Wavelet analysis for modeling suspended sediment discharge

Hafzullah Aksoy, Tanju Akar and N. Erdem Ünal

Technical University of Istanbul, Civil Engineering Faculty, Division of Hydraulics, Maslak 34469 Istanbul, Turkey.
Tel.: + 90 212 2856577. Fax: + 90 212 2856587. E-mail: *haksoy@itu.edu.tr*

Received 16 April 2002; accepted in revised form 2 April 2003

Abstract Wavelets, functions with zero mean and finite variance, have recently been found to be appropriate tools in investigating geophysical, hydrological, meteorological, and environmental processes. In this study, a wavelet-based modeling technique is presented for suspended sediment discharge time series. The model generates synthetic series statistically similar to the observed data. In the model in which the Haar wavelet is used, the available data are decomposed into detail functions. By choosing randomly from among the detail functions, synthetic suspended sediment discharge series are composed. Results are compared with those obtained from a moving-average process fitted to the data set.

Keywords Data generation; stochastic modeling; suspended sediment discharge; wavelet analysis

Nordic Hydrology **Vol 35 No 2 pp 165–174** © IWA Publishing 2004

Introduction

Flow in a river transports sediment to river reservoirs through the existing river channel. Sediment is classified as bed load and suspended load, depending on the mode of transportation. Bed load is that fraction of sediment transported intermittently near the riverbed, whereas suspended load is the remaining fraction transported continuously within the flow. Suspended load is measured by sampling from river flow. The most common practice is to correlate the suspended load to the river discharge. The fraction of bed load can be estimated by the traditional bed load equations, as continuous sampling is not an easy task for bed material. This fraction is added to the suspended load to obtain the total sediment load of the river. The load can also be given in terms of volume. Once the annual total sediment volume accumulated in the river reservoir is known, the dead storage volume is obtained, in common practice, simply by multiplying the accumulation by the design life of the reservoir.

The dead storage volume is of great importance in the design of river reservoirs. It is the volume that accommodates the sediment which accumulates during the lifetime of the reservoir. Its underestimation may shorten the life of the reservoir, whereas its overestimation leads to undesirable costs. The dead storage volume can be determined by empirical approaches (Bogardi 1974). Traditional sediment transport equations can also be used (Garde and Ranga Raju 1977). Monitoring and sampling are other solutions for the determination of the dead volume. Remote sensing and geographic information system technologies are now available (Baban and Yusof 2001). However, statistical analysis and stochastic modeling techniques such as autoregressive moving-average (ARMA) models (Box *et al.* 1994) are still attractive methods, used in hydrology for generating streamflow discharge time series. These techniques are also used separately for sediment discharge, as simultaneous streamflow and sediment discharge records are not available for comparatively long periods in developing countries such as Turkey, preventing the development and application of statistical relationships between streamflow and sediment. For example, a Bayesian approach was developed by Szidarovszky *et al.* (1976) for generating distribution functions of the sediment yield in ephemeral streams. Phien and Arbhabhirama (1979) performed a statistical analysis of the sediment accumulated in reservoirs, and Phien (1981) correlated sediment load and river flow sequences in order to determine the expected value and variance of the accumulation. Tingsanchali and Lal (1992) developed a combined deterministic–stochastic model for generating daily sediment concentrations from daily discharges, and Skoklevski and Velickov (1998) analyzed suspended load transportation by using stochastic methods.

This study aims at simulating annual mean suspended sediment discharge series by using wavelet functions that have zero mean and finite variance. The concept of wavelets in its present form was first proposed by Morlet (1981) and Grossman and Morlet (1984). Wavelets were then employed for signal transmission practices in electronic engineering (Daubechies 1988; Mallat 1989). Wavelets were also used in geophysics (Kumar and Foufoula-Georgiou 1993; Szilagyi *et al.* 1999; Kulkarni 2000). Recently, they have been found to be useful in hydrology, for synthetic data generation. For example, Feng (1998) and Smith *et al.* (1998) generated synthetic flow time series. Bayazit and Aksoy (2001) and Bayazit *et al.* (2001) used wavelets for the same purpose. Aksoy (2001) generated wavelet-based synthetic sequent-peak algorithms for determining the storage capacity of river reservoirs.

