



Threshold Autoregression, Limit Cycles and Cyclical Data

H. Tong; K. S. Lim

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 42, No. 3. (1980), pp. 245-292.

Stable URL:

<http://links.jstor.org/sici?sici=0035-9246%281980%2942%3A3%3C245%3AATALCAC%3E2.0.CO%3B2-Q>

Journal of the Royal Statistical Society. Series B (Methodological) is currently published by Royal Statistical Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/rss.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

Threshold Autoregression, Limit Cycles and Cyclical Data

By H. TONG and K. S. LIM

Department of Mathematics, University of Manchester Institute of Science and Technology

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION on Wednesday, March 19th, 1980, Professor P. WHITTLE in the Chair]

SUMMARY

The notion of a limit cycle, which can only exist in a non-linear system, plays the key role in the modelling of cyclical data. We have shown that the class of threshold autoregressive models is general enough to capture this notion, a definition of which in discrete time is proposed. The threshold value has an interesting interpretation. Simulation results are presented which demonstrate that this new class of models exhibits some well-known features of non-linear vibrations. Detailed analyses of several real data sets are discussed.

Keywords: THRESHOLD AUTOREGRESSION; LIMIT CYCLE; CYCLICAL DATA; NON-LINEAR AUTOREGRESSION; TIME IRREVERSIBILITY; THRESHOLD AUTOREGRESSIVE/MOVING AVERAGE MODELS; NON-LINEAR VIBRATIONS; JUMP RESONANCE; AMPLITUDE-FREQUENCY DEPENDENCY; SUB-HARMONICS; HIGHER HARMONICS; CANADIAN LYNX; MINK AND MUSKRAT; PREDATOR-PREY; WOLF'S SUNSPOT NUMBERS; RAINFALL-RIVERFLOW; EVENTUAL FORECASTING FUNCTION; STABILITY; AKAIKE'S INFORMATION CRITERION; HOUSEHOLDER TRANSFORMATIONS

1. INTRODUCTION

It may be said that the era of linear time series modelling began with such linear models as Yule's autoregressive (AR) models (1927), first introduced in the study of sunspot numbers. In the past five decades or so, we have seen remarkable successes in the application of linear time series models in diverse fields, e.g. Box and Jenkins (1970), and the recent Nottingham International Time Series Conference in March 1979. These successes are perhaps rather natural in view of the significant contributions of linear differential equations in all branches of science. In particular, as far as a one-step-ahead prediction is concerned, a linear time series model is often quite adequate.

However, just as a linear differential equation is totally inadequate as a tool to analyse more intricate phenomena such as limit cycles, time irreversibility, amplitude-frequency dependency and jump resonance, a linear time series model should give place to a much wider class of models if we are to gain deeper understanding into the structure of the mechanism generating the observed data. For example, no linear Gaussian model can explain properly the saw-tooth cycles apparent in the Canadian lynx data (see, for example, discussion of papers by Campbell and Walker, 1977, and Tong, 1977a), and many riverflow data (see, for example, Lawrance and Kottegoda, 1977).

The new era of *practical* non-linear time series modelling is, without doubt, long overdue. In this paper, we describe the theory and practice of a new class of non-linear time series models which are based on the idea of piece-wise linearization. Sections 6 and 9 of this paper are due to both authors while the other sections are due to the first author.

We propose the following requirements for our non-linear time series models, in order of preference :

- (i) statistical identification of an appropriate model should not entail excessive computation;
- (ii) they should be general enough to capture some of the non-linear phenomena mentioned previously;

- (iii) one-step-ahead predictions should be easily obtained from the fitted model and, if the adopted model is non-linear, its overall prediction performance should be an improvement upon the linear model;
- (iv) the fitted model should preferably reflect to some extent the structure of the mechanism generating the data based on theories outside statistics;
- (v) they should preferably possess some degree of generality and be capable of generalization to the multivariate case, not just in theory but also in practice.

Before describing a newly introduced class of non-linear time series models, it may serve us well in recalling some elementary, yet important, properties in the theory of non-linear differential equations or non-linear systems. Here, no stochastic element is involved and only those properties relevant to later exposition are included.

2. NON-LINEAR DIFFERENTIAL EQUATIONS

(i) By definition, the *principle of superposition* does not hold in the non-linear case. In addition, the notion of a “complementary function” and a “particular integral” ceases to be meaningful here.

(ii) Unlike a stable linear system, in which the output (i.e. the solution of the differential equation) dies away when the input is “switched off”, the output of a stable non-linear system may contain sustained oscillations which persist in the absence of input.

To illustrate this, let x_1 and x_2 denote the numbers of two species. Kolmogorov (see, for example, Minorsky, 1962, p. 69) has considered the general system of non-linear differential equations,

$$\frac{dx_1}{dt} = \alpha_1(x_1, x_2)x_1, \quad \frac{dx_2}{dt} = \alpha_2(x_1, x_2)x_2, \quad (2.1)$$

where α_1 and α_2 are continuous functions of x_1 and x_2 with continuous first derivatives. Under very general conditions, he has shown that sustained oscillations (of relatively small amplitude) prevail. It is instructive to quote the following words of Minorsky (1962) in his discussion of the above phenomenon, in which a “common sense” picture of a state of equilibrium is supplemented by relatively small fluctuations :

“Topologically this . . . is precisely a stable limit cycle in the (x_1, x_2) plane onto which wind the spiral trajectories from the outside as well as from the inside. The outside spiral trajectories are those which characterise the establishment of the biological phenomenon and the limit cycle is its representation in a stationary state. . . . As far as is known, no experimental verification of these results has been made so far. If this is done eventually and the Kolmogorov theory is confirmed, this will give valuable information regarding the actual biological probabilities involved in the co-existence of the two species.”

As has been touched on by Tunncliffe-Wilson (1977), limit cycles will play a central role in the modelling of cyclical data. We may write equation (2.1) in the following form,

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \quad (2.2)$$

where the over-dot denotes the time derivative, $\mathbf{x} = (x_1, x_2)^T$, is called the *state vector*,† and

$$\mathbf{A} = \begin{bmatrix} \alpha_1(x_1, x_2) & 0 \\ 0 & \alpha_2(x_1, x_2) \end{bmatrix},$$

where, for greater generality, we may sometimes allow α_1 and α_2 to be discontinuous. The (x_1, x_2) -plane is sometimes referred to as the *phase plane* (or the *state space* in higher dimensional

† ^T denotes transpose.

cases). As an example of the phase plane, Fig. 1 represents that of the following non-linear differential equations from the output of an analogue simulation†

$$\frac{dP}{dt} = \begin{cases} -2(P(t)-5) & \text{if } H(t) < 15, \\ 0.5(P(t)-5) & \text{if } H(t) \geq 15, \end{cases} \quad (2.3a)$$

$$\frac{dH}{dt} = \begin{cases} (H(t)-8) & \text{if } P(t) < 10, \\ -2(H(t)-8.4) & \text{if } P(t) \geq 10. \end{cases} \quad (2.3b)$$

Note that the spiral trajectories do not wind, from the outside, into a (singular) point, but they eventually go round and round closed loops, leaving an interior region untraversed, demonstrating the existence of a limit cycle. Note also that as functions of t , $P(t)$ and $H(t)$ are both *periodic* after the transients have died out. In (2.3), the limit cycle is self-excited, while in

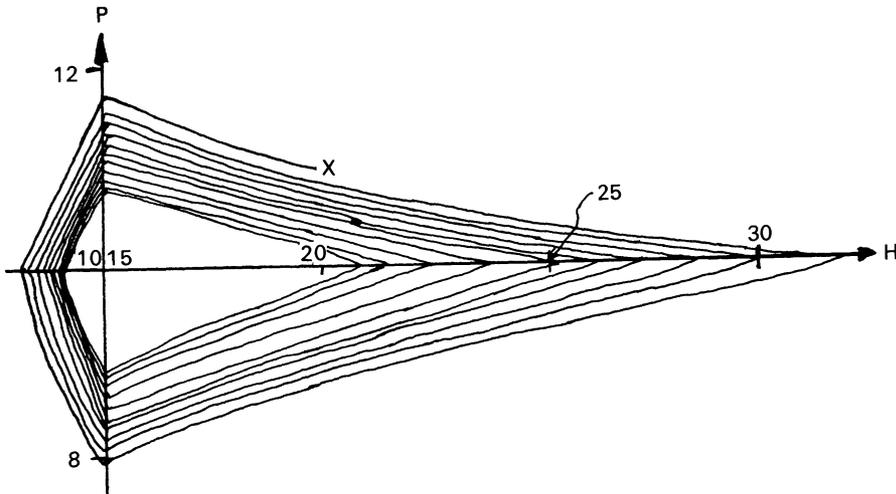


FIG. 1. Simulated phase plane of a continuous time TAR, initial point being denoted by a cross.

some other cases limit cycles may require a certain input to excite them, e.g. in a grandfather clock. For further discussion of the many important properties of the class of piece-wise linear differential equations, see, for example, Aizerman (1963, Ch. V), which refers to the contributions of the Russian school of non-linear vibrations, consisting of A. A. Andronov, F. R. Gantmakher, M. A. Aizerman and others.

(iii) Unlike a linear system, in which the “amplitude” and “frequency” of the output (signal) are functionally independent, the frequency domain analysis (sometimes called the harmonic analysis) of a non-linear system is much more complex. Non-linear vibration engineers have introduced notions such as “amplitude-frequency dependency”, “jump resonance” and others.

3. A LIMIT CYCLE IN DISCRETE TIME

The discussion in continuous time of the last section is only relevant in so far as it gives us a reference frame for developing non-linear time series models in discrete time. This situation is not unlike that in which Yule (1927) first developed his celebrated AR models.

In this paper, we focus on the notion of a limit cycle, leaving the mathematical formulations of the other notions for a non-linear system for future developments. We will, however, indicate how the latter notions manifest themselves in the data through some numerical examples in Section 6.

† The unpublished M.Sc. dissertation by Mr P. K. Wong of UMIST (1978) may be consulted for more similar examples.

For each integer n , let \mathbf{x}_n denote a k -dimensional (state) vector, satisfying the equation

$$\mathbf{x}_n = \mathbf{f}(\mathbf{x}_{n-1}). \tag{3.1}$$

Definition 3.1. A k -dimensional vector \mathbf{x}^* is called a *limit point* if there exists an \mathbf{x}_0 , not equal to \mathbf{x}^* , such that starting with n equal to zero, \mathbf{x}_n tends to \mathbf{x}^* component-wise, as n tends to infinity.

Let \mathcal{C} denote the set of k -dimensional vectors \mathbf{c}_i (of finite Euclidean norm), $i = 1, \dots, T$, T being a positive integer $\leq \infty$.

Definition 3.2. \mathcal{C} is called a *limit cycle of period T* if

(i) $\exists \mathbf{x}_0 \notin \mathcal{C}$ such that starting with n equal to zero, \mathbf{x}_n will ultimately fall into \mathcal{C} as n increases;

(ii) $\mathbf{c}_i = \mathbf{f}(\mathbf{c}_{i-1}), \quad i = 2, 3, \dots, T,$

$\mathbf{c}_{T+i} = \mathbf{c}_i, \quad i = 1, 2, \dots,$ and

(iii) T is the smallest such positive integer.

If, in addition, the assertion of (i) holds on replacing \mathbf{x}_0 by any point ($\notin \mathcal{C}$) in its neighbourhood, then \mathcal{C} is called a *stable limit cycle of period T*.

We shall introduce the notion of a *fractional period* later. A limit cycle of infinite period is sometimes referred to as a *chaotic state* (Li and Yorke, 1975).

It is important to note that a surprisingly complicated structure can arise from a simple non-linear function \mathbf{f} , in the recursive relation of equation (3.1), even when k is equal to one. We refer to Li and Yorke (1975) and May (1976) for some remarkable examples. Of particular note is the result in the former paper which states that a cycle of period 3 implies a chaotic state for almost every \mathbf{x}_0 (in the case $k = 1$), if \mathbf{f} is *continuous*.

The following example is instructive :

Example 3.1.

$$x_n = \begin{cases} 4x_{n-1} & \text{if } |x_{n-1}| \leq \frac{1}{2}, \\ \frac{1}{2}x_{n-1} & \text{if } |x_{n-1}| > \frac{1}{2}. \end{cases} \tag{3.2}$$

This simple example is a special case of the one given by Tong (1977b), and it admits limit cycles of period 3 with the ‘‘ascension time’’ being shorter than the ‘‘descension time’’.

We describe a general extension of (3.2) in the next Section.

4. THRESHOLD AUTOREGRESSIVE MODELS IN DISCRETE TIME

A threshold autoregressive model in discrete time (TAR) was first mentioned in Tong (1977b) and reported briefly in Tong (1978, 1980a). A fuller account was available for private circulation in an unpublished report by Tong in 1978. We now give a more systematic description here.

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

Definition 4.1. $\{\mathbf{X}_n; J_n\}$ is said to be a *general TAR* if

$$\mathbf{X}_n = \mathbf{B}^{(J_n)} \mathbf{X}_n + \mathbf{A}^{(J_n)} \mathbf{X}_{n-1} + \boldsymbol{\varepsilon}_n^{(J_n)} + \mathbf{C}^{(J_n)}, \tag{4.1}$$

where, for $J_n = j$, $\mathbf{A}^{(j)}$ and $\mathbf{B}^{(j)}$ are $k \times k$ (non-random) matrix coefficients, $\mathbf{C}^{(j)}$ is a $k \times 1$ vector of constants, and $\{\boldsymbol{\varepsilon}_n^{(j)}\}$ is a k -dimensional strict white noise sequence of independence random vectors with a diagonal covariance matrix. It is also assumed that $\{\boldsymbol{\varepsilon}_n^{(j)}\}$ and $\{\boldsymbol{\varepsilon}_n^{(j')}\}$ are independent for $j \neq j'$.

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

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

$$\mathbf{R} = R_1 \cup R_2 \cup \dots \cup R_l, \quad \text{say}$$

where

$$R_i = (r_{i-1}, r_i].$$

(A) Writing $\mathbf{X}_n = (X_n, X_{n-1}, \dots, X_{n-k+1})^T$,

$$\mathbf{A}^{(j)} = \begin{bmatrix} a_1^{(j)} & a_2^{(j)} & \dots & a_{k-1}^{(j)} & a_k^{(j)} \\ \mathbf{I}_{k-1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (\text{a companion matrix}),$$

$$\mathbf{B}^{(j)} = \mathbf{0}, \quad \boldsymbol{\varepsilon}_n^{(j)} = (\boldsymbol{\varepsilon}_n^{(j)}, 0, \dots, 0), \quad \mathbf{C}^{(j)} = (a_0^{(j)}, 0, \dots, 0),$$

and $R_j^{(k)} = \mathbf{R} \times \mathbf{R} \times \dots \times \mathbf{R} \times R_j \times \mathbf{R} \times \dots \times \mathbf{R}$ is the cylinder set in the cartesian product of k real lines, on the interval R_j with d th coordinate space (d some fixed integer belonging to $\{1, 2, \dots, k\}$), and setting $J_n = j$ if $\mathbf{X}_{n-1} \in R_j^{(k)}$, we have

$$X_n = a_0^{(j)} + \sum_{i=1}^k a_i^{(j)} X_{n-i} + \varepsilon_n^{(j)}, \quad (4.2)$$

conditional on $X_{n-d} \in R_j$; $j = 1, 2, \dots, l$. Since $\{J_n\}$ is now a function of $\{\mathbf{X}_n\}$ itself, we call the univariate time series $\{X_n\}$ given by equation (4.2) a *self-exciting threshold autoregressive model of order $(l; k, \dots, k)$ or SETAR $(l; k, \dots, k)$* where k is repeated l times. If, for $j = 1, 2, \dots, l$,

$$a_i^{(j)} = 0 \quad \text{for } i = k_j + 1, \quad k_j + 2, \dots, k,$$

then we call $\{X_n\}$ a *SETAR $(l; k_1, k_2, \dots, k_l)$* . We call r_1, \dots, r_{l-1} the *thresholds*. Note that an SETAR $(1; k)$ is just a *linear AR model of order k* .

(B) (X_n, Y_n) is called an *open loop threshold autoregressive system* with $\{X_n\}$ as the observable output and $\{Y_n\}$ as the observable input, if

$$X_n = a_0^{(j)} + \sum_{i=1}^{m_j} a_i^{(j)} X_{n-i} + \sum_{i=0}^{m_j'} b_i^{(j)} Y_{n-i} + \varepsilon_n^{(j)}, \quad (4.3)$$

conditional on $Y_{n-d} \in R_j$; ($j = 1, \dots, l$), where $\{\varepsilon_n^{(j)}\}$; $j = 1, \dots, l$, 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 $(l, (m_1, m_1'), \dots, (m_l, m_l'))$* .

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

5. SOME PERSPECTIVES

The essential idea underlying the class of threshold autoregressive models is the piece-wise linearization of non-linear models over the state space by the introduction of the thresholds $\{r_0, r_1, \dots, r_l\}$; these models are *locally linear*. Similar ideas were used by Priestley (1965), Priestley and Tong (1973) and Ozaki and Tong (1975), in the analyses of non-stationary time series and time dependent systems, in which local stationarity was the counterpart of our present local linearity.

Of course, local linearity abounds in many practical situations. Indeed, if this were not the case, linear time series analysis could not have survived this long. For example, it may be argued that many real systems are non-linear only in so far as they exhibit the phenomenon of *saturation*. Another example is an electrical relay which is a particularly simple piece-wise linear system. In fact, by some co-ordinate transformations of the relay systems, a general class of piece-wise linear differential equations has been established and studied in depth. (See, for example, Aizerman, 1963.) Tong (1980a) has argued, on physical and biological grounds, for the adoption of piece-wise linear models in the analysis of the Canadian lynx data and Wolf's sunspot numbers. Sugawara (1952, 1961), Todini and Wallis (1977) and Chander in an unpublished Ph.D. thesis (1965) have studied the rainfall-riverflow system from the standpoint

of piece-wise linearization. Other related works include Fujishige and Sawaragi (1974), Robinson and Sworder (1974), Rishel (1975), Waltman and Butz (1977), Jacobs and Lewis (1978), Haggan and Ozaki (1980) and Ozaki (1980).

Although at a deeper level, our general TAR must be considered at present as an *ad hoc* class of non-linear time series models, in a certain sense this class is not without some generality. Consider a general first order non-linear autoregressive process, NLAR(1), of the form

$$X_n = f(X_{n-1}) + \epsilon_n \tag{5.1}$$

It seems intuitively clear that subject to general conditions on f , an NLAR(1) may be approximated arbitrarily closely by a TAR. We may argue heuristically as follows. Suppose that $f(x)$ is continuous in a closed interval $[x', x'']$. It is well known that $f(x)$ is uniformly continuous in $[x', x'']$ and that the Weierstrass theorem ensures that $f(x)$ may be approximated arbitrarily closely by $\hat{f}(x)$, where

$$\hat{f}(x) = f(x_{(i_{k-1})}) + \alpha_k x, \tag{5.2}$$

for

$$x \in [x_{(i_{k-1})}, x_{(i_k)}], \quad k = 1, \dots, l,$$

where

$$x_{(i_0)} = x', \quad x_{(i_l)} = x'',$$

and the partition

$$[x', x''] = [x', x_{(i_1)}] \cup [x_{(i_1)}, x_{(i_2)}] \cup \dots \cup [x_{(i_{l-1})}, x'']$$

is defined depending on the degree of accuracy of the approximation required. Therefore we have obtained an SETAR($l; 1, 1, \dots, 1$) to the NLAR, with thresholds $\{x_{(i_1)}, x_{(i_2)}, \dots, x_{(i_{l-1})}\}$. For an NLAR(k),

$$X_n = f(X_{n-1}, \dots, X_{n-k}) + \epsilon_n, \tag{5.3}$$

we may re-write it in vector notation

$$\mathbf{X}_n = \tilde{\mathbf{f}}(\mathbf{X}_{n-1}) + \boldsymbol{\epsilon}_n, \tag{5.4}$$

where

$$\mathbf{X}_n = (X_n, X_{n-1}, \dots, X_{n-k+1})^T, \\ \boldsymbol{\epsilon}_n = (\epsilon_n, \underbrace{0, \dots, 0}_{k-1})$$

and

$$\tilde{\mathbf{f}}(\mathbf{X}_{n-1}) = (f(X_{n-1}, \dots, X_{n-k}), X_{n-1}, \dots, X_{n-k+1})^T.$$

A vector version of the Weierstrass theorem will then establish a general TAR approximation of an NLAR(k) under general conditions on f .

A more challenging problem concerns the following non-linear Markovian system (NMS) :

$$\begin{aligned} \mathbf{X}_n &= \mathbf{f}(\mathbf{X}_{n-1}) + \boldsymbol{\epsilon}_n \\ \mathbf{Y}_n &= \mathbf{g}(\mathbf{X}_n), \end{aligned} \tag{5.5}$$

where \mathbf{X}_n and \mathbf{Y}_n are a k -dimensional unobservable vector and a q -dimensional ($q \leq k$) observable vector respectively, and $\boldsymbol{\epsilon}_n$ defines a zero mean stationary k -dimensional strict white noise sequence and is independent of \mathbf{X}_{n-1} .

Suppose that \mathbf{g} is a partition preserving mapping from \mathbf{R}^k to \mathbf{R}^q , $k \geq q$, in the sense that $\{\mathbf{g}(R_i^{(k)})\}$ defines a partition of \mathbf{R}^q for every partition $\{R_i^{(k)}\}$ of \mathbf{R}^k , where for any set A ,

$$\mathbf{g}(A) = \{\mathbf{y}; \mathbf{g}(\mathbf{x}) = \mathbf{y}, \mathbf{x} \in A\}.$$

It seems plausible that an NMS with a partition preserving mapping g may be arbitrarily closely approximated by a TAR under general conditions on f . The problem arises as to the characterization of the class of such mappings. In the case of $k = q$, we know that it contains at least one element, the identity mapping.

