International Journal of Control
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713393989

Modelling and analysis of non-linear time series

S. Chen [a]; S. A. Billings [a]
[a] Department of Control Engineering, University of Sheffield, Sheffield, U.K

Online Publication Date: 01 December 1989
To cite this Article: Chen, S. and Billings, S. A. (1989) 'Modelling and analysis of non-linear time series', International Journal of Control, 50:6, 2151 - 2171
To link to this article: DOI: 10.1080/00207178908953491
URL: http://dx.doi.org/10.1080/00207178908953491

INT. J. CONTROL, 1989, VOL. 50, NO. 6, 2151–2171

Modelling and analysis of non-linear time series

S. CHEN† and S. A. BILLINGS†

The modelling of non-linear time series is reviewed and new results are introduced by employing some ideas from the identification of non-linear control systems. Both global and local conditions for stationarity and invertibility are established for the general non-linear time-series model and it is shown how these results provide a framework for time-series estimators. Methods of computing multistep-ahead predictions are studied and the usefulness of polynomial models is discussed.

1. Introduction

The theory and practice of linear time-series analysis is now a well-established field (see e.g. Box and Jenkins 1976). Most processes encountered in the real world, however, are non-linear to some extent, and in many practical applications non-linear models may be required in order to achieve an acceptable predictive accuracy. For over a decade time series analysis has moved towards the non-linear area and various non-linear time-series models, such as bilinear models (Granger and Andersen 1978 a), linear and non-linear threshold autoregressive models (Tong 1983, Ozaki 1981), exponential autoregressive models (Ozaki 1985), state-dependent models (Priestley 1980) and doubly stochastic models (Tjøstheim 1986), have been developed. Practical applications have shown that non-linear models cannot only provide a better fit to the data but can also reveal rich dynamic behaviour such as limit cycles and bifurcations, which cannot be captured by linear models.

The present study reviews existing non-linear time-series models. By adopting the results from the realization of stochastic control systems, a unified representation, namely the NARMA (Non-linear AutoRegressive Moving Average) model, is introduced. Various existing non-linear time-series models can then be viewed as particular parametrizations of the NARMA model. Threshold time-series models are discussed in a separate section because of their special structure.

Stationarity conditions for the general non-linear time series model are analysed next. Unlike linear models, which are either stationary or non-stationary, stationarity of a non-linear time series model can be a local property. The novel definitions of global and local stationarity are therefore introduced. As a dual to stationarity, general invertibility conditions are also investigated, and the new concepts of global and local model-invertibility or m-invertibility are proposed. It is shown how these new concepts provide important implications for the estimation of polynomial time-series models, and this provides a new perspective on previous results that questioned the usefulness of such models (Granger and Andersen 1978 a). The computation of multi-step ahead predictions and the application of some of the concepts to the annual sunspot numbers is also included.

Received 11 May 1989.

† Department of Control Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, U.K.

0020-7179/89 $3.00

2. The NARMA time-series model

It is well known that input–output descriptions that expand the current output in terms of past inputs and outputs provide models that represent a broad class of non-linear stochastic control systems, and a unified representation called the NARMAX model has been derived under some mild assumptions (Leontaritis and Billings 1985 a, b, Chen and Billings 1989 b). Adopting this approach for time-series modelling leads to the following general model:

$$y(t) = f(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e)) + e(t) \tag{1}$$

where $\{y(t)\}$ is a time series, $\{e(t)\}$ a strictly white-noise process and $f(\cdot)$ some non-linear function. The model (1) is about as far as one can go in terms of specifying a general finite-dimensional non-linear relationship (Priestley 1980) and can be referred to as the NARMA model because of its resemblance to the ARMA model

$$y(t) = a_0 + \sum_{i=1}^{n_y} a_i y(t-i) + \sum_{i=1}^{n_e} b_i e(t-i) + e(t) \tag{2}$$

The derivation of (1) can follow the same procedure as the NARMAX model by regarding $e(t)$ as the system 'input' (for details see Leontaritis and Billings 1985 a, b). An interpretation of (1) is that $y(t)$ is separated into two components. The part of $y(t)$ that can be predicted from the past is given by the conditional expectation

$$E[y(t) \mid y(t-1), y(t-2), \ldots] = f(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e)) \tag{3}$$

or equivalently

$$E[y(t) \mid e(t-1), e(t-2), \ldots] = f(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e)) \tag{4}$$

(recalling from (1) that $y(t-i)$ is a function of $e(t-1), e(t-2), \ldots$) and the unpredictable part $e(t)$ defined as the innovation. Before using (1) in non-linear time-series modelling and analysis, the form of the non-linear function $f(\cdot)$ must be given and, in doing so, various existing time-series models are obtained.

2.1. *The state-dependent model*

Priestley (1980) developed an interesting form for $f(\cdot)$ and called the resulting time series model a *state-dependent model.* If

$$\mathbf{x}_t = (e(t-n_e+1), \ldots, e(t), y(t-n_y+1), \ldots, y(t))^{\mathrm{T}} \tag{5}$$

is regarded as the 'state vector' at time t then the state-dependent model can be written as

$$y(t) = \mu(\mathbf{x}_{t-1}) + \sum_{i=1}^{n_y} \phi_i(\mathbf{x}_{t-1}) y(t-i) + \sum_{i=1}^{n_e} \psi_i(\mathbf{x}_{t-1}) e(t-i) + e(t) \tag{6}$$

Many existing time-series models can be interpreted as special cases of (6) by choosing particular forms for $\phi_i(\cdot)$ and $\psi_i(\cdot)$.

Linear model

By taking $\mu(\mathbf{x}_{t-1})$, $\phi(\mathbf{x}_{t-1})$ and $\psi_i(\mathbf{x}_{t-1})$ as constants for all i, (6) reduces to the ARMA model (2).

Bilinear model

The general bilinear model is given by (Granger and Andersen 1978 a, Subba Rao and Gabr 1984)

$$y(t) = a_0 + \sum_{i=1}^{n_y} a_i y(t-i) + \sum_{i=1}^{n_e} b_i e(t-i) + \sum_{i=1}^{n_e}\sum_{j=1}^{n_y} c_{ij} e(t-i)y(t-j) + e(t) \quad (7)$$

and this can be viewed as a special case of (6) by taking $\mu(\mathbf{x}_{t-1})$ and $\phi_i(\mathbf{x}_{t-1})$ as constants and putting

$$\psi_i(\mathbf{x}_{t-1}) = b_i + \sum_{j=1}^{n_y} c_{ij} y(t-j)$$

Exponential model

The exponential AR model of Ozaki (1985) takes the form

$$y(t) = \sum_{i=1}^{n_y} \{\alpha_i + \beta \exp[-y^2(t-1)]\} y(t-i) + e(t) \quad (8)$$

and this can be obtained by setting $\mu(\mathbf{x}_{t-1}) = 0$, $\psi_i(\mathbf{x}_{t-1}) = 0$ and

$$\phi_i(\mathbf{x}_{t-1}) = \alpha_i + \beta_i \exp[-y^2(t-1)]$$

In the present study the more general exponential ARMA model defined by

$$y(t) = \sum_{i=1}^{n_y} \{\alpha_i + \beta_i \exp[-y^2(t-k_i)]\} y(t-i) + \sum_{i=1}^{n_e} \{\gamma_i + \eta_i \exp[-y^2(t-m_i)]\} e(t-i) + e(t) \quad (9)$$

is introduced, which is again a special case of (6).

