Geziraj. of eng. & applied sci.4 (1) : 70 — 94 (2009)

Nonlinear Time Series Models:

An Application on Amount of Water Flow of Blue Nile River Measured at Eldaim Station

Merdi Ahmed Orsud[1] and Bassam Younis Ibrahim[2]

[1]Faculty of Mathematical & Computer Sciences, University of Gezira, Medani, Sudan.
[2]Sudan University of Science and Technology.

ABSTRACT

This paper had undertaken nonlinear time series modelling and in particular, discussed wavelet smoothing technique to decompose the time series into a wavelet smoothed component and a random component. The random component was then modelled by an appropriate linear ARIMA process or diagonal pure bilinear process.

Before smoothing technique was applied, flow data was tested for linearity and then filtered. By investigating the plot of the third cumulant, it was found that diagonal pure bilinear process of order two was the best for the data sets under study. diagonal Pure bilinear of order two model was fitted to time series data set based on the mean daily Blue Nile River flow variable at Eldaim Station, (during the period January 2005 to December

2006) using wavelet smoothing technique. A simulation technique was performed to find the appropriateness of the model by comparing its performance with the actual time series data.

The wavelet smoothing technique demonstrated an attractive technique to model such a time series data.

Key Words: ARIMA; Bilinear time series; Cumulant; Wavelet filtering.

INTRODUCTION

There are many techniques used to decompose non-stationary time series data into their components, one of these techniques used by Brockwell and Davis (1996). That technique stands on passing the series through a low pass filter. The filter allow the proposed modeling approach based on wavelet smoothing. The main idea was to pass the nonlinear series through a wavelet filter which separates the series into two components: a deterministic smoothed version of the series and a random component. In hydrology the ability to model the average daily river flow for rivers plays an important role in the prediction of possible disasters such as floods.

The objective of this paper was to simplify the mathematical structure of the series and remove any trend so that the random component could be represented by a less complicated model. The expectation was that, linear integrated autoregressive moving average model and a simplest model from the conditional mean models, namely Diagonal Pure Bilinear model, would be sufficient for representing the random component (Oyet, 2000). We are attracted to Diagonal Pure Bilinear model from the conditional mean models because of its simple mathematical structure. In this way, the problem of selecting and fitting a suitable nonlinear model, from large number of possible ones, to the series was avoided partially. Also an attempt to study nonlinear time series models and applied it on one life data series, namely the mean daily Blue Nile river flow at Eldaim Station, to explain the suitable model for Blue Nile river flow series and then to use fitted model in forecasting.

Before using data in series in any formal modeling it would be necessary to test the time series for linearity. The statistical approach and methods adopted would establish whether the time series might be linear or nonlinear (Tong, 1990; Fan and Yao, 2003).

Data ware passed through wavelet filter to decompose series into smoothed version and random component. Autoregressive Integrated Moving Average process or Diagonal Pure Bilinear process was fitted to random component. To check whether the wavelet smoothing techniques along with using the Autoregressive Integrated Moving Average process or Diagonal Pure Bilinear process to model random component sustains specific properties of the original process, the properties were studied by simulating time series based on random component.

In the following some information about Autoregressive Integrated Moving Average (ARIMA) and Diagonal Pure Bilinear (*DPBL*) processes was given.

Autoregressive Integrated Moving Average (ARIMA) Model

In many practical situation the assumption of stationarity of time series is too restrictive, in that their plots was frequently show some kind of trend in the mean and possibly in the variance (Harvey, 1981).

Fortunately, most of the non-stationary series encountered could be transformed into stationary series. Such series are termed homogeneous non-stationary (Pindyak & Rubinfeles, 1981). Homogeneous non-stationary series transformed into stationary by successive differencing (Box & Jenkins, 1976); i.e., by considering ∇X_t , $\nabla^2 X_t$,. where $\nabla =$ l- B is the difference operator. A generalization of the Autoregressive Moving Average (ARMA) models incorporating the above type of non-stationarity is given by the class of autoregressive integrated moving average (ARIMA) process and is defined as follows

If d is a non-negative integer, then X_t is said to be an *ARIMA*(p, d, q) process if

$$\emptyset\,(B)\nabla^{d} X_t = \theta(B)\varepsilon_t \;; \qquad (1)$$

Where $\emptyset(.)$ and $\theta\,(.)$ are stationary AR and invertible MA polynomials of degree p and q respectively.