The method in this study was developed by Bayazit and Aksoy (2001). It uses the Haar wavelet – the simplest wavelet. The idea behind the method is to decompose the available data into their details first, and then to randomly add them up into one single unit to obtain annual suspended sediment discharge. The following sections, after a summary of the method, give the results of an application to a 32-year annual suspended sediment discharge series taken from the Juniata River at Newport in Pennsylvania, USA, together with conclusions.

Method

A continuous function with zero mean and finite variance is called a wavelet (Rao and Bopardikar 1998). An infinite number of functions can qualify as wavelets. Some examples of wavelets are the Morlet, Mexican hat, Shannon and Meyer types. A simpler wavelet is the Haar wavelet (Figure 1), defined as:

$$\psi(t) = \begin{cases} 1 & 0 \le t \le 1/2 \\ -1 & 1/2 \le t \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

A discrete wavelet transform of a function $f(t)$ is defined as:

$$\mathrm{d}(k,l) = \int_{-\infty}^{\infty} f(t) 2^{-k/2} \psi(2^{-k}t - l)\mathrm{d}t \tag{2}$$

where k is a scale variable and l is a translation variable, with both being integers ($k > 0$ means stretching, and $k < 0$ means contracting the wavelet, whereas l is its translation in time). The term $\psi(2^{-k}t - l)$ is called a wavelet function, and corresponds with windows of various widths at various instants of time. They are scaled and shifted versions of the mother wavelet. The inverse transform is:

$$f(t) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \mathrm{d}(k,l) 2^{-k/2} \psi(2^{-k}t - l) \tag{3}$$

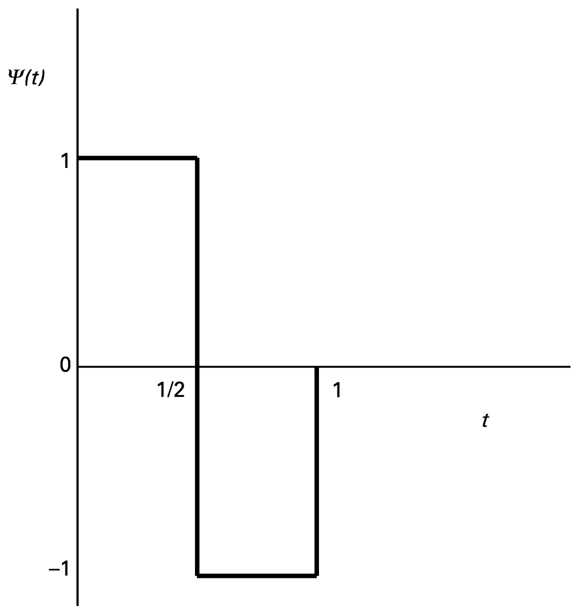

Figure 1 Haar wavelet

Multi-resolution analysis is the basis for the method. It decomposes a signal, and then reconstructs it. In this study, the Haar wavelet is used, due to its simplicity. Therefore, decomposition of a signal (multi-resolution analysis) with the Haar wavelet is considered and detailed below (Rao and Bopardikar 1998).

For a certain value of k, let us define $f_k(t)$ as the average of $f(\tau)$ over an interval of size 2^k:

$$f_k(t) = \frac{1}{2^k} \int_{2^k l}^{2^k(l+1)} f(\tau)\mathrm{d}\tau \quad 2^k l < t < 2^k(l+1) \tag{4}$$

The resolution will decrease as k increases. A change in the data resolution with a change in k, the resolution level, can be seen in the upper part of Figure 2, in which the average of a sample time series taken at different resolution levels, according to Eq. (4), is shown. Note that the data sample used in Figure 2 has 16 elements. Note also the increase in ordinates of $f_k(t)$ with a decrease in k.