Next, consider the recursive relation,

$$x_n = f(x_{n-1}, \dots, x_{n-k}) \quad (n \geq 0, k < \infty), \tag{5.6}$$

where f is such that $|x_j| < \infty$ for all $j < \infty$. It is well known that unless some “stability” condition is placed on f , the recursion will diverge. One example is a polynomial in $x_{n-1}, x_{n-2}, \dots, x_{n-k}$. However, in practice, we can usually circumvent this problem by introducing some inbuilt restrictions on the range space of f .

Let d be a pre-fixed integer chosen from $\{1, 2, \dots, k\}$, $k < \infty$. We agree to set

$$x_i = 0, \forall i < 0.$$

Definition 5.1. Let f be a point transformation from \mathbf{R}^k to \mathbf{R} given by (5.6). Let S be a finite interval of \mathbf{R} . f_S is said to be a *stabilizer of f induced by S* if it has the following properties :

- (i) $f_S(x_{n-1}, \dots, x_{n-k}) = x_n$,
- (ii) $x_{n-d} \in S \Rightarrow f_S(x_{n-1}, \dots, x_{n-k}) = f(x_{n-1}, \dots, x_{n-k})$, and
- (iii) $x_{n-d} \notin S \Rightarrow f_S(x_{n-1}, \dots, x_{n-k}) = c$, $c \notin S$ and $|c| < \infty$.

Theorem 5.1. f_S defines a stable recursion in the sense that

$$|f_S(x_{n-1}, \dots, x_{n-k})| < \infty, \quad \forall (x_{n-1}, \dots, x_{n-k}) \in \mathbf{R}^k \quad \text{and all } n.$$

Proof. Denote the row vector (x_n, \dots, x_{n-k+1}) by \mathbf{x}_n . We agree to call x_i an outlier if $x_i \notin S$. Suppose that \mathbf{x}_{n_0} is the first vector with its first component x_{n_0} being an outlier. Under the recursion f_S ,

$$\mathbf{x}_{n_0+d} = (c, x_{n_0+d-1}, \dots, x_{n_0+d-k+1}).$$

Obviously, $\forall n \geq n_0 + d$, \mathbf{x}_n has at least one component equal to c . It now remains to be shown that the number of components of \mathbf{x}_n equal to c is monotonically nondecreasing as n increases to infinity. There are two possibilities subsequent to \mathbf{x}_{n_0+d} . One possibility is that no more outliers will occupy the first component, except for the recurring c , in which case f_S defines a stable recursion. The other possibility is that a new outlier will occupy the first component in addition to the recurring c . Because each outlier will subsequently produce one further component equal to c , we have proved by induction that the number of components equal to c is monotonically non-decreasing. Hence, there exists an $M < \infty$, such that for all $n \geq M$,

$$|f_S(x_{n-1}, \dots, x_{n-k})| < \infty, \quad \forall (x_{n-1}, \dots, x_{n-k}) \in \mathbf{R}^k.$$

Therefore, by the finiteness of M ,

$$|f_S(x_{n-1}, \dots, x_{n-k})| < \infty, \quad \forall (x_{n-k}, \dots, x_{n-k}) \in \mathbf{R}^k \quad \text{and all } n.$$

This completes the proof of the theorem.

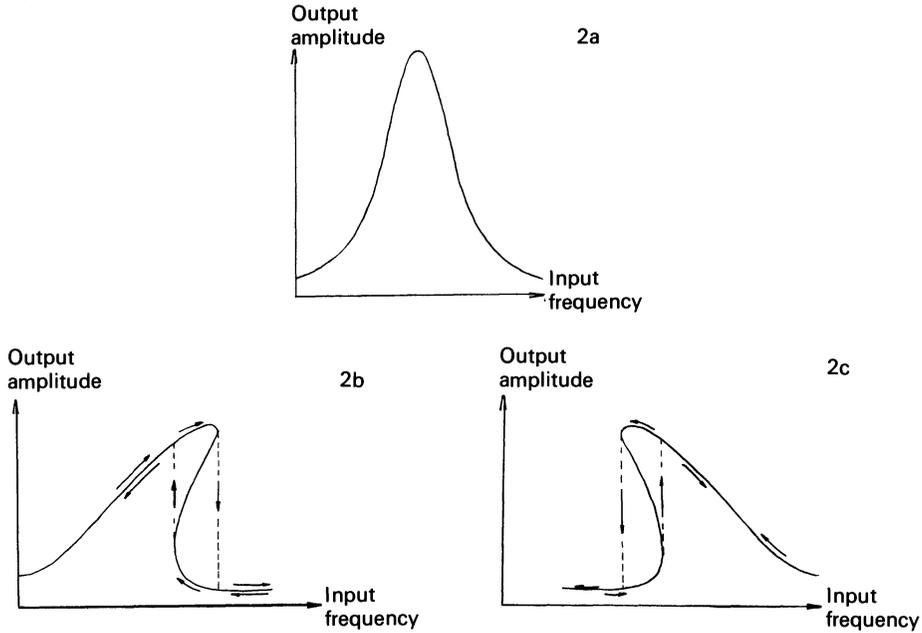
It is interesting to note that f_S corresponds to a *threshold* model.

Finally, the question of stationarity for an NLAR(1), as well as its marginal and conditional distributions, has been studied by Jones (1978), who has applied Tweedie’s (1975) general results concerning ergodicity of a Markov chain over a *general* state space to NLAR(1), and has also indicated possible extension to NLAR(k), $k \geq 1$. It is possible to show that a sufficient condition for the SETAR $\{\mathbf{X}_n\}$ described in case (A) of Section 4 to be ergodic in the sense of Tweedie (1975) is that the maximum eigenvalue of $\mathbf{A}^{(j)T} \mathbf{A}^{(j)}$ is strictly less than unity, $j = 1, \dots, l$, and the $e_n^{(j)}$ ’s have absolutely continuous distributions. This is obtained by applying Corollary 5.2 of Tweedie (1975) with $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$ for a $k \times 1$ vector \mathbf{x} , and $\|\mathbf{A}\|^2 =$ maximum eigenvalue of $\mathbf{A}^T \mathbf{A}$ for a $k \times k$ matrix \mathbf{A} . However, it is interesting to investigate if the eigenvalue requirement could be weakened.

6. TAR MODELS AND NON-LINEAR VIBRATIONS

We now describe some simulation results which demonstrate that TAR models exhibit interesting features well known in non-linear vibrations.

(1) *Jump resonance*. It is well known that, unlike a linear system, the output amplitude may have a "resonance jump" at different frequencies depending on whether the input frequency (of constant amplitude) is monotonically increasing or monotonically decreasing. (See Figs 2a, 2b and 2c.)



Figs 2a, 2b, 2c. Jump resonance.

The time plots of Figs 3a and 3b clearly show that our SETAR can capture this engineering notion. The engineering terminology of a "hard spring" and a "soft spring" is an indication of the mode of the "restoring force" of the system. Figs 3a and 3b correspond respectively to the SETAR(2; 9, 3), $d = 5$ and SETAR(2; 3, 8), $d = 6$ given below. (White noise inputs are replaced by sinusoids in this exercise.)

$$X_n = \begin{cases} 0.4655 + 1.1448X_{n-1} - 0.4801X_{n-2} + 0.1273X_{n-3} \\ \quad - 0.3580X_{n-4} + 0.2565X_{n-5} - 0.0781X_{n-6} - 0.0493X_{n-7} \\ \quad + 0.2186X_{n-8} + 0.0526X_{n-9} + input & \text{if } X_{n-5} \leq 3.05, \\ 1.1940 + 1.1181X_{n-1} - 0.5017X_{n-2} - 0.0594X_{n-3} + input & \text{if } X_{n-5} > 3.05, \end{cases} \quad (6.1a)$$

$$X_n = \begin{cases} 1.3003 + 1.3243X_{n-1} - 0.7023X_{n-2} - 0.0750X_{n-3} + input & \text{if } X_{n-6} \leq 3.31, \\ 0.2004 + 1.2112X_{n-1} - 0.6971X_{n-2} + 0.6191X_{n-3} \\ \quad - 1.0178X_{n-4} + 0.9967X_{n-5} - 0.7688X_{n-6} + 0.6119X_{n-7} \\ \quad - 0.0551X_{n-8} + input & \text{if } X_{n-6} > 3.31. \end{cases} \quad (6.1b)$$

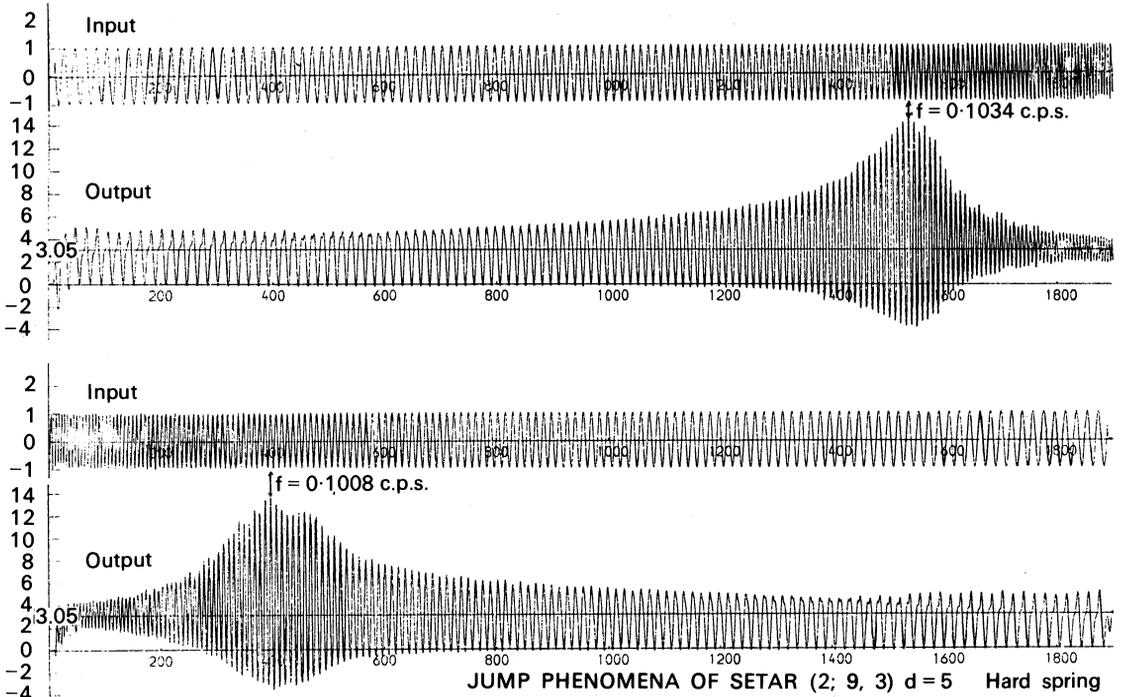


FIG. 3a. Jump resonance, hard spring.

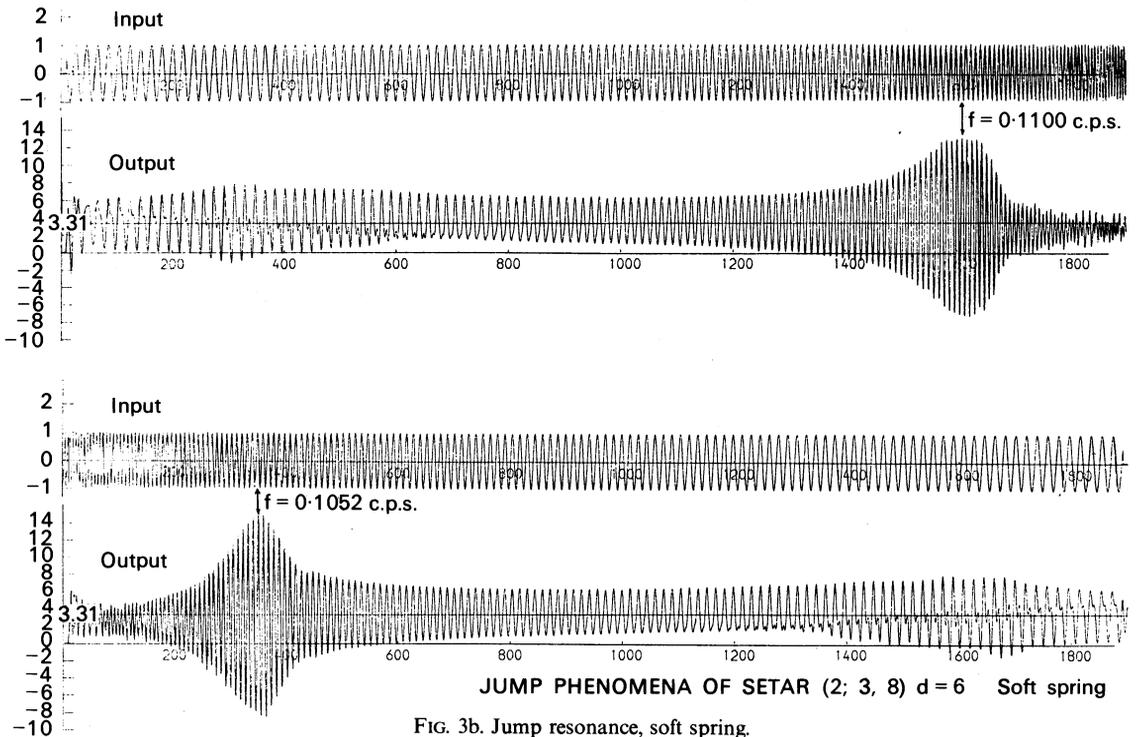


FIG. 3b. Jump resonance, soft spring.

It is also well known that the output amplitude of a non-linear system may have a resonance jump at different amplitudes depending on whether the input amplitude (of constant frequency) is monotonically increasing or monotonically decreasing. Fig. 4 corresponds to the time plots of the following threshold model :

$$X_n = \begin{cases} X_{n-1} + 2(Y_n - Y_{n-1}) & \text{if } |Y_{n-1} - Y_{n-2}| > 10, \\ X_{n-1} + 0.1(Y_n - Y_{n-1}) & \text{if } |Y_{n-1} - Y_{n-2}| \leq 10. \end{cases} \quad (6.2)$$

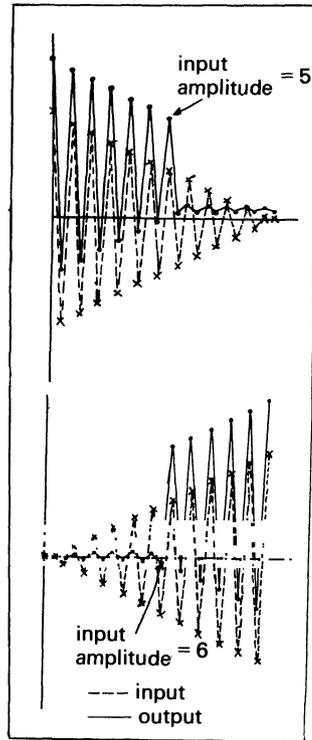


FIG. 4. Jump resonance.

(2) *Amplitude-frequency dependency.* It is well known that, unlike a linear system, the output signal may show different frequencies of oscillations for different amplitudes. The time plots of Figs 5a and 5b correspond respectively to the two SETAR (2; 3, 3), $d = 1$ given by equations (6.3) and (6.4) respectively :

$$X_n = \begin{cases} 1.6734 - 0.8295X_{n-1} + 0.1309X_{n-2} - 0.0276X_{n-3} + \varepsilon_n^{(1)} & \text{if } X_{n-1} > 0.5, \\ 1.2270 + 1.0516X_{n-1} - 0.5901X_{n-2} - 0.2149X_{n-3} + \varepsilon_n^{(2)} & \text{if } X_{n-1} \leq 0.5, \quad \text{var } \varepsilon_n^{(i)} = 0.003^2, \quad i = 1, 2, \end{cases} \quad (6.3)$$

$$X_n = \begin{cases} 0.15 + 0.85X_{n-1} + 0.22X_{n-2} - 0.70X_{n-3} + \varepsilon_n^{(1)} & \text{if } X_{n-1} \leq 3.05, \\ 0.30 - 0.80X_{n-1} + 0.20X_{n-2} + 0.70X_{n-3} + \varepsilon_n^{(2)} & \text{if } X_{n-1} > 3.05, \quad \text{var } \varepsilon_n^{(i)} = 0.003^2, \quad i = 1, 2. \end{cases} \quad (6.4)$$

Note that Fig. 5a shows the tendency of high frequency of oscillations when the amplitudes are high. Fig. 5b shows the reverse tendency.

- (3) *Limit cycles.* Quite a few figures showing limit cycles for SETAR will be given in Section 9.
- (4) *Subharmonics.* By a subharmonic it is usually meant an output oscillation at a fraction of

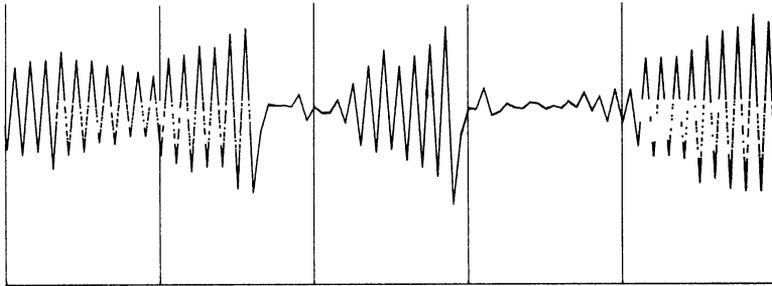


FIG. 5a. Time plots of (6.3): amplitude-frequency dependency, high (low) amplitudes having high (low) frequencies.

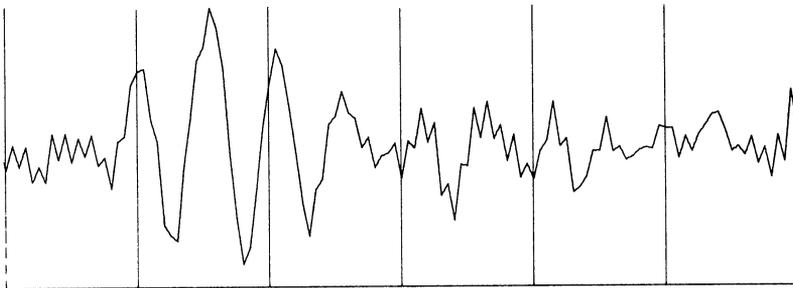


FIG. 5b Time plots of (6.4): amplitude-frequency dependency, high (low) amplitudes having low (high) frequencies.

the input oscillation frequency. The time plots of Fig. 6 correspond to the following simple SETAR(3; 0, 1, 0) with a periodic input $\{Y_n\}$:

$$X_n = \begin{cases} 2X_{n-1} + Y_n & \text{if } |X_{n-1}| \leq 2, \\ Y_n & \text{if } |X_{n-1}| > 2, \end{cases} \quad \text{where } Y_n = \begin{cases} -1 & \text{if } n \text{ is odd,} \\ 1 & \text{if } n \text{ is even.} \end{cases} \quad (6.5)$$

(5) *Higher harmonics.* By a higher harmonic it is usually meant an output oscillation at a multiple of the input oscillation frequency. The time plots of Fig. 7 correspond to the following simple TARSO model model with a periodic input $\{Y_n\}$:

$$X_n = \begin{cases} -(2 + \sqrt{2}) Y_n - (1 + \sqrt{2}) & \text{if } -1 < Y_n \leq -1/\sqrt{2} \\ -\sqrt{2} Y_n - 1 & \text{if } -1/\sqrt{2} < Y_n \leq 0 \\ \sqrt{2} Y_n - 1 & \text{if } 0 < Y_n \leq 1/\sqrt{2} \\ (2 + \sqrt{2}) Y_n - (1 + \sqrt{2}) & \text{if } 1/\sqrt{2} < Y_n \leq 1 \end{cases} \quad (6.6)$$

7. LIMIT CYCLES, LIMIT POINTS AND SETAR

In Section 2(ii) we described a limit cycle as one possible mode of oscillations of a system when the input is "switched off". This motivates the following definition of a *limit cycle* of a stochastic system in which the input may consist of some random and some deterministic components.

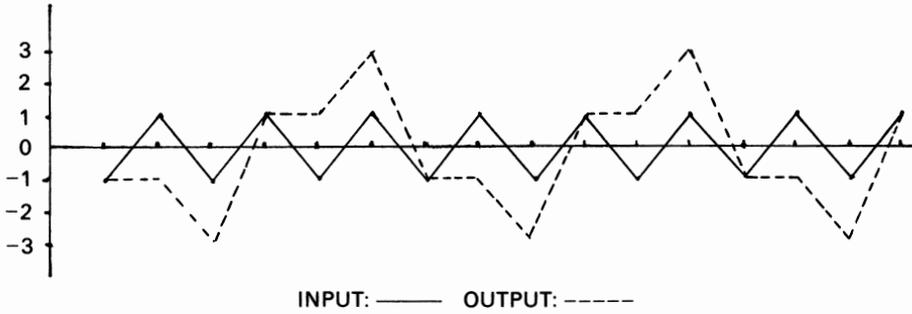


FIG. 6. Sub-harmonics.

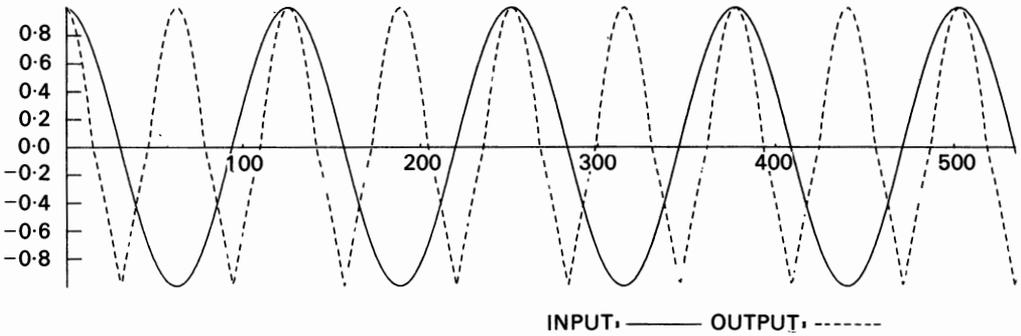


FIG. 7. Higher harmonics.