Other models such as the threshold AR model (Tong and Lim 1980) can also be interpreted in a similar manner. Threshold time-series models are discussed in § 3.

It can be seen that the state-dependent model provides a useful framework for non-linear time-series modelling. It does, however, create a problem of how to identify the functional form of $\mu(\mathbf{x}_{t-1})$, $\phi_i(\mathbf{x}_{t-1})$ and $\psi_i(\mathbf{x}_{t-1})$. Realizing that a linear approximation

$$\left.\begin{aligned} \phi_i(\mathbf{x}_t) &= \phi_i^{(0)} + \mathbf{x}_i^{\mathrm{T}} \mathbf{a}_i \\ \psi_i(\mathbf{x}_t) &= \psi_i^{(0)} + \mathbf{x}_i^{\mathrm{T}} \mathbf{b}_i \end{aligned}\right\} \quad (10)$$

does not have the generality required, Priestley (1980) suggested that $\mathbf{a}_i$ and $\mathbf{b}_i$ should each take the form of a random walk. While this endows the state-dependent model with considerable flexibility, it complicates the identification procedure (see e.g. Priestley 1980, Haggan *et al.* 1984). Notice that this interpretation of the state-dependent model brings it very close to the doubly stochastic time-series model of Tjøstheim (1986).

2.2. *The polynomial model*

The first-order approximation of (10) does not provide sufficient accuracy to approximate the general NARMA model (1). A straightforward solution would be to use a higher-order approximation, and this yields the polynomial ARMA model

$$y(t) = p(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e)) + e(t) \quad (11)$$

where $p(\cdot)$ is a polynomial function of degree l. It is obvious that the model (11) can be viewed as a particular parametrization of (1). While several authors have suggested this form of model as a natural extension of the ARMA model, Leontaritis and Billings (1985) provided a rigorous derivation to justify the use of this description.

While the polynomial model often produces higher predictive accuracy in practical applications, it has been criticized for being explosive (Granger and Andersen 1978 a, Ozaki 1985). It is clear that if $y(t)$ and $e(t)$ are unrestricted then the model (11) will almost certainly become explosive. This, however, does not imply that polynomial models are not very useful in modelling non-linear time series whose underlying process is stable and non-divergent. This aspect is further discussed in § 4.

2.3. *The rational model*

The rational difference-equation model was introduced by Sontag (1979) following a study of polynomial response maps for deterministic non-linear control systems, and an extension to stochastic control systems was given by Billings and Chen (1989). Consider initially $e(t)$ to be a physical input. If the response function $F_i(e(t), e(t-1), \ldots)$ of the underlying process is polynomial and has a finite-dimensional state-space realization then the underlying process satisfies a rational time-series model as a consequence of Sontag's results:

$$y(t) = \frac{p_1(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e))}{p_2(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e))} + e(t) \tag{12}$$

where $p_1(\cdot)$ and $p_2(\cdot)$ are polynomial functions of degrees l_1 and l_2 respectively. Again, the model (12) can be considered as another parametrization of (1).

Note that the conditions for the existence of such a rational model are very mild and the model (12) is not valid only at points where $p_2(\cdot) = 0$. It can immediately be seen that the polynomial time-series model (11) is included in (12). However, the rational time-series model is no longer linear in the parameters.

2.4. *The linear-in-the-parameter model*

A class of non-linear time-series models that are linear in the parameters are worthy of special attention because rich linear identification techniques can readily be applied to this class of models. A general linear in the parameters model takes the form

$$y(t) = \sum_{i=1}^{n_q} z_i(t)\theta_i + e(t) \tag{13}$$

where the θ_i are parameters and the $z_i(t)$ are 'regressors', which are some transformations of lagged $y(t)$ and $e(t)$:

$$z_i(t) = z_i(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e)) \tag{14}$$

For the ARMA model (2) the $z_i(t)$ are a constant and past values of $y(t)$ and $e(t)$. If the cross-products of $y(t-i)$ and $e(t-j)$ are further included, (13) represents the general bilinear model (7). If the $z_i(t)$ are chosen to be monomials of lagged $y(t)$ and $e(t)$, a polynomial model is obtained. For an exponential ARMA model some of the $z_i(t)$ take the form $\exp[-y^2(t-k_j)]\, y(t-j)$ and $\exp[-y^2(t-m_j)]\, e(t-j)$. The choice of 'regressors' is, however, not restricted to these cases. In other applications, for example, $z_i(t)$ may take the forms $\log[y^2(t-j)]$, $\tanh[e(t-j)]$ etc. In practice,

physical knowledge of the underlying process can often be used to help in selecting the forms for $z_i(t)$.

The model (13) is often referred to as a *linear* or *pseudolinear regression model* depending on whether the $z_i(t)$ include lagged $e(t)$. The flexibility in the choice of $z_i(t)$ allows a wide class of non-linear time series to be represented by (13), and unified computer software can therefore be developed to identify such models.

3. The threshold NARMA model

If a non-linear system displays significantly different dynamic characteristics in different regions, a single model may not adequately represent the process, and a series of models may therefore be required in order to capture the different dynamic behaviour of the system. Threshold models are particularly useful in such a situation and have been widely used in modelling non-linear random vibrations (Tong and Lim 1980, Ozaki 1981).

A general threshold non-linear model can be described as

$$\begin{aligned} y(t) = f^{(i)}(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e)) \\ + e(t) \quad \text{if } \mathbf{y}_{t-1} \in \mathbf{R}^{(i)}, \quad i = 1, \ldots, q \end{aligned} \tag{15}$$

where

$$\mathbf{y}_{t-1} = (y(t-1), \ldots, y(t-n_y))^{\mathrm{T}} \tag{16}$$

the $\mathbf{R}^{(i)}$ are given regions of n_y-dimensional euclidean space, and the $f^{(i)}(\cdot)$ are some non-linear functions. The model (15) may be referred to as the TNARMA (Threshold NARMA) model. Some special examples of (15) are now given.

