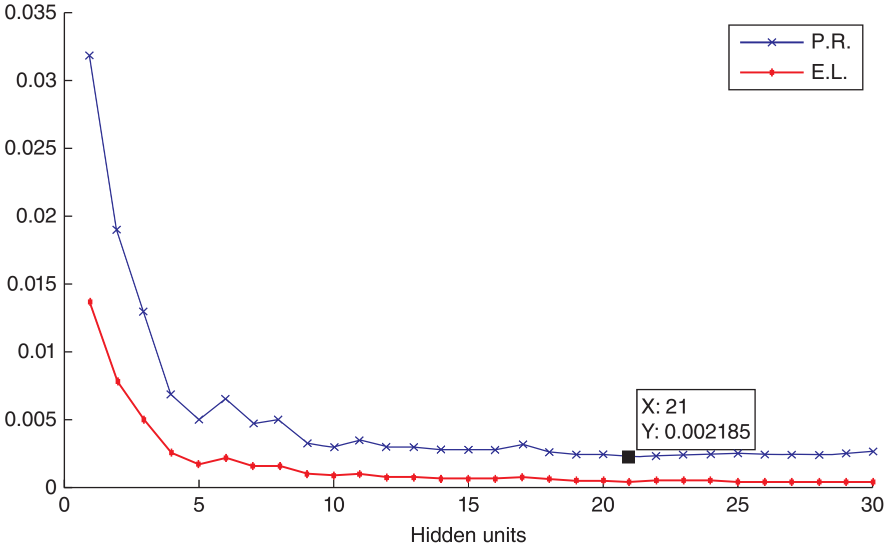

Figure 10.1 *Prediction risk (P.R.) and empirical loss (E.L.) for the first 30 hidden units for the Mackey–Glass equation.*

MODEL SELECTION

In this section the appropriate number of hidden units is determined by applying the model selection algorithm presented in Chapter 4. In Figure 10.1 the prediction risk and empirical loss for the Mackey–Glass equation are presented. The prediction risk was estimated up to a maximum of 35 hidden units. To estimate the prediction risk and find the optimal number of hidden units, the bootstrap method was used.

A close inspection of Figure 10.1 reveals that the empirical loss decreases (almost) monotonically as the complexity of the network increases. As expected, a better fit is obtained as more hidden units are used; however, the generalization ability does not necessarily increase. On the other hand, the prediction risk decreases (almost) monotonically until a minimum is reached and then will start to increase (almost) monotonically. The minimum value of the prediction risk, 0.002185, is obtained when a wavelet network with 21 hidden units is used. Hence, 21 hidden units were selected for the construction of the wavelet network model.

INITIALIZATION AND TRAINING

In the first two steps, the training set and correct topology of the wavelet network were selected. Next, the wavelet network can be constructed and trained. The training data set consists of 982 pairs. First, the wavelet network must be initialized. The

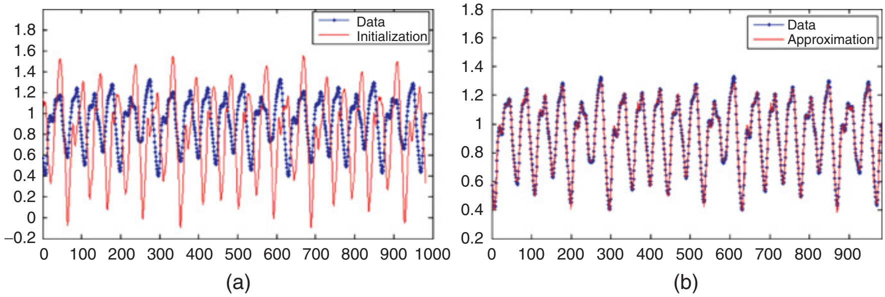

Figure 10.2 *(a) Initialization of the wavelet network using the backward elimination method and 21 hidden units; (b) approximation of the wavelet network after the training phase for the Mackey–Glass equation.*

backward elimination method was used for initialization of the set of parameters $w = (w_i^{[0]}, w_j^{[2]}, w_{\lambda+1}^{[2]}, w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]})$ of the wavelet network. Our results in Chapter 3 indicate that the BE method significantly outperforms alternative methods. A wavelet basis is constructed by scanning the first four levels of the wavelet decomposition of the data set.