Definition 7.1. A stochastic model

$$\mathbf{X}_n = \mathbf{f}(\mathbf{X}_{n-1}, \varepsilon_n^{(i)}, \varepsilon_{n-1}^{(i)}, \dots, \varepsilon_{n-p_i}^{(i)}, u_n^{(i)}, \dots, u_{n-p_i}^{(i)}, i = 1, \dots, q), \tag{7.1}$$

where $\{\mathbf{X}_n\}$ is an observable k -dimensional time series and for $i = 1, \dots, q$, $\{\varepsilon_n^{(i)}\}$ is an unobservable one-dimensional time series and $\{u_n^{(i)}\}$ is a one-dimensional deterministic sequence, is said to admit a limit cycle if, by denoting $\hat{\varepsilon}_n^{(i)}$ as the conditional expectations of $\varepsilon_n^{(i)}$ given $\mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \dots$, (assumed to exist),

$$\tilde{\mathbf{X}}_n = \mathbf{f}(\tilde{\mathbf{X}}_{n-1}, \hat{\varepsilon}_n^{(i)}, \hat{\varepsilon}_{n-1}^{(i)}, \dots, \hat{\varepsilon}_{n-p_i}^{(i)}, \mathbf{0}, i = 1, \dots, q) \tag{7.2}$$

(where the vector of zeros is of dimension $p_i + 1$) induces a recursive relation in $\tilde{\mathbf{X}}_n$, say,

$$\tilde{\mathbf{X}}_n = \mathbf{g}(\tilde{\mathbf{X}}_{n-1}), \tag{7.3}$$

which has a limit cycle in the sense of Definition 3.2.

Note that if $\varepsilon_n^{(i)}$ is a zero mean random variable and independent of $\mathbf{X}_{n-1}, \mathbf{X}_{n-2}, \dots$ for each i , then

$$\tilde{\mathbf{X}}_n = \mathbf{f}(\tilde{\mathbf{X}}_{n-1}, \mathbf{0}), \tag{7.4}$$

where the zero vector of dimension $K = \sum_{i=1}^q (p_i + p_i + 2)$. In what follows we assume this independence.

Now, for simplicity, consider a (stable) SETAR(3; k, k, k) with $a_0^{(i)} = 0; i = 1, 2, 3$ and $-\infty < r_1 < 0 < r_2 < \infty$. In this case, equation (7.3) takes the specific form

$$\tilde{\mathbf{X}}_n = \mathbf{A}^{(j)} \tilde{\mathbf{X}}_{n-1}, \text{ if } \tilde{\mathbf{X}}_{n-1} \in R_j^{(k)}, j = 1, 2, 3. \tag{7.5}$$

Let $\rho(\mathbf{A})$ denote the modulus of the maximum eigenvalue of the matrix \mathbf{A} . Suppose that

$$\rho(\mathbf{A}^{(1)}) < 1, \rho(\mathbf{A}^{(2)}) > 1, \rho(\mathbf{A}^{(3)}) < 1. \tag{7.6}$$

The only stationary solution of the equation

$$\mathbf{x} = \mathbf{A}^{(j)} \mathbf{x} \quad \text{if } \mathbf{x} \in R_j^{(k)}, \quad j = 1, 2, 3, \quad (7.7)$$

is the zero vector, which belongs to $R_2^{(k)}$. However, that $\rho(\mathbf{A}^{(2)})$ is strictly greater than unity implies that this solution cannot be stable, i.e. there is no stable limit point. On the other hand, the system is stable. Therefore the only stable solutions are periodic, i.e. limit cycles. The extension to an SETAR ($l; k_1, \dots, k_l$) with $a_0^{(i)} = 0; i = 1, \dots, l$, and $0 \in R_i$ for some i not equal to 1 or l , is straightforward. However, the problem of the theoretical classification of solutions, into the number of admissible limit points and limit cycles, for a general SETAR, in terms of the coefficients $a_i^{(j)}$'s, is not completely solved. In practice, this is not necessarily a serious drawback because once an SETAR model has been fitted, we can always check numerically whether it admits a limit cycle with the current observation being the initial point \mathbf{x}_0 . We develop this point in Section 8.

8. STATISTICAL IDENTIFICATION

Given a finite record, a linear autoregressive (AR) model can be very easily fitted by efficient computational algorithms such as Levinson-Durbin's or the Householder transformation. (For discussion of the former, see, for example, Box and Jenkins, 1970, and of the latter see, for example, Golub, 1965.)

For the fitting of a general non-linear autoregressive model, the above techniques would no longer be suitable, and a much more time-consuming search algorithm would be necessary. However, in view of its piece-wise linearity, a threshold model can still be fitted by the efficient method of Householder transformations. The Levinson-Durbin method cannot be applied here in view of the lack of "Toeplitzian property" of the TAR.

We give only a description of a statistical method of identification. Sampling properties of the estimates of parameters are not included but an application of the recent results of Klimko and Nelson (1978) may prove fruitful. A Gaussian assumption is made on all the white noise sequences. This enables us to write down the likelihood function and derive the maximum likelihood estimates of the unknown parameters, much in the same way as in the linear AR case. It is easy to check that the Jacobian of the transformation from the white noise terms to the observations is unity. The initial part of our identification procedure is based on Akaike's Information Criterion (Akaike, 1973), denoted by AIC, which, for each specified threshold model, takes the generic form,

$$\text{AIC}(k) = N \ln(\text{RSS}/N) + 2k, \quad (8.1)$$

where RSS is the residual sum of squares of the fitted model, based on maximum likelihood estimates of the defining parameters, N is the "effective number of observations" (to be explained later) and k is the number of independent parameters of the model. Equation (8.1) is, of course, strictly speaking, valid only when the "end effects" of the likelihood function are negligible, as are usually assumed in this kind of analysis. (See, for example, Bartlett, 1966, p. 271.) We sometimes normalize the AIC by dividing it by N .

We describe, in some detail, one *computational* procedure implementing the proposed AIC identification for the class of SETAR ($2; k_1, k_2$). Other classes may be considered in a similar way. First, let d and L be prefixed, where L is the maximum order to be entertained for each of the l piece-wise linear AR models. The choice of L is subjective and usually depends on the sample size. It may be allowed to be different for different regions R_i , but, for the convenience of description, we have set them to be all the same here. (In our program the more flexible alternative is adopted. The programs are obtainable from the authors upon request.) Let n_0 be the maximum of d, L . Let $\{x_1, x_2, \dots, x_n\}$ denote the observed data and t_q the sample 100 q th percentile. Suppose that we agree to use $\{t_{0.30}, t_{0.40}, t_{0.50}, t_{0.60}, t_{0.70}\}$ as a set of potential candidates for the estimation of r_1 , the threshold value. Note that this choice is, of course, arbitrary but convenient, and may be changed if necessary. For each choice of t_q , we re-arrange the data set into two sub-sets and set up two sub-systems of linear equations, one for R_1 and the other for R_2 .

The following is a typical example. Suppose $x_{n_0-d+1}, x_{n_0-d+4}, x_{n_0-d+5}, \dots$ are less than or equal to t_q , and the others are greater than t_q . Then

$$\begin{bmatrix} x_{n_0+1} \\ x_{n_0+4} \\ x_{n_0+5} \\ \dots \end{bmatrix} = \begin{bmatrix} 1 & x_{n_0} & x_{n_0-1} & \dots \\ 1 & x_{n_0+3} & x_{n_0+2} & \dots \\ 1 & x_{n_0+4} & x_{n_0+3} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \\ a_2^{(1)} \\ \dots \end{bmatrix}, \text{ i.e. } \mathbf{x}_1 = \mathbf{A}_1 \boldsymbol{\theta}_1, \text{ say,} \quad (8.2a)$$

$$\begin{bmatrix} x_{n_0+2} \\ x_{n_0+3} \\ x_{n_0+6} \\ \dots \end{bmatrix} = \begin{bmatrix} 1 & x_{n_0+1} & x_{n_0} & \dots \\ 1 & x_{n_0+2} & x_{n_0+1} & \dots \\ 1 & x_{n_0+5} & x_{n_0+4} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} a_0^{(2)} \\ a_1^{(2)} \\ a_2^{(2)} \\ \dots \end{bmatrix}, \text{ i.e. } \mathbf{x}_2 = \mathbf{A}_2 \boldsymbol{\theta}_2, \text{ say.} \quad (8.2b)$$

We may obtain estimates of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ by Householder transformations of the matrices \mathbf{A}_1 and \mathbf{A}_2 respectively. For each fixed t_q and d , we use AIC to determine the orders of the two piecewise linear AR's, k_1 and k_2 . Specifically, \hat{k}_1 is the minimum AIC estimate of k_1 , i.e.

$$\text{AIC}(\hat{k}_1) = \min_{0 \leq k_1 \leq L} \{N_1 \ln(\text{RSS}_1(k_1)/N_1) + 2(k_1 + 1)\}, \quad (8.3)$$

where N_1 is the number of elements in \mathbf{x}_1 and $\text{RSS}_1(k_1)$ is the residual sum of squares $\|\mathbf{x}_1 - \mathbf{A}_1 \hat{\boldsymbol{\theta}}_1\|^2$. Here $\hat{\boldsymbol{\theta}}_1$ is the least squares estimate of $\boldsymbol{\theta}_1$, assuming a k_1 th order AR model, and $\|\cdot\|$ denotes the Euclidean norm of a vector. \hat{k}_2 is obtained in a similar way.

Recalling that the computation is fixed at t_q , we may write

$$\text{AIC}(t_q) = \text{AIC}(\hat{k}_1) + \text{AIC}(\hat{k}_2), \quad (8.4)$$

because $\varepsilon_m^{(1)}$'s and $\varepsilon_n^{(2)}$'s are independent of each other. Next, we allow t_q to vary over a pre-selected set of t_q 's and minimize the $\text{AIC}(t_q)$ over this set. That value of t_q , \hat{r}_1 say which is such that

$$\text{AIC}(\hat{r}_1) = \min_{\{t_q\}} \{\text{AIC}(t_q)\}, \quad (8.5)$$

is adopted as our current estimate of r_1 , the threshold value, and the \hat{k}_1, \hat{k}_2 corresponding to this \hat{r}_1 our estimates of k_1 and k_2 . Therefore, the minimum AIC model adopted for the fixed value of d is SETAR(2; \hat{k}_1, \hat{k}_2) with threshold \hat{r}_1 . In all the above searching stages, the *total effective number* of observations remains the same, namely $n - n_0$, while the effective number in each region is smaller. (Care should be taken to ensure that they are sufficiently large.)

Finally, we have to search over d for a set of pre-selected positive integers. The different choices of d may alter n_0 and hence $n - n_0$. In order to get some cross-comparison between the $\text{AIC}(\hat{r}_1)$'s for the different choices of d , we normalize the former. Thus, for each d , we write

$$\text{AIC}(d) = \text{AIC}(\hat{r}_1)/(n - \max\{d, L\}), \quad (8.6)$$

where $\text{AIC}(\hat{r}_1)$ is defined in (8.5) for this choice of d . After this last search stage over d , we have completed the minimum AIC identification, which will give us estimates of $d, r_1, k_1, a_i^{(1)}; i = 0, \dots, k_1, a_i^{(2)}; i = 0, \dots, k_2, \text{var } \varepsilon_j^{(1)}$ and $\text{var } \varepsilon_j^{(2)}$.

To complement the final stage of the identification, namely that of d , we also compute the so-called *eventual forecasting function*, $\text{eff}(d)$, for each d . Specifically, for each fixed d , we go through all the afore-mentioned search stages, ending with a minimum AIC estimated SETAR(2; \hat{k}_1, \hat{k}_2) with threshold value \hat{r}_1 . Using the observed data and the fitted model, we may easily obtain the one-step-ahead prediction of X_{n+1} , because the observed value of X_{n-d} determines in which R_i region it falls. Denote this predicted value of X_{n+1} by x_{n+1} . Now, pretending that this x_{n+1} was the observed value of X_{n+1} , we may repeat the same calculation and obtain x_{n+2} , etc. The plot of x_{n+m} , $m = 1, 2, \dots$, against m is, in fact, just a convenient way of visualizing the "systematic part" of the fitted SETAR(2; \hat{k}_1, \hat{k}_2) model, given the observed data. It should not, however, be confused

with the more-than-one-step-ahead prediction function. The $\text{eff}(d)$ should therefore either tend to a constant or a periodic function, unless an unstable model has been fitted. The former indicates a limit point and the latter a limit cycle. By comparing the $\text{eff}(d)$ for the different choices of d , we may be able to form some subjective judgement as to the preferred choice.

Yet another complementary technique we sometimes find useful in the final stage is that based on a kind of pseudo-cross-validation. We delete the last 10 per cent of observations, say, in the identification procedure, and then compare the one-step-ahead prediction errors on using the fitted model to forecast these deleted observations. Suppose that with d equal to d_0 the total of the prediction errors is a minimum. We then repeat the identification procedure with the complete data set, with d fixed at d_0 . If the fitted model using the complete data set does not differ much from that based on the incomplete data set, then adopt the former as our final model with d equal to d_0 .

A final check is obtained by studying the fitted residuals and the one-step-ahead prediction errors. The plotting of these is routine in our computer package.

9. TAR MODELS FOR REAL DATA

(A) *The Canadian lynx data (1821-1934)*. This set of data has been analysed extensively by many statisticians. (See, in particular, Campbell and Walker, 1977; Tong, 1977a, and Bhansali, 1979.) We now list what we regard as significant features of these data as follows:

- (i) obvious cycles of approximately 10 years with varying amplitudes;
- (ii) the rise period, from a local minimum to the next local maximum, exceeding the descent period, from a local maximum to the next local minimum, thereby showing time irreversibility.

The proposed identification procedure has enabled us to select the following SETAR (2; 8,3) model as our model for the data which has been logarithmically transformed (to the base 10):

$$X_n = \begin{cases} 0.5239 + 1.0359X_{n-1} - 0.1756X_{n-2} + 0.1753X_{n-3} \\ \quad - 0.4339X_{n-4} + 0.3457X_{n-5} - 0.3032X_{n-6} + 0.2165X_{n-7} \\ \quad + 0.0043X_{n-8} + \varepsilon_n^{(1)} & \text{if } X_{n-2} \leq 3.1163, \\ 2.6559 + 1.4246X_{n-1} - 1.1618X_{n-2} - 0.1094X_{n-3} + \varepsilon_n^{(2)} & \text{if } X_{n-2} > 3.1163, \end{cases} \quad (9.1)$$

where $\text{var } \varepsilon_n^{(1)} = 0.0255$, $\text{var } \varepsilon_n^{(2)} = 0.0516$. (The pooled mean sum of squares of residuals = 0.0360.)

Fig. 8 shows that the eff is an asymmetric periodic function of period ten years (counting minimum year to minimum year *inclusively*), i.e. model (9.1) has a limit cycle of period 9 years as determined by Definition 7.1. The rise and descent periods are six and three, respectively. The limit cycle may be generated from (2.6226, 2.8945, 3.2523, 3.4601, 3.4257, 3.2281, 2.9793, 2.7884, 2.6639). It is interesting to note that a similar limit cycle can be picked up even by fitting a SETAR to just 80 of the 114 observations. The fact that the threshold value depends on X_{n-2} is particularly interesting in view of its implications of a lead-lag relation of approximately 2 years between the lynx population and its prey (cf. Bulmer, 1975). We will consider this point again in Example C.

Tong's AR(11) model (Tong, 1977b) and Campbell-Walker's harmonic-component-plus-AR(2) models (Campbell and Walker, 1977) have been recognized to be inconclusive owing to their linearity. (See Tong, 1980a, and the discussion of the above papers.) Threshold models certainly seem to offer exciting possibilities here. (See also Haken, 1978, p. 9.) The estimated threshold at about 3.1 gives us a rough idea of the critical lynx population in its co-existence with their prey. Figs 9a and 9b show the gain spectra of the fitted model, corresponding to

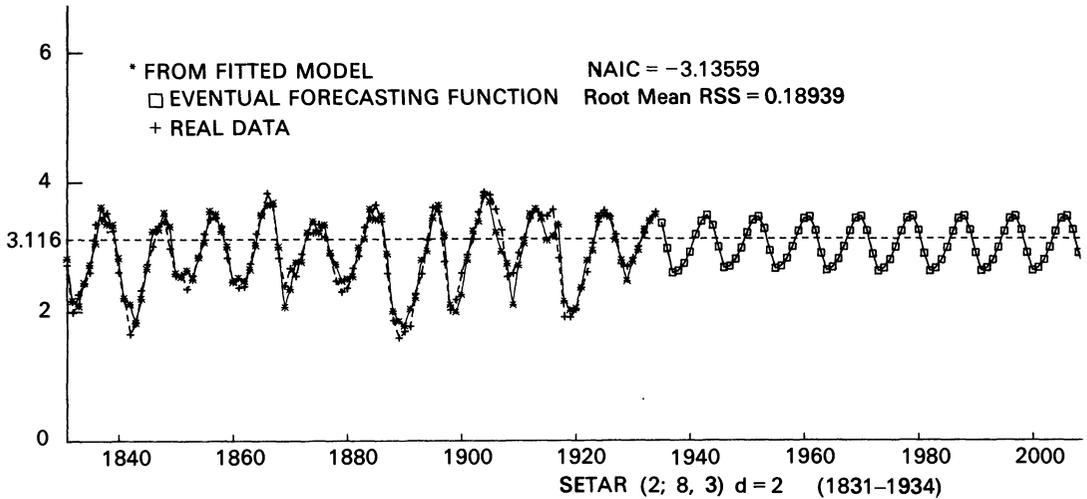
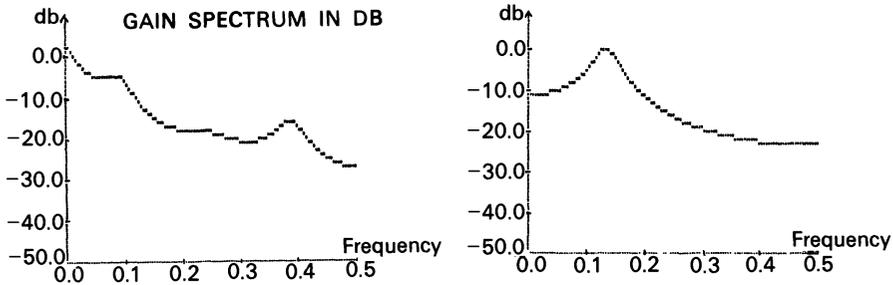


FIG. 8. SETAR for lynx data.



FIGS 9a, 9b. Gain spectra for lynx data.

$X_{n-2} > 3.1163$ and $X_{n-2} \leq 3.1163$ respectively. They appear to peak at different frequencies which might be interpreted as indicating some "amplitude-frequency dependency".

Tong (1980a) has compared the one-step-ahead predictions based on the linear models mentioned in the last paragraph with those based on a SETAR. In particular, the SETAR (2; 6, 3), $d = 2$, fitted to the years 1821-1920 (*op. cit.*) reduces the root-mean-square-error of one-step-ahead predictions (RSME) by 10 per cent when compared with the AR(12) reported in Tong (1977b, p. 466).

At this point we may anticipate a predator-prey system behind the whole scene, for the modelling of which our TARSC may offer interesting possibilities. Unfortunately, we have been unable to obtain reasonably "clean" snowshoe rabbit data in the Hudson Bay area of the same period of time. Some other "dirty" rabbit data of (probably) not exactly the same region were extracted from MacLulich (1937) and discussed in an unpublished report by Tong, which did not give any definite conclusion.

(B) *Sunspot data.* In his discussion of Morris' analysis of the sunspot data, Priestley (1977) has noted that a threshold AR model may be appropriate.

The following SETAR(2; 4, 12) is fitted to Wolf's sunspot numbers $\{X_t; t = 1700, \dots, 1920\}$.

$$X_t = \begin{cases} 10.5440 + 1.6920X_{t-1} - 1.1592X_{t-2} + 0.2367X_{t-3} + 0.1503X_{t-4} + \varepsilon_t^{(1)} & \text{if } X_{t-3} \leq 36.6, \\ 7.8041 + 0.7432X_{t-1} - 0.0409X_{t-2} - 0.2020X_{t-3} + 0.1730X_{t-4} \\ \quad - 0.2266X_{t-5} + 0.0189X_{t-6} + 0.1612X_{t-7} - 0.2564X_{t-8} \\ \quad + 0.3195X_{t-9} - 0.3891X_{t-10} + 0.4306X_{t-11} - 0.0397X_{t-12} + \varepsilon_t^{(2)} & \text{if } X_{t-3} > 36.6, \end{cases} \quad (9.2)$$

where $\text{var } \varepsilon_t^{(1)} = 254.64$, $\text{var } \varepsilon_t^{(2)} = 66.80$. (The pooled mean sum of squares of residuals = 153.71.)

Fig. 10 shows the fitted residuals, the one-step-ahead predictions and the eff. Note that the eff is a periodic function of a 31-year period, consisting of 3 local maxima and 3 local minima, i.e. 3 "local cycles". The local cycles are *asymmetric* with rise (descent) periods being 4 (6), 4 (6), 4 (7). We may regard 31/3 as a "fractional period" of the sunspot cycle. We note that the asymmetry of these cycles runs in a reversed direction to that of the lynx. Fig. 11 shows the "high" and "low" gain spectra, which tend to be related to the empirical observation that the skewness of the sunspot cycles depends on their amplitudes. Logarithmic and square-root transformations of the data have been tried but we have not observed any obvious advantage in this case.