Box and Jenkins methodology was used for identifying, estimating and diagnosing ARIMA models.

Bilinear models

The BL(p,q, P,Q) model is defined by

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{i=1}^{P}\sum_{j=1}^{Q} \beta_{ij} X_{t-i}\varepsilon_{t-j} + \varepsilon_t \quad (2)$$

where $\emptyset_i$, θ_j and β_{ij} are constants.

It is called bilinear model because it is linear in the X' s given errors and linear in the errors given X' s, but it is not linear with respect to both of them (Rao and Da Silva, 1992).

This model is an extension of the (linear) ARMA model obtained by adding the nonlinear term $\sum_{i=1}^{P}\sum_{j=1}^{Q}\beta_{ij}\ X_t -i^{\varepsilon}t - j$ to the right hand side. When $\beta_{ij} =\ 0$ for $i \neq j$ the model is called Diagonal Bilinear (DBL) model (Tong, 1990).

If the DBL model written as

$$X_t = \sum_{i=1}^{P}\sum_{j=1}^{Q}\beta_{ij} X_{t-i}\varepsilon_{t-j} + \varepsilon_t \qquad (3)$$

the model is called Diagonal Pure Bilinear (DPBL) model (Rao, 1981). The properties of the DPBL model have been studied by Oyet (2000).

Third Order Moments and Cumulants

Before models ware fitted to the random component W(t) of the series X_t using the DPBL(q) process given by equation (3), the order of the process would be determined using the third order moments and cumulants. If third order stationary was assumed, the third order cumulants might depend only on kl and k_2 for all admissible integers t, k_l and k_2 was given by

$$C(k_1, k_2) = m(k_1, k_2) - \mu[R(k_1) + R(k_2) + R(k_1 - k_2)] - \mu^3$$

where $m(K_1, K_2) = E(X_t \; X_{t+k} \; X_{t+k2})$. It had been shown that the cumulants $C(k_1, k_2)$ of real valued process X_t have the following symmetric relationship:

$$C(k1, k2) = C(k_2, k_l) = C(-k_1, K_2 - k_l) = C(k_l - K_2, - K_2).$$

From this relation, once the value in the upper half of the first quadrant of the Euclidean plane are known, then all the values of $C(K_1, k_2)$ are defined.

Table (1): C(l, k2) Pattern for Arbitrary q

k_2 \ k_1	1	2	3	...	q	q+1	q+2	q+3	...
1		NZ	NZ	NZ	...	NZ	NZ	0	0
	...								

Where NZ represents a nonzero entry.

Form Oyet (2000) $C(k1, k2) = 0$ for $K_1 \leq l$, $K_2 - K_1 > q$ and $k_l > q$, $k_2 - k_1 \geq q$ and $C(k_l, k_2)$ nonzero when $k_2 > k_1$. Based on these results, it had been shown that the third order cumulants define a pattern in the upper half of the first quadrant of the klk2 plane as shown in table (1). This pattern was then easily extended to the entire Euclidean plane from the symmetric relationship satisfied by the cumulants. A useful pattern then, for detecting the order of a DPBL(q) as can be seen in table (l) is: $C(k_l, k_2) = 0$, for $K_2 = q + 2, q + 3, \ldots$ and nonzero elsewhere for an arbitrary value of q.

To determine the order of the model the behavior of the standardized cumulants had been investigated as given by equation (4)

$$p(l,k2) = \frac{C(1,k2)}{C(0,0)} \qquad (4)$$

for a given finite sample time series Xt satisfying equation (3). The third order cumulants in equation (4) was estimated by

$$C(k_1,k_2) = \frac{1}{n-k_1-k_2} \sum_{t=1}^{n-k_1-k_2} (X_t - \bar{X})(X_{t+k_1} - \bar{X})(X_{t+k_2} - \bar{X})$$

The order of the best model will be $k^*_{2\,-1}$, where k^*_2 is the value of k_2 at which $p(l, k_2)$ cuts-off. The cut-off point refers to the point at which $p(l, k_2) = 0$. Since the sample estimates of $p(l, k_2)$ will not be exactly zero, standardized cumulants trace was used, a plot of the absolute values of $\hat{p}(1, k2)$ versus k_2, to determine the point where $\hat{P}(1, k_2)$ cuts off. Therefore, K_2^* would be the value of k_2 where the standardized cumulants trace begins to stabilize and hence , the order of the model would be K_2^* - 1.