The initial wavelet basis consists of 254 wavelets. However, not all wavelets in the wavelet basis contribute to the approximation of the original time series. The wavelets that contain fewer than six sample points of the training data in their support were removed. The truncated basis contains 116 wavelet candidates. The remaining wavelets were ranked and the "best" 21 wavelets were used for construction of the wavelet network.

First, we want to test if the initialization provides a good approximation of the Mackey–Glass function. To do so, the MSE between the initial approximation and the underlying function is computed. After the initialization the MSE was 0.220581, and the initialization needed 0.45 second to finish. The underlying function and the initialization of the wavelet network is shown in Figure 10.2a. A closer inspection of the figure reveals that the initialization is very good. The wavelet network converged after only 7277 iterations. The training stopped when the minimum velocity, 10^{-4}, of the training algorithm was reached. In this case the minimum velocity was increased slightly, from 10^{-5} to 10^{-4}, to avoid very large training times. The MSE error after training is 0.000299, and the total amount of time needed to train the network (initialization and training) was 39.2 seconds. $\bar{R}^2 = 99.30\%$, POCID = 89.60%, and IPOCID = 91.23%. The initialization of the WN and the final approximation after the training phase are presented in Figure 10.2.

MODEL ADEQUACY

Before proceeding to prediction of the time series out-of-sample, the model adequacy of the wavelet network will be studied. The n/p ratio is 5.06, indicating that each parameter of the network corresponds to five values. To avoid overfitting in problems

TABLE 10.1 Residual Testing[a]

	Parameter	p-Values
n/p Ratio	5.06	
Mean	0.0000	
Median	−0.0011	
S. dev.	0.0173	
DW	0.6238	0.0000
LB Q-stat.	954.3906	0.0000
JB stat.	25.7340	0.0000
KS stat.	15.0600	0.0000
R^2	99.44%	
$\bar{R}^2$	99.30%	

[a]S. dev., standard deviation; DW, Durbin–Watson; LB, Ljung–Box; KS, Kolmogorov–Smirnov.

where only a small number of observations are available, it is useful to look at the n/p ratio.

In a closer examination of the residuals we found that the mean of the residuals is zero with a standard deviation of 0.0173. The normality hypothesis is rejected as well as the hypothesis that the residuals are uncorrelated. This is not a problem for wavelet networks since, unlike linear models, normal distribution for the residuals was not assumed. Finally, the fitting of the wavelet network to the data is very good, with $R = 99.44\%$ and $\bar{R} = 99.30\%$. The results are reported analytically in Table 10.1.

The various error criteria are reported in Table 10.2. A close inspection of the table confirms our previous results that the fit is very good. The MSE is only 0.0003, while the NMSE and RMSE are only 0.0056 and 0.0173, respectively. Similarly, the maximum absolute error is 0.0673. Finally, the MAPE and SMAPE are 1.60% and 0.80%, respectively.

The estimated parameters of the regression between the target values and the network output are presented in Table 10.3. A close inspection reveals that the parameter b_0 is not statistically different from zero while the parameter b_1 is not statistically different from 1 at significance level 0.05. Moreover, the linear regression is statistically significant according to the F-statistic. Figure 10.3 is a scatter plot between the target values and the network output. Finally, as evident from Table 10.4, the change in direction metrics is very high. More precisely, POCID, IPOCID, and POS are 89.60%, 91.23%, and 100%, respectively.