Using a method due to Ozaki and Tong (1975), Akaike (1978) has shown that the sunspot data are better modelled as non-stationary over a long period, although they may be regarded as stationary over a shorter period. Some of the non-stationarity must be due to the introduction

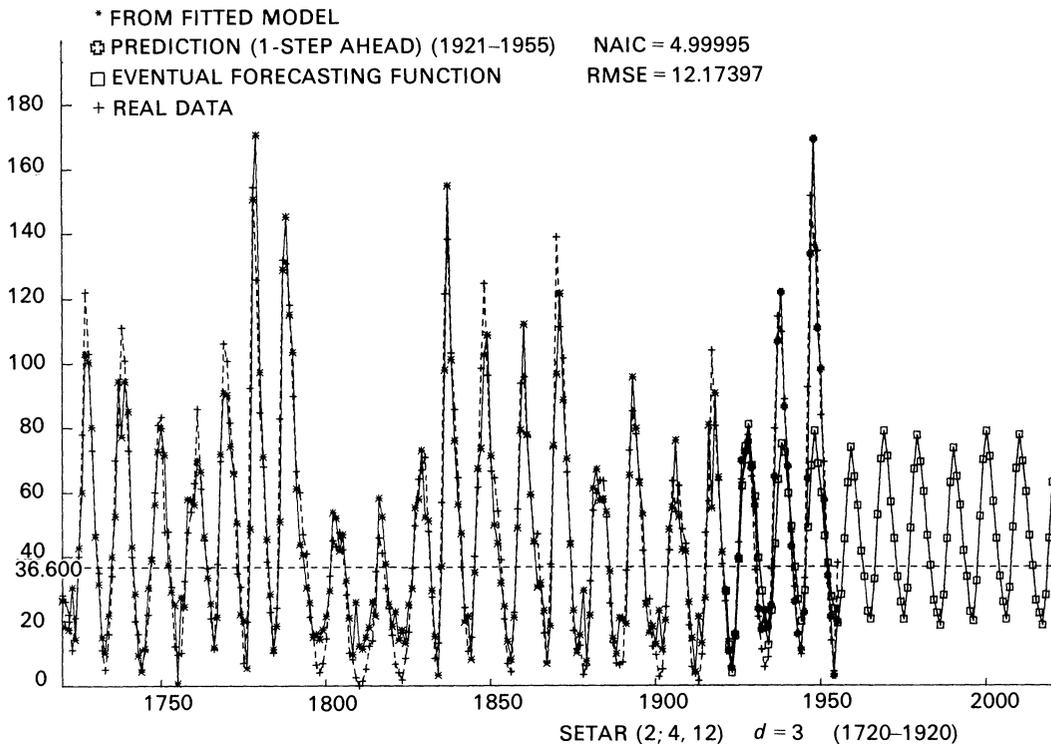
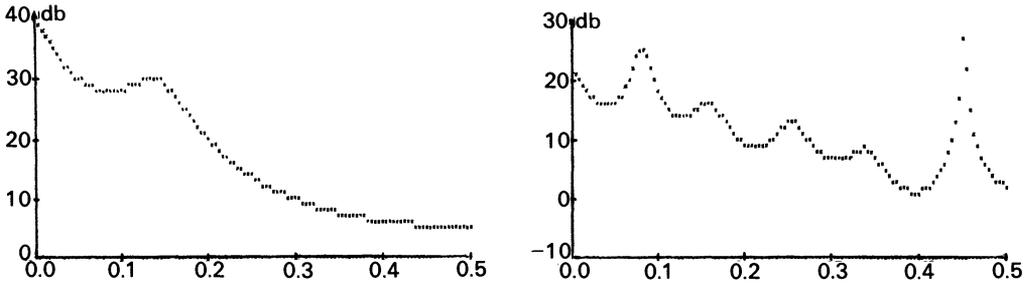


FIG. 10. SETAR for sunspot data 1700-1920.



FIGS 11a, 11b. Gain spectra for sunspot data 1700–1920.

of the photographic recording technique towards the later part of the record. We therefore look at the data from 1837 to 1924 more closely and the following SETAR(2; 4, 2) model is fitted :

$$X_t = \begin{cases} 25.2120 + 0.9820X_{t-1} - 0.0377X_{t-2} - 0.6378X_{t-3} + 0.2454X_{t-4} + \varepsilon_t^{(1)} & \text{if } X_{t-5} \leq 47.4, \\ 0.3585 + 0.7569X_{t-1} - 0.0531X_{t-2} + \varepsilon_t^{(2)} & \text{if } X_{t-5} > 47.4, \end{cases} \quad (9.3)$$

where $\text{var } \varepsilon_t^{(1)} = 231.030$, $\text{var } \varepsilon_t^{(2)} = 63.075$. (The pooled mean sum of squares of residuals = 157.819.)

We note that the one-step-ahead predictions for the period up to 1944 are reasonable but deteriorate rapidly from then on thereby suggesting some non-stationarity of the sunspot data.

(C) *Mink–muskrat data* (1767–1849), from Jones (1914). Bulmer (1974, 1975), Jenkins (1975) and Chan and Wallis (1978) have attempted to explain the predator–prey relation of animal population data such as the mink–muskrat by means of essentially *linear* models. In contrast to these approaches, and motivated by Section 2(ii), we have fitted the following *non-linear* time

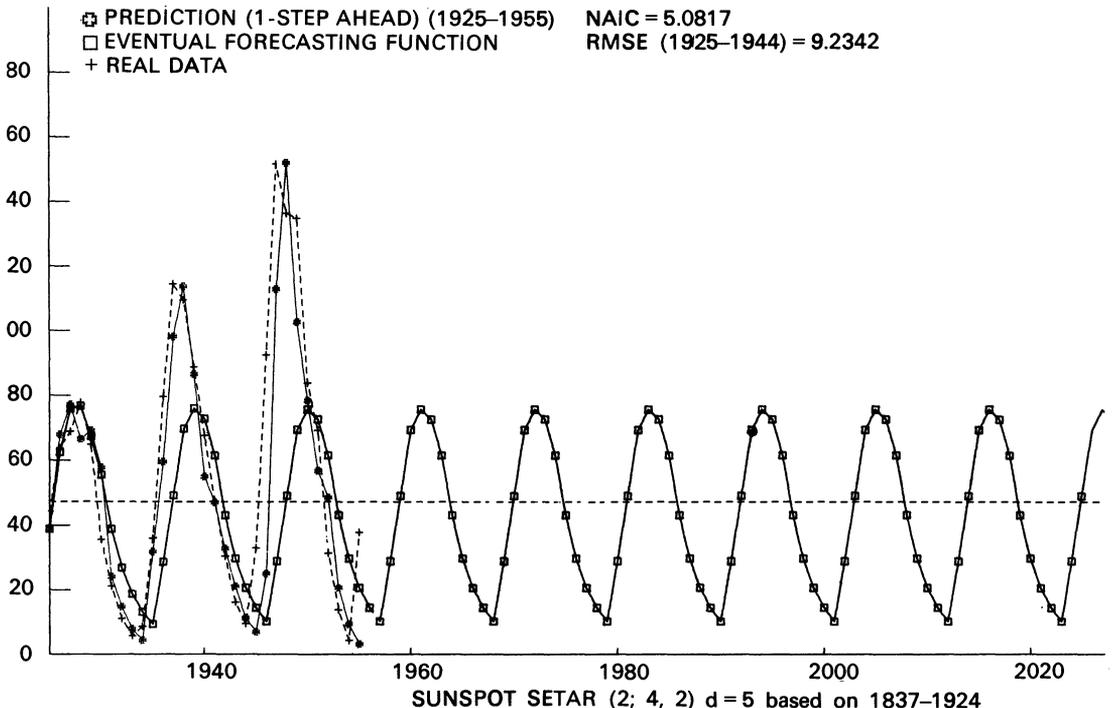
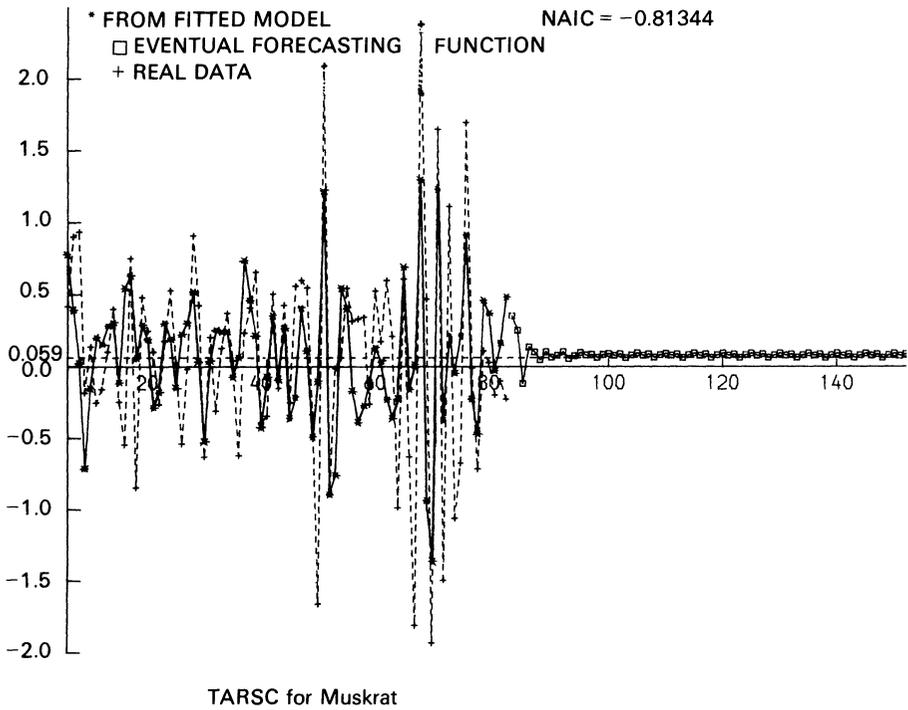
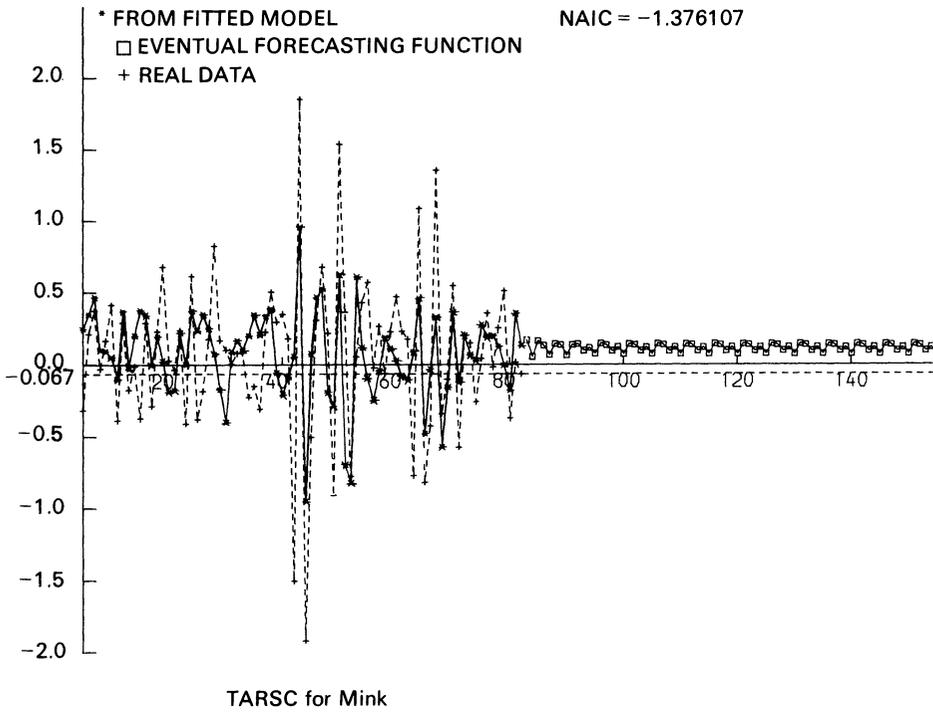


FIG. 12. SETAR for sunspot data 1837–1924.



FIGS 13, 14. TARSC for mink and muskrat data.

series model, specifically a TARSC model, to the mink and muskrat data (to the base e) of 1767–1849 after first differencing the logarithmically transformed data. We denote them by P_t and H_t , respectively :

$$P_t = \begin{cases} 0.1345 - 0.5988P_{t-1} + 0.0391H_{t-1} + \varepsilon_t^{(1)} & \text{if } H_{t-2} \leq 0.0592, \\ 0.2326 - 0.5272P_{t-1} + 0.1047H_{t-1} - 0.6445P_{t-2} + 0.1002H_{t-2} + \varepsilon_t^{(2)} & \text{if } H_{t-2} > 0.0592, \end{cases} \quad (9.4a)$$

$$H_t = \begin{cases} 0.4405 - 0.3867H_{t-1} - 0.3465P_{t-1} + \eta_t^{(1)} & \text{if } P_{t-2} \leq -0.0672, \\ 0.1976 - 0.4967H_{t-1} - 0.1608P_{t-1} - 0.3516H_{t-2} - 0.3802P_{t-2} + 0.0150H_{t-3} + \eta_t^{(2)} & \text{if } P_{t-2} > -0.0672, \end{cases} \quad (9.4b)$$

where $\text{var } \varepsilon_t^{(1)} = 0.2907$, $\text{var } \varepsilon_t^{(2)} = 0.1506$ (pooled value = 0.2170), $\text{var } \eta_t^{(1)} = 0.4616$, $\text{var } \eta_t^{(2)} = 0.3073$ (pooled value = 0.3588).

This model seems to lend some support for a predator–prey model in this case. The fitted threshold values are also interesting and seem to give some support for the approximate 2 year lead-lag relationship between the “muskrat cycle” and the “mink cycle” noted by Bulmer (1975). Note also the signs of coefficients of H_{t-2} and P_{t-2} in (9.4). Indeed, this fitted model has a limit cycle of period 5 years. The mink and muskrat effs show periodic functions with opposite skewness. (See Figs 13 and 14.) This is again what one might expect in a predator–prey situation, adding yet further support to the predator–prey hypothesis. (See Fig. 15.) The fact that the mink limit cycle is wholly above the threshold value while the muskrat limit cycle oscillates about the threshold value seems to be tentatively related to Bulmer’s conclusion that the muskrat cycles drive the mink cycles and not the other way round.

However, this example has also revealed the difficulty of bivariate TAR time series modelling to *very short* data sets. The desire to keep the number of parameters to a reasonable level has led to a rather high residual variance. Bearing this in mind, we must emphasise the tentative nature of the model (9.4), which cannot be taken as giving conclusive evidence in support of the predator-prey hypothesis. On the other hand, the limitation of a linear model in this respect is well known. (See, for example, Tong, 1980a.)

(D) *Kanna riverflow and rainfall data* (daily record of year 1956). It was Sugawara’s tank model (1961) for the analysis of the riverflow–rainfall relation which led Tong (1977b, 1978, 1980a) to the formulation of the threshold models. It therefore seems appropriate that we should conclude our case studies with a hydrological example.

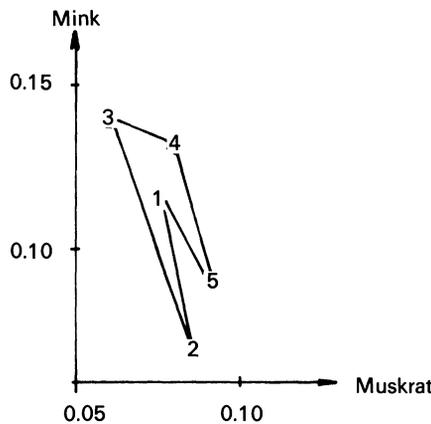


FIG. 15. Phase diagram for mink–muskrat.

* FROM FITTED MODEL
Δ PREDICTION (1-STEP AHEAD) (281-366) NAIC = -6.03804
+ REAL DATA RMSE = 0.03055

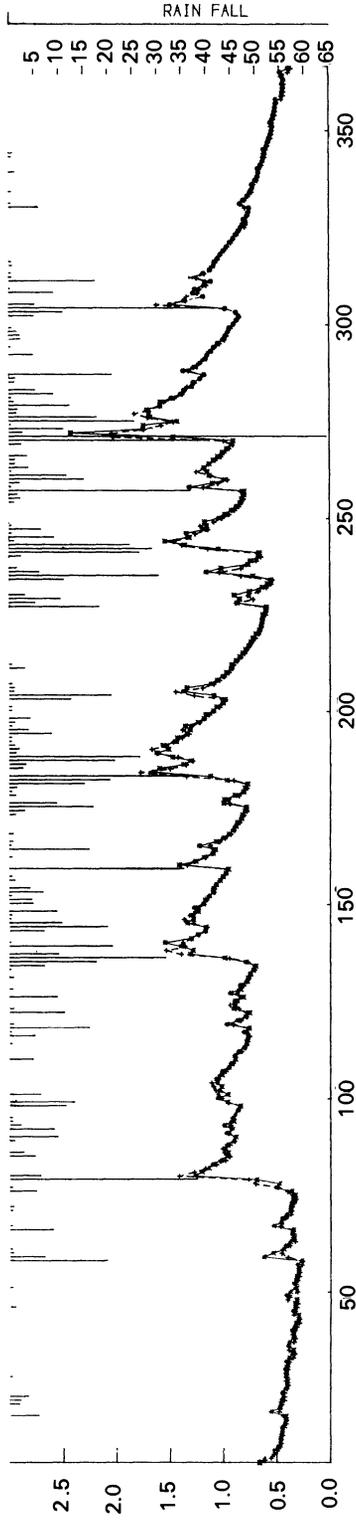


FIG. 16. TARSO(2; (5, 4), (2, 2)) for Kanna riverflow (in mm/day) in the year 1956. The vertical lines give daily record of rainfall (in mm/day) with the vertical scale denoted on the right-hand margin.

The Kanna River is a river with a small catchment area (under 1000 km²) in Japan. Seasonal variations of Japanese rivers are quite regular due to the rather well-defined rainy season there. The ground soil is also rarely dry. It is, therefore, reasonable to expect that most of the cyclical variation of the riverflow data can be explained by that of the rainfall data if only the transformation from the latter to the former is adequately modelled. As a result, we may treat the latter (denoted by Y_t) as an instrumental time series and fit a TARSO model to the former (denoted by X_t , after a logarithmic transformation to the base 10).

The following model is fitted, using only record of the first 280 days.

$$X_t = \begin{cases} 0.0185 + 0.9992X_{t-1} + 0.0065Y_{t-1} - 0.1519X_{t-2} - 0.0017Y_{t-2} \\ \quad + 0.1236X_{t-3} - 0.0004Y_{t-3} - 0.0295X_{t-4} - 0.0014Y_{t-4} \\ \quad + 0.0065X_{t-5} + \varepsilon_t^{(1)} & \text{if } Y_{t-1} \leq 4.6000, \\ 0.1281 + 0.5044X_{t-1} + 0.0146Y_{t-1} + 0.2767X_{t-2} + 0.0014Y_{t-2} + \varepsilon_t^{(2)} \\ \quad & \text{if } Y_{t-1} > 4.6000, \end{cases} \quad (9.5)$$

where $\text{var } \varepsilon_t^{(1)} = 0.0012$, $\text{var } \varepsilon_t^{(2)} = 0.0173$ (pooled variance = 0.0047).

Based on this fitted model (9.5), we have obtained one-step-ahead predictions of the next 86 days, and Fig. 16 represents an 18 per cent reduction in the RMSE when compared with the linear model. We would suggest that the TARSO models could be useful for the purpose of synthetic hydrology. However, a practically more important problem is the modelling of the rainfall, which so far seems to have bedevilled time series analysts! The solution of this difficult problem will pave the way for a long-range forecasting of floods.

10. SOME DISCUSSION

Through our practical experiences in applying the threshold models to real data, we are led to believe that this new class of models offers exciting potential in the analysis of cyclical data and opens up new vistas. However, much work remains to be done and we would just mention a few areas.

Following the same idea as in Ozaki and Tong (1975), we can partition the time axis suitably so as to arrive at a class of locally stationary TAR models. For example, the rainfall–riverflow relationship may change in an obvious way between the summer seasons and the winter seasons for some rivers. We have some encouraging results in a non-stationary TARSO modelling of the River Cam data, which will be reported elsewhere.

We are certainly conscious of the possible shortcomings in using the minimum AIC method in our model identification. We have made it clear in our proposed procedure, and we emphasize once again, that this method is not the only tool we have used, although our experiences have led us to believe that it can give us good service, provided we use it sensibly. For example, we have been particularly cautious when the minimum AIC method selects a model whose parametric dimensionality is near to the maximum possible dimension entertained. (See, for example, Shimizu, 1978.) It seems that the latest Bayesian extension of the minimum AIC method developed by Akaike (1979) holds out the possibility of a more sophisticated procedure. Briefly, we may treat $\exp(-\frac{1}{2}\text{AIC}(k))$ as the “likelihood” of the k th order model from which we may obtain the posterior distribution over the class of models under consideration, prior being some reasonably simple distribution, say, proportional to $(k+1)^{-1}$. A Bayesian model may then be obtained by averaging the class of models under consideration with respect to the posterior distribution.

Of course, in principle there is no difficulty in extending our TAR by including the moving average terms, obtaining a TARMA. We have as yet insufficient practical experience in the identification of a TARMA, the main difficulty being the computer time consideration. Another

possibly useful direction of extension is to allow the $\mathbf{A}^{(j)}$'s and $\mathbf{B}^{(j)}$'s of (4.1) to be functions of \mathbf{X}_{n-1} , which includes the *piece-wise polynomial* approximation. (See also Tong, 1980a.)