THE RESULTS

Wavelet Smoothing

The time series X_t is then allowed to pass through linear wavelet filter which decomposes X_t into a non random wavelet smoothed version $\hat{\eta}$(t) and random component W(t). W(t) is remainder of the time series X_t after $\hat{\eta}$ (t) has been removed. It has some autocorrelation structure, but the underlying aspect of W(t) is that it is still a time series and moreover, it has a simpler structure than X_t.

The series X_t is constructed from linear combinations of $\eta^{(j)}$ at various levels of j given by

$$X_t = \eta^{(\emptyset)} + \eta^{(0)} + \eta^{(1)} + \ldots + \eta^{(m)} + \ldots = \eta\,(t\,,m) + W(t) \qquad (5)$$

where $\eta^{\emptyset}$ is a multiple of a scaling function $\emptyset$(t) and $\eta^{(j)}$ is a linear combination of 2^j dilated and translated versions of a mother function $\psi(t)$. The linear combination $\eta\ (t\ ,m)$ could be written in terms of $\emptyset(t)$ and $\psi^{-j,k}(t) = 2^{j/2}\ \psi\ (2^{j}t - k)$ as

$$\eta(t;m) = d\phi(t) + \sum_{j=0}^{m} \sum_{k=0}^{2^{j}-1} c_{jk} N \psi^{-j,k}(t) \qquad (6)$$

The completed work would involve the Daubechies wavelet system generated by $_5\emptyset^{(t)}$ and $_5\psi^{(t)}$. Along with equation (6) the series X_t is given by

$$X_t = \sum_{j=1}^{N} q_j(t)\omega_j + W(t) \qquad \text{where} \quad N = 2^{m+1} \qquad (7)$$

The components of the Nxl vector $\omega = (\omega\ 1, \ldots, \omega\ N)'$are the filter coefficients $\{d, c_{jk}\}$ which would be determined from n realizations of the time series X_t. The vector $q = (q_I(t), \ldots, q_N(t))'$ is comprised of the wavelet system chosen for the filtering process. From equation (7) the nonlinear time series is X_t broken down into and described by two components. The first component is nonrandom wavelet smoothed version and the second is a random process. (Strang, 1989; Antoniadis, et al 1994; Morettin, 1997).

Estimation

Without loss of generality, the space of all possible values of t had been normalized to the [0, 1] interval. Given n realizations of nonlinear time series X_t then, from Oyet (2000), the smoothed version was evaluated as

$$\hat{\eta}(r;m) = \int_0^1 h(r;t) X_t \upsilon(t) d\zeta(t) \tag{8}$$

where

$$h(r;t) = \mathbf{q}'(r)\mathbf{B}'\mathbf{q}(t); \qquad \mathbf{B} = \mathbf{B}(\upsilon,\zeta) = \int_0^1 \mathbf{q}(t)\mathbf{q}'(t)\upsilon(t)d\zeta(t)$$

And ξ(t) is the empirical distribution function of $\{ti\}_i^n = 1$. It could be established that $\hat{\eta}$(r) is unbiased with variance

$$V(\hat{\eta}(r;m)) = \frac{R_w(0)}{n}\mathbf{q}'(r)\mathbf{B}^{-1}\mathbf{D}_1\mathbf{B}^{-1}\mathbf{q}(r) + 2\mathbf{q}'(r)\mathbf{B}^{-1}\mathbf{D}_2\mathbf{B}^{-1}\mathbf{q}(r)$$

Where

$$\mathbf{D}_1 = \int_0^1 \mathbf{q}(t)\mathbf{q}'(t)\upsilon^2(t)d\zeta(t)$$

$$\mathbf{D}_2 = \int_0^1 \int_{(t>s)} \mathbf{q}(t)\mathbf{q}'(s)\upsilon(t)\upsilon(s)R_w(t-s)d\zeta(t)d\zeta(s)$$

and R_W (.) is the autocovariance function of W(t). At this point, it might be determined how to choose the most appropriate values for υ(t). The approach taken by Oyet (2000) was to choose υ(t) such that the Integrated Variance of $\hat{\eta}$ (r) denoted by IV($\hat{\eta}$ (r)) is minimized. The integrated variance of $\hat{\eta}$(r) is given by