The difference between the successive averages $f_{k-1}(t)$ and $f_k(t)$ is defined as the detail function:

$$g_k(t) = f_{k-1}(t) - f_k(t) \tag{5}$$

The middle part of Figure 2 shows the detail functions calculated by Eq. (5) for different resolution levels. Note, from Eq. (4), that $f_4(t) = 0$ for all t. It can be seen that:

$$f(t) = \sum_{k=-\infty}^{\infty} g_k(t) \tag{6}$$

Comparing Eq. (3) with Eq. (6), Bayazit and Aksoy (2001) concluded that multi-resolution analysis using the detail functions was identical to wavelet decomposition with a Haar wavelet. According to Eq. (6), the original signal is obtained when all the detail functions are summed up. At the bottom of Figure 2, $f(t)$, the sum of the four detail functions by Eq. (6) is

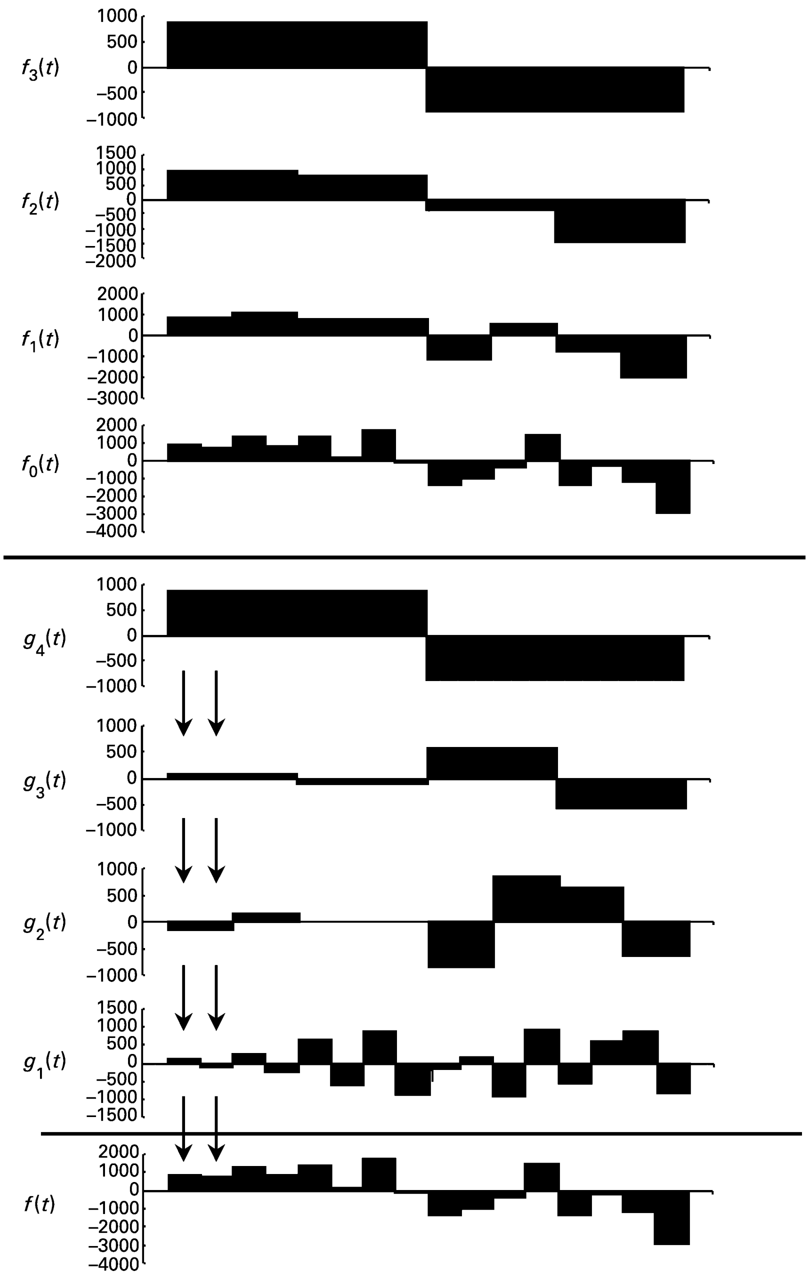

Figure 2 Decomposition and reconstruction of a data sequence

seen, and it represents the original data, $f_0(t)$. Eq. (6) is the basis for the generation algorithm explained below.

