감조하천의 홍수위 예측에 있어서 한계자기회귀모형의 응용
Application of Genetic Threshold Auto-regressive Model to Forecast Flood for Tidal River

Chen Guo xin* · 안선복** · 고진석*** · 지홍기****
Chen Guo xin · An Shan fu · Ko Jin seok · Jee Hong kee

요 지

한계자기회귀모형(TAR)을 응용하여 동시에 해조와 홍수의 영향을 받을 때 삽교천 감조구간의 삽교호수위관측소의 월 최고수위를 예측하는 모형을 구축하였으며, 모형구축과정에서 유전알고리즘으로 한계값과 자기회귀계수의 매개변수를 최적화한다. 계산결과 한계자기회귀모형은 감조하천의 비선형성특성을 모의 할 수 있으며, 예측의 정확도와 예측성능의 안정성을 확보할 수 있다. 연구결과 유전한계자귀회귀모형으로 감조하천구간의 월 최고수위를 예측하는 것이 가능하며, 또한 감조하천구간에서 기타 수문요소의 비선형성 서열예측 중에서도 광범한 실용가치가 있다고 본다.

핵심용어 : 한계자기회귀모형(TAR), 감조하천구간, 월 최고수위, 유전알고리즘

1. Introduction

River water level forecasting is one of important components of hydrology forecasting, based on a great deal of data of hydrology, meteorology, geomorphology etc. For unsteady flows, generally St. Venant equation was been established to simulation. However, not only it is difficult for water level forecasting model established by St. Venant equation to solve and correct real-time but also need river topographical data which are difficultly obtained, because the water level series of the tidal reaches often exhibit some well-known intricatefeatures of nonlinear vibrations such as jump, random and harmonic due to influencing by many uncertainty factors of tides and runoff. As the development of statistical theory and time series theory since the 1970's, we have seen remarkable successes in the application of time series model to dynamic data simulation and forecasting in diverse fields including in hydrology, e.g. ARIMA model presented by Box and Jenkins supplying powerful mathematics tool for dealing with linear time series. But linear time series model is totally inadequate as a tool to analyze more intricate nonlinear phenomena, therefore, the study and application of nonlinear time series model have been seriously made more and more. Among this process, H. Tong presented threshold auto-regressive model (TAR) (H. Tong, 1980), which can effectively reflect nonlinear dynamic system charactered by limit cycles, cyclical data, jump resonance, time irreversibility and amplitude-frequency dependency. TAR model is one of ideal model for resolving model construction and forecasting of more intricate cyclic nonlinear time series, and ensures the forecasting accuracy and adaptability by controlling threshold values. Based on TAR model is quite adequate to reflect nonlinear feature, this paper adopted it for

* 정회원·영남대학교 대학원·박사과정·E-mail: gxchen1201@hotmail.com
** 정회원·영남대학교 대학원·박사과정·E-mail: asf0290@hanmail.com
*** 정회원·영남대학교 대학원·박사과정·E-mail: springtime@yu.ac.kr
**** 정회원·영남대학교 건설환경공학부·교수·E-mail: hkjee@yu.ac.kr

water level simulation of tidal reaches and researches the feasibility and simulation results of its application to tidal reaches.

2. TAR model

The general form of TAR model presented by H. Tong is

$$x_t = \begin{cases} \varphi_0^{(1)} + \varphi_1^{(1)} x_{t-1} + \cdots + \varphi_{p_1}^{(1)} x_{t-p_1} + a_t^{(1)}, -\infty < x_{t-d} < \bar{x}_1, \\ \varphi_0^{(2)} + \varphi_1^{(2)} x_{t-1} + \cdots + \varphi_{p_2}^{(2)} x_{t-p_2} + a_t^{(2)}, \bar{x}_1 \le x_{t-d} < \bar{x}_2, \\ \cdots\cdots \\ \varphi_0^{(r)} + \varphi_1^{(r)} x_{t-1} + \cdots + \varphi_{p_r}^{(r)} x_{t-p_r} + a_t^{(r)}, \bar{x}_{r-1} \le x_{t-d} < +\infty \end{cases} \tag{2.1}$$

Where $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{r-1}$ are termed threshold values; d is delay step number $(d \ge 1)$ r is threshold interval number; $\varphi_1^{(i)}, \cdots, \varphi_{p_i}^{(i)} (i = 1, \cdots, r)$ is autoregressive coefficients of the i th interval; $p_i (i = 1, \cdots, r)$ is the order number of the AR model at the i th interval; $\{a_t^{(i)}\} (i = 1, \cdots, r)$ is independent white noise series.

