AN APPLICATION OF THE THRESHOLD AUTOREGRESSION PROCEDURE TO CLIMATE ANALYSIS AND FORECASTING

Shi Jiuen (史久恩)

Beijing Institute of Meteorology, Beijing

Zhou Qinfang (周琴芳)

Beijing Meteorological Center, State Meteorological Administration, Beijing

and *Xiang Jingtian* (项静恬)

Institute of Applied Mathematics, Academia Sinica, Beijing

Received June 28, 1984

ABSTRACT

In this paper a nonlinear method of time series analysis—threshold autoregressive (TAR) model in discrete time is used. The TAR procedure consists of four parts: model building, statistical identification, parameter estimation and forecasting.

The object of this study is to estimate monthly total precipitation of Shanghai and Beijing by using open loop TAR model. We can see that the trend of forecasting is in agreement with observations.

I. INTRODUCTION

Meteorological data such as temperature, pressure and precipitation reflect the situation of some aspects of the atmospheric state. On one hand, they have their own evolution rules; and on the other hand, they are influenced by other factors. It is hard to get the determinate expression related to their own evolution. Hence, it is of importance to make studies on them by taking them as random series. Good results can be obtained by adopting time series analysis method[1] in analysing random series in which there is a certain kind of relationship between each other. However, in practice, it is so complicated that linear model can not describe them.

In this paper, a nonlinear method of time series analysis—threshold autoregressive (TAR) model in discrete time[2,3] is used. In recent years, this model becomes rather perfect and it is very effective in application.

The basic idea of the TAR model is to deal with nonlinear systems by using the piecewise linearization. Owing to the threshold control, the recursion stability is ensured, and limit cycle concept is introduced into the nonlinear random systems. The TAR model can effectively describe nonlinear fluctuation, explain various kinds of the stable periodical cycles in nature, and reflect the catastrophe similar to jump resonance[4].

The purpose of this paper is to make studies on monthly precipitation series by using TAR model. The TAR procedure consists of four parts: model building, statistical identification, parameter estimation and forecasting. The stress is put on the studies on general definition of TAR, three special cases, open loop TAR model and their application to forecasting.

II. TAR MODELS IN DISCRETE TIME

Let $\{X_n\}$ be a k-dimensional time series and, for each n, let J_n be an observable random variable, taking integer values $(1, 2, \cdots, l)$.

Definition: $\{X_n; J_n\}$ is said to be general TAR if

$$\mathbf{X}_n=\mathbf{B}^{(J_n)}\mathbf{X}_n+\mathbf{A}^{(J_n)}\mathbf{X}_{n-1}+\varepsilon_n^{(J_n)}+\mathbf{C}^{(J_n)}, \tag{1}$$

where, for $J_n=j$, $A^{(j)}$ and $B^{(j)}$ are $k\times k$ (non-random) matrix coefficients, $C^{(j)}$ is a $k\times l$ vector of constants, and $\{\varepsilon_n^{(j)}\}$ is a k-dimensional strict white noise sequence of independent random vectors with a diagonal covariance matrix. It is also assumed that $\{\varepsilon_n^{(j)}\}$ and $\{\varepsilon_n^{(j')}\}$ are independent of each other for $j\neq j'$.

We now single out a few interesting special cases of the general TAR for further development.

First, let $\{r_0, r_1, \cdots, r_l\}$ denote a linearly ordered subset of the real numbers, such as $r_0, r_1, \cdots, r_l$, where r_0 and r_l are taken to be $-\infty$ and $+\infty$ respectively. They define a partition of the real axis R, i. e.

$$\mathbf{R}=\mathbf{R}_1\cup\mathbf{R}_2\cup\cdots\cup\mathbf{R}_l$$

where $\mathbf{R}_i=(r_{i-1}, r_i)$. Eq. (1) can be usually reduced to the following three patterns of model:

1. *Self-Exciting TAR Model*

Writing

$$\mathbf{X}_n=\{X_n,\ X_{n-1},\cdots,X_{n-k+1}\}^{\tau}$$

$$\mathbf{A}^{(j)}=\left[\begin{array}{cccc|c} a_1^{(j)} & a_2^{(j)} & \cdots & a_{k-1}^{(j)} & a_k^{(j)} \\ \hline & & \mathbf{I}_{k-1} & & 0 \end{array}\right].$$

$$\mathbf{B}^{(j)}=0,$$

$$\varepsilon_n^{(j)}=(\varepsilon_n^{(j)},\ 0,\cdots,0),$$

$$\mathbf{C}^{(j)}=(a_0^{(j)},\ 0,\ \cdots,\ 0),$$

and $R_j^{(k)}=\mathbf{R}\times\mathbf{R}\times\cdots\times R_j\times\cdots\times\mathbf{R}$ is the cylinder set in the Cartesian, product of k real lines, at the interval R_j with dth coordinate space (d is some fixed integer and $1\leqslant d\leqslant k$), and setting $J_n=j$ if $X_{n-1}\in R_j^{(k)}$, we can reduce Eq. (1) into

$$X_n=a_0^{(j)}+\sum_{i=1}^{k}a_i^{(j)}X_{n-i}+\varepsilon_n^{(j)}, \qquad \text{when } X_{n-d}\in R_j\ .$$

$$(j=1,2,\cdots l,) \tag{2}$$

Since $\{J_n\}$ is now a function of $\{X_n\}$ itself, we call the univariate time series $\{X_n\}$ given by Eq. (2) a self-exciting TAR model of order $(d, l; k, \cdots, k)$ or SETAR $(d, l; k, \cdots, k)$, where k is repeated l times. If $a_i^{(j)}=0$, for $j=1, 2, \cdots, l$ and $i=k_j+1,\ k_j+2, \cdots, k$, then we call $\{X_n\}$ a SETAR $(d,\ l;\ k_1,\ k_2,\ \cdots,\ k_l)$ and $\{r_1, r_2, \cdots, r_{l-1}\}$ the thresholds.

Note that a SETAR $(d, l; k)$ is just a linear AR model of order k.

2. *Open-Loop TAR Model*

$\{X_n, Y_n\}$ is called an open-loop TAR system with $\{X_n\}$ as the observable output and $\{Y_n\}$ the observable input, if

$$X_n = a_0^{(j)} + \sum_{i=1}^{m} a_i^{(j)} X_{n-i} + \sum_{i=0}^{m'} b_i^{(j)} Y_{n-i} + \varepsilon_n^{(j)}, \quad \text{when } Y_{n-d} \in R_j$$

$$(j = 1, 2, \cdots, l) \qquad (3)$$

where $\{\varepsilon_n^{(j)}\}$ with $j = 1, \cdots, l$, are strict white noise sequences with zero mean and finite variances and each being independent of $\{Y_n\}$. The l white noise sequences are assumed to be independent of one another. We denote this system by TARSO $[d, l; (m_1, m'_1), \cdots, (m_l, m'_l)]$.

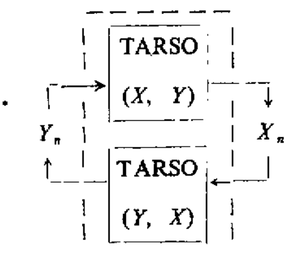
The flow diagram of TARSO model can be drawn as follows

3. *Closed-Loop TAR Model*

$\{X_n, Y_n\}$ is called a closed-loop TAR system, or TARSC, if both $\{X_n, Y_n\}$ and $\{Y_n, X_n\}$ are TARSO. We assume that all the stationary white noise sequences involved are independent of one another.

We denote this system by TARSC $[d, l; d', l'; (m_1, m'_1), \ldots, (m_l, m'_l); (s_1, s'_1), \cdots, (s_l, s'_l)]$.

The flow diagram of TARSC model can be drawn as follows



III. TARSO MODELS FOR REAL DATA

1. *TARSO Model and Forecasting of Monthly Total Precipitation Time Series of Shanghai*

Let $\{X_t\}$ be a monthly total precipitation time series of Shanghai and $\{Y_t\}$ be the mean kinetic energy at 500 hPa over the Northern Hemisphere.

The following model is fitted by using the records in the first 300 months starting from 1951.

$$X_t=\begin{cases}(1)\ -0.070-30.382X_{t-1}+24.488Y_{t-1}+55.723Y_{t-2}-33.994Y_{t-3}\\ \quad +33.067Y_{t-4}-21.055Y_{t-5}+59.027Y_{t-6}-9.243Y_{t-7},\ \text{if}\\ \quad X_{t-5}\leqslant 78.7;\\ (2)\ 0.058-0.110X_{t-1}-3.917X_{t-2}+13.139Y_{t-1}-30.564Y_{t-2}\\ \quad +81.479Y_{t-3}-115.310Y_{t-4}+183.548Y_{t-5}-71.957Y_{t-6}\\ \quad +32.849Y_{t-7},\ \text{if}\ X_{t-5}>78.7.\end{cases}$$

Based on TARSO [5, 2; (1,7), (2,7)] model, we have obtained one-step-ahead predictions of the next 3 months, and Table 1 represents the results of independent test forecast (in mm) during June 1976 to August 1976.