The self-exciting threshold linear model of Tong (1983) can obviously be generalized to the non-linear case:

$$\begin{aligned} y(t) = f^{(i)}(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e)) \\ + e(t) \quad \text{if } r_{i-1} < y(t-d) \leqslant r_i, \quad i = 1, \ldots, q \end{aligned} \tag{17}$$

where the r_i are thresholds satisfying

$$-\infty = r_0 < r_1 < \ldots < r_{q-1} < r_q = +\infty \tag{18}$$

and d $(1 \leqslant d \leqslant n_y)$ is a physical time lag. The model (17) represents a class of TNARMA models that are important from the point of view of practical identification. If the $f^{(i)}(\cdot)$ are chosen to be polynomial functions then (17) includes the threshold non-linear model given by Lai and Hsieh (1988). If these polynomial functions are further independent of lagged $e(t)$ then (17) becomes the threshold non-linear model of Ozaki (1981). The present study will restrict attention to those $f^{(i)}(\cdot)$ that are linear in the parameters. Such TNARMA models can be described as

$$y(t) = \sum_{j=1}^{n_\theta^{(i)}} z_j^{(i)}(t)\theta_j^{(i)} + e(t) \quad \text{if } r_{i-1} < y(t-d) \leqslant r_i, \quad i = 1, \ldots, q \tag{19}$$

where the $z_j^{(i)}(t)$ are given non-linear functions as defined in (14).

4. Stationarity

Stationarity conditions for the linear model (2) are well known. If the roots of

$$A(\bar{z}) = 1 - \sum_{i=1}^{n_y} a_i \bar{z}^i = 0 \tag{20}$$

are all outside the unit circle in the $\tilde{z}$-plane then the stochastic process generated by (2) is stationary. The stationarity of a linear model is a 'global' property in the sense that the model is either stationary or non-stationary—the stationarity of the model does not depend on the statistical properties of $e(t)$ apart from $e(t)$ being stationary, and no amplitude restriction is required. This can be viewed from another angle. A linear system is either stable or divergent, and the stability does not depend on the initial conditions or the amplitude of input. On the other hand, the stability of a non-linear system is generally dependent on the initial conditions and input amplitudes. An investigation of stationarity for a general non-linear time-series model is a very complicated matter. Only in a few very simple cases can the conditions for the linear case be directly extended. Consider the simple rational model

$$y(t) = \frac{0{\cdot}99 + e^2(t-1)}{1 + 0{\cdot}1y^2(t-1) + e^2(t-1)} y(t-1) + e(t) \tag{21}$$

This model is stationary because

$$0 < \frac{0{\cdot}99 + e^2(t-1)}{1 + 0{\cdot}1y^2(t-1) + e^2(t-1)} < 1$$

The stationarity conditions for the first-order NAR model

$$y(t) = f(y(t-1)) + e(t) \tag{22}$$

have been investigated by Jones (1976). In particular, if $|f(y)| < |y|$, that is $f(\cdot)$ is a contraction mapping, the process is certainly stationary. For the exponential AR model

$$\begin{aligned} y(t) = {} & \{1{\cdot}95 + 0{\cdot}23 \exp[-y^2(t-1)]\} y(t-1) \\ & - \{0{\cdot}96 + 0{\cdot}24 \exp[-y^2(t-1)]\} y(t-2) + e(t) \end{aligned} \tag{23}$$

which Ozaki (1985) has shown is stationary, it is interesting to study the transient phase of the process. Assume that $e(t)$ has a small amplitude and $y(0) = y(-1) = 0$. The model starts in the non-stationary region and $y(t)$ is explosive until the amplitude of $y(t)$ becomes large enough. Then the process settles down to the stationary region. If $e(t)$ has a larger amplitude, the process will reach the stationary area more quickly.

In general, a non-linear process can be stationary when $y(t)$ and $e(t)$ are in a region, non-stationarity is introduced when they move to another area. In the following the results of Lin *et al.* (1989) are used to show that under certain conditions the non-linear time series generated by (1) is asymptotically stationary and converges in the mean-square sense to a stationary time series. The definitions of global and local stationarity are then introduced. First define $n = \max\{n_y, n_e\}$ and rewrite (1) as

$$y(t) = f(y(t-1), \ldots, y(t-n), e(t-1), \ldots, e(t-n)) + e(t) \tag{24}$$

Assumption 1

It is assumed that the function $f(\cdot)$ in (24) has the following properties.

(i) The non-linear equation

$$y = f(y, \ldots, y, 0, \ldots, 0) \tag{25}$$

has at least one solution. Denote a solution of (25) by $\bar{y}$.

(ii) For any $\rho > 0$ and $\sigma > 0$ the collection of mappings

$$\begin{aligned}
&\{f(\cdot, y_2, \ldots, y_n, e_1, \ldots, e_n) \mid y_2, \ldots, y_n \in \mathbf{S}_1(\bar{y}, \rho); e_1, \ldots, e_n \in \mathbf{S}_2(0, \sigma)\} \\
&\qquad \vdots \\
&\{f(y_1, \ldots, y_{n-1}, \cdot, e_1, \ldots, e_n) \mid y_1, \ldots, y_{n-1} \in \mathbf{S}_1(\bar{y}, \rho); e_1, \ldots, e_n \in \mathbf{S}_2(0, \sigma)\} \\
&\{f(y_1, \ldots, y_n, \cdot, e_2, \ldots, e_n) \mid y_1, \ldots, y_n \in \mathbf{S}_1(\bar{y}, \rho); e_2, \ldots, e_n \in \mathbf{S}_2(0, \sigma)\} \\
&\qquad \vdots \\
&\{f(y_1, \ldots, y_n, e_1, \ldots, e_{n-1}, \cdot) \mid y_1, \ldots, y_n \in \mathbf{S}_1(\bar{y}, \rho); e_1, \ldots, e_{n-1} \in \mathbf{S}_2(0, \sigma)\}
\end{aligned} \tag{26}$$

are equi-Lipschitz-continuous in a mean-square metric on $\mathbf{S}_1(\bar{y}, \rho)$ and $\mathbf{S}_2(0, \sigma)$ with Lipschitz constants $a_1(\rho, \sigma), \ldots, a_n(\rho, \sigma)$ and $b_1(\rho, \sigma), \ldots, b_n(\rho, \sigma)$ respectively, where

$$\mathbf{S}_1(\bar{y}, \rho) = \{y \mid \|y - \bar{y}\| \leqslant \rho\}, \quad \mathbf{S}_2(0, \sigma) = \{e \mid \|e\| \leqslant \sigma\} \tag{27}$$

The norm $\|\cdot\|$ is a mean-square norm, and $y_1, \ldots, y_n \in \mathbf{S}_1(\bar{y}, \rho)$ means

$$y_1 \in \mathbf{S}_1(\bar{y}, \rho), \ldots, y_n \in \mathbf{S}_1(\bar{y}, \rho)$$