Finally, in the case of linear models, the notion of a *state* has been fully developed and is identified with a set of *observable* basis vectors of the *predictor space* of Akaike (1974). This fundamental notion gives a precise mathematical meaning to the information reduction process expressed by the linear ARMA model under the *only* assumption of finiteness of the dimension of the predictor space. Such a notion is lacking in the non-linear case. In this respect, the TAR (or the TARMA) models, as well as all other known classes of non-linear time series models, must be regarded at present as *ad hoc* (Akaike, private communication). We would argue that the formulation of this fundamental notion will be a most challenging and urgent problem for the next stage of development in non-linear time series modelling. Towards this end, it seems that a topological approach might offer some insight. Now, let \mathcal{X} denote a separable metric space generated by $\hat{X}_1, \hat{X}_2, \dots$, the metric being the mean square norm. Here

$$\hat{X}_i = E[X_i | X_0, X_{-1}, \dots], \quad i = 0, 1, 2, \dots$$

We call \mathcal{X} the *general predictor space*. (If the \hat{X}_i 's are linear in X_0, X_{-1}, \dots , then this general predictor space reduces to the predictor space of Akaike.) We now follow the rigorous definition of the *dimension* of a separable metric space given by Menger and Urysohn. (See, for example, Hurewicz and Wallman, 1941, p. 24.) Now, a fundamental theorem in *dimension theory* (*op. cit.* p. 52) shows that if \mathcal{X} has dimension $n (< \infty)$ then among the totality of continuous real-valued functions defined over \mathcal{X} , there is a set of $2n + 1$ (but not any fewer) functions $\xi_1, \xi_2, \dots, \xi_{2n+1}$ (the co-ordinate functions), which form a *basis*, in the sense that every continuous real-valued function f defined on \mathcal{X} is expressible in the form

$$f = g(\xi_1, \dots, \xi_{2n+1}),$$

where g is a continuous function of $2n + 1$ variables. We may identify $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_{2n+1})$ as a *state vector*, which seems to offer possibilities of further developments towards a fuller understanding of the structural aspects of non-linear time series models.

ACKNOWLEDGEMENTS

We are most grateful to Dr H. Akaike, Mr N. Komura, Mr T. Ozaki and Professor M. B. Priestley for their assistance, discussions and criticisms during the formative years of the present work. The two referees' very careful scrutinies of the paper and helpful suggestions have led to a considerably improved version.

Our thanks also to Mr P. K. Wong of UMIST for supplying us with Fig. 1.

Dr Tong's research was supported in part by a grant from the Science Research Council of the United Kingdom.

REFERENCES

- AIZERMAN, M. A. (1963). *Theory of Automatic Control*. New York: Pergamon.
- AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd Int. Symp. on Inf. Theory* (B. N. Petrov and F. Csaki, pp. 267-281. Budapest: Akademiai Kiado.
- (1978). Stochastic theory of minimal realisations. *I.E.E.E. Trans. Auto. Control*, **AC-19**, 667-674.
- (1978). On the likelihood of a time series model. *The Statistician*, **27**, 217-235.
- (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, **66**, 237-242.
- BARTLETT, M. S. (1966). *Stochastic Processes*, 2nd ed. Cambridge: Cambridge University Press.
- BHANSALI, R. (1979). A mixed spectrum analysis of the lynx data. *J. R. Statist. Soc. A*, **142**, 199-209.
- BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- BULMER, M. G. (1974). A statistical analysis of the 10 year cycle in Canada. *J. Anim. Ecol.*, **43**, 701-718.
- (1975). Phase relations in the 10 year cycle. *J. Anim. Ecol.*, **44**, 609-621.
- CAMPBELL, M. J. and WALKER, A. M. (1977). A survey of statistical work on the MacKenzie River series of annual Canadian lynx trappings for the years 1821-1934, and a new analysis. *J. R. Statist. Soc. A*, **140**, 411-431; Discussion 448-468.

- CHAN, W. Y. T. and WALLIS, K. F. (1978). Multiple time series modelling: another look at the mink-muskrat interaction. *Appl. Statist.*, **27**, 168–175.
- FUJISHIGE, S. and SAWARAGI, Y. (1974). Optimal estimation for continuous systems with jump process. *I.E.E.E. Trans. on Auto. Control*, **AC-19**, 225–228.
- GOLUB, G. (1965). Numerical methods for solving linear least square problems. *Numerische Mathematik*, **7**, 206–216.
- HAGGAN, V. and OZAKI, T. (1980). Amplitude-dependent AR model fitting for non-linear random vibrations. *Biometrika*, **67**,
- HAKEN, H. (1978). *Synergetics: An Introduction*, 2nd ed. Heidelberg: Springer.
- HUREWICZ, W. and WALLMAN, H. (1941). *Dimension Theory*. New Jersey: Princeton University Press.
- JACOBS, P. A. and LEWIS, P. A. W. (1978). Discrete time series generated by mixtures—I: correlational and runs properties. *J. R. Statist. Soc. B*, **40**, 94–105.
- JENKINS, G. M. (1975). The interaction between the muskrat and mink cycles in North Canada. In *Proc. of the 8th Int. Biometric Conference* (Constanța, Romania, August 1974) (L. C. A. Corsten and T. Postelnicu, eds).
- JONES, D. A. (1978). Non-linear autoregressive processes. *Proc. Roy. Soc. London. A*, **360**, 71–95.
- JONES, J. W. (1914). *Fur-Farming in Canada*, 2nd ed. Ottawa: Commission of Conservation.
- KLIMKO, L. A. and NELSON, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.*, **6**, 629–642.
- LAWRANCE, A. J. and KOTTEGODA, N. T. (1977). Stochastic modelling of riverflow time series (with Discussion). *J. R. Statist. Soc. A*, **140**, 1–47.
- LI, T.-Y. and YORKE, J. A. (1975). Period three implies chaos. *Amer. Math. Monthly*, **82**, 988–992.
- MACLULICH, D. A. (1937). *Fluctuations in the Number of the Varying Hare (Lepus americanus)*. University of Toronto Studies No. 43, Biol. series. Toronto: University of Toronto Press.
- MAY, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, **261**, No. 5560, 459–467.
- MINORSKY, N. (1962). *Non-linear Oscillations*. New York: Van Nostrand.
- OZAKI, T. (1980). Non-linear time series models for non-linear random vibrations. *J. Appl. Prob.* (to appear).
- OZAKI, T. and TONG, H. (1975). On fitting of non-stationary autoregressive models in time series analysis. In *Proc. 8th Hawaii Int. Conf. on System Sciences*, pp. 225–226. North Hollywood: Western Periodicals.
- PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes", *J. R. Statist. Soc. B*, **27**, 204–237.
- (1977). Discussion of papers by Campbell *et al.* *J. R. Statist. Soc. A*, **140**, 448–450.
- PRIESTLEY, M. B. and TONG, H. (1973). On the analysis of bivariate non-stationary processes. *J. R. Statist. Soc. B*, **35**, 153–166, 179–188.
- RISHEL, R. (1975). Control of systems with jump Markov disturbances. *I.E.E.E. Trans. Auto. Control*, **AC-20**, 241–244.
- ROBINSON, V. G. and SWORDER, D. D. (1974). A computational algorithm for design of regulators for linear jump parameter systems. *I.E.E.E. Trans. on Auto. Control*, **AC-19**, 47–49.
- SHIMIZU, R. (1978). Entropy maximisation principle and selection of the order of an autoregressive Gaussian process. *Ann. Inst. Stat. Maths.*, **30**, 263–270.
- SUGAWARA, M. (1952). On the method of deriving the daily discharge of the River Koza from the daily precipitation (in Japanese). *Res. Memo of Inst. of Stat. Maths., Tokyo*, Vol. **8**, No. 10.
- (1962). On the analysis of run-off structure about several Japanese rivers. *Jap. J. Geophys.*, **2**, 1–76.
- TODINI, E. and WALLIS, J. R. (1977). Using CL for daily or longer period rainfall—run-off modelling. In *Math. Models for Surface Water Hydrology* (C. Cirane, L. Marine and D. Wallis, eds). London: Wiley.
- TONG, H. (1977a). Some comments on the Canadian lynx data. *J. R. Statist. Soc., A*, **140**, 432–436, 448–468.
- (1977b). Discussion of a paper by A. J. Lawrance and N. T. Kottegoda. *J. R. Statist. Soc. A*, **140**, 34–35.
- (1978). On a threshold model. In *Pattern Recognition and Signal Processing* (C. H. Chen, ed.). The Netherlands: Sijthoff and Noordhoff.
- (1980a). A view on non-linear time series model building. *Time Series* (O. D. Anderson, ed.). Amsterdam: North-Holland.
- TUNNICLIFFE-WILSON, G. (1977). Discussion on papers by Campbell *et al.* *J. R. Statist. Soc. A*, **140**, 455–456.
- TWEEDIE, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Processes and Their Applications*, **3**, 383–403.
- WALTMAN, P. and BUTZ, E. (1977). A threshold model of antigen-antibody dynamics. *J. Theor. Biol.*, **65**, 499–512.
- YULE, G. U. (1927). On the method of investigating periodicities in distributed series with special reference to Wolfer's sunspot numbers. *Phil. Trans. Roy. Soc. London, A*, **226**, 267–298.

DISCUSSION OF THE PAPER BY DR TONG AND MS LIM

Dr C. CHATFIELD (Bath University): I would like to congratulate the authors on making a substantial contribution to non-linear time-series modelling. I particularly welcome the fact that the paper combines new theoretical work with a number of practical examples using real data.

The authors are certainly right in suggesting that the time is ripe to look at alternatives to linear time-series models. The newcomer to non-linear models would do well to start by reading Granger and Andersen (1978) and Priestley (1978). These references and Subba Rao (1979) introduce an alternative class of models called *bilinear* models. The TAR models have the useful properties that they are locally linear and that they admit limit cycles, but in some other respects I find bilinear models more appealing. I hope that it will not be too long before bilinear and TAR models can be compared on real data. In particular I hope the authors can tell us how their model for the sunspot series compares with the bilinear models fitted by Granger and Andersen and by Subba Rao.

The first general question that might be asked in respect of non-linear modelling is: "How can we tell if a given time series is non-linear?" or "How can we decide if it is worth trying to fit a non-linear model?" The answer does not appear to be easy. In particular it is no use fitting a linear model, carrying out the usual diagnostic checks (such as looking at the autocorrelation function of the residuals), and hoping that these will indicate non-linearity because they won't. The tests, which are based on second-order properties, are designed to see if the "best" linear model has been fitted and not to indicate non-linearity. Indeed Granger and Andersen (1978) have shown that one can find a linear model and a bilinear model with the same second-order properties, and so they suggest looking at the second-order properties of $\{X_t^2\}$ as well as $\{X_t\}$ in order to distinguish between a linear and a bilinear model. More generally one might look at moments of $\{X_t\}$ which are higher than second-order, and the bispectrum is one possibility. In their examples, the authors have tried a TAR model because, for one reason or another, the "best" linear model was felt to be inadequate. For example, in the lynx data, the time "going up" systematically exceeds the time "coming down". What other features should we be looking for? Can the authors suggest a more general tactic for detecting non-linearity?

Let me now turn to the requirements listed by the authors in the introduction. Firstly they say that statistical identification should not entail excessive computation. Reading Section 8, I formed the impression that the computational problems are very much harder than those in both the linear and bilinear cases, so that it is not clear if the first requirement is satisfied. I would like to ask the authors how much more computing time is typically required to fit a TAR model.

Another sensible requirement proposed by the authors is that the overall prediction performance should be an improvement upon the linear model. Here I must confess to being a little disappointed. The reduction in RMSE is only 10 per cent for the lynx data, though 18 per cent for the riverflow data. Would the authors give us similar comparisons for the other two examples? Would the improvement be more substantial if predictions were compared for more than one step ahead? The extra complexity of fitting TAR models can, of course, only be justified by a substantial reduction in RMSE and/or by additional insight into the process mechanism.

Despite my queries and suggestions for future research, which are inevitable in any good read paper, I would like to conclude by saying how much I have enjoyed today's paper, and I have great pleasure in proposing the vote of thanks.

Dr G. TUNNICLIFFE WILSON (University of Lancaster): This paper strikes a welcome balance between theory and applications, but I confess to being more impressed by the latter. Following a tradition of empirical modelling, the authors have recognized features in the data which are not explained by linear models and have sought model extensions that *are* successful in representing these features.

Fundamentally, they use different linear models for different parts of the data, and I admire their ingenuity in demonstrating by simulation examples that TARs have sufficient potential for their task. The success so far demonstrated in practical applications suggests strongly that these models "approximate to the truth". Perhaps this success needs explaining as much from a data-analytic viewpoint as by investigating the theoretical properties of the models. For example in linear modelling a stable autoregression is *ensured* by almost all fitting processes. Is there a

similar law to ensure the stability and cyclical properties of fitted TARs? I would like to know more about the failures which the authors have decently buried.

My main concern upon reading the theoretical part of the paper, was that according to their own arguments the authors should have used thresholds in all the predicting variables of their autoregressions. How, therefore, have they managed to achieve success using a threshold in one variable only? Most of the applications are to series with strong cycles so the predicting variables do not wander over the whole of their possible range, but are effectively confined to a closed orbit in this space. The choice of a single threshold in one lagged variable is effectively a means of defining the two parts of this orbit, or equivalently of the cycle, over which different linear predictors may most profitably be constructed. I would suggest that if more thresholds were used, then the choice of threshold variable would not be so critical, but the direction in which a threshold was crossed would become important.

I would expect worthwhile improvements to follow from attempts to better define the state of the system producing the cycle. A second order system is likely, so that a "level" and "slope" measurement should adequately represent the state. I believe that classical time series operations such as smoothing to remove noise, and filtering to correct for trends and low frequency modulating effects could be useful in extracting these measurements. This approach recognises that stochastic effects may enter in many different ways, and whilst in linear models all the components may be gathered into one ARIMA model with no loss of information, for non linear models it may be best to decompose the series so as to extract the basic cycle. This cycle should be predictable using a non-linear function of the two state variables only—possibly linearized at different points of the cycle. Forecasts of the original series could then be resynthesized from the components.

With their emphasis on producing a simple prediction formula, TARs may be failing to exploit the evident structure of many cyclical series.

The models which have been presented to us this evening may have to be refined in many ways, but a good start to empirical non-linear modelling has been made and the authors should be congratulated for their perseverance with TARs. I have much pleasure in seconding the vote of thanks.

The vote of thanks was passed by acclamation.

Dr R. J. BHANSALI (Department of Computational and Statistical Science, University of Liverpool): I would like to extend my congratulations to the authors on an interesting paper. Although considerable work on the development of the sampling properties of the identification methods proposed in Section 8 still needs to be done, the threshold autoregressive models appear to offer novel possibilities for the modelling of practical time series.

Apart from the applications to biological and other physical time series discussed in the paper, I might mention commodity price series as a possible class of Economic time series where applications of these models may be useful, in particular for describing the Cob-web phenomena—that is, cycles arising because of the interaction between price and production of agricultural commodities. The inadequacy of the Random Walk hypothesis (Labys and Granger, 1970) for the modelling of monthly Cocoa price series, 1949–73, is discussed by Beenstock and Bhansali (1980), who have suggested that within the class of linear autoregressive models, a second-order model provides a better fit to the changes of Cocoa prices. However, over the forecasting period of July 1974–July 1977, the second-order model provides only a modest improvement in the predictions of the future cocoa prices. The need for fitting a non-linear model is indicated by an examination of the residuals obtained after fitting the second order model. These are found not to be approximated by the Normal distribution, though the Laplace distribution provides a better fit.

I was also interested to note the authors' rather pragmatic attitude towards the usefulness of Akaike's information criterion for the identification of time series models. This pragmatic

attitude appears to be in marked contrast to the almost religious attitude adopted earlier by Dr Tong in his analysis of the lynx data.

Dr M. G. BULMER (Oxford University): I should like to comment on the biological interpretation of the lynx and mink–muskrat data discussed in Section 9. The authors suggest that the lynx cycle is driven by a predator–prey interaction between the lynx and the snowshoe hare. There is good biological evidence that this is not the case. The hare cycle almost certainly drives the lynx cycle, but direct assessment of the impact of lynx predation on hare populations shows that it is too weak to be capable of causing the hare cycle. It has been suggested that the hare cycle is due to a plant–herbivore interaction (the hare being the predator and its plant food the prey). This situation might have been inferred from the periodogram of the lynx, which should be symmetrical about its peak value (which it is not) if the lynx–hare interaction drives the cycle, whereas it will exhibit a red shift (as observed) if the hare drives the lynx (Bulmer, 1978).

For the mink–muskrat data the authors fit a model which has a limit cycle of period 5 years. All previous authors have agreed that both mink and muskrat have a periodicity of about $9\frac{1}{2}$ years, the same as the lynx. A possible (though rather speculative) explanation is based on the facts that horned owls eat both hares and mink, and that mink eat muskrat. Thus the hare cycle drives an owl cycle, which drives a mink cycle, which drives a muskrat cycle. The observed phase lags are consistent with this explanation, mink being in phase with hares and muskrat two years earlier.

In conclusion, I must admit that I am rather doubtful of the gain in understanding which is likely to result from fitting the type of model developed in this paper. I would give higher priority to the fourth of the five requirements proposed in the Introduction.

Dr E. KHABIE-ZEITOUNE (North East London Polytechnic): It is my pleasure to congratulate the authors for a most stimulating paper. The non-stationary threshold models raise some challenging problems.

I would like to put forward the thesis that the juggling with the AIC criterion in this paper might one day be thought of as a preliminary identification/estimation method, only paving the way towards a fully fledged maximum likelihood estimation, applicable to a class of, say, non-linear SETARMA models.

The authors mention that the Levinson–Durbin procedure is not available for non-Toeplitzian block covariance matrices. Perhaps I should state here that this procedure has been generalized to deal with the inversion of “Toeplitzian” and “non-Toeplitzian” block covariance matrix Γ under mild conditions, with the computation of “generalized partial autocorrelations” (unpublished paper). This generalization leads to some very interesting results: If $\{\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_n}\}$ is a set of random p -vectors with covariance matrix $\Gamma = (\gamma_{i,j})$, $\gamma_{i,j}$ being the covariance matrix of (X_{t_i}, X_{t_j}) , then one can compute p^2 -matrix coefficients $\alpha_{n,h} = \alpha_{n,h}(\{\gamma_{i,j}\})$, dependent on $\gamma_{i,j}$'s, such that the set of random vectors $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, defined by

$$\mathbf{X}_{t_i} = \sum_{j=1}^i \alpha_{i,j} \mathbf{Y}_j \quad (1)$$

is uncorrelated, and such that

$$\mathbf{X}_{(n)}^T \Gamma^{-1} \mathbf{X}_{(n)} = \sum_{i=1}^n \mathbf{Y}_i^T \mathbf{Y}_i \quad (2)$$

with

$$\mathbf{X}_{(n)}^T = (\mathbf{X}_{t_1}^T, \dots, \mathbf{X}_{t_n}^T).$$

A computationally feasible methodology for exact maximum constrained likelihood estimation of model parameters can now be put forward for a number of models of stochastic

processes, both stationary or otherwise, in a unified approach. This methodology, which can be embedded into a computer program, requires small memory storage. It isolates correlational properties from model properties: given a model with parameters to be estimated, compute autocovariances of, and their first- and second-order derivatives with respect to, model parameters; then compute “partial autocovariances”, then $\alpha_{i,j}$ ’s and hence Y_j ’s from (1); further, under some probability density assumption $f(y)$ for Y_j ’s (an additional assumption of independence might be made here, not needed in the Gaussian case) compute the “exact” likelihood of, and derivatives with respect to, model parameters; finally maximize locally using constrained non-linear optimization or Newton–Raphson routines (the necessary Kuhn–Tucker conditions may be written down).

The above method may be successfully applied to the following models:

(i) Stationary time series

ARMA (no problem for starting values of autocovariances recursions);

Random phase (nonlinear: $X_t = A \cos(\omega t + \varepsilon)$)

DARMA of Jacobs and Lewis;

RARMA (unpublished; \Leftrightarrow ARMA with random orders and coefficients, all coefficients independent apart from AR ones which can be dependent, similarly MA coefficients).

(ii) Non-stationary processes

ARMA/RARMA (problems there: more unknown parameters than data values);

Processes with independent increments;

SETARMA ???

The method will be illustrated by reference to a SETAR(1) model, the problem being that of the computation of the autocovariances from the model parameters.

Consider the following threshold model: $X_t = U_t X_{t-1} + \varepsilon_t$, $E(X_t) = 0$, ε_t being white noise, with the random variable

$$U_t = \sum_{l=1}^L \phi^{(l)} \mathbf{1}\{X_{t-d} \in R_l\},$$

$\mathbf{1}\{\cdot\}$ being the indicator function of the event $\{\cdot\}$. This model can be written

$$X_t = \phi^{(l)} X_{t-1} + \varepsilon_t \quad \text{if } X_{t-d} \in R_l.$$

If $\text{Prob}(X_{t-d} \in R_l)$ is independent of t , then the difficulty I am going to mention will not arise. However, when it cannot be assumed that $\text{Prob}(X_{t-d} \in R_l)$ is constant with respect to t , then after some algebraic manipulation, one can show that

$$\gamma_{t,t+k} = E(X_t X_{t+k}) = \sum_{l=1}^L \phi^{(l)} \gamma_{t,t+k-1}^{(l)} \pi_t^{(l)}, \tag{3}$$

where

$$\gamma_{t,t+k-1}^{(l)} = \text{cov}[(X_t, X_{t+k-1}) | \text{under model } \Gamma],$$

and

$$\pi_t^{(l)} = \text{Prob}(\mathbf{1}\{X_{t-d} \in R_l\} = 1) = \text{Prob}(\mathbf{1}\{\sum_{h=1}^{t-d} \alpha_{t-d,h} Y_h \in R_l\} = 1).$$

If one assumes further that Y_1, \dots, Y_n are independent, then

$$\pi_t^{(l)} = \int \dots \int f(y_1) \dots f(y_n) dy_1 \dots dy_n \mathbf{1}\{\sum_{h=1}^{t-d} \alpha_{t-d,h} Y_h \in R_l\}$$

This results in a high order non-linear system of equations (3) to be solved in order to obtain Γ .

If each of the l AR(1) models, $l = 1, \dots, L$ is stationary, then $\gamma_{i,t+k-1}^{(l)} \equiv \gamma_{k-1}^{(l)}$ can be computed without difficulty. Further, under the Gaussian assumption, X_t is then strictly stationary and hence $\pi_t^{(l)} = \pi_t$, independent of t . System (3) then shows that the $\gamma_{t,t+k} \equiv \gamma_k$ does not depend on t , and hence the SETAR(1) process is stationary. Moreover, the exact likelihood can be computed without difficulty. Now the likelihood maximisation will assign the same values to the parameters $\pi^{(l)}$, even if one considers another partition $R'_1 \cup \dots \cup R'_L$, such that $\text{Prob}(X_{t-d} \in R) \equiv \text{Prob}(X_{t-d} \in R'_l)$, $l = 1, \dots, L$. Hence no information on the choice of the partition is provided there by likelihood. The AIC criterion is irrelevant as both partitions have the same number of parameters. In this respect, the authors' split of the data into subsets, though heuristic, is invaluable for identifying the preferred partition. The generalization of these ideas will be presented elsewhere.