The basic concept of TAR model is based on the idea that the threshold values $\bar{x}_i (i = 1, \cdots, r-1)$ are introduced to observation series $\{x_t\}$ which are divided into r number intervals according to threshold values. The series $\{x_t\}$ are allocated to different interval based on values of $\{x(t-d)\}$ in the condition of the delay step number d. Then $x(t)$ falling into different intervals is described by corresponding AR model. Thus the r number AR models integrally construct nonlinear dynamic time series $\{x(t)\}$.

Because TAR model is of the nonlinear time series model which is based on piece-wise linearization, the construction of the model follows the parameter estimation method and test criteria as AR model, such as least square method and Akaike Information Criterion (AIC). However, different from construction of AR model focused on the parameter estimation and model test, TAR model mainly considers the identifications of the threshold interval number r, threshold values $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{r-1}$ and delay step number d. Theoretically, the process of TAR construction is a problem of multi-dimension optimal method for $r, \bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{r-1}$, and d. If the AIC is adopted for determination of model order $p_1, \cdots, p_r$ of diverse threshold intervals and object function of $r, \bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{r-1}$ and d, the object function value A can be expressed $A = f(p_1, \cdots, p_r, d, \bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{r-1})$. When the A value is minimum, the optimal parameter is obtained. Because there are considerable computations, a preferable method that is genetic algorithm was introduced to optimize parameters of TAR model.

3. Application of TAR model to forecasting water level at tidal reaches

Based on monthly maximum water level series $\{x(t) \quad t = 1, 2, \cdots, 132\}$ of Sapgyoho hydrology gauging station in Korea from 1988 to 1998, the TAR model was established and tested by monthly maximum water level series from 1999 to 2001. Model-building steps are as follow.

1) The delay step number d of TAR model and model order $p_i (i = 1, 2, \cdots, r)$ of AR model in threshold interval I can be determined with the technique of auto-correlation analysis. The auto-correlation coefficient of delay step number k of series $\{x_t\}$ can be given by formula:

$$R(k)=\frac{\sum_{t=k+1}^{n}(x(t)-\bar{x})(x(t-k)-\bar{x})}{\sum_{t=1}^{n}(x(t)-\bar{x})^2} \quad (3.1)$$

Where $\bar{x}=\frac{1}{n}\sum_{t=1}^{n}x(t)$, n is sample size of $\{x_t\}$, $k=1,2,\cdots,n_k,n_k<\frac{n}{10}or\frac{1}{4}$. The autocorrelation coefficients of the former 20 orders are shown in table 1.

Table 3.1. Autocorrelation coefficients of monthly maximum water level

k	1	2	3	4	5	6	7	8	9	10
R(k)	0.4	0.1	-0.07	-0.14	-0.27	-0.38	-0.29	-0.2	0.03	0.07
k	11	12	13	14	15	16	17	18	19	20
R(k)	0.33	0.46	0.3	0.13	0.04	-0.09	-0.36	-0.37	-0.29	-0.11

According to results of table 3.1, autocorrelation coefficients of R(1) and R(12) are distinct, therefore delay 1 step and 12 step can be adopted for autoregressive coefficients of AR model in threshold intervals. Delay step number d of TAR model is 1 in this case.

2) Optimal ranges of threshold values of TAR model were determined by experience distribution graph of time series data $\{x_t\}$.