Definition 1

For any t and $N \geqslant -t$ define the series

$$y_N(t) = F_N^t(y(-N-1), \ldots, y(-N-n), e(t-1), \ldots, e(-N-n)) + e(t) \tag{28}$$

where

$$\begin{aligned}
&F_N^t(y(-N-1), \ldots, y(-N-n), e(t-1), \ldots, e(-N-n)) \\
&\quad = F_{N-1}^t(f(y(-N-1), \ldots, y(-N-n), e(-N-1), \ldots, e(-N-n)) + e(-N), \\
&\qquad y(-N-1), \ldots, y(-N-n+1), e(t-1), \ldots, e(-N-n+1))
\end{aligned} \tag{29}$$

with

$$y(-N-1) = \ldots = y(-N-n) = \bar{y} \tag{30}$$

and

$$\begin{aligned}
&F_{-t}^t(y(t-1), \ldots, y(t-n), e(t-1), \ldots, e(t-n)) \\
&\qquad = f(y(t-1), \ldots, y(t-n), e(t-1), \ldots, e(t-n))
\end{aligned} \tag{31}$$

Note that $y_N(t)$ is simply the time series $y(t)$ defined in (24) with the specific initial condition (30).

Theorem 1

Let Assumption 1 hold and assume that for any $\rho > 0$ and $\sigma > 0$ there exist $0 < \rho_1 \leqslant \rho$ and $0 < \sigma_1 \leqslant \sigma$ such that $\mu(\rho_1, \sigma_1) < 1$, where $\mu(\rho_1, \sigma_1)$ is the maximum eigenvalue of the matrix $(\mathbf{A}^T\mathbf{A})^{1/2}$ with

$$\mathbf{A} = \left[\begin{array}{c|ccc} 0 & & \mathbf{I} & \\ \hline a_n(\rho_1, \sigma_1) & a_{n-1}(\rho_1, \sigma_1) & \ldots & a_1(\rho_1, \sigma_1) \end{array}\right] \tag{32}$$

Then there exist $0 < \rho_2 \leqslant \rho_1$ and $0 < \sigma_2 \leqslant \sigma_1$ such that

(i) the non-linear time series $y(t)$ defined in (24), with initial conditions $y(-N-1), \ldots, y(-N-n) \in \mathbf{S}_1(\bar{y}, \rho_2)$ and restriction $e(t), \ldots, e(-N-n) \in \mathbf{S}_2(0, \sigma_2)$, is bounded in a mean-square metric; that is, $y(t) \in \mathbf{S}_1(\bar{y}, \rho_1)$, $t \geqslant -N$;

(ii) with the restriction $e(t), \ldots, e(N-n) \in \mathbf{S}_2(0, \sigma_2)$, for any t and $N \geqslant -t$, the sequence $\{y_N(t)\}_{N=-t}^{\infty}$ is a Cauchy sequence and the time series

$$y^*(t) = y^*(e(t), e(t-1), \ldots) = \lim_{N \to \infty} y_N(t) \tag{33}$$

is well defined;

(iii) the time series $y^*(t)$ defined in (33) is stationary;

(iv) the time series $y(t)$ defined in (24) with $y(0), \ldots, y(-n+1) \in \mathbf{S}_1(\bar{y}, \rho_2)$ and $e(t), \ldots, e(-n+1) \in \mathbf{S}_2(0, \sigma_2)$ converges in the mean-square sense to $y^*(t)$.

Proof

See Lin *et al.* (1989). This theorem is a simplified version of Theorem 1 in that paper.

Assertion (i) of Theorem 1 may be interpreted as meaning that 'bounded input' leads to 'bounded output' in the mean-square sense. Note that in assertion (iii) of Theorem 1 the stationarity of the noise process $e(t)$ is implicitly assumed (in fact, $e(t)$ is assumed to be independently identically distributed). The key requirement in Theorem 1 is that

$$\mu(\rho_1, \sigma_1) < 1 \tag{34}$$

Given $\rho_2 > 0$ and $\sigma_2 > 0$, Theorem 1 states that the time series $y(t)$ defined in (24) is stationary if initial series values belong to $\mathbf{S}_1(\bar{y}, \rho_2)$, the noise process $e(t)$ is bounded in $\mathbf{S}_2(0, \sigma_2)$, and there exist $\rho_1 \geqslant \rho_2$ and $\sigma_1 \geqslant \sigma_2$ such that (34) holds. It can immediately be seen that, given a non-linear function $f(\cdot)$, ρ_2 and σ_2 may not be arbitrarily chosen if the requirement (34) is to be satisfied. This suggests the following definition.

Definition 2

If ρ_2 and σ_2 in Theorem 1 can be arbitrarily large then the time series defined in (24) is globally stationary; otherwise, it is locally stationary on $\mathbf{S}_1(\bar{y}, \rho_2)$ and $\mathbf{S}_2(0, \sigma_2)$.

The reason for the choice of the words 'global' and 'local' is as follows. The process generated by a globally stationary model is always stationary for all kinds of distributed $e(t)$ as long as $e(t)$ is stationary. That is, the stationarity is globally valid on the normed space of all the stationary distributions. A locally stationary model, on the other hand, can only generate a stationary process on a region of this normed space, that is on $\mathbf{S}_2(0, \sigma_2)$. As an example of global stationarity, consider the exponential ARMA model (9). If the roots of

$$\tilde{A}(\tilde{z}) = 1 - \sum_{i=1}^{n_y} \alpha_i \tilde{z}^i = 0$$

all lie outside the unit circle in the $\tilde{z}$-plane then the time series defined by (9) is globally stationary (see Ozaki 1985).

4.1. *Local stationarity of polynomial models*

Local stationarity is not a rare phenomenon among non-linear time-series models, and a typical example is the polynomial model (11). For models that are linear in lagged $y(t)$, conditions for local stationarity can easily be derived. For example, consider

$$y(t) = \alpha e^2(t-1)y(t-1) + e(t), \quad \alpha \neq 0$$

Because

$$\|y(t)\| \leqslant |\alpha| \, \|e^2(t-1)\| \, \|y(t-1)\| + \|e(t)\|$$

this model is stationary if

$$|\alpha| \, \|e^2(t-1)\| < 1$$

However,

$$\|e^2(t-1)\| \geqslant \|e(t-1)\| \, \|e(t-1)\|$$

if the mean-square norm is chosen as $\|e\| = (E[e^2])^{1/2}$, $\|e(t-1)\| = \sigma_e$, where σ_e^2 is the variance of $e(t)$. The model is therefore stationary if

$$\sigma_e^2 < |\alpha|^{-1}$$

It is seen that σ_2 can be chosen as

$$0 < \sigma_2 < |\alpha|^{-1/2}$$

The choice of ρ_2 is arbitrary. Note that no restriction on the amplitude $|e(t)|$ is imposed in this case. For more general polynomial models that are non-linear in lagged $y(t)$, however, additional amplitude restrictions on $e(t)$ and initial series values may be required in order for such models to be locally stationary. This implies that the distribution of $e(t)$ should be truncated, otherwise there might be no ρ_2 and σ_2 such that (34) holds.