Let us consider a data sample of size $M = 2^K$, where K is a positive integer ($K = 4$ for the sequence in Figure 2), taken from a stochastic process $f(t)$ with zero mean: $f(1), f(2), \ldots, f(M)$. Define the sample $f_k(i)$ ($k = 0, 1, \ldots, K; i = 1, \ldots, M$) consisting of averages of 2^k successive elements of the sample. $f_0(i)$ is the original sample and $f_K(i)$ is a sample of all zeros, since the average of M elements is zero. The detail function $g_k(t)$ has a sample consisting of M elements, given by Eq. (4) for $k = 1, 2, \ldots, K$.

Thus, for each element f_i of the original sample, there are K detail function values, $g_k(i)$, corresponding with different resolutions. Choosing from M elements for each $g_k(t)$ randomly, and then summing them up by Eq. (6), a simulated value is obtained for $f(t)$ as:

$$f(j) = \sum_{k=-1}^{K} g_k(j) \tag{7}$$

where j is the index for generated elements. The generation algorithm is given step by step as follows (Bayazit and Aksoy 2001), and is illustrated in Figure 3 for $K = 4$.

(1) In order to obtain the first element of the series ($j = 1$), g_k values ($k = 0, 1, \ldots, K$) are chosen from M values randomly, and summed to obtain f_1 (Figure 3).
(2) The second element ($j = 2$) is generated by choosing for each k, g_k coming just after the g_k values chosen in the first step. The f_2 is obtained by the summation of these (Figure 3).
(3) Data generation is continued in this way for the desired number of times: For the generation of each element f_j, the detail function values right next to those of the previous step $j - 1$ at each resolution level are used.

Wavelet analysis in the form presented in this study is a non-parametric method. The advantage of using a non-parametric method for data generation is that it is not necessary to choose a distribution for the stochastic process and estimate its parameters. A drawback with the method is that the skewness of the data cannot be preserved. Bayazit *et al.* (2001) therefore concluded that the algorithm could not reproduce the coefficient of skewness for non-normal series. Simulated series have skew coefficients that vary widely, tending to zero on average as the number of simulations increases, which may be explained by the central

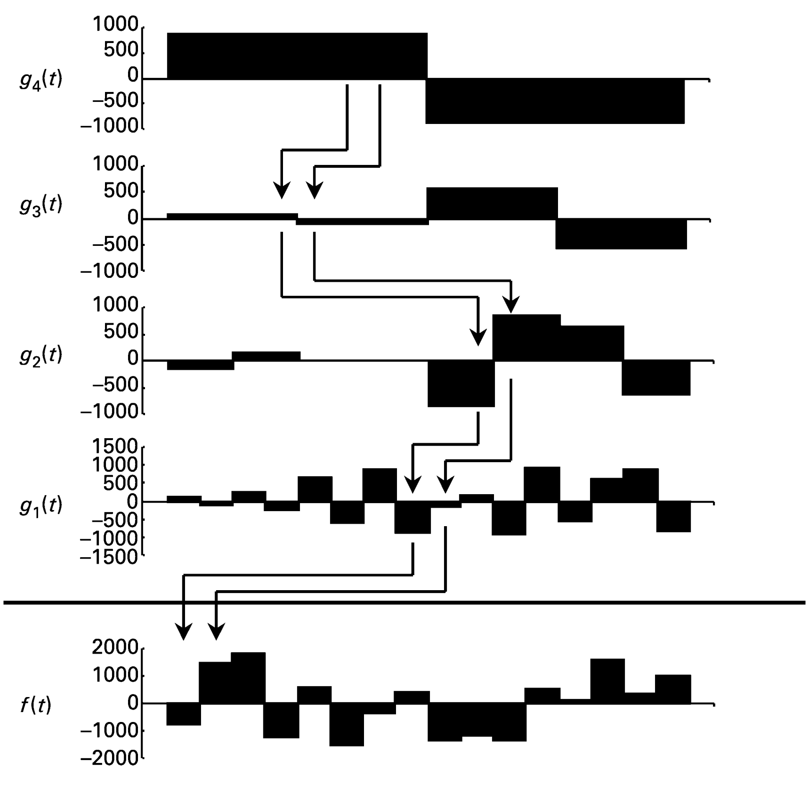

Figure 3 Construction of a synthetic data sequence

limit theorem. Non-normal series can be normalized by a suitable transformation, simulated by wavelet analysis, and then transformed back. This procedure approximately reproduces the coefficient of skew of the original series.