$$IV(\hat{\eta}(r;m)) = \int_0^1 V(\hat{\eta}(r;m))dr = \frac{R_w(0)}{n} tr\{\mathbf{D}_1\mathbf{H}^{-1}\} + 2tr\{\mathbf{D}_2\mathbf{H}^{-1}\}$$

$$= \frac{R_w(0)}{n}\int_0^1 \|\mathbf{H}^{-1/2}\mathbf{q}(t)\|^2\, \upsilon(t)d\zeta(t) \tag{9}$$

$$+ 2\int_0^1\int_{(t>s)} \mathbf{q}'(s)\mathbf{H}^{-1}\mathbf{q}(t)\upsilon(t)\upsilon(s)R_w(t-s)d\zeta(t)d\zeta(t)$$

where $\mathbf{H} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{A} = \int_0^1 \mathbf{q}(r)\mathbf{q}'(r)dr$.

The IV given in equation (9) can be minimized by finding an appropriate weight function υ (t). The approach taken by Oyet (2000) is to search for an absolutely continuous measure which minimizes the IV loss function by allowing the measure ξ to be extended to the space of all distribution functions. The weight suggested is

$$\upsilon_0(t;u) = \frac{u}{\| \mathbf{A}^{1/2}\mathbf{q}(t) \|}; \qquad u_Q = \int_0^1 \| \mathbf{A}^{1/2}\mathbf{q}(t) \| \, dt \quad (10)$$

Then υ_o $(t;u_Q)$ minimizes equation (8) under the constraints that m(t) υ (t) I and $\int_0^1 m\ (t)dt = 1$.

Model Building

Before building a time series model for the daily amounts of water flow of Blue River Nile at Eldaim Station from the 1/1/2005 to 31/12/2006 the original series plot was inspected. This would help selecting appropriate transformations for the series. Figure (1) show the plot of the original series.

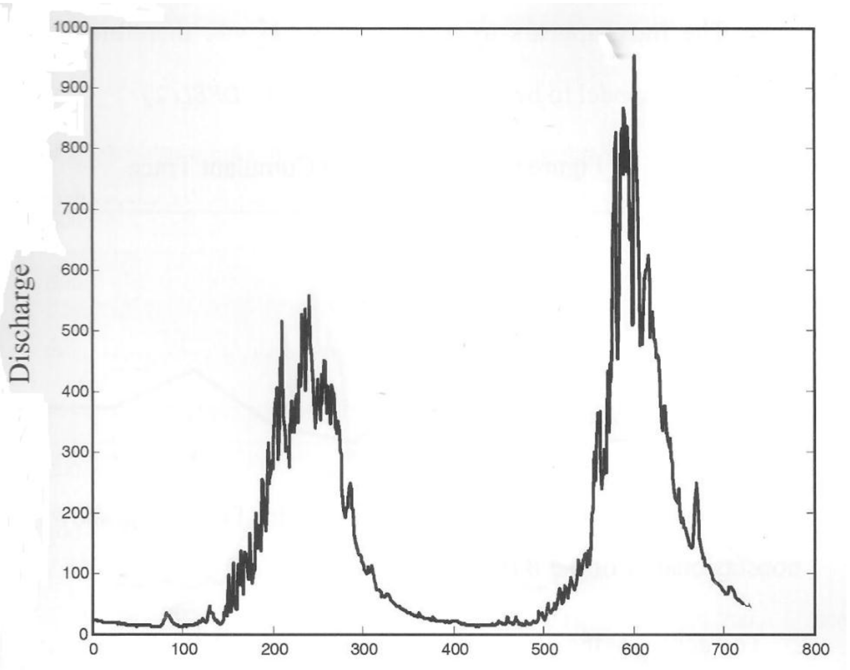
Linearity Test

Mc Lead and Li test was applied to test the linearity of the given data set . The test statistics for lag 2, 4, 6, 8, 10, and 16 are shown in table (2).