I have difficulty in interpreting the event $\{ \text{the observation } x_{t-d} \in R_l \}$. If it means that $X_{t-d} \in R_l$ conditional upon information I_{t-d} available prior to $(t-d)$, and also information from $(t-d+1)$ to $(t-1)$, then $\pi_t^{(l)}$ depends on t and the difficulty remains.

Professor M. B. PRIESTLEY (University of Manchester Institute of Science and Technology): Tonight's paper is one of a group which have appeared recently on non-linear time series models, and which I feel represent a significant advance in the methodology of time series. We now have several classes of "tractable" non-linear models (e.g. bilinear, threshold autoregressive and exponential autoregressive) which have been shown to be capable of providing good fits to a wide variety of data, and which possess more interesting structural properties than the conventional linear models. The basic idea underlying the TAR models is that, when we abandon linear models, we should look first at models which are "locally" linear. However, in this context the term "local" does not refer to a neighbourhood of a particular time point—rather, it refers to a particular region of the "state space" of the process. (The former notion is related to "non-stationarity", rather than "non-linearity", and there is an interesting form of duality between these two concepts.) For the AR(k) process,

$$X_t + a_1 X_{t-1} + \dots + a_k X_{t-k} = \varepsilon_t,$$

the evolution of the process is determined by $(X_{t-1}, X_{t-2}, \dots, X_{t-k})$ (together with the future ε_t 's), these k quantities acting as "initial conditions" in determining the solution of the above difference equation from time t onwards. Consequently, the "state" of the process at time t is represented by the k -dimensional vector, $\mathbf{x}_t^T = (X_t, X_{t-1}, \dots, X_{t-k+1})$, and the most general form of locally linear AR(k) model would be one in which the coefficients were all functions of \mathbf{x}_{t-1} , i.e. would take the form

$$X_t + a_1(\mathbf{x}_{t-1}) X_{t-1} + \dots + a_k(\mathbf{x}_{t-1}) X_{t-k} = \varepsilon_t. \quad (*)$$

We may refer to this as a general "state-dependent model". Although this type of model can be put into the form of the authors' equation (5.4), their "piecewise linearization" approach would involve the partition of the k -dimensional space, R^k , into a multitude of "small" regions in each of which the coefficients (a_1, \dots, a_k) were assumed to take constant values. Such an approach would be quite horrendous from a computational point of view, and the authors' way round this difficulty is to assume that the coefficients depend on only *one component* of \mathbf{x}_{t-1} , namely X_{t-d} (d being some specified integer, $1 \leq d \leq k$).

There is, however, an alternative way of dealing with general state-dependent models, which I will now indicate very briefly. The simplest form of functional dependence of the coefficients on the state-vector is that in which each a_i is a linear function of \mathbf{x}_{t-1} , i.e.

$$a_i(\mathbf{x}_{t-1}) = a_i^{(0)} + \mathbf{x}_{t-1}^T \boldsymbol{\beta}_i, \quad \text{say.}$$

This assumption is quite restrictive, but we may relax it by allowing the "gradients", $\boldsymbol{\beta}_i$, to be themselves state-dependent, so that the (a_i) are then only locally linear functions of \mathbf{x}_{t-1} . If we do

this we are then faced with the problem of specifying the functional form of the (β_i) , but we can obviate this difficulty by simply letting the β_i "wander" over time, i.e. we allow $\beta_i = \beta_i^{(t)}$ to depend purely on the time parameter. The basic idea now is to let the $(\beta_i^{(t)})$ wander in the form of a *random walk*, i.e. to set $\beta_i^{(t)} = \beta_i^{(t-1)} + v_t$, where the (v_t) are independent zero mean random variables with variance matrix, Σ_v , say. The estimation procedure then determines, for each t , those values of $(\beta_i^{(t)})$ which, roughly speaking, minimise the discrepancy between X_{t+1} and its predictor, \tilde{X}_{t+1} , computed from the model. The estimation procedure is thus based on a sequential type of algorithm, similar in nature to the *Kalman filter* algorithm, and it leads to coefficients (a_i) which are "locally optimal" in the sense that they provide the best "local predictor" for the next observation. (The "smoothness" of the (a_i) as functions of x_{t-1} is controlled by the ratio of $\|\Sigma_v\|$ to σ_ϵ^2 .) Once we have determined suitable values of the (a_i) over a range of time points we can plot these as functions of the corresponding state vectors, and then, using some form of multidimensional smoothing (e.g. via "splines" or the "kernel" method), we can build up a graphical picture of the functional form of the (a_i) . Thus, for a TAR model the (a_i) should appear as "ridges" of step-functions, depending only on one component of x_{t-1} .

The general state-dependent model (*) includes, as special cases, the TAR and exponential autoregressive models, and, by adding moving average terms, it can also accommodate bilinear models (see Priestley, 1979).

As far as threshold models are concerned, the authors have given a convincing demonstration of their applicability to a wide range of data, and their modelling fitting expertise is certainly most impressive. As the authors show, these models can give rise to some fascinating features (such as limit cycles and jump phenomena), and they will, I am sure, stimulate much interest in this new and rapidly growing area of time series analysis.

Dr B. W. SILVERMAN (University of Bath): It would be interesting to know whether any connections can be made between the threshold models discussed tonight and the ideas of catastrophe theory, which might well give rise to models with piecewise behaviour of the kind described. Certainly the electric relay can be viewed in these terms. Models based on catastrophe theory would be attractive from the point of view of the authors' criteria (iv) and (v), while any relations with the authors' methodology would help with the fitting of catastrophe-theoretic ideas to real data.

Mr E. J. GODOLPHIN (Royal Holloway College): I would like to join the other discussants in congratulating the authors on an interesting and thought-provoking paper.

I have two questions to ask the authors, the first of which is about the R_j s defined in Section 4, which seem to be best regarded either as random variables or possibly as deterministic but unknown quantities to be derived from the available data. Am I right in thinking that these quantities are likely to be considerably more important to the specification of the model than even the various sets of autoregressive parameters themselves? If this is so, I wonder if the authors could say a little more about the properties of their threshold estimates beyond the comments made in Section 8? For example, in one of the authors' examples the number 4.6 appears in equation (9.5); but how useful an estimate is that?

Secondly, in Section 8 the authors also refer to the eventual forecast function which they adopt for specification purposes. Have the authors considered obtaining a functional form for the eventual forecast function for variable lead times? I am thinking, for example, about results which would parallel a paper of my own (Godolphin, 1975) which deals with the non-stationary linear case, including seasonal models. If it were possible to compare the different kinds of functional forms for these eventual forecast functions with those for the linear models, this might provide an interesting way of exploring the manner in which the authors have succeeded in generalizing the linear case.

Dr. D. A. JONES (Institute of Hydrology, Crowmarsh Gifford, Wallingford, Oxon OX10 8BB): Recently a number of wide classes of non-linear time-series models have been proposed: autoregressive models, a class of which are discussed by the authors, and bilinear models (Granger and Andersen, 1978; Granger, 1978; Priestley, 1978). An example of the use of a smooth non-linear autoregressive structure, as opposed to one which is sectionally linear, appears in O'Connell and Jones (1979). These developments make it appropriate to question whether time-series models are necessary.

A "model" here means a complete probabilistic description of a series (apart from certain parameters). The answer will be different depending on the purpose of the data analysis. I am mainly thinking of problems where forecasts are to be constructed. Follow-up questions concern whether the model properties that are used in practice can be replaced by methods which are not model-dependent, and whether checks of the complete structure of models are actually available. A question which possibly encompasses these is whether linear *models* are actually used at present. It can be argued that relatively little use is made of linear models, as opposed to linear forecasts, on the basis that standard techniques involving ARMA structures are concerned essentially with forecasts rather than models. A question of a different character is whether discrete-time models are realistic. Should not all real processes be thought of as evolving continuously in time, even if at a rather basic level?

Much of time-series analysis is directed towards constructing forecasts. In this situation a forecasting rule can be fitted directly, rather than fitting a model. A class of possible rules for estimating y_t from y_{t-1}, y_{t-2}, \dots is first defined in terms of a number of parameters θ : let $\hat{y}_t(\theta)$ be the value of the forecast of y_t and let $\hat{Y}_t(\theta)$ be used when y_{t-1}, y_{t-2} are treated as random variables. Given observations (y_1, \dots, y_T) on a random process, the rule is fitted as follows:

- (a) choose a loss function: squared-error loss is used here for convenience,
- (b) for any θ define

$$s_T(\theta) = T^{-1} \sum_{t=1}^T (y_t - \hat{y}_t(\theta))^2,$$

- (c) find $\hat{\theta}_T$ such that $s_T(\hat{\theta}_T) \leq s_T(\theta)$ for all θ ,
- (d) use $\hat{y}_t(\hat{\theta}_T)$ to forecast y_t , and $s_T(\hat{\theta}_T)$ as the estimated mean square error of the forecast.

Under weak conditions on the forecast rule and the process $\{Y_t\}$,

$$\lim_{T \rightarrow \infty} s_T(\theta) = \sigma(\theta) = E[\{Y_t - \hat{Y}_t(\theta)\}^2]$$

and $\hat{\theta}_T \rightarrow \theta^0$, where $\sigma(\theta^0) \leq \sigma(\theta)$ for all θ . Thus the best rule out of the chosen class is found. Note that the forecast need not be a one-step-ahead forecast, and that no explicit model needs to be assumed. The above is just a fit *via* minimization of the sum of squares of forecast errors, as used in the authors' paper.

The authors of this paper have shown that, for various data sets, certain non-linear forecasts are better than linear forecasts. It is a very large step to go from these forecasts to writing down a model in terms of impulses which are both independent and Gaussian, as seems to be implied in the paper. It is only too easy to interpret a fitted forecast $\hat{x}_t = \alpha x_{t-1}^2$, say, as meaning that a model $x_t = \alpha x_{t-1}^2 + \varepsilon_t$ ($\varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$) is being used, but this should always be avoided. A quadratic forecasting rule, as used by Cox (1977), is a perfectly valid choice, whereas the supposedly corresponding model would usually be rejected out of hand.

Dr H. AKAIKE (The Institute of Statistical Mathematics, Tokyo): Strictly speaking, every real system is non-linear and non-stationary. Thus, when Dr Tong and Ms Lim try to generalize their TAR model it inevitably leads to the blurring of the nature of their model.

The authors note that the era of practical non-linear time series modelling is long overdue. Actually the modelling of each particular non-linearity was the key to our success in

implementing computer controls of cement rotary kilns and supercritical thermal power plants (Otomo, Nakagawa and Akaike, 1972; Nakamura and Akaike, 1979). Our experience with these systems suggest the importance of identifying a variable which characterizes the dynamics of a system. In Sugawara's tank model this variable is generated with the aid of an imaginary system of underground reservoirs. In these examples it was the analysis of the physical characteristic of each system that led to the choice of a particular "conditioning variable".

The variable J_n of TAR is an example of such a "conditioning variable", but since we are not told how practically to identify the variable we must resort to the examples. The examples of Canadian lynx and sunspot data show that the conditioning variables characterize the beginnings of the downward and upward paths of one cycle of oscillation. This observation clarifies why the model does not work well with the Mink-muskrat data, where the periodicity is not so clear. In the case of the Kanna riverflow data, again the input series is used to identify the upward and downward paths of the riverflow.

If these simple observations can capture the essence of these examples, what is the use of the elaborate generalization of TAR?

Professor D. R. COX (Imperial College, London): The authors' account of non-linear models, and in particular threshold autoregressive models, is very valuable. It is interesting and important to see the kinds of qualitative behaviour that simple systems of this kind can produce. I am, however, extremely uneasy at the analysis of the data in Section 9. For instance, I can see that it is interesting to show that (9.1) has limit cycles, but are the authors claiming that fitting 14 (or really more) parameters in this way to 114 observations tells us anything about what is "really happening"? Is the mechanical use of AIC, or any other criterion, a good idea: perhaps there are much simpler models that give nearly as good a value of AIC? Once the need for an irreversible process is clear, the possibilities are so rich that in the absence of strong guidance from theory, graphical or other preliminary analysis to establish the approximate form of dependencies present seems very desirable. Finally, in terms of prediction, how does the authors' model compare with the much simpler, although explosive, model I reported in the discussion of Campbell and Walker (1977)?

Professor K. W. HIPEL (University of Waterloo) and Professor A. I. MCLEOD (University of Western Ontario): We fully concur with Dr Tong and Ms Lim when they state that "The new era of practical non-linear time series modelling is, without doubt, long overdue." The authors should be commended for not only describing the theory for a new class of non-linear time series models but also for presenting procedures for model identification and efficient estimation of the model parameters.

When modelling hydrologic time series that are measured at short periods of time such as hourly or daily time intervals, the fitted stochastic models must take into account unique non-linear properties of the data that are caused by complex physical processes. For example, when precipitation falls on a river basin this causes the flows at a given location in a river to increase while after the precipitation has ceased the flows return to their former levels. This ascension-recession behaviour of the hydrograph of flow *versus* time makes the modelling of daily and hourly flows an arduous task. Comprehensive appraisals regarding research in stochastic hydrology have stressed the need for flexibility in modelling this type of phenomenon (see, for example, Lawrance and Kottogoda (1977), Kibler and Hipel (1979) and Hipel and McLeod (1980)) and recently some new non-linear models have been examined. Some of these models include the non-linear autoregressive model of O'Connell and Jones (1979) which is based upon the theoretical work of Jones (1978), the model of Yakowitz (1973) which is similar to the model examined by O'Connell and Jones (1979), and also the non-parametric Markov model of Yakowitz (1979). In addition, the bilinear model described by Granger and Andersen (1978) and also Priestley (1978) may be useful in hydrology. It may be instructive to compare the

models in today's paper to some of the aforementioned models in order to judge which models would be most appropriate to use in practice.

Some hydrologists are also concerned with the types of stochastic models that are employed to model data that are available at longer time periods such as weekly, monthly or yearly intervals. This is because the correlation structure for river flows that are relatively high may be different from the correlation pattern for low flows. Possible physical reasons for this behaviour include the manner in which the carrying capacity of the river channel varies, depending upon the level of the river, and the way in which different water table levels can affect the base flow of a river. The non-linear models of today's paper may prove to be effective for modelling phenomena of this type. Although short memory ARMA models have been shown to provide a statistical explanation for the Hurst phenomenon when modelling yearly geophysical time series (Hipel and McLeod, 1978), perhaps it may be worthwhile to determine if non-linear models provide a significant improvement over linear models when modelling annual data.

When modelling annual sunspot numbers from 1700 to 1960, McLeod *et al.* (1977) found that after taking a square root transformation of the data, the most appropriate ARMA model to fit to the transformed series is a constrained AR(9) model with the autoregressive parameters from lags three to eight left out of the model. In (9.2) and (9.3), Dr Tong and Ms Lim present their piecewise linear autoregressive models for modelling specified portions of the sunspot number series. Would a data transformation and perhaps omitting some of the less significant parameters from their models, help to lessen the values of the AIC?

Professor MITUAKI HUZII (Tokyo Institute of Technology, Dept. of Information Sciences): This paper gives us new ideas and methods for modelling non-linear systems. I think it will be an interesting problem to investigate the statistical properties of the process defined by (4.1) or (4.2). The reason is as follows:

- (i) The likelihood function of the observations depends on the condition $\{X_{n-d} \in R_j, j = 1, 2, \dots, l\}$. So, if we intend to examine the statistical properties of the maximum likelihood estimates of the unknown parameters, we have to know the properties of the process.
- (ii) When $\{R_j\}$ or $\{\gamma_0, \gamma_1, \dots, \gamma_l\}$ are unknown, we have to give a method for estimating these values. For this, the statistical properties of the process will be needed.

Dr I. T. JOLLIFFE (University of Kent at Canterbury): I have four questions on this interesting and useful paper.

The first concerns forecasts for more than one step ahead. It is stressed in Section 8 that the *eventual forecast function* (eff) is a plot of one-step-ahead predictions. However, with cyclical data it will often be required to forecast several time periods (at least one full cycle) into the future. How can such forecasts best be made?

A related point is that in practice different cycles in the same series will often have very different amplitudes and different cycle lengths; the sunspot data supply a good example, since amplitudes vary widely and the average cycle length was shorter in the first half of this century than in earlier periods. Can such variability be captured or, even better, forecast using the authors' models?

Thirdly, the authors' real data series are assumed to be stationary for all or part (sunspot data) of their length. Do the authors have any suggestions for dealing with non-stationary series?

My final question concerns the impressive range of behaviour exhibited by the examples of the threshold models given in Tables 3–7. How difficult was it to construct these examples, and are there other types of behaviour for which the authors failed to find models within their class?

Mr T. OZAKI (The Institute of Statistical Mathematics, Tokyo): This paper is of particular interest to me as it is closely related with some of my own work. I have the following three comments:

First, the set of threshold AR models is not general enough to include linear AR models, if (iii) of Definition 5.1 is assumed. If the proposed set of non-linear models and their identification method were appropriately defined we would expect to get a linear model for a linear process and a stable non-linear model for a stable non-linear process. My experience suggests that the linear threshold AR model often fails to satisfy this requirement (see Ozaki (1979a)).

Secondly, although the authors stress the importance of limit cycles in non-linear time series modelling, they did not give clear explanation of its mechanism for the threshold AR models. Extensive discussions of the limit cycle of non-linear time series models are given in Ozaki (1980), Haggan and Ozaki (1979) and Ozaki (1979c), based on the van der Pol equation, where the "shift back to center" property, which makes the process stationary ergodic in the sense of Tweedie (1975), was realized by making the instantaneous characteristic roots of the model state-dependent.

Thirdly, I would like to mention that the *jump resonance* and *amplitude-frequency dependency*, which are well known in relation to the *Duffing equation*, are also discussed extensively in Ozaki and Oda (1977), Haggan and Ozaki (1979) and Ozaki (1979b). If the authors would try to develop an analysis of the mechanism of the phenomena, these papers would be useful references. In these papers the "amplitude-dependent restoring property" of the Duffing equation is realized by the amplitude-dependence of the arguments of the instantaneous characteristic roots of the non-linear time series model.

Professor P. M. ROBINSON (University of Surrey): The threshold models are an interesting and important class, particularly in view of the connections that Tong and Lim have established with the theory of non-linear vibrations. Once the decision is made to forgo the great simplicity of the linear model, however, one is confronted with an embarrassingly wide choice of non-linear ones, even though many have yet to come under close scrutiny. This has recently led me to investigate a nonparametric approach to non-linear time series analysis. Let the conditional distribution of X_n given X_{n-1}, X_{n-2}, \dots , have expectation $f(X_{n-1})$, so that one has the general NLAR (1) (equation (5.1)). Instead of assuming a specific form for $f(x)$ we could estimate $f(x)$, for any given x , by

$$f(x) = \left(\sum_{n=1}^N w_{nN}(x) \right)^{-1} \sum_{n=1}^N w_{nN}(x) X_n \quad (1)$$

The non-negative weight functions $w_{nN}(x)$ can depend on X_1, \dots, X_N , and will generally give greatest weight to n -values for which X_{n-1} is close to x ; one possibility is $w_{nN}(x) = 1$, $|x - X_{n-1}| \leq c$; $= 0$, $|x - X_{n-1}| > c$. The estimator (1) can be shown to minimize a certain loss function; alternative loss functions will produce, for example, better robustness properties. Non-parametric regression estimators such as (1) have been considered previously, but in the case of independent, non-time series, observations; a recent reference is Stone (1977). Note that we could use the same type of approach to estimate other features of the distribution of X_n conditional on X_{n-1} . Extension is also possible to the general NLAR (p), for $p > 1$. Examination of the estimated $f(x)$ could provide a test of linearity or suggest a class of non-linear model for subsequent parametric analysis. If Tong and Lim's threshold model is selected, the non-parametric estimate may be of some use in its identification and estimation.

Dr T. SUBBA RAO (University of Manchester Institute of Science and Technology): The class of threshold autoregressive models proposed by the authors is definitely very useful, but sometimes other types of non-linear models fitted to the real time series may lead to models with fewer parameters. One class of such models—bilinear time series models—has recently been

extensively studied by Granger and Andersen (1978) and Subba Rao (1979). The advantage of this model is that it is possible to obtain theoretical expressions for the moments, spectra, etc., whereas such things are not always possible in quite a number of non-linear models. Using the estimation methods described by Subba Rao (1979), bilinear models are fitted by Mr M. M. Gabr and myself to sunspot data and the Canadian lynx data (the details of which will be reported elsewhere).

Sunspot data

The first 246 observations are used for fitting, and the fitted model is

$$\begin{aligned}
 X_t - 1.209X_{t-1} + 0.502X_{t-2} - 0.173X_{t-9} \\
 = 5.891 - 0.0098X_{t-2}e_{t-1} + 0.0103X_{t-8}e_{t-1} - 0.0048X_{t-8}e_{t-3} \\
 + 0.0016X_{t-3}e_{t-2} + 0.0014X_{t-4}e_{t-7} + e_t.
 \end{aligned}$$

The mean sum of squares of residuals is 141.18 and the AIC value is 1186.2.

The one-step-ahead predicted values from the bilinear model together with the true values are given in Table D1.

TABLE D1

<i>t</i>	247	248	249	250	251	252	253	254	255	256
True values	92.6	151.6	136.3	134.7	83.9	69.4	31.5	13.9	4.4	38.0
Predicted values	77.9	130.0	149.8	119.8	86.2	51.4	38.9	18.8	3.3	25.7

Since the authors have not actually tabulated their predicted values, it is not possible to compare the performance.