TABLE 10.2 Error Criteria[a]

Md.AE	MAE	MaxAE	SSE	RMSE	NMSE	MSE	MAPE	SMAPE
0.0111	0.0134	0.0673	0.2935	0.0173	0.0056	0.0003	1.60%	0.80%

[a] Md.AE, median absolute error; MAE, mean absolute error; MaxAE, maximum absolute error; SSE, sum of squared errors; RMSE, root mean squared error; NMSE, normalized mean squared error; MSE, mean squared error; MAPE, mean absolute percentage error; SPAME, symmetric mean absolute percentage error.

TABLE 10.3 Regression Statistics[a]

	Parameter	*p*-Values	S.E.	*T*-Stat.
b_0	−0.0040	0.0774	0.0023	−1.7676
b_1	1.0044	0.0000	0.0024	417.6064
$b_1 = 1$ Test	1.0044	0.0687	0.0024	1.8223
R^2	99.44%			
F	174,391.7457	0.0000		
DW	0.6310	0.0000		

[a]S.E., squared error; DW, Durbin–Watson.

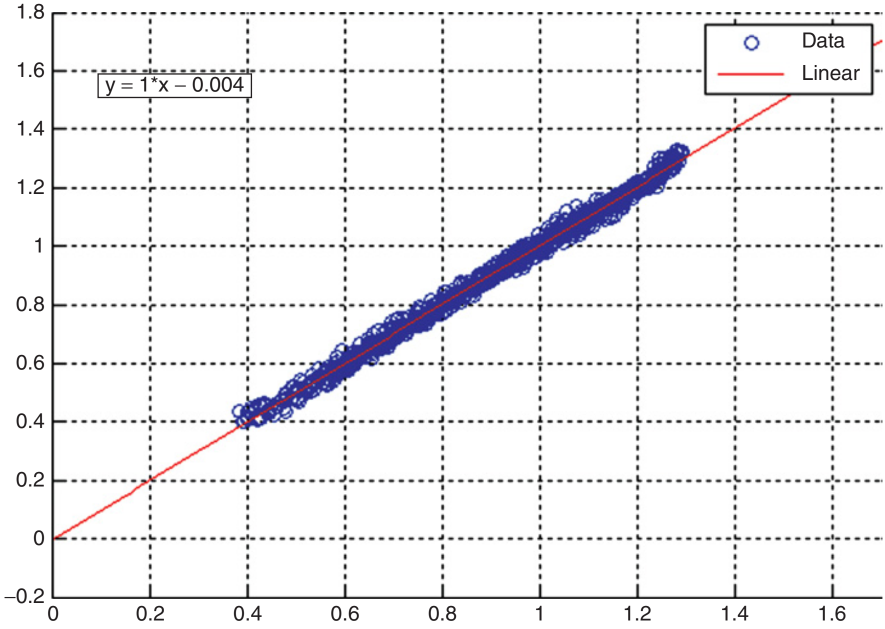

Figure 10.3 *Scatter plot of the wavelet network's fit versus the target values.*

TABLE 10.4 Change-in-Direction Metrics[a]

POCID	IPOCID	POS
89.60%	91.23%	100%

[a]POCID, prediction of change in direction; IPOCID, independent prediction of change in direction; POS, prediction of sign.

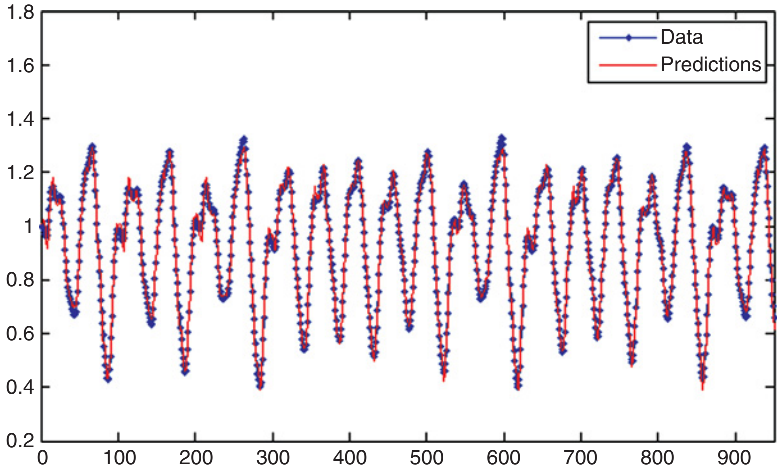

Figure 10.4 *Out-of-sample prediction of the wavelet network using 21 hidden units.*