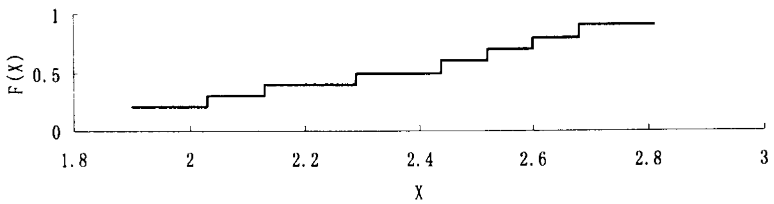

Figure 3.1. Experience distribution graph of observation X - F(X)

Figure 3.1 shows the experience distribution graph of monthly maximum water level series. If r=2, the threshold value corresponds to the frequency 0.5, and if r=3, the threshold values correspond to frequency 0.33 and 0.67 (Wang Wensheng, 2001). Based on experience distribution graph, threshold value range of the corresponding frequency o.5 approximately falls into interval (2.12, 2.15).

3) In the last step, the threshold values $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_{r-1}$ and autoregressive coefficients $\varphi_1^{(i)},\cdots,\varphi_{p_i}^{(i)}(i=1,\cdots,r)$ were optimized.Genetic algorithm was adopted to synchronously optimize threshold values and autoregressive coefficients based on the criterion of the model fitted error minimum. Considering that the method of minimum first power is more robust than method of minimum squares for estimation of model parameters (Jin Juliang, 2001), the optimal criterion function is based on minimum fitted absolute deviation.

$$\min f(\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_{r-1};\varphi_0^{(i)},\cdots,\varphi_{p_i}^{(i)})=\sum_t|\hat{x}_t-x_t| \quad (3.2)$$

Thus the objection function was obtained and parameters were optimized with genetic algorithm by real encoding. The population size is 200; probability of crossover is 0.9; probability of mutation is 0.001; Iterative times are 1000 times. The results are:

$\varphi_0^{(1)} = 1.95, \varphi_1^{(1)} = 0.08, \varphi_{12}^{(1)} = 0.04; \varphi_0^{(2)} = 2.57, \varphi_1^{(2)} = 0.23, \varphi_{12}^{(2)} = 0.13; \bar{x}_1 = 2.13$

4. Model test and conclusion

The monthly maximum water levels were forecasted with the TAR model established from above. Error analysis of the monthly maximum water level series from 1999 to 2001 in Sapgyoho hydrology gauging station were listed in the table 4.1.

Table 4.1 Percentage of the absolute values of forecasting error falling into diverse intervals (%)

interval	[0,0.1]	[0,0.2]	[0,0.3]	[0,0.4]	[0,0.5]	[0,0.6]
percentage	33.33	47.22	72.22	86.11	97.22	100

From the results oftable 4.1, the index values of TAR forecasting falled into the permissible error ranges, which expounded the robust performance of TAR forecasting. Therefore, the TAR model based on genetic algorithm for water level forecasting of tidal rivers is feasible and has high adaptability to forecast other hydrological nonlinear time series in tidal rivers. The results of forecasting effect can be seen from figure 4.1.

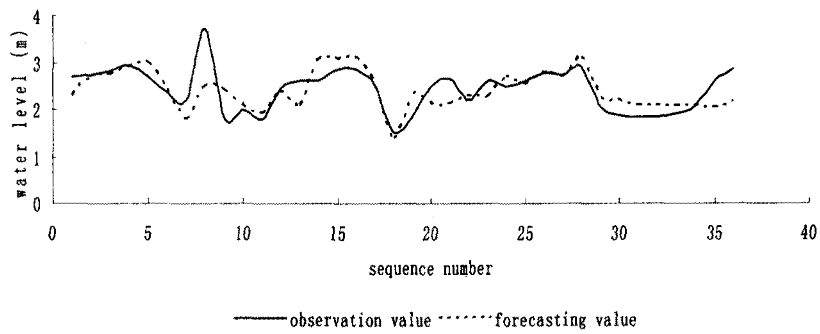

Figure 4.1 Forecasting monthly maximum water levels by TAR model

References

1. H. Tong, K. S. Lim. 1980. Threshold Autoregression, Limit Cycles and Cyclical Data. Journal of the Royal Statistical Society [J]. Series B (Methodological), Vol. 42, No. 3., pp. 245-292.
2. Wang Wensheng, 2001. Application of threshold autoregressive model in hydrologic stochastic simulation [J]. Sichuan water power, vol.20: 47-50.
3. Jin Juliang, et al. 2001. Genetic threshold autoregressive model for predicting annual runoff [J]. Sichuan water power, vol.20: 22-24.