Using polynomial models in simulation with gaussian noise excitation $e(t)$ can therefore often lead to explosive behaviour. This is because if the noise is really gaussian-distributed, it may occasionally have a large amplitude value resulting from the tails of the gaussian distribution (although the possibility of $e(t)$ taking such a value is very small). When this occurs, the model output may jump out of the stability region, causing divergence of the future model-output values. A simple illustration is

$$y(t) = \alpha y^2(t-1)y(t-1) + e(t), \quad 0 < \alpha < 1$$

Assume that at time $t = k$ the realization of $e(k)$ takes a very large value, resulting in $\alpha y^2(k) \gg 1$. From $t > k$, $y(t)$ becomes divergent unless at some future time the net result of $\alpha y^3(t-1) + e(t)$ happens to become small again.

Limiting the noise amplitude is not a mild restriction. For example, it rules out the gaussian noise process. The gaussian assumption is, however, an idealization of many real situations, and for most real systems the noise may well have limited amplitude values. It is of course an advantage that a model is globally stationary or locally stationary without the restriction on noise amplitude. Unfortunately the general polynomial model does not belong to this class.

4.2. *Usefulness of polynomial models*

The explosive nature is a disadvantage of the polynomial model. This does not mean, however, that the polynomial model has little value in modelling non-linear time series when the underlying process is stable and non-divergent. In practice, for stable and non-divergent time series, polynomial models often fit the observed time-series values better and produce more accurate predictions than linear models.

The better fit to data can be explained by non-linear approximation theory. Most processes encountered in the real world are non-linear and can generally be represented by the NARMA model (1). A linear model is only a first-order approximation of $f(\cdot)$ that is valid in a small region around a choice of $\bar{y}$. If the non-linearity is severe, a linear approximation may be totally inadequate. It is also known from systems theory that the class of bilinear models is not sufficient to approximate all non-linear functions $f(\cdot)$ in a region around $\bar{y}$ (see e.g. Fliess and Normand-Cyrot 1982). The set of polynomial models is, however, sufficient to approximate (1) within an arbitrary accuracy in a region around $\bar{y}$ (see e.g. Chen and Billings 1989 b). Given an accuracy requirement, the region in which a polynomial model holds is usually much larger than that for a linear model.

Using a fitted polynomial model to simulate the underlying process may often result in divergent behaviour. It is, however, quite another matter to compute h-step-ahead predictions using a polynomial model. This is because in the latter case observed past time-series values are used in the calculation. For a stationary non-linear time series generated from an underlying process within a stable region, provided that h is not too large, the prediction calculated based on observed time-series values should remain stable. As an illustration, consider the stable first-order NAR process of (22), where the function $f(\cdot)$ for this model is shown in Fig. 1. It is expected that a linear model

$$y(t) = \alpha y(t-1) + e(t), \quad |\alpha| < 1 \tag{35}$$

will provide a poor fit to the data generated from such an underlying process, and a polynomial model

$$y(t) = p(y(t-1)) + e(t) \tag{36}$$

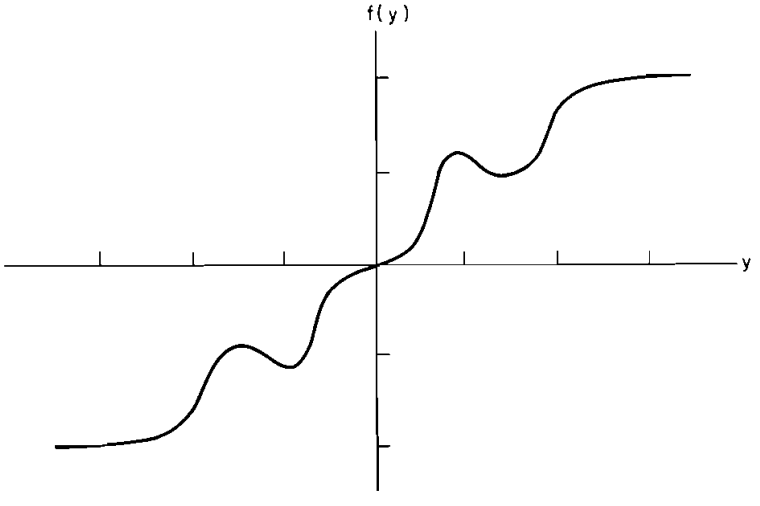


Figure 1. A non-linear function.

will produce a much better fit to the data. Simulation using these two models under gaussian noise assumptions will produce a stationary model output from the linear model (35) and almost certainly a divergent model output from the polynomial model (36). This does not imply, however, that this polynomial AR model cannot be used to model the given underlying process and to predict future time-series values. In fact, it fits the data better, and the one-step- and two-step-ahead predictions based on (36),

$$\left.\begin{aligned} y(t+1\,|\,t) &= p(y(t)) \\ y(t+2\,|\,t) &= E[p(y(t+1))\,|\,t] = E[p(p(y(t)) + e(t+1))\,|\,t] \end{aligned}\right\} \tag{37}$$

(see § 5.2) are certainly stable. In general, provided that h is not too large, the h-step-ahead prediction based on (36) will be stable. Since (36) is a better approximation to the given underlying non-linear process than (35), it is obvious that predictions based on (36) will be more accurate.

Two of the most important criteria for non-linear time-series models are

(i) parsimonious structure so that statistical identification entails no excessive computation;

(ii) multistep-ahead predictions should be easily obtained from the fitted model and the overall predictive performance should be an improvement upon the linear model.

The polynomial time-series model, or at least the polynomial AR model, is seen to satisfy both these requirements. The linear-in-the-parameters structure of such a model enables the direct application of many linear identification techniques. Modelling non-linear time series using polynomial models is therefore practical and useful. Of course the explosive nature of the polynomial model should be carefully monitored. A possible way to alleviate this problem is to employ a threshold structure with both non-linear and linear model sets. For example, for the above first-order NAR process an even better model is

$$y(t) = \begin{cases} p(y(t-1)) + e(t) & \text{if } |y(t-1)| \leqslant r \\ a_0 + e(t), & \text{if } |y(t-1)| > r \end{cases} \tag{38}$$

A more general example of a threshold polynomial model is

$$y(t) = \begin{cases} p(y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e)) + e(t) & \text{if } |y(t-d)| \leqslant r \\ a_0 + \sum_{i=1}^{n_y} a_i y(t-i) + \sum_{i=1}^{n_e} b_i e(t-i) + e(t) & \text{if } |y(t-d)| > r \end{cases} \tag{39}$$

which can be globally stationary.