PREDICTING THE EVOLUTION OF THE CHAOTIC MACKEY–GLASS TIME SERIES

In this section the performance of the wavelet network out-of-sample is evaluated. The training data set consists of 982 pairs and the out-of-sample data set consists of 950 pairs. These additional data were not used in training the wavelet network. The ability of the wavelet network to forecast the evolution of the chaotic Mackey–Glass equation is presented in Figure 10.4. Examining the figure, we can conclude that the wavelet network has very good generalization and forecasting ability, since the values predicted for the wavelet network are very close to the real target values of the Mackey–Glass equation.

Various statistics for the out-of-sample residuals are presented in Table 10.5. The residuals exhibit the same dynamics as in the training sample. The mean is close to zero and the standard deviation is 0.0186. The normality hypothesis is rejected as

TABLE 10.5 Out-of-Sample Residual Testing[a]

	Parameter	p-Values
n/p Ratio	4.89	
Mean	0.0001	
Median	–0.0002	
S. dev.	0.0186	
DW	0.5755	0.0000
LB Q-stat.	750.1360	0.0000
JB stat.	46.4359	0.0000
KS stat.	14.7923	0.0000
R^2	99.32%	
$\bar{R}^2$	99.15%	

[a]S. dev., standard deviation; DW, Durbin–Watson; LB, Ljung–Box; KS, Kolmogorov–Smirnov.

TABLE 10.6 Out-of-Sample Error Criteria.[a]

Md.AE	MAE	MaxAE	SSE	RMSE	NMSE	MSE	MAPE	SMAPE
0.0110	0.0142	0.0699	0.3294	0.0186	0.0068	0.0003	1.65%	0.83%

[a]Md.AE, median absolute error; MAE, mean absolute error; MaxAE, maximum absolute error; SSE, sum of squared errors; RMSE, root mean squared error; NMSE, normalized mean squared error; MSE, mean squared error; MAPE, mean absolute percentage error; SPAME, symmetric mean absolute percentage error.

TABLE 10.7 Out-of-Sample Regression Statistics[a]

	Parameter	p-Values	S.E.	T-Stat.
b_0	−0.0061	0.0177	0.0026	−2.3767
b_1	1.0067	0.0000	0.0027	374.0211
$b_1 = 1$ Test	1.0067	0.0135	0.0027	2.4751
R^2	99.33%			
F	139,891.7574	0.0000		
DW	0.5867	0.0000		

[a]S.E., squared error; DW, Durbin–Watson.

well as the hypothesis that the residuals are uncorrelated. Finally, the accuracy of the predictions of the wavelet network is very good with $R = 99.32\%$ and $\bar{R} = 99.15\%$. The results are reported analytically in Table 10.5.

Furthermore, the very good predictive power of the wavelet network is verified by the various error criteria in Table 10.6. The MSE in the out-of-sample data set is only 0.000347, and the NMSE and RMSE are only 0.00686, and 0.0186, respectively. Similarly, the maximum absolute error is 0.0699. Finally, MAPE and SMAPE are 1.65% and 0.83%, respectively. Additional error criteria are reported in Table 10.6.

Our results so far indicate that the wavelet network proposed can accurately forecast the evolution of a chaotic time series such as the Mackey–Glass equation.