Rao and Bopardikar (1998) are referred to for detailed information on multi-resolution analysis of wavelet functions. Particular details on multi-resolution analysis of the Haar wavelet can be found in Bayazit and Aksoy (2001) and Bayazit *et al.* (2001).

Application

The wavelet-based algorithm was applied to an annual mean suspended sediment discharge series, 32 years in length, taken from the Juniata River at Newport in Pennsylvania, USA (USGS station number 01567000). The basin area is 8,687 km^2. Thirty-two years of observation – from 1952 to 1983 – were used. Characteristics of the annual mean suspended sediment discharge at the station are given in Table 1, and the corresponding time series is plotted in Figure 4. As seen from Table 1, the data set has a non-symmetric distribution with a skewness coefficient of about 3. However, it was stressed in the previous section that the proposed wavelet-based algorithm is valid only for data sets with a skewness coefficient of zero. Therefore, the data set was transformed by using $y = x^{\theta}$, where x is the original data and y is its transformed value. This is known as the Box and Cox (1964) transformation. Exponent θ in the transformation was set to 0.0526 to obtain zero skewness. The data set, however, is independent.

Two thousand synthetic series, each 32 years in length, were generated by the wavelet algorithm. The results are given in Table 2. The mean is preserved. Average standard deviation is only 8% lower than the observed value. The coefficient of skewness obtained after inverse transformation is lower than the observed counterpart. The relative error between the two coefficients is rather high. However, the uncorrelated structure of the data set was preserved in the generated series, although correlation coefficients of individual series vary over rather wide ranges. Maximum and minimum sediment discharge values averaged for 2000 series are comparable with their observed counterparts. The average of the

Table 1 Characteristics of observed annual mean suspended sediment discharge series

Average (tonnes/day)	Standard deviation (tonnes/day)	Skewness coefficient	Autocorrelation coefficient			Maximum (tonnes/day)	Minimum (tonnes/day)
			*r*1	*r*2	*r*3		
677.2	460.2	2.895	0.042	0.037	−0.122	2688	153.6

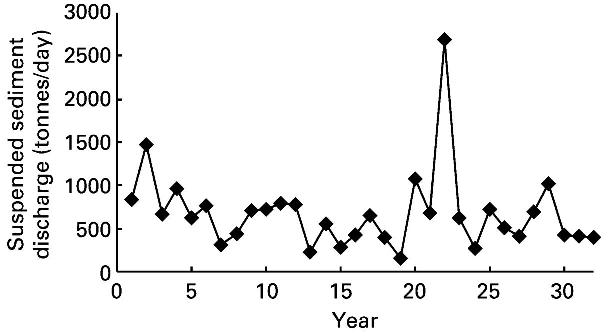

Figure 4 Observed annual mean suspended sediment discharge series

Table 2 Characteristics of generated data (wavelet)

	Average (tonnes/ day)	Standard deviation (tonnes/day)	Skewness coefficient	Autocorrelation coefficient			Maximum (tonnes/ day)	Minimum (tonnes/ day)
				*r*1	*r*2	*r*3		
Ave.	677.6	422.3	1.683	−0.043	−0.027	0.017	2094	182.3
Min.	618.7	232.1	−0.122	−0.418	−0.458	−0.474	1081	79.37
Max.	777.2	919.0	5.288	0.360	0.497	0.734	5583	304.0

maxima was found to be lower than that of the observed series, and the average of minima was found to be higher.

The wavelet algorithm was compared to a first-order moving-average process [MA(1)] fitted to the data set. Again, 2000 series were generated. Results of the MA(1) process are listed in Table 3, from which it is seen that the average is perfectly preserved. The average standard deviation is 10% lower than the observed value. A less skewed distribution was obtained. The correlation structure of the data set, however, was preserved: i.e. no correlation was found in the generated series. The average of the maxima was found to be lower and the average of the minima was found to be higher than those of the observed series.