5. Prediction and m-invertibility

One of the most important uses for time-series models is to provide forecasts or predictions. In order to use the general non-linear time-series model (1) for prediction, it is necessary to estimate the unobserved system noise sequence $\{e(t)\}$ based on the observed time-series values $\{y(t)\}$ and the given model. This is also crucial in the identification of time-series models with a non-linear MA part. Granger and Andersen (1978 b) introduced a generalized definition of invertibility as follows. Assume that the system is modelled exactly by (1) and that all $y(t)$ are known. Let a prediction error or

residual sequence $\{\varepsilon(t)\}$ be generated from

$$\varepsilon(t) = y(t) - f(y(t-1), \ldots, y(t-n_y), \varepsilon(t-1), \ldots, \varepsilon(t-n_e)) \tag{40}$$

with an arbitrary initial condition

$$\varepsilon(t) = \tilde{\varepsilon}_p, \quad t = -n_e + 1, \ldots, 1, 0 \tag{41}$$

Then the model (1) is said to be invertible if

$$E[(e(t) - \varepsilon(t))^2] \to 0 \quad \text{as } t \to \infty \ . \tag{42}$$

A slightly modified version was given by Hallin (1980). The above definition of invertibility for time series is not very suitable for practical situations because the system is often much more complex than the models that are estimated. A more realistic definition of invertibility called *model-invertibility* or *m-invertibility* has been introduced to deal with this aspect (Chen and Billings 1989 a).

Let us adopt the non-linear-filtering point of view and regard the estimation of $\{\varepsilon(t)\}$ as an inverse problem where the data-generating mechanism is the given model, $y(t)$ becomes the system 'input' and $\varepsilon(t)$ the system 'output'. Note that there is no requirement for the underlying process producing $y(t)$ to be modelled exactly by the given model. In fact, the model may not be anything related to the generation of $y(t)$. An inverse theorem of Theorem 1 is given as follows.

Assumption 2

It is assumed that the function $f(\cdot)$ in (24) has the following properties:

(i) the non-linear equation

$$\varepsilon = -f(0, \ldots, 0, \varepsilon, \ldots, \varepsilon) \tag{43}$$

has at least one solution; we denote a solution of (43) by $\bar{\varepsilon}$;

(ii) if $\mathbf{S}_1(\bar{y}, \rho)$ and $\mathbf{S}_2(0, \sigma)$ in Assumption 1 are replaced by $\mathbf{S}_1(0, \rho)$ and $\mathbf{S}_2(\bar{\varepsilon}, \sigma)$ respectively then the second condition (ii) in Assumption 1 holds.

Definition 3

For any t and $N \geqslant -t$, similarly to Definition 1, define a series $\varepsilon_N(t)$ as the series generated by

$$\varepsilon(t) = y(t) - f(y(t-1), \ldots, y(t-n), \varepsilon(t-1), \ldots, \varepsilon(t-n)) \tag{44}$$

with the specific initial condition

$$\varepsilon(-N-1) = \ldots = \varepsilon(-N-n) = \bar{\varepsilon} \tag{45}$$

Theorem 2

Let Assumption 2 hold and assume that for any $\rho > 0$ and $\sigma > 0$ there exist $0 < \rho_1 \leqslant \rho$ and $0 < \sigma_1 \leqslant \sigma$ such that $\bar{\mu}(\rho_1, \sigma_1) < 1$, where $\bar{\mu}(\rho_1, \sigma_1)$ is the maximum eigenvalue of the matrix $(\mathbf{B}^T\mathbf{B})^{1/2}$ with

$$\mathbf{B} = \left[\begin{array}{c|c} 0 & \mathbf{I} \\ \hline b_n(\rho_1, \sigma_1) & b_{n-1}(\rho_1, \sigma_1) \quad \ldots \quad b_1(\rho_1, \sigma_1) \end{array}\right] \tag{46}$$

Then there exist $0 < \rho_2 \leqslant \rho_1$ and $0 < \sigma_2 \leqslant \sigma_1$ such that

(i) the inverse non-linear time series $\varepsilon(t)$ defined in (44), with initial conditions $\varepsilon(-N-1), \ldots, \varepsilon(-N-n) \in \mathbf{S}_2(\bar{\varepsilon}, \sigma_2)$ and restriction $y(t), \ldots, y(-N-n) \in \mathbf{S}_1(0, \rho_2)$, is bounded in a mean-square metric; that is, $\varepsilon(t) \in \mathbf{S}_2(\bar{\varepsilon}, \sigma_1)$, $t \geqslant -N$;

(ii) with the restriction $y(t), \ldots, y(-N-n) \in \mathbf{S}_1(0, \rho_2)$ for any t and $N \geqslant -t$, the sequence $\{\varepsilon_N(t)\}_{N=-t}^{\infty}$ is a Cauchy sequence and the time series

$$\varepsilon_*(t) = \varepsilon_*(y(t), y(t-1), \ldots) = \lim_{N \to \infty} \varepsilon_N(t) \tag{47}$$

is well defined;

(iii) if $y(t)$ is stationary then the time series $\varepsilon_*(t)$ defined in (47) is stationary;

(iv) the inverse time series $\varepsilon(t)$ defined in (44) with $\varepsilon(0), \ldots, \varepsilon(-n+1) \in \mathbf{S}_2(\bar{\varepsilon}, \sigma_2)$ and $y(t), \ldots, y(-n+1) \in \mathbf{S}_1(0, \rho_2)$ converges in the mean-square sense to $\varepsilon_*(t)$.

Proof

This theorem is a dual to Theorem 1.

The key requirement in Theorem 2 is that

$$\bar{\mu}(\rho_1, \sigma_1) < 1 \tag{48}$$

Given $\rho_2 > 0$ and $\sigma_2 > 0$, for any two sequences $\{\varepsilon^{[i]}(t)\}$ $i = 1, 2$ with two different initial conditions satisfying

$$\varepsilon^{[i]}(t) \in \mathbf{S}_2(\bar{\varepsilon}, \sigma_2), \quad t = -n+1, \ldots, 0, \quad i = 1, 2 \tag{49}$$

if the process $y(t)$ is bounded in $\mathbf{S}_1(0, \rho_2)$ and there exist $\rho_1 \geqslant \rho_2$ and $\sigma_1 \geqslant \sigma_2$ such that (48) holds then Theorem 2 ensures that $\varepsilon^{[i]}(t)$, $i = 1, 2$, are bounded in $\mathbf{S}_2(\bar{\varepsilon}, \sigma_1)$ and they both converge in the mean-square sense to the same process. As a consequence,

$$E[(\varepsilon^{[1]}(t) - \varepsilon^{[2]}(t))^2] \to 0 \quad \text{as } t \to \infty \tag{50}$$

This becomes the definition of m-invertibility given by Chen and Billings (1989 a). In particular, if it is further assumed that the underlying process is exactly described by the given model then (50) is equivalent to (42). That is, the estimated sequence $\varepsilon(t)$ converges in the mean-square sense to the unobserved noise process $e(t)$. The m-invertibility condition in this case coincides with the definition given by Granger and Andersen (1978 b). As in the case of stationarity, the definitions of global and local m-invertibility are now introduced.