Canadian lynx data

We now consider the Canadian lynx data, discussed by the authors in Section 9. The data is logarithmically transformed. The bilinear model is fitted to the first hundred observations, and the fitted model is

$$\begin{aligned}
 X_t - 0.8845X_{t-1} + 0.1699X_{t-2} + 0.1271X_{t-4} - 0.5514X_{t-10} + 0.5280X_{t-11} \\
 = 1.117 - 0.1653X_{t-8}e_{t-10} - 0.0970X_{t-5}e_{t-8} + 0.0922X_{t-1}e_{t-1} + e_t.
 \end{aligned}$$

This model has nine parameters and the mean sum of squares of residuals is 0.0329 and the AIC is -283.577 which are considerably less than the values obtained by the authors. The one-step-ahead predictors for the next fourteen observations together with the true values are given in Table D2.

TABLE D2

<i>t</i>	101	102	103	104	105	106	107
True values	2.360	2.601	3.054	3.386	3.553	2.468	3.187
Predicted values	2.442	2.756	2.897	3.135	3.411	3.512	2.922
<i>t</i>	108	109	110	111	112	113	114
True values	2.723	2.686	2.821	3.000	3.201	3.424	3.531
Predicted values	2.706	2.583	2.844	2.966	3.159	3.299	3.415

Since the models we are fitting are non-linear models, it is interesting to fit the models to the original data rather than the transformed data. It is known that the predictors obtained for the

original data from the transformed data are highly biased. Now the question is why the authors did not consider fitting the threshold model to the original Canadian lynx data. The bilinear model fitted to the first 100 observations for the *original* Canadian lynx data is

$$\begin{aligned} X_t - 0.941X_{t-1} + 0.455X_{t-2} + 0.102X_{t-5} - 0.238X_{t-9} \\ = 612.68 - 0.00256X_{t-8}e_{t-10} - 0.000375X_{t-1}e_{t-7} \\ + 0.000164X_{t-2}e_{t-5} + 0.000142X_{t-7}e_{t-2} \\ + 0.000049X_{t-5}e_{t-4} + e_t. \end{aligned}$$

The mean sum of squares of residuals is 439075.2 and AIC = 1189.3.

The one-step-ahead predictors, together with their true values, are given in Table D3. Except for two values, the predictors are reasonably good.

TABLE D3

<i>t</i>	101	102	103	104	105	106	107
True values	229.0	399.0	1132.0	2432.0	3574.0	2945.0	1537.0
Predicted values	127.0	953.9	1147.2	2295.1	2931.4	2775.2	1653.3
<i>t</i>	108	109	110	111	112	113	114
True values	529.0	485.0	662.0	1000.0	1590.0	2657.0	3396.0
Predicted values	561.8	391.6	620.3	775.4	1474.9	2187.2	3252.5

It may be pointed out here that the test proposed by Subba Rao and Gabr (1981) has shown that the log (Canadian lynx data) is linear although it is not Gaussian.

An alternative non-linear model which may be useful for representing "cyclical data" is

$$Y(t) = \alpha \cos \{ \theta t + \phi + \beta X(t) \}$$

where α , θ and β are constants, ϕ is a uniform random variable and $\{X(t)\}$ is a strictly stationary process. (ϕ and $\{X(t)\}$ are independent.) This model is used for representing frequency modulated signals. If $\{X(t)\}$ is Gaussian with correlation coefficient $\gamma(t)$, then it can be shown (Hannan, 1970, p. 85) that the auto-covariance function of $Y(t)$ is

$$E\{Y(t)Y(t+s)\} = \frac{1}{2}\alpha^2 e^{-\beta^2(1-\gamma(s))} \cos(\theta s)$$

which shows that the covariance function is a harmonic function with decreasing amplitude as the lag s increases. The sample autocovariance functions of both the sunspot data and the Canadian lynx data are harmonic functions with decreasing amplitudes suggesting that the above model may also be very appropriate. The statistical analysis of this model is under investigation.

The AUTHORS replied later, in writing, as follows.

Dr Silverman's suggestion is most exciting, although we did speculate upon some "catastrophic" connection in the original version of the paper! A fuller exploration of the TAR-modelling/catastrophe relation is now available in Tong (1980b) and here we give only a *brief* indication as follows. The most famous catastrophe is the so-called cusp-catastrophe, which is characterized by the five *qualitative* features of *bimodality*, *inaccessibility*, *hysteresis* (*limit cycle*), *sudden jumps* and *divergence*. (See, for example, Zeeman, 1977.) It seems that the lynx data exhibit these features (Tong, 1980b) and that our SETAR model is really an execution of the cusp-catastrophic paradigm. (See Fig. D1.) Table D5 shows that the fitted SETAR(2; 8, 3) model of (9.1) has captured some of the probabilistic structure of the lynx data (cf. Table D4). The *bimodality* of some of the fitted conditional distributions is particularly interesting. The "crater"

TABLE D4
Bivariate histogram of log lynx data

$X(N)$ $X(N-2)$	0.2 -0.4	0.6 -0.8	1.0 -1.2	1.4 -1.6	1.8 -2.0	2.2 -2.4	2.6 -2.8	3.0 -3.2	3.4 -3.6	3.8 -4.0	4.2 -4.4	4.6 -4.8	5.0 Total								
5.0-	0	0	0	0	0	0	0	0	0	0	0	0	0								
4.8-	0	0	0	0	0	0	0	0	0	0	0	0	0								
4.6-	0	0	0	0	0	0	0	0	0	0	0	0	0								
4.4-	0	0	0	0	0	0	0	0	0	0	0	0	0								
4.2-	0	0	0	0	0	0	0	0	0	0	0	0	0								
4.0-	0	0	0	0	0	0	0	0	0	0	0	0	0								
3.8-	0	0	0	0	0	0	0	1	1	0	0	0	3								
3.6-	0	0	0	0	0	2	2	0	0	1	0	0	5								
3.4-	0	0	0	0	2	1	2	2	1	3	3	1	20								
3.2-	0	0	0	0	1	1	2	2	1	1	4	1	14								
3.0-	0	0	0	0	0	0	1	0	0	0	5	0	8								
2.8-	0	0	0	0	1	2	1	1	0	2	6	0	13								
2.6-	0	0	0	0	1	1	0	3	1	3	2	0	12								
2.4-	0	0	0	1	0	3	2	2	3	0	1	0	15								
2.2-	0	0	0	0	0	0	3	2	3	1	0	0	9								
2.0-	0	0	0	0	1	0	1	1	0	0	0	0	4								
1.8-	0	0	0	0	1	1	1	0	0	0	0	0	5								
1.6-	0	0	0	0	0	2	1	0	0	0	0	0	3								
1.4-	0	0	0	0	1	0	0	0	0	0	0	0	1								
1.2-	0	0	0	0	0	0	0	0	0	0	0	0	0								
1.0-	0	0	0	0	0	0	0	0	0	0	0	0	0								
0.8-	0	0	0	0	0	0	0	0	0	0	0	0	0								
0.6-	0	0	0	0	0	0	0	0	0	0	0	0	0								
0.4-	0	0	0	0	0	0	0	0	0	0	0	0	0								
0.2-	0	0	0	0	0	0	0	0	0	0	0	0	0								
Total	0	0	0	1	3	5	4	9	13	12	13	8	14	22	5	3	0	0	0	0	112

TABLE D5
Bivariate histogram from fitted SETAR(2; 8, 3), d = 2

X(N) (N-2)	0.2	0.6	1.0	1.4	1.8	2.2	2.6	3.0	3.4	3.8	4.2	4.6	5.0	Total
	-0.4	-0.8	-1.2	-1.6	-2.0	-2.4	-2.8	-3.2	-3.6	-4.0	-4.4	-4.8	5.0	
5.0-	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.8-	0	0	0	0	0	1	0	0	0	0	0	0	0	2
4.6-	0	0	0	0	1	0	1	0	0	1	0	0	0	5
4.4-	0	0	0	2	1	4	3	4	2	3	1	0	0	31
4.2-	0	1	0	5	9	8	15	11	20	10	15	6	3	114
4.0-	0	1	2	7	13	20	28	33	42	41	36	32	15	306
3.8-	0	3	2	9	19	32	34	50	78	66	68	49	28	609
3.6-	0	1	5	6	14	22	40	58	66	102	104	106	127	958
3.4-	0	1	7	8	14	17	32	52	62	80	104	162	159	1285
3.2-	1	0	1	5	14	28	34	44	65	117	142	212	242	1353
3.0-	0	0	2	4	10	11	23	43	46	70	113	142	226	1283
2.8-	0	0	5	9	16	31	38	60	77	126	190	218	182	1127
2.6-	0	1	3	6	10	23	26	46	79	142	162	131	122	853
2.4-	0	1	2	8	15	43	66	97	119	110	101	45	19	652
2.2-	0	3	7	3	5	14	28	59	70	100	85	75	51	529
2.0-	0	0	2	4	16	35	54	65	50	53	36	14	7	344
1.8-	0	1	5	12	23	33	39	40	36	22	19	10	4	245
1.6-	0	1	3	4	20	22	19	14	7	1	1	0	0	147
1.4-	0	0	2	6	10	12	15	11	8	9	2	2	0	79
1.2-	0	0	1	3	8	5	11	5	1	4	5	1	0	44
1.0-	0	1	3	2	3	5	0	0	1	0	0	0	0	15
0.8-	0	1	2	0	1	2	1	0	0	2	2	0	0	12
0.6-	0	0	1	1	1	1	0	0	1	0	0	0	0	4
0.4-	0	1	0	0	0	0	0	0	0	0	0	0	0	1
0.2-	1	4	12	15	44	79	147	245	344	529	652	853	1126	9998
Total	1	4	12	15	44	79	147	245	344	529	652	853	1126	9998

† Based on 10 000 point simulations.

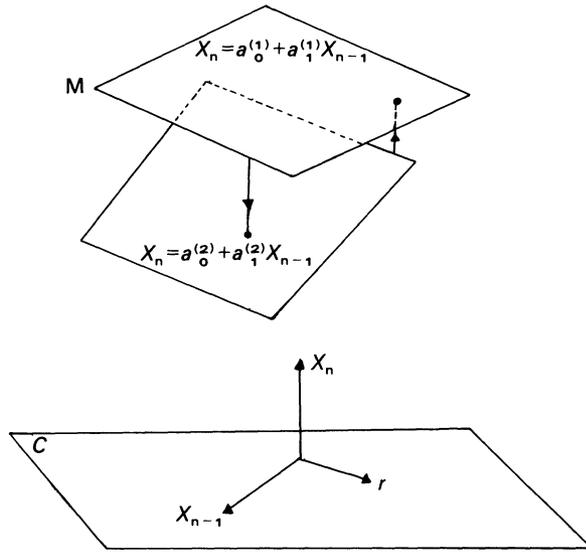


FIG. D1. Catastrophe paradigm for SETAR(2; 1, 1).

shape of the fitted bivariate distribution is certainly related to the concept of *perturbed limit cycles*. We have found simulation exercises with decreasing noise variance quite instructive. Not only would the connection with catastrophe theory *elevate TAR modelling above the ad hoc status*, it could also suggest ways of refining our models. In fact, for the lynx data, different threshold values may be used depending on whether the data are going up or down. This idea is also related to Dr Tunnicliffe-Wilson's suggestion of incorporating a "slope" in TAR modelling. As a result of some preliminary investigation, the following more refined SETAR is fitted to the first 100 lynx observations (logarithmically transformed):

$$X_n = \begin{cases} 0.5382 + 1.0602X_{n-1} - 0.2547X_{n-2} + 0.1598X_{n-3} - 0.3626X_{n-4} \\ \quad + 0.2100X_{n-5} - 0.2201X_{n-6} + 0.2753X_{n-7} - 0.0264X_{n-8} + \varepsilon_n^{(1)} \\ \quad \text{if } \{X_{n-1} - X_{n-2} \geq 0 \text{ and } X_{n-2} \leq 3.4\} \\ \quad \text{or } \{X_{n-1} - X_{n-2} \leq 0 \text{ and } X_{n-2} \leq 3.3\}, \\ 0.6354 + 1.6359X_{n-1} - 1.1985X_{n-2} + 0.3032X_{n-3} + \varepsilon_n^{(2)} \\ \quad \text{if } \{X_{n-1} - X_{n-2} \geq 0 \text{ and } X_{n-2} > 3.4\} \\ \quad \text{or } \{X_{n-1} - X_{n-2} \leq 0 \text{ and } X_{n-2} > 3.3\}, \end{cases} \quad (A1)$$

where $\text{var } \varepsilon_n^{(1)} = 0.0322$, $\text{var } \varepsilon_n^{(2)} = 0.0537$. (The pooled mean sum of squares of residuals = 0.0383.) This new model seems to give encouraging results when compared with the SETAR model reported in Tong (1980a). More details about this refined SETAR class will be reported by Lim and Tong elsewhere. The catastrophe paradigm has highlighted the significance of the *piecewise* aspect of our approach and relegated the actual choice of the class of submodels to a secondary position. Thus, as we have indicated in Tong (1980a), threshold polynomial AR models, threshold bilinear models and threshold exponential AR models, etc. are all ripe for exploitation. A connection between TAR modelling and catastrophe theory implies a connection between TAR models and non-linear vibrations, because the latter are often most elegantly explained in the language of catastrophe theory. Viewed in this light, our demonstrations in Section 6 are really quite trivial and natural, although we are indeed most delighted with the generally favourable reception of them. Mr Ozaki's approach to random

vibrations is different from ours. Firstly, in his references the notion of a *limit cycle in discrete time* has never been defined, all his discussion being confined to the *classical (deterministic) continuous time* case, typified by references to the Van der Pol equation. We must confess that our earlier discussion of limit cycles in terms of a predator-prey system (e.g. Tong, 1978) suffered from a similar defect. In Tong and Pemberton (1980), we have drawn attention to the possible pitfalls of this approach. We are also not clear in what way the so-called "instantaneous characteristic roots" of a non-linear AR model are relevant to the development of TAR models. Specifically, the following exponential AR(k) (EXPAR(k)) model has been proposed:

$$X_n = \sum_{j=1}^k a_j(X_{n-1}) X_{n-j} + \varepsilon_n \quad (\text{A2})$$

where ε_n is the usual white noise with variance σ^2 and

$$a_j(x) = \phi_j + \pi_j e^{-\gamma x^2} \quad (\gamma \geq 0). \quad (\text{A3})$$

Now, the so-called "instantaneous characteristic roots" of (A2) are just another way of looking at the "frequencies" of the spectral peaks of the "instantaneous spectral density function" for X_n

$$f_n(\omega) = \sigma^2 / [2\pi |1 - \sum_j a_j(X_{n-1}) e^{-ij\omega}|^2]. \quad (\text{A4})$$

These "frequencies" are functions of the "amplitude" X_{n-1}^2 (a random variable). Whilst we sympathize with the use of this kind of *physical consideration* to deliver a new class of non-linear time series models, we prefer not to treat these "frequencies" as the most dominant *physical notion*. For, if (A2) defines a stationary ergodic stochastic process, it has a *unique spectrum* and therefore *unique peak frequencies* (or a *unique proper frequency* when $k = 2$) for all n . Then, what are the amplitude-dependent frequencies? Incidentally, Atkinson and Caughey (1968a, b) have discussed continuous time first-order SETAR with emphasis on the spectrum. We would also prefer retaining our original terminology of SETAR to Mr Ozaki's own preference of "linear threshold AR".

Dr Tunnicliffe-Wilson and Mr Ozaki have raised the important question of stability. In fact, there are at least two types of stability, one systematic with the innovation absent and the other stochastic with the innovation present. Let us consider the former case now. In the linear case, local stability implies global stability. More specifically, if a linear model is stable over the dynamical range, say S , of the observations, then its extrapolation beyond S would not cause any problem. However, in the non-linear case, this is not and indeed cannot be expected to be so. There is simply no information contained in the data about the (non-linear) behaviour beyond S . Engineers have long recognized this point and the famous saturation system is a fine example. Our Definition 5.1 and Theorem 5.1 is a modest attempt to formalize this recognition for our purpose of model building and is useful not only for TAR models but also for EXPAR, bilinear (BL) and polynomial AR (PAR) models, etc. Naturally, the final product is necessarily a *threshold model* and whether it is a good model or not is a different matter. Using this artifact, we can also avoid what seems to us a rather unnatural distinction between a model based forecast and a non-model based forecast discussed by Dr Jones. For, the parameters involved in his forecast rule inevitably define a model (albeit not necessarily unique) whether we like it or not. Professor Cox's lynx model can be thus stabilized. Of course, if a model is already globally stable then there is no need for invoking the stabilization. Thus, Mr Ozaki has really missed the point here. We are also puzzled by the second part of his first comment because the *sole* SETAR model quoted in Ozaki's research memo (1979a) is at variance with the original SETAR model fitted by us, which is now in print (Tong, 1980a). Incidentally, we would just mention that the EXPAR model for the lynx data reported in Haggan and Ozaki (1980b, p. 67) appears to be explosive in our simulation studies *when the innovation is present*. The fact that some of the "characteristic roots" of the purely ϕ -operator are virtually on the unit circle might account for this stochastic instability. (Coefficients are correct to two decimal places as reported in their paper). Yet, our stabilization

could still have come to its rescue. It might be wise to re-examine the appropriateness of fitting a *symmetric* process, such as EXPAR (a fact which he has correctly recognized), to *asymmetric* data such as the lynx.

Many discussants have commented on our identification and fitting of SETAR models. The general view seems to be favourable with the significant exception of Professor Cox. It is a little amusing to see that he finds our use of AIC *mechanical* while, at the other extreme, Dr Bhansali has gone almost as far as labelling one of us a heretic! Perhaps Professor Cox's concern about the number of parameters might be alleviated if some of the intermediate terms are suppressed as in subset AR models. (See, for example, Tong, 1977a.) It is gratifying to note that all the three classes of non-linear time series models mentioned in Professor Priestley's contribution have relied on AIC for their *practical* identification. Of course, this does not imply that AIC is the only tool for this job. In fact, we would concur with Professor Cox and Dr Tunnicliffe-Wilson on the importance of graphical analysis and other classical time series operations. We do have some experience in the use of univariate and bivariate histograms, simulation studies, scatter diagrams and sample regression functions etc. as aids for our model identification and detecting non-linearity, some of which have been indicated in Tong (1980b) and more details will be reported by Lim and Tong elsewhere. We would even argue that, since a *non-linear* time series model is a transformation of the probability distributions of the (unobservable) input white noise process to those of the (observable) output process, it is not so much the residual sum of squares (RSS) or the number of parameters (p), but rather the general shape of the distributions (univariate, bivariate, trivariate, etc.) which is of paramount importance. In practice, we would caution against attaching too much significance to RSS and p as means of comparing *different classes* of models because of the almost inevitably different optimization algorithms and the often different number of effective observations. After all, the most important purpose of fitting a *non-linear* model is to gain a *better* understanding of the probabilistic structure underlying the data. At this point, we should appreciate the greater insight gained from moving away from *linear Gaussian* models. We are also interested in Dr Khabie-Zeitoune's suggestion of the feasibility of performing exact maximum likelihood computation using small memory storage.

Whilst Professor Priestley seems to be striving for greater flexibility, Dr Akaike seems to be urging us to go in quite a different direction. Rather than blurring the essential feature of our model, we feel that the introduction of the indicator variable J_n is similar to the facility provided by a modern vari-focal lens in photography (from where?) which enables us to focus on the more interesting aspects of reality. We should have made it clearer that the definition of J_n need not always be restricted to the few specific suggestions we have given. *It can be much more general.* Catastrophe consideration has suggested model (A1). The threshold model of neuron firing (see, for example, Brillinger and Segundo, 1979) suggests a J_n dependent on $\sum_{j=1}^p a_j X_{n-j}$ for some p and some a_j 's. In a private communication, Dr G. Gudmundsson has suggested a particularly interesting case for hydrological application in which the precipitation is alternately in the form of rain and snow. Professors Hipel and McLeod have also given us valuable suggestions for hydrological applications. The several engineering references quoted in our paper suggest a J_n which follows a Markov chain. In economic applications, Chien and Chan (1979) and Dr Bhansali's discussion might also lead to some useful suggestions. Professor Priestley's striving for greater flexibility is undoubtedly interesting and we look forward to seeing some real applications. We suspect that Professor Robinson's non-parametric regression is not unrelated to one aspect of Professor Priestley's approach. This type of exercise is more appealing to us than Dr Chatfield's suggestion of comparing the different forecasts, because we know only too well the kind of tangle the latter might lead to even in the linear case. However, to satisfy Dr Chatfield's expressed wish (and no doubt that of many others) we give Tables D6, D7 and D8, although we must caution against any general inference from them. For the purpose of discrimination, we think that it is also important to know more about the different types of distributions to which the different classes of non-linear models can give rise. Dr Jones' work (1978) would be very valuable here. A Ph.D. student at UMIST, Mr J. Pemberton, has obtained

TABLE D6
One-step-ahead predictions of log lynx (fitting period 1821-1920)

Year	Predicted values												
	Real Data	(1) SETAR	(2) AR(12)	(3) Moran	(4) Moran-BL	(5) C-W1	(6) C-W2	(7) C-W3	(8) Subset BL	(9) Cox	(10) EXPAR	(11) TPAR	(12) (A1)
1921	2-3598	2-3109	2-455	2-4504	2-5059	2-5511	2-4895	2-4262	2-442	3-3514			2-3415
1922	2-6010	2-8077	2-8099	2-8099	2-8369	2-8745	2-8768	2-9018	2-756	2-6910			2-8051
1923	3-0539	2-9106	2-899	2-8974	2-9589	2-9412	3-0180	3-0945	2-897	2-8823			3-0012
1924	3-3860	3-3703	3-231	3-3495	3-3003	3-1803	3-2790	3-2942	3-135	3-3626			3-4171
1025	3-5532	3-5875	3-388	3-4676	3-4578	3-2164	3-3053	3-2368	3-411	3-5282	3-4191		3-3691
1026	3-4676	3-4261	3-332	3-4465	3-4226	3-1788	3-2145	3-1633	3-512	3-4662	3-4074		3-5093
1927	3-1867	3-0936	3-007	3-1966	3-1907	3-0345	3-0009	3-0425	2-922	3-1365	2-7771		3-0762
1928	2-7235	2-7706	2-688	2-8666	2-8694	2-8719	2-7789	2-8519	2-706	2-8033	2-5928		2-7699
1929	2-6857	2-4217	2-428	2-4307	2-4715	2-6442	2-5343	2-5311	2-583	2-4470	2-4626		2-5724
1930	2-8209	2-7644	2-765	2-7357	2-6585	2-9063	2-8005	2-7258	2-844	2-7925	2-5572		2-7168
1931	3-0000	2-9397	2-984	2-9554	2-9336	3-1378	3-0968	3-0607	2-966	2-9959	2-8777		2-9222
1932	3-2014	3-2462	3-217	3-1036	3-0913	3-2538	3-3013	3-3571	3-159	3-1493	3-2096		3-2367
1933	3-4244	3-3701	3-365	3-2490	3-2217	3-2840	3-3999	3-4653	3-299	3-2985	3-2802		3-3765
1934	3-5310	3-4468	3-503	3-4077	3-3598	3-2851	3-4181	3-3966	3-415	3-4532	3-3080		3-5602
MSE		0-0144	0-018	0-0168	0-0204	0-0371	0-0232	0-0297	0-018	0-0093	0-0415	0-0395	0-0093
Var*	0-3287	0-3897	0-3632	0-3151					0-1932				0-2961

* Var = Var $X_{t,r}$. For the real data, it is based on the fitting period. Other entries are based on simulated samples of 10 000 points from the fitted models, with transients removed. The same convention is adopted in tables D7 and D8.

some results in this respect. Partly as an answer to Dr Chatfield's first question we would suggest that qualitative analyses such as bimodality, skewness, etc., are vital for deciding whether a linear Gaussian model is adequate or not. Incidentally, Rosenblatt (1979) might be relevant for Professor Robinson's examination of the estimated non-parametric regression as a test of linearity.