When the results of wavelet algorithm are compared with those of the moving-average process, it is seen that both methods are very good at replicating the average of the observed suspended sediment discharge series as an average of 2000 series. When generated series were taken individually, it is seen that the moving-average process has a wider range for the average. Standard deviation is almost the same in both methods, although it is closer to the observed value in the wavelet algorithm. Its range is, however, wider in the moving-average process. Both methods were found to be less effective in preserving the non-symmetric shape of the distribution. This is a result of the transformation – not of the methods – as the skewness coefficients of the simulated series were zero before the inverse transformation was applied. Annual suspended sediment discharges were found to show no correlation. The independent structure was obtained, on average, in both methods, with a wider range in the moving-average process for the first- and second-order correlation coefficients. A higher maximum for the maxima and a lower minimum for the minima were obtained for both methods. Wider ranges for those characteristics were obtained when using the moving-average process.

Figure 5 shows the similarity between the distributions of the observed and simulated annual suspended sediment discharge series. The distributions of the wavelet algorithm and moving-average process are based on the average of 2000 series. It is seen that the shape of the distribution is preserved by both methods. The same mode of the observation was generated with a relatively higher frequency. The positively skewed distribution is obtained in both methods. The cumulative frequency diagrams of maximum, mean, and minimum

Table 3 Characteristics of generated data (MA)

	Average (tonnes/ day)	Standard deviation (tonnes/ day)	Skewness coefficient	Autocorrelation coefficient			Maximum (tonnes/ day)	Minimum (tonnes/ day)
				*r*1	*r*2	*r*3		
Ave.	677.1	410.9	1.485	−0.089	−0.022	−0.025	1984	190.1
Min.	418.9	135.4	−0.096	−0.521	−0.447	−0.457	767.3	67.30
Max.	1007	1676	5.183	0.571	0.585	0.501	9934	356.7

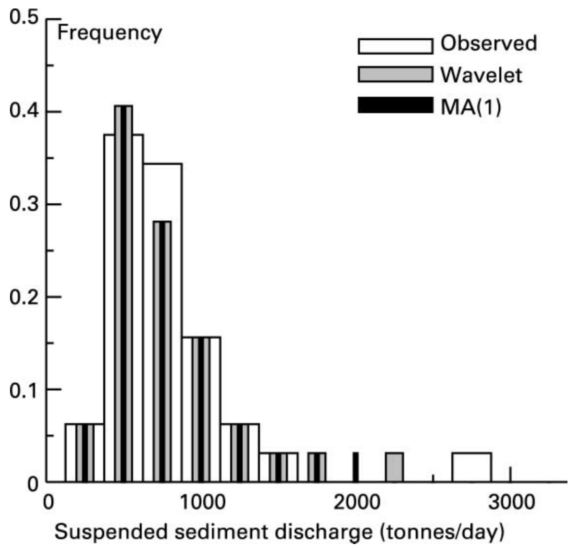

Figure 5 Frequency histogram of observed and simulated suspended sediment discharge series

annual suspended sediment discharges of 2000 series are plotted in Figure 6, from which it is seen that the moving-average process gives a wider range of simulations. Like the flow duration curve, the suspended sediment discharge duration curve can also be plotted as in Figure 7, where five series are compared. Curve (1) is the duration curve of the 32-year observation. Curve (2) is the duration curve of the average of 2000 series simulated by the wavelet algorithm, and curve (3) is that simulated by the moving-average process. These two curves match the observed duration curve perfectly. Curves (4) and (5) consist of 64,000 suspended sediment discharges (2000 series, each 32 years in length) simulated by wavelet and moving-average, respectively. The curves match the observed duration curve as well as curves (2) and (3) taken on average. However, as expected, higher maxima and lower minima are observed for very small and very large percentages of time exceeded.