Table 1. Forecast Verification on Independent Data (in mm)

	June	July	August
Prediction (P)	139.3	137.3	176.4
Observation (O)	139.1	142.7	232.4
|P−O|	0.2	5.4	56.0
Climate Mean	178.0	144.1	134.6

2. *TARSO Model and Forecasting of Monthly Total Precipitation Time Series of Beijing.*

Let $\{X_t\}$ be the monthly rainfall data of Beijing, and $\{Y_t\}$ be the mean kinetic energy at 500 hPa.

The following nonlinear time series model is fitted by using 300 months starting from 1951.

$$X_t=\begin{cases}(1)\ 0.126-0.132X_{t-1}-0.223X_{t-2}+0.929X_{t-3}+102.722X_{t-4}\\ \quad -31.271Y_{t-1}-25.709Y_{t-2}-2.822Y_{t-3}+62.845Y_{t-4}\\ \quad +60.163Y_{t-5}+72.005Y_{t-6}-243.160Y_{t-7},\ \text{if}\ X_{t-4}\leqslant 38.1;\\ (2)\ -0.109-20.655X_{t-1}+24.280Y_{t-1}-68.050Y_{t-2}+12.204Y_{t-3}\\ \quad +137.792Y_{t-4}+103.062Y_{t-5}-107.265Y_{t-6}+10.842Y_{t-7},\ \text{if}\\ \quad X_{t-4}>38.1.\end{cases}$$

We have obtained TARSO [4, 2; (4, 7), (1,7)] model, and Table 2 shows the results of independent test forecast during July 1976 and August 1976.

Table 2. Forecast Verification of Monthly Rainfall of Beijing (in mm)

	July	August
P	170.7	188.1
O	173.9	284.8
|P−O|	3.2	96.7
Climate Mean	210.5	172.2

Form the above table we can see that the trend of forecasting is in agreement with observations.

IV. DISCUSSION

(1) So far as a stable linear system is concerned, ARMA (p, q) model (autoregressive

moving average) and statistical identification can be built according to a certain method when limited data $\{X_n\}$ (namely, the limited length of samples) is available. As for non-linear system, it is impossible to find a general method for model-building. A special non-linear method is built only in special limitation. The target described by TAR model is nonlinear system, but it is based on the piecewise linearization of nonlinear models. Therefore, TAR model can be built in accordance with the means for linear autoregressive model-building. One of the advantages of TAR model is that the essential difficulties in model-building can be avoided.

(2) The authors made climate analysis and forecasting of monthly precipitation and monthly mean temperature by using TAR model. The practice shows that the forecasts made with TARSO are better than those with SETAR. The good results in forecasting the tendency of climate parameter series are obtained. However, as regards quantitative forecasts, it is not so desirable to forecast extremely abnormal situation.

(3) A minor change in selecting the thresholds for TARSO has been made, namely, $Y_{n-d} \in R_j$ in Eq. (3) is replaced by $X_{n-d} \in R_j$. The satisfactory result seems to support doing so. It is proposed that on the basis of TAR model, experiments will be made by adopting TAR moving average model (TARMA)[5]. We anticipate that thus-obtained results may be even better.

REFERENCES

[1] Box, G. E. P. and Jenkins, G. M., *Time Series Analysis, Forecasting and Control*. Holden-Day, San-Francisco, 1970.

[2] Tong, H., On a threshold model, in *Pattern Recognition and Signal Processing* (Chen, C. H. ed.). Sijthoff and Noordhoff, The Netherlands, 1978.

[3] Tong, H., and Lim, K. S., *The Journal of the Royal Statistical Society, series B (Methodological)*, **42** (1980), 245—292.

[4] Tong, H., *Catastrophe Theory and Threshold Autoregressive Modelling*. Tech. Rep. No. 125, Dept. of Mathematics, UMIST, 1980.

[5] 项静恬等, 数学的实践与认识, 1985, 2: 63—67