The estimated parameters of the regression between the target values and the network output are presented in Table 10.7. The scatter plot between the out-of-sample target values and the forecasted values of the wavelet network is presented in Figure 10.5. In both cases it is clear that the values forecast and the target values are similar. Moreover, the linear regression is statistically significant according to the F-statistic. Finally, as shown in Table 10.8, the change-in-direction metrics are very high. More precisely, POCID, IPOCID, and POS are 87.36%, 90.41%, and 100%, respectively, indicating that the wavelet network can predict with great accuracy the changes in the direction of the chaotic dynamic system.

CONFIDENCE AND PREDICTION INTERVALS

After the wavelet network is constructed and trained, it can be used for prediction. However, in many applications, and especially in finance, risk managers may be more interested in predicting intervals for future movements of the underlying function than

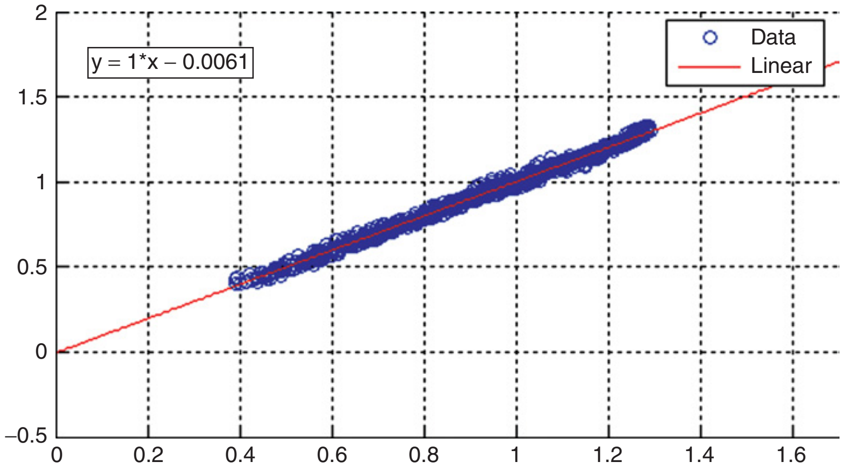

Figure 10.5 *Scatter plot between the wavelet network's output and the target values.*

simply point estimates. Hence, confidence and prediction intervals can be constructed. In this section both confidence and prediction intervals are constructed using the balancing method. Using the BS method, 200 training samples are created and divided into eight groups. In each group the average output of the wavelet networks is estimated. Next, 1000 new bootstrapped samples are created for the eight average outputs to estimate the model variance given by (7.25). Then the confidence intervals are estimated with a level of significance a of 5%. Unlike the previous case studies and examples, in this case no noise is added to the underlying function. Moreover, the network fitting and prediction are very good, as mentioned in the preceding section, with $\bar{R}^2 = 99.30\%$ in-sample and $\bar{R}^2 = 99.15\%$ out-of-sample. As a result, it is expected that the variance $\sigma_p^2 = \sigma_m^2 + \sigma_\varepsilon^2$ will be very small.

Figure 10.6 presents the confidence intervals and the true underlying function, which is the Mackey–Glass equation. Since the confidence intervals are very narrow, for clarity only one section is shown in Figure 10.6. It is clear that the underlying function is always between the confidence intervals.

In addition, in Figure 10.7 the prediction intervals for the out-of-sample data set together with the data read and the average forecast of the wavelet network for the 200 bootstrapped samples are presented. PICP = 98.8%.

TABLE 10.8 Out-of-Sample Change in Direction Metrics[a]

POCID	IPOCID	POS
87.36%	90.41%	100.00%

[a] POCID, prediction of change in direction; IPOCID, independent prediction of change in direction; POS, prediction of sign.