Definition 4

If σ_2 and ρ_2 in Theorem 2 can be arbitrarily large, the model (24) is called *globally m-invertible*; otherwise it is locally m-invertible on $\mathbf{S}_2(\bar{\varepsilon}, \sigma_2)$ and $\mathbf{S}_1(0, \rho_2)$.

The interpretations of the words 'global' and 'local' are similar to those in the case of stationarity. $\varepsilon(t)$ can always be computed from a globally invertible model regardless of the statistical properties of $y(t)$, while the ability to compute $\varepsilon(t)$ from a

locally invertible model depends upon the statistical properties of $y(t)$. An example of a globally invertible model is the ARMA model (2) with the roots of

$$B(\tilde{z}) = 1 + \sum_{i=1}^{n_e} b_i \tilde{z}^i = 0$$

all lying outide the unit circle in the $\tilde{z}$-plane. Invertibility (corresponding to local m-invertibility in Definition 4) conditions for certain types of bilinear time-series models have been investigated previously (see e.g. Granger and Andersen 1978 b, Quinn 1982, Subba Rao and Gabr 1984), and these conditions can be extended to similar types of polynomial models that are linear in lagged $\varepsilon(t)$:

$$y(t) = p_0(y(t-1), \ldots, y(t-n_y)) + \sum_{i=1}^{n_e} p_i(y(t-1), \ldots, y(t-n_y))\varepsilon(t-i) + \varepsilon(t) \tag{51}$$

where $p_i(\cdot)$, $i = 0, 1, \ldots, n_e$ are polynomials of finite degree (Chen and Billings 1989 a, b). As a simple example, consider the model

$$y(t) = \alpha y^2(t-1)\varepsilon(t-1) + \varepsilon(t), \quad \alpha \neq 0$$

This model is invertible if

$$\alpha^2 \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E[y^4(t)] < 1 \tag{52}$$

(see Chen and Billings 1989 a). If $y(t)$ is stationary, (52) becomes

$$\alpha^2 E[y^4(t)] < 1 \tag{53}$$

and the model in this case is locally m-invertible on $\mathbf{S}_2(0, \sigma_2)$ and $\mathbf{S}_1(0, \rho_2)$ with

$$\left.\begin{aligned} 0 < \rho_2 < |\alpha|^{-1/2} \\ 0 < \sigma_2 < \infty \end{aligned}\right\} \tag{54}$$

where the norm is chosen as $\|y\| = (E[y^2])^{1/2}$.

For the polynomial model that is non-linear in lagged $\varepsilon(t)$, if the amplitude of $y(t)$ and the initial values for $\varepsilon(t)$ are not limited, ρ_2 and σ_2 in general do not exist and the model is not even locally invertible. For example, consider

$$y(t) = \alpha\varepsilon^3(t-1) + \varepsilon(t), \quad 0 < \alpha < 1 \tag{55}$$

If $|y(t)|$ is not small enough, this model is not invertible because the generation of $\varepsilon(t)$ is explosive. The invertibility of the general polynomial ARMA model (11) therefore depends upon particular realizations of $y(t)$. For some observed time-series sequences, $\{\varepsilon(t)\}$ may be computed using models like (55), and in other cases the generation of $\varepsilon(t)$ using models like (55) is divergent.

5.1. *Invertibility and identification of non-linear MA models*

M-invertibility plays an important role in identifying models with a non-linear MA part (Chen and Billings 1989 a). Let the model (1) be parametrized with a parameter vector θ of dimension n_θ:

$$y(t) = f(y(t-1), \ldots, y(t-n_y), \varepsilon(t-1, \theta), \ldots, \varepsilon(t-n_e, \theta); \theta) + \varepsilon(t, \theta) \tag{56}$$

where $\theta \in D_M$ with D_M a subset of n_θ-dimensional Euclidean space. Identification consists of selecting a model (a parameter vector θ) within D_M that best describes the

recorded data. It is obvious that, in order to proceed with the identification, there must exist a subset D_{M_s} of D_M, for each $\theta \in D_{M_s}$ so that $\varepsilon(t, \theta)$ can be calculated using (56) with given $\{y(t)\}$. That is, there must exist a non-empty D_{M_s} such that the model (56) is at least locally m-invertible on some $\mathbf{S}_2(\bar{\varepsilon}, \sigma_2)$ and $\mathbf{S}_1(0, \rho_2)$. It has previously been shown that the convergence of recursive identification algorithms depends crucially upon m-invertibility (Chen and Billings 1989 a).

For polynomial models that are linear in lagged $\varepsilon(t, \theta)$, such D_{M_s}, $\mathbf{S}_1(0, \rho_2)$ and $\mathbf{S}_2(\bar{\varepsilon}, \sigma_2)$ exist, although their explicit expressions are usually difficult to obtain. For polynomial models with terms that are non-linear in lagged $\varepsilon(t, \theta)$, the situation is much more complex. A general observation is that, unless the amplitude of $y(t)$ is restricted within a certain bound, D_{M_s} does not exist and it is impossible to identify terms like $\theta_i \Pi_{j=1}^{M} \varepsilon^{k_j}(t-j, \theta)$ using algorithms that require the calculation of $\varepsilon(t, \theta)$. How small $|y(t)|$ should be in order to guarantee the existence of D_{M_s} is generally impossible to say, and is dependent on particular examples. The authors' experience in identifying non-linear control systems has shown that, in some cases, terms like $\varepsilon(t-i, \theta)\varepsilon(t-j, \theta)$ can be identified without difficulty, while in other examples the generation of $\varepsilon(t, \theta)$ become explosive and the identification procedure fails. This confirms the above analysis.

5.2. *Computing multistep-ahead predictions*

In this subsection the computation of multistep-ahead predictions is considered. Given a model and observed time-series values up to t, the task is to compute $y(t+h|t)$, a prediction of $y(t+h)$. It is well known that

$$E[(y(t+h) - y(t+h|t))^2] \tag{57}$$

is minimized if and only if

$$y(t+h|t) = E[y(t+h)|t] \tag{58}$$

where $E[y(t+h)|t]$ is the conditional expectation of $y(t+h)$ based on $y(k)$, $k \leqslant t$.