Table (2): McLeod and Li Linearity Test for Blue Nile River Flow data

Lag	Q	Q-critical	Hid
2	1434.954	5.991	Reject
4	2790.106	9.488	Reject
6	4068.053	12.592	Reject
8	5289.689	15.507	Reject
10	6455.478	18.307	Reject
16	9455.638	26.296	Reject

Figure (1): The Plot of the Daily Water Flow at Eldaim Station



Day

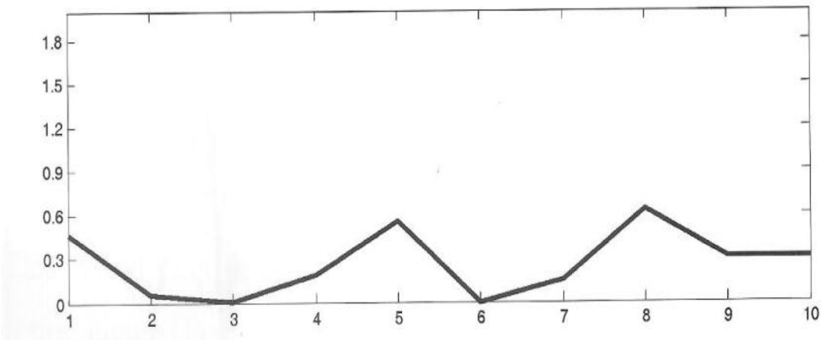
For all lags in table (2) the results provide significant evidence that the river flow time series data was non-linear.

Model Fitting to Flow Data

Daily River flow data ware passed through wavelet filter to decompose series into smoothed version η(t;m) and random component W(t).

The standardized cumulants trace in figure (2) was based on W(t). The trace appears to stabilize after $k_2^* = 3$, therefore the order of the model to be fitted is DPBL(k_2^* -1) = DPBL(2)

Figure (2): Standardized Cumulant Trace



The plot of the random component in Figure (3) shows nonstasionarity of the W(t) series.

To achieve the stationarity the series W(t) had been transformed by taking first difference. Figure (4) shows the plotting of the series W(t) after taking the first difference. The plotting of the series shown stationarity after first difference.

Figure (3): The Plot of Random Component W(t)

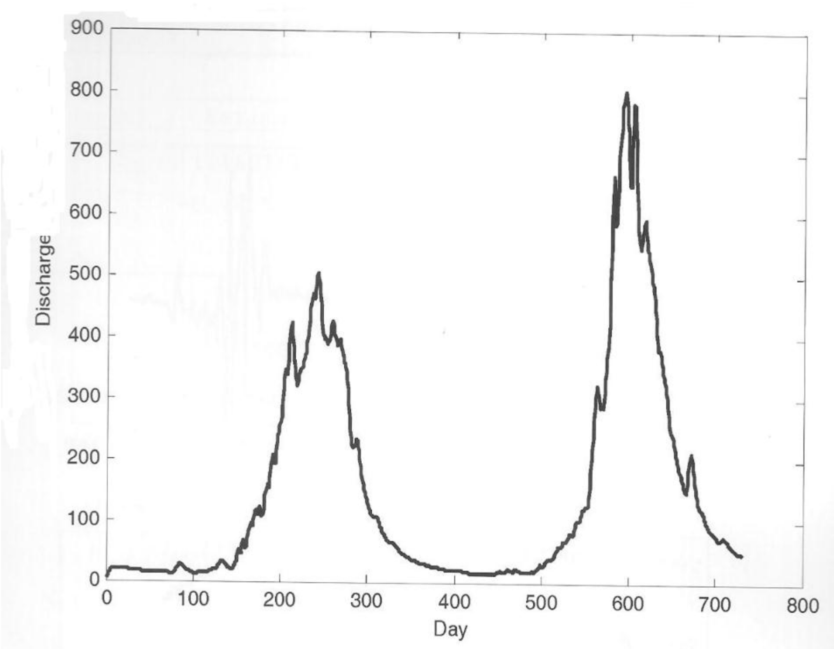


Figure (4): Plot of the Series W(t) After First Difference

Using procedure Model in SAS statistical package, which is a procedure of nonlinear estimation that requires initial values, the parameter estimates had been obtained.

Diagonal Pure Bilinear model given by equation (3), was fitted to time series after the wavelet smoothed version had been removed. i. e. The model was fitted to the time series W(t).