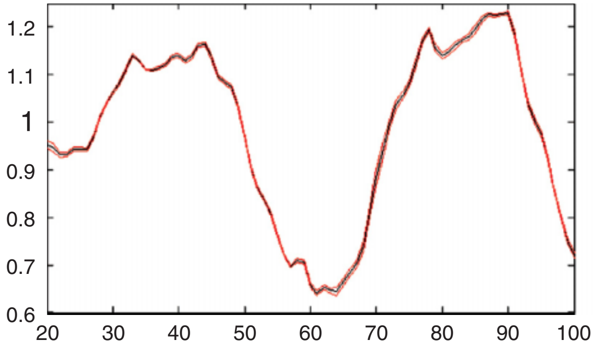

Figure 10.6 *In-sample confidence intervals (light gray line) together with the Mackey–Glass equation (black line) using the balancing method.*

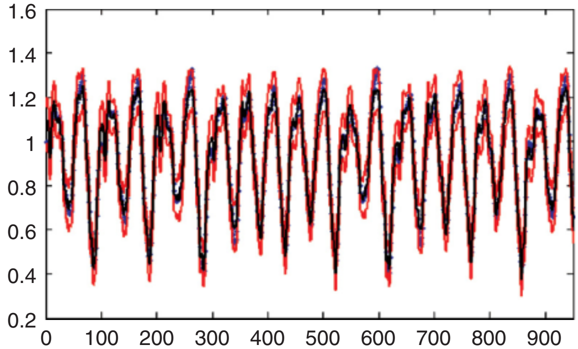

Figure 10.7 *Out-of-sample prediction intervals (light gray lines) together with the Mackey–Glass equation (gray line) and the network approximation (black line) using the balancing method (PICP = 98.8%).*

CONCLUSIONS

In this chapter the ability of a wavelet network to learn and predict the dynamics of a chaotic system was tested using the Mackey–Glass equation. The chaotic Mackey–Glass differential delay equation is recognized as a benchmark problem that has been used and reported by a number of researchers for comparing the learning and generalization ability of different models.

The objective was to predict $\upsilon = 50$ steps ahead using four past samples: $x_{n-\upsilon}, x_{n-\upsilon-6}, x_{n-\upsilon-12}$, and $x_{n-\upsilon-18}$. Following the framework proposed in this book, a wavelet network with 21 hidden units was constructed. The optimal topology of the network was selected by estimating the minimum prediction risk criterion using the bootstrap method. The wavelet network model was initialized by applying the backward elimination method.

The trained wavelet network provided a very good fit to the data. Moreover, the performance of the wavelet network in the out-of-sample data was very good. The accuracy of the predictions of the wavelet network was very good, with $R = 99.32\%$ and $\bar{R} = 99.15\%$.

Our results indicate that the wavelet network can predict the future evolution of the chaotic system accurately as well as predict the change in the direction of the chaotic Mackey–Glass dynamic system.

REFERENCES

Cao, L., Hong, Y., Fang, H., and He, G. (1995). "Predicting chaotic time series with wavelet networks." *Physica, Ser D*, 85, 225–238.

Chen, Y., Yang, B., and Dong, J. (2006). "Time series prediction using a local linear wavelet neural wavelet." *Neurocomputing*, 69, 449–465.

Hsu, W., and Tenorio, M. F. (1992). "Plastic network for predicting the Mackey–Glass time series." *International Joint Conference on Neural Networks*, Baltimore, MD.

Inoue, H., Fukunaga, Y., and Narihisa, H. (2001). "Efficient hybrid neural network for chaotic time series prediction." In *Artificial Neural Networks—ICANN* Springer-Verlag, New York, 712–718.

Iyengar, S. S., Cho, E. C., and Phoha, V. V. (2002). *Foundations of Wavelet Networks and Applications*. CRC Press, Grand Rapids, MI.

Mackey, M. C., and Glass, L. (1977). "Oscillation and chaos in physiological control systems." *Science*, 197(4300), 287–289.

Platt, J. (1991). "A resource-allocating network for function interpolation." *Neural Computation*, 3, 213–225.

Yingwei, L., Sundararajan, N., and Saratchandran, P. (1997). "A sequential learning scheme for function approximation using minimal radial basis function neural networks." *Neural Computation*, 9(2), 461–478.