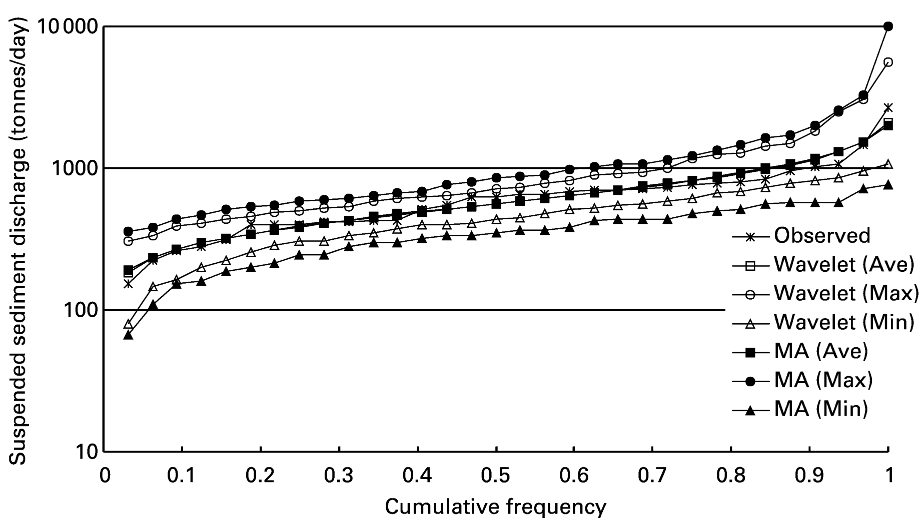

Figure 6 Cumulative frequency diagram of the observed and generated suspended sediment discharge series

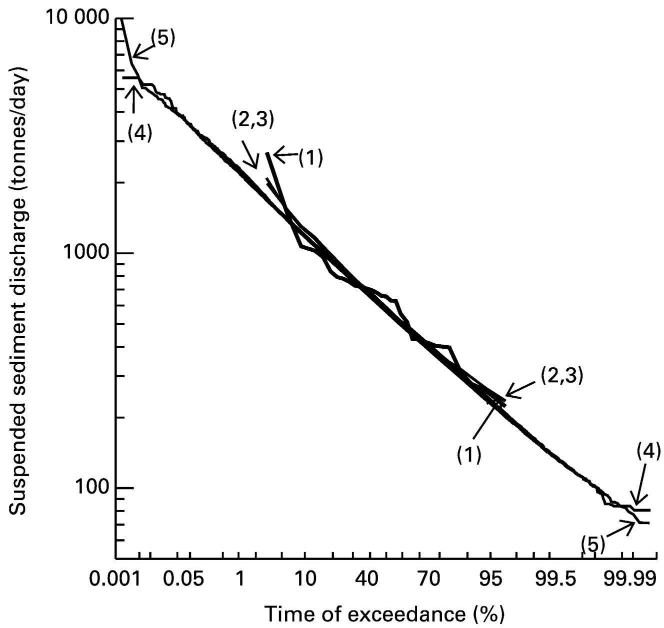

Figure 7 Suspended sediment discharge duration curves: (1) observed; (2) wavelet (average of 2000 series); (3) MA(1) (average of 2000 series); (4) wavelet (32-year 2000 series); (5) MA(1) (32-year 2000 series)

Conclusions

A newly developed synthetic data generation approach based upon wavelet theory is proposed in this study. The approach first decomposes a data series into its details, and then reconstructs those details randomly, to end up with a synthetic data series. The Haar wavelet is chosen for its simplicity. Other wavelets can also be considered for decomposition of the data series that might improve the results. Application of the approach to an annual mean suspended sediment discharge series, and its comparison to a moving-average process fitted to the data set, suggest that the approach can be used as a synthetic data generation tool. When the results of this study are combined with those from earlier applications of the proposed approach (Bayazit and Aksoy 2001; Bayazit *et al.* 2001; Aksoy 2001), it is concluded that the wavelet-based approach is an alternative to the traditional AR and MA type models and/or their seasonal or periodic versions for synthetic generation of hydrological data sets.

Acknowledgments

The authors thank Randy S. Parker of the US Geological Survey for providing data, and two anonymous reviewers for their constructive comments.