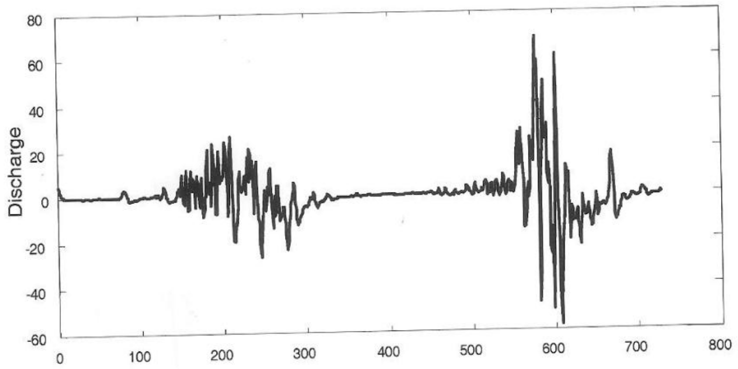


Table (3): Summary of Residual Errors

The MODEL Procedure

Nonlinear Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	R-Square	Adj R-Sq
X	3	724	24075.8	33.2538	0.7032	0.7024

This table lists the sum of squared errors (SSE), the mean square error (MSE), and R2 and adjusted R2 statistics.

Table(4): Parameter Estimates

Nonlinear Parameter Estimates				
		Approx		Approx
Parameter	Estimate	Std Err	t Value	Pr > It I
$\theta 1$	1 . 160161	0.0326	35.60	< 0.0001
$\theta 2$	- 0.48053	0.0326	-14.75	< .0001
c	0.01442	0.2139	0.07	0.9463

For flow time series data were fitted to DPBL (2) and the estimated coefficients of the model from table (4) are θ_1 = 1.160161, θ_2 = - 0.48053 and c = 0.014472.

Therefore the fitted model is

$$W(t) = 0.047387 + 1.160161\, W_{t-l}\, e_{t-1} - 0.48053\, W_{t-2}\, e_{t-2} + e_t$$

Simulation Study

This stage consider whether the wavelet smoothing technique which is using the Diagonal Pure Bilinear process to model W(t) sustains specific properties of the original data, mainly the mean, standard deviation and maximum value. The properties were studied by simulating n = 2000 time series data based on W(t). Then the wavelet smoothed version combined with the simulated random linear component to obtain the following time series

Next, mean, standard deviation and maximum values were be calculated for each of the four samples from the n = 2000 and compare those sampling properties to that of the original series. Finally, this process was investigated to simulate the original time series by overlaying the time series plot of the original series with simulated series.

Table (5): Sampling Properties of X_t, mean ($\tilde{X}_t$), St-Dev ($\tilde{X}_t$) and Max(Xt)

	Min	Median	Mean	Max	St - dev
X_t	11.23	61.69	151.6738	954.88	189.633
Mean ($\tilde{X}_t$)	144.07972	155.3927	154.2671	162.20316	7.5395581
St.Dev ($\tilde{X}_t$)	176.62307	185.1076	184.6215	191.64778	6.4281428
Max ($\tilde{X}_t$)	920.55761	921.7349	921.7349	922.91210	1.3593664

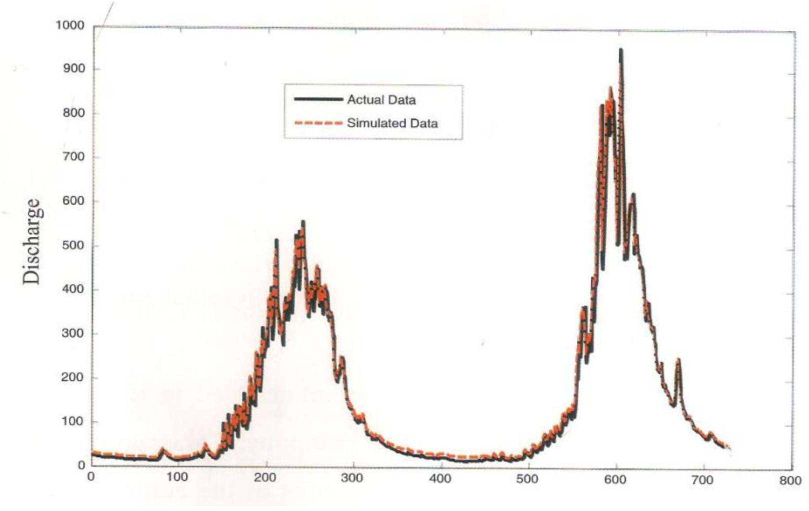
Based on the result in table (5) for the mean, standard deviation and maximum value of the simulation study, the procedure used sustained these properties. The original series X_t had a mean of 155.6738 while the mean of the means of ($\tilde{X}_t$) was 154.2671. The standard deviation of X_t equal 189.6336 while for ($\tilde{X}_t$) the mean of the standard deviations was 184.6215.