For the linear model it is easy to calculate the optimal prediction of $y(t+h)$, and for simple bilinear models the optimal prediction of $y(t+h)$ can be obtained after some algebraic operations (see Granger and Andersen 1978 a). For simple polynomial models the optimal prediction can similarly be obtained based on conditional-expectation operations. A simple example is

$$y(t) = a_1 y^2(t-1) + a_2 y(t-1)y(t-2) + e(t)$$

Since

$$y(t+1) = a_1 y^2(t) + a_2 y(t)y(t-1) + e(t+1)$$

the optimal one-step-ahead prediction is

$$\begin{aligned} y(t+1|t) &= a_1 E[y^2(t)|t] + a_2 E[y(t)y(t-1)|t] + E[e(t+1)t] \\ &= a_1 y^2(t) + a_2 y(t)y(t-1) \end{aligned}$$

Similarly, since

$$y(t+2) = a_1 y^2(t+1) + a_2 y(t+1)y(t) + e(t+2)$$

the optimal two-step-ahead prediction is

$$\begin{aligned} y(t+2|t) &= a_1 E[y^2(t+1)|t] + a_2 E[y(t+1)y(t)|t] + E[e(t+2)|t] \\ &= a_1 E[y^2(t+1)|t] + a_2 E[y(t+1)|t]y(t) \\ &= a_1 E[y^2(t+1)|t] + a_2 y(t+1)|t)y(t) \end{aligned}$$

Note that

$$\begin{aligned} E[y^2(t+1)|t] &= E[\{a_1 y^2(t) + a_2 y(t)y(t-1) + e(t+1)\}^2|t] \\ &= E[\{a_1 y^2(t) + a_2 y(t)y(t-1)\}^2|t] \\ &\quad + 2E[\{a_1 y^2(t) + a_2 y(t)y(t-1)\}e(t+1)|t] + E[e^2(t+1)|t] \\ &= \{a_1 y^2(t) + a_2 y(t)y(t-1)\}^2 + \sigma_e^2 = y^2(t+1)|t) + \sigma_e^2 \end{aligned}$$

Therefore the optimal two-step-ahead prediction is

$$y(t+2|t) = a_1 y^2(t+1|t) + a_2 y(t+1)|t)y(t) + a_1\sigma_e^2$$

In a practical implementation σ_e^2 would usually be replaced by an estimate $\hat{\sigma}_e^2$.

For complex polynomial models the computation of optimal h-step-ahead predictions can become cumbersome as h increases and may even be prohibitive. A practical solution is to replace terms like $E[y^k(t+i)|t]$ and $E[y(t+i)y(t+j)y(t+k)|t]$ by terms like $y^k(t+i|t)$ and $y(t+i|t)y(t+j|t)y(t+k|t)$ respectively. The prediction computed by this kind of substitution is of course suboptimal because in general

$$\left.\begin{aligned} E[y^k(t+i)|t] &\neq (E[y(t+i)|t])^k \\ E[y(t+i)y(t+j)y(t+k)|t] &\neq E[y(t+i)|t]E[y(t+j)|t]E[y(t+k)|t] \end{aligned}\right\} \quad (59)$$

where $i, j, k \geqslant 1$. A suboptimal two-step-ahead prediction for the model (36) is for example $y(t+2|t) = p(p(y(t)))$. Note that the computation of optimal multistep-ahead predictions using a general bilinear model is also by no means an easy task and may require some kind of suboptimal solution in practice.

As an illustration that polynomial models can produce better predictions than linear models, the time series of the annual sunspot numbers for the years 1700–1955 are considered. These 256 observations are listed in Appendix A1 of Tong (1983). Gabr and Subbat Rao (1981) fitted the following linear and bilinear models to the first 221 observations:

(*a*) *full AR model*

$$\begin{aligned} y(t) &= 1{\cdot}2163y(t-1) - 0{\cdot}4670y(t-2) - 0{\cdot}1416y(t-3) + 0{\cdot}1691y(t-4) \\ &\quad - 0{\cdot}1473y(t-5) + 0{\cdot}0543y(t-6) - 0{\cdot}0534y(t-7) + 0{\cdot}0667y(t-8) \\ &\quad + 0{\cdot}1129y(t-9) + 0{\cdot}1900y_m + e(t) \end{aligned} \quad (60)$$

(*b*) *subset AR model*

$$y(t) = 1{\cdot}2496y(t-1) - 0{\cdot}5510y(t-2) + 0{\cdot}1450y(t-9) + 0{\cdot}1564y_m + e(t) \quad (61)$$

(*c*) *subset bilinear model*

$$\begin{aligned} y(t) = {} & 1{\cdot}5012y(t-1) - 0{\cdot}7670y(t-2) + 0{\cdot}1152y(t-9) + 6{\cdot}8860 \\ & - 0{\cdot}01458y(t-2)e(t-1) + 0{\cdot}006312y(t-8)e(t-1) \\ & - 0{\cdot}007152y(t-1)e(t-3) + 0{\cdot}006047y(t-4)e(t-3) \\ & + 0{\cdot}003619y(t-1)e(t-6) + 0{\cdot}004334y(t-2)e(t-4) \\ & + 0{\cdot}001782y(t-2)e(t-2) + e(t) \end{aligned} \tag{62}$$

where the mean y_m of $y(t)$ for the first 221 observations is 43·481. In the current study the following polynomial AR model was fitted to the first 221 observations:

(*d*) *subset polynomial AR model*

$$\begin{aligned} y(t) = {} & 1{\cdot}1171y(t-1) - 0{\cdot}091963y(t-2) + 0{\cdot}38299y(t-9) - 0{\cdot}000028084y^3(t-1) \\ & + 0{\cdot}0044321y(t-1)y(t-8) - 0{\cdot}00016203y(t-2)y(t-5)y(t-8) \\ & - 0{\cdot}0025043y^2(t-9) + 0{\cdot}000095842y(t-1)y(t-5)y(t-8) \\ & + 0{\cdot}000011992y(t-5)y^2(t-7) - 0{\cdot}21815y(t-3) \\ & + 0{\cdot}0000095906y(t-2)y(t-3)y(t-4) + e(t) \end{aligned} \tag{63}$$

These four models were then used to calculate the predictions of the last 35 observations. The optimal prediction formulae for the linear models (60) and (61) can easily be written down. For the polynomial AR model (63), although it is easy to derive the optimal formulae for one-step- and two-step-ahead predictions, the optimal formulae become very complicated for $h > 2$, and therefore suboptimal predictions discussed above were used for the model (63) in actual computation. Using the bilinear model (62) to compute optimal predictions also