References

Aksoy, H. (2001). Storage capacity for river reservoirs by wavelet-based generation of sequent-peak algorithm. *Wat. Resour. Manag.*, **15**(6), 423–437.

Baban, S.M.J. and Yusof, K.W. (2001). Modelling soil erosion in tropical environments using remote sensing and geographical information systems. *Hydrol. Sci. J.*, **46**(2), 191–198.

Bayazit, M. and Aksoy, H. (2001). Using wavelets for data generation. *J. Appl. Stat.*, **28**(2), 157–166.

Bayazit, M., Onoz, B. and Aksoy, H. (2001). Nonparametric streamflow simulation by wavelet or Fourier analysis. *Hydrol. Sci. J.*, **46**(2), 623–634.

Bogardi, J. (1974). *Sediment Transport in Alluvial Streams*, Akademiai Kiado, Budapest.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *J. Stat. Soc.*, **B26**, 511–552.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time Series Analysis, Forecasting and Control* (3rd edn), Prentice-Hall, Englewood Cliffs, New Jersey.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Commun. Pure & Appl. Math.*, **41**, 909–996.

Feng, G. (1998). A method for simulation of periodic hydrologic time series using wavelet transform, stochastic models of hydrological processes and their applications to problems of environmental preservation. In: *Proceedings, NATO Advanced Research Workshop,23–27, November 1998, Moscow, Russia*, Water Problems Institute, Moscow.

Garde, R.J. and Ranga Raju, K.G. (1977). *Mechanics of Sediment Transportation and Alluvial Stream Problems*, Wiley Eastern, New Delhi.

Grossman, A. and Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, **15**(4), 723–736.

Kulkarni, J.R. (2000). Wavelet analysis of the association between the southern oscillation and the Indian summer monsoon. *Int. J. Climatol.*, **20**, 89–104.

Kumar, P. and Foufoula-Georgiou, E. (1993). A multicomponent decomposition of spatial rainfall fields, 1. Segregation of large- and small-scale features using wavelet transforms. *Wat. Resour. Res.*, **29**(8), 2515–2532.

Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

Morlet, J. (1981). Sampling theory and wavelet propagation. *Proceedings, 51st Annual Meeting of the Society for Exploration Geophysics, Los Angeles, USA*.

Phien, H.N. (1981). Reservoir sedimentation with correlated inflows. *J. Hydrol.*, **53**, 327–341.

Phien, H.N. and Arbhabhirama, A. (1979). A statistical analysis of the sediment volume accumulated in reservoirs. *J. of Hydrol.*, **44**, 231–240.

Rao, R.M. and Bopardikar, A.J. (1998). *Wavelet Transforms, Introduction to Theory and Applications*, Addison-Wesley, Reading, MA.

Skoklevski, Z. and Velickov, S. (1998). Suspended load transportation process within Vardar River basin in the Republic of Macedonia. In: *XIXth Conference of the Danube Countries on Hydrological Forecasting and Hydrological Bases of Water Management, Osijek, Croatia, 15–19 June 1998*, D. Geres and D. Trninic (Eds.), Hrvatske Vode, Osijek, pp. 717–727.

Smith, L.C., Turcotte, D.L. and Isacks, B.C. (1998). Streamflow characterization and feature detection using a discrete transform. *Hydrol. Processes*, **12**, 233–249.

Szidarovszky, F., Yakowitz, S. and Krzysztofowicz, R. (1976). A Bayes approach for simulating sediment yield. *J. of Hydrol. Sci.*, **3**(1–2), 33–44.

Szilagyi, J., Parlange, M.B., Katul, G.G. and Albertson, J.D. (1999). An objective method for determining principal time scales of coherent eddy structures using orthonormal wavelets. *Adv. Wat. Resour.*, **22**(6), 561–566.

Tingsanchali, T. and Lal, N.K. (1992). A combined deterministic–stochastic model of daily sediment concentrations in a river. In: *Proceedings of the Sixth IAHR Symposium on Stochastic Hydraulics, Taipei*, J.T. Kuo (Ed.), A.A. Balkema, Rotterdam, pp. 221–228.