The maximum of X_t equal 954.88 while the mean of maxim of ($\tilde{X}t$) equal 921.7349.

Figure (5) displays time series plot of both the actual series X_t and simulated series ($\tilde{X}t$) . From the plot it is observed that the simulated series model fits the original series effectively.

Figure (5): Plot of Actual Time Series With Overlaid Simulated Diagonal Pure Bilinear Time Series Plot.

The effectiveness of the model procedure used in this stage based on two components: the first component was ability of fitted model to keep up sampling properties of the actual series, and the second is to model capacity to simulate the actual time series. From the above the sampling properties and the simulated plot for Blue Nile flow data gave evidence that the procedure was very effective in modeling time series data of river flow.



Day

CONCLUSION

In many studies, time series data considered to follow linear process when analyzed, and this is best approach provided that the assumption of linearity is correct. A problem arises when researcher assumes the time series follows a linear process but in fact it is nonlinear.

In life many time series data are nonlinear, in any study when this non linearity is ignored the fitted models will have no meaning. River flow time series data (January 2005 to December 2006), were considered for this paper and tested for linearity and exhibited nonlinearity.

This paper fited time series model to the river flow variable based on an approach involved decomposing the time series into two component a deterministic smoothed version and a random component, using wavelet smoothing techniques, where the random component was assumed to follow a Diagonal Pure Bilinear process.

Simulation studies were conducted to identify whether the fitted models behave in similar manner to the actual time series. Sampling properties, mean, standard deviation and maximum value were calculated and compared to the original series sampling properties. Also a plot which consisted of simulated series and original series was constructed.

The wavelet smooth techniques was effectively modeled for the mean daily river flow time series data. The results from simulated data were almost identical to their respective original data series.

Research in modeling river flow data is important and continuous processes in hydrology field . For the water flow amounts, measured at Eldaim station, and on the basis of the results of modeling, it is reasonable to recommend that the hydrologists in water Resources Department to use the obtained model to predict or control water flow at Eldaim Station. In future to aid hydrologists in making a proper predictions it is recommended to apply techniques on too large than two years river flow data series. Due to its simplicity, feasibility and flexibility of linear ARIMA model process, it is recommended to use it if possible in analyses of time series.

REFERENCES

Antoniadis, A., et al. (1994), "Wavelet Method for Curve Estimation", *Journal of the American Statistical Associa-tion*, 89, 340 - 1352.

Box, G. E. P. , and Jenkins, G. M. (1976), "Time Series Analysis: Forecasting and Control", San Francisco: Holden — Day.

Brockwell, P. Jo, and Davis, R. A. (1996), Introduction to Time Series and Forecasting, New York: Springer-Verlag.

Fan, J. and Yao, Q. (2003), "Nonlinear Time Series: Nonparame-tric and Parametric Methods", Springer-verlag, New York, Inc.

Harvey, A. C. (1981), " Time Series Models", Oxford: Allan.

Morettin, P. A. (1997), "Wavelets in Statistics", Resenhas, 3, 211-272.

Oyet, A. J. (2000) , "Nonlinear Time Series Modeling: Order Identification and Wavelet Filtering", Technical Report, Department of Mathematics and Statistics, Memorial University of New Found Land.

Pindyck, Roberts. , and Rubinfeled, Daniel L (1981). "Econometric Models and Economic Forecasting", Tokyo: Mc Graw. Hill, Inc.

Subba Rao, T. (1981), "On The Theory of Bilinear Time Series Models", Journal of The Royal statistical Society (Series B), 43, 244 — 255.

Subba Rao T. and Da Silva, Eduarda A. (1992), "Identification of Bilinear Time Series Models BL (p, 0, p, Statistica Sinica, 2, 465 — 478.

Strang, G. (1989), "Wavelet and Dilation Equations: A Brief Introduction", SIAM Review, 31, 614 - 627.

Tong, H. (1990), "Nonlinear Time Series: A Dynamical System Approach", Oxford University Press, Oxford.