Dr Bhansali, Mr Godolphin and Professor Huzii are right in suggesting the need for a more thorough study of the sampling properties of the parameter estimates. We will report some results elsewhere. There can be no definite answer to Mr Godolphin's first question. For the lynx data the answer is affirmative and for the Kanna data the answer is negative. The number 4.6 in equation (9.5) merely indicates the state of very low rainfall. As for his second question, which is related to ones raised by Dr Chatfield and Dr Jolliffe, our answer is no, but we are currently

- (1) See Tong (1980a, p. 54).
- (2) See Tong (1977, p. 466).
- (3) See Tong (1977, p. 468) for reference of P. A. P. Moran's AR(2) with fitting period 1821-1934.
- (4) See Tong (1977, p. 454, under item Dr T. Subba Rao). Fitting period is 1821-1934.
- (5), (6) and (7). See M. J. Campbell and A. M. Walker (1977, especially equations (4.13) and (4.19) and p. 462). See also Tong (1980a). Fitting periods are all 1821-1934.
- (8) See Dr T. Subba Rao's contribution to the discussion of this paper.
- (9) See Tong (1977, p. 453, under item Professor D. R. Cox) with regression coefficients 0.345173, 1.099411, 0.120404, 0.116176, -0.383841 obtained by us.
- (10) See unpublished M.Sc. dissertation by Mr M. C. Wong (1980), University of Manchester, who gave the following parameter estimates:

i	1	2	3	4	5	
$\hat{\phi}_i$	1.097	-0.444	0.376	-0.330	0.257	
$\hat{\pi}_i$	-0.265	-0.151	-0.525	-0.091	-0.021	
i	6	7	8	9	10	11
$\hat{\phi}_i$	-0.247	0.245	-0.232	0.197	0.290	-0.341
$\hat{\pi}_i$	0.140	-0.601	0.293	0.151	-0.069	

$\exp\{-2.45X_{n-i}^2\}$ is associated with $\hat{\pi}_i, i = 1, 2, \dots, 10$. His fitting period is 1821-1924 and he has followed Haggan and Ozaki (1980b) for notation.

- (11) See above dissertation. He has followed Ozaki (1979a) for notation. $T = 1.02$.

i	1	2	3	4	5	
$\hat{\phi}_i$	0.822	-0.593	-0.188	-0.417	0.219	
$\hat{\pi}_i$	0.236	0.160	0.549	0.083	0.027	
i	6	7	8	9	10	11
$\hat{\phi}_i$	-0.083	-0.366	0.094	0.374	0.229	-0.329
$\hat{\pi}_i$	-0.130	0.556	-0.301	-0.176	0.050	

- (12) See equation (A1) of our reply.

working on his suggestion. Many-step-ahead predictions are non-trivial for the non-linear case. Whether a 10 per cent reduction in RMSE is considered substantial in the case of one-step-ahead predictions should presumably depend on the relative inadequacy of the linear models and the relative importance of the forecasts.

TABLE D7
One-step-ahead predictions of sunspot numbers (fitting period 1770–1869)

Year	Real data	Predicted values		
		(1) AR (2)	(2) BL	(3) SETAR
1870	139.0	92.664	94.240	95.308
1871	111.2	158.681	157.054	133.366
1872	101.6	71.551	109.128	79.982
1873	66.2	78.193	76.328	82.609
1874	44.7	34.765	39.099	44.418
1875	17.0	30.005	33.503	32.384
1876	11.3	6.249	7.262	6.884
1877	12.4	18.376	20.562	9.377
1878	3.4	24.110	23.818	31.721
1879	6.0	10.481	3.589	1.922
1880	32.3	20.765	20.968	18.998
1881	54.1	56.342	59.799	62.342
1882	59.7	68.466	69.462	61.432
1883	64.7	60.079	57.866	46.870
1884	63.5	61.832	63.699	53.870
1885	52.2	58.623	58.976	52.468
1886	25.4	42.667	41.630	39.907
1887	13.1	12.737	6.936	13.756
1888	6.8	14.800	17.236	8.754
1889	6.3	14.814	13.282	23.483
Var	1385.2	1541.5	2107.5	1275.4
MSE for 20 point predictions		346.6	293.4	267.6
MSE for 10 point predictions		622.6	507.5	422.1

Key to Table D7

(1) $X_n = 14.70 + 1.425X_{n-1} - 0.731X_{n-3} + \varepsilon_n$, $\text{var } \varepsilon_n = 228$. See Granger and Andersen (1978, p. 86).

(2) $X_n = 14.70 + 1.425X_{n-1} - 0.731X_{n-2} + \varepsilon_n$, where $\varepsilon_n = -0.0222\varepsilon_{n-2} - \eta_{n-1} + 0.202\varepsilon_{n-1} + \eta_n$, $\text{var } \eta_n = 197$. See Granger and Andersen (1978, p. 86).

(3)
$$X_n = \begin{cases} 5.2659 + 1.8891X_{n-1} - 1.5289X_{n-2} + 0.3039X_{n-3} \\ \quad + 0.3387X_{n-4} + \varepsilon_n^{(1)} & \text{if } X_{n-3} < 36.6, \\ 0.3900 + 1.1366X_{n-1} - 0.3645X_{n-2} + 0.0524X_{n-3} + \varepsilon_n^{(2)} & \text{if } X_{n-3} > 36.6, \end{cases}$$

where $\text{var } \varepsilon_n^{(1)} = 154.88$, $\text{var } \varepsilon_n^{(2)} = 94.00$ (pooled variance = 121.73).

As for computation time, we can only blame ourselves for giving such a detailed description of our identification procedure in Section 8, which has undoubtedly given Dr Chatfield the wrong impression. In fact, it has taken our CDC 7600 computer twelve seconds for the complete SETAR identification, as described in Section 8, of the lynx data. In a private communication, Dr Tunnicliffe-Wilson has indicated the feasibility of using GENSTAT for fitting TAR models which should make TAR modelling more readily available.

TABLE D8
 One-step-ahead predictions of sunspot numbers (fitting period from 1700 to 1920 for model (1) and from 1700 to 1945 for all others)

Year	Data	Predicted values				
		(1) SETAR	(2) BL(3,4)	(3) BL	(4) EXPAR	(5) TPAR
1921	26.1	29.182				
1922	14.2	10.236				
1923	5.8	5.294				
1924	16.7	15.728				
1925	44.3	39.363				
1926	63.0	69.650				
1927	69.0	72.138				
1928	77.8	75.358				
1929	64.9	67.323				
1930	35.7	55.433				
1931	21.2	23.508				
1932	11.1	16.952				
1933	5.7	22.958				
1934	8.7	17.707				
1935	36.1	24.472				
1936	79.7	64.559				
1937	114.4	106.468				
1938	109.6	121.579				
1939	88.8	86.124				
1940	67.8	67.788				
1941	47.5	42.865				
1942	30.6	25.737				
1943	16.3	15.686				
1944	9.6	10.921				
1945	32.2	22.277				
1946	92.6	64.050	61.099	77.9	59.4681	58.8084
1947	151.6	133.464	126.314	130.0	127.7654	127.8114
1948	136.3	169.018	135.483	149.8	123.5977	127.2716
1949	134.7	110.580	85.781	119.8	98.7405	96.2994
1950	83.9	97.849	64.032	86.2	97.8016	96.7381
1951	69.4	57.312	30.662	51.4	67.7415	68.0585
1952	31.5	33.665	67.742	38.9	63.0849	63.0153
1953	13.9	21.137	58.286	18.8	7.5928	9.3160
1954	4.4	2.767	37.504	3.3	5.9210	6.2002
1955	38.0	19.766	0.095	25.7	11.3457	11.4411
MSE*		148.205	1173.744	164.75	506.640	515.334

The prediction period 1921–1955 consists of 3 fairly representative cycles of different amplitudes.

- (1) See equation (9.2) of this paper. $\text{Var } X_n = 1340.3$ (c.f. 1168.9 of the observed.)
- (2) See Subba Rao (1979). $\text{Var } X_n = 1 \times 10^{58}$ (c.f. 1155.1 of the observed.)
- (3) See Dr T. Subba Rao's discussion of this paper. $\text{Var } X_n = 1059.2$ (c.f. 1155.1 of the observed.)
- (4) See note (10) of Table A3. $\hat{\gamma} = 0.000168$.

$$\begin{array}{l} \hat{\phi}_i \quad 0.789 \quad -0.170 \quad -0.053 \quad 0.166 \quad -0.034 \quad -0.078 \quad 0.113 \\ \hat{\pi}_i \quad 0.802 \quad -0.402 \quad -0.252 \quad -0.120 \quad -0.182 \quad 0.273 \end{array}$$

- (5) See note (11) of Table A3. $\hat{T} = 96.1$.

$$\begin{array}{l} \hat{\phi} \quad 1.717 \quad -0.655 \quad -0.318 \quad 0.010 \quad -0.246 \quad 0.240 \quad 0.109 \\ \hat{\pi}_i \quad -0.008 \quad 0.004 \quad 0.002 \quad 0.001 \quad 0.002 \quad -0.003 \end{array}$$

* A linear AR (10) fitted to 1700–1945 has MSE = 482.0.

Finally, we come to the analysis of the real data in Section 9. Dr Bulmer has clarified some of the doubts we did have regarding the lynx-hare hypothesis. However, he has not stated what model he would use for the lynx. Plainly he cannot retain his AR plus harmonic component model. Our current view is that empirical evidence seems to support a cusp-catastrophe model in which the amount of food in the present year is the one control parameter and the population density of recent years the other. Our SETAR model may then be regarded as a statistical expression of this cusp-catastrophe model. (See our discussion in the first paragraph and for more details see Tong, 1980b.) Dr Subba Rao has fitted bilinear models to the lynx data, for both the log transformed data and the original data. Our simulation studies suggest that BL models give *skew, unimodal* bivariate distributions. (Gaussian white noise is assumed throughout.) It seems clear that it is the linear AR part of the BL models which "explains" the cyclicity of the data; the bilinear terms probably account for the skewness of the probability distributions. We conjecture that the non-existence of limit cycles of BL models (see, for example, Brockett, 1977) implies that a BL process has, under general conditions, unimodal joint distributions. Despite these remarks, it is noteworthy that Dr Subba Rao has apparently succeeded in making BL modelling a practical proposition. His *subset* BL models represent an important step in this direction because a *full* BL model usually consists of too many parameters for efficient computation. Now, regarding his point about transformation, besides making the usual Gaussian assumption of the white noise more plausible, a logarithmic transformation might also have some stabilising effect. (See, for example, Rosenblatt, 1971, p. 164.) In fact, our simulation studies suggest that his BL model for the *original* lynx data tends to have a rather wide dynamical range, with a substantial proportion on the negative side extending beyond $-20\,000$. Our simulated sample of 10 000 data has a mean 1450 and a variance 1.4×10^{10} which may be compared with observed values of 1528 and 2.662×10^7 respectively.

Dr Bulmer seems to have overlooked the fact that our analysis of the mink-muskrat data is for the period of 1767-1849 and a *first differencing* operation is applied to both the log transformed mink data and the log transformed muskrat data. The observed limit cycle of period five years is probably due to the high-pass filtering property of a differencing operation. The following TARSC model is now fitted to the 1848-1909 data, which were used by Jenkins (1975) and Chan and Wallis (1978). (We had some difficulty in obtaining these data previously.) Let $X_n = \ln(\text{number of mink in year } 1847 + n)$, $Y_n = \Delta \ln(\text{number of muskrats in year } 1848 + n)$.

$$X_n = \begin{cases} 8.1624 + 0.3437X_{n-1} + 0.4510Y_{n-1} + 0.0696X_{n-2} - 0.0713Y_{n-2} \\ \quad - 0.4119X_{n-3} + 0.5353Y_{n-3} + 0.2228X_{n-4} + \eta_n^{(1)} \\ \quad \text{if } Y_{n-5} \leq -0.0443, \\ 5.4058 + 0.5266X_{n-1} + 0.4653Y_{n-1} + 0.3631X_{n-2} - 0.2820Y_{n-2} \\ \quad - 0.2207X_{n-3} + 0.2009Y_{n-3} - 0.1585X_{n-4} + \eta_n^{(2)} \\ \quad \text{if } Y_{n-5} > -0.0443, \end{cases}$$

where $\text{var}(\eta_n^{(1)}) = 0.0369$, $\text{var}(\eta_n^{(2)}) = 0.0234$ pooled variance = 0.0282),

$$Y_n = \begin{cases} 2.9045 - 0.0212Y_{n-1} - 0.6994X_{n-1} - 0.3047Y_{n-2} + 0.4254X_{n-2} \\ \quad + 0.0485Y_{n-3} + \varepsilon_n^{(1)} \quad \text{if } X_{n-5} \leq 10.9616, \\ 5.8527 + 0.3032Y_{n-1} - 0.5387X_{n-1} - 0.1289Y_{n-2} + \varepsilon_n^{(2)} \\ \quad \text{if } X_{n-5} > 10.9616, \end{cases}$$

where $\text{var}(\varepsilon_n^{(1)}) = 0.0385$, $\text{var}(\varepsilon_n^{(2)}) = 0.0841$ (pooled variance = 0.0589). This fitted model has a 10 year period limit cycle with six ascension years and four descension years for the mink, and four ascension years, three descension years and then two ascension years and one descension year for the muskrat.

Given the limited data length and noisiness of the data, our TARSC model seems reasonably successful. It also seems to be one of the very few real examples of bivariate non-linear time series models.

We must admit that we are a little disappointed with the results of all the non-linear time series models, including SETAR, BL, EXPAR and TPAR, which have been fitted to the sunspot numbers. The very large number of sunspot numbers near the minimum is the main source of difficulty. The other source of difficulty is the well-known inhomogeneity of the data. One feature has come to light during our simulation studies which concerns the full BL (3, 4) model reported in Subba Rao (1979). The AR operator there has one pair of complex roots in the unstable region (\hat{a}_3 should read -0.27).

REFERENCES IN THE DISCUSSION

- ATKINSON, J. D. and CAUGHEY, T. K. (1968a). Spectral density of piecewise linear first order systems excited by white noise. *Int. J. Non-linear Mech.*, **3**, 137–156.
- (1968b). First order piecewise linear systems with random parametric excitation. *Int. J. Non-linear Mech.*, **3**, 399–411.
- BEENSTOCK, M. and BHANSALI, R. J. (1980). Analysis of cocoa price series by autoregressive model fitting techniques. *J. Agric. Econ.*, **31**, 237–242.
- BRILLINGER, D. R. and SEGUNDO, J. P. (1979). Empirical examination of the threshold model of neutron firing. *Biol. Cybern.*, **35**, 213–220.
- BROCKETT, R. W. (1977). Convergence of Volterra series of infinite intervals and bilinear approximations. In *Non-linear Systems and Applications* (V. Lakshmikantham, ed.), pp. 39–46. New York: Academic Press.
- BULMER, M. G. (1978). The statistical analysis of the ten year cycle. In *Time Series and Ecological Processes* (H. H. Shugart, ed.), pp. 141–153. Philadelphia: SIAM. (SIAM-SIMS Conference No. 5.)
- CHIEN, M. J. and CHAN, L. (1979). Non-linear input–output model with piecewise affine coefficients. *J. Econ. Theory*, **21**, 389–410.
- COX, D. R. (1977). Discussion on papers by Campbell *et al.* *J. R. Statist. Soc. A*, **140**, 453–454.
- GODOLPHIN, E. J. (1975). A direct basic form for predictors of autoregressive integrated moving average processes. *Biometrika*, **62**, 483–496.
- GRANGER, C. W. J. (1978). New classes of time series models. *The Statistician*, **27**, 237–253.
- GRANGER, C. W. J. and ANDERSEN, A. P. (1978). *Introduction to Bilinear Time Series Models*. Gottingen: Vandenhoeck and Ruprecht.
- HAGGAN, V. and OZAKI, T. (1980b). Amplitude-dependent exponential AR model fitting for non-linear random vibrations. In *Time Series* (O. D. Anderson, ed.), pp. 57–71. Amsterdam: North Holland.
- HANNAN, E. J. (1970). *Multiple Time Series*. New York and London: Wiley.
- HIPPEL, K. W. and MCLEOD, A. I. (1978). Preservation of the rescaled adjusted range. 2, simulation studies using Box–Jenkins models. *Water Resources Research*, **14**(3), 509–516.
- HIPPEL, K. W. and MCLEOD, A. I. (1979). Perspectives in stochastic hydrology. In *Time Series* (O. D. Anderson, ed.), pp. 73–102. Amsterdam: North-Holland.
- KIBLER, D. F. and HIPPEL, K. W. (1979). Surface water hydrology. *Rev. Geophys. and Space Phys.*, **17**(6), 1186–1209.
- LABYS, W. C. and GRANGER, C. W. J. (1970). *Speculation, Hedging and Commodity Price Forecasts*. D. C. Heath and Co., Lexington, Mass.
- MCLEOD, A. I., HIPPEL, K. W. and LENNOX, W. C. (1977). Advances in Box–Jenkins modelling. 2. Applications. *Water Resources Research*, **13**(3), 577–586.
- NAKAMURA, H. and AKAIKE, H. (1979). Use of statistical identification for optimal control of a supercritical thermal power plant. In *Identification and System Parameter Estimation* (R. Isermann, ed.). Oxford: Pergamon.
- O'CONNELL, P. E. and JONES, D. A. (1979). Some experience with the development of models for the stochastic simulation of daily flows. In *Inputs for Risk Analysis in Water Systems* (E. A. McBean, K. W. Hipel and T. E. Unny, eds), pp. 287–314. Fort Collins, Colorado: Water Resources Publications.
- OTOMO, T., NAKAGAWA, T. and AKAIKE, H. (1972). Statistical approach to computer control of cement rotary kilns. *Automatica*, **8**, 35–48.
- OZAKI, T. and ODA, H. (1977). Non-linear time series model identifications by Akaike's Information Criterion. In *Information and Systems* (B. Dubuisson ed.). Oxford: Pergamon.
- OZAKI, T. (1979a). Non-linear threshold AR models for non-linear random vibrations, Research Memo. No. 157, Institute of Statistical Mathematics, Tokyo. (To appear in *J. of Appl. Prob.*)
- (1979b). Statistical analysis of Duffing process through non-linear time series models, Research Memo. No. 151, Institute of Statistical Mathematics, Tokyo. (To appear in *J. of Appl. Mechanics*.)
- (1979c). Statistical analysis of perturbed limit cycle processes through non-linear time series models, Research Memo. No. 158, Institute of Statistical Mathematics, Tokyo. (Also submitted for publication.)

- PRIESTLEY, M. B. (1978). Non-linear models in time series analysis. *The Statistician*, **27**, 159–176
- (1979). On a general class of non-linear time series models. *Bull. Int. Statist.*, **42**, to appear.
- ROSENBLATT, M. (1979). *Markov Processes: Structural and Asymptotic Behaviour*. Berlin: Springer-Verlag.
- (1977). Linearity and non-linearity in time series: prediction. *Bull. Int. Statist. Inst.*, **42**, to appear.
- STONE, C. J. (1977). Consistent non-parametric regression (with Discussion). *Ann. Statist.*, **5**, 595–645.
- SUBBA RAO, T. (1979). On the theory of bilinear time series models—II. Technical Report No. 121, Department of Mathematics, UMIST.
- SUBBA RAO, T. and GABR, M. M. (1981). A test for linearity of stationary time series. Submitted to *Appl. Statist.*
- TONG, H. (1980b). Catastrophe theory and threshold autoregressive modelling. Technical Report No. 125, Dept. of Mathematics, UMIST. (Abstract in *Resumé des Communications*, Journées de Statistique, Toulouse, 19–22 May 1980.)
- TONG, H. and PEMBERTON, J. (1980). On stability and limit cycles of non-linear autoregression in discrete time. *Cahiers du CERO, Bruxelles*, **22**, No. 2, 137–148.
- YAKOWITZ, S. J. (1973). A stochastic model for daily river flows in an arid region. *Water Resources Research*, **9**(5), 1271–1285.
- (1979). A non-parametric Markov model for daily river flow. *Water Resources Research*, **15**(5), 1035–1043.
- ZEEMAN, E. C. (1977). *Catastrophe Theory: Selected Papers 1927–1977*. Mass.: Addison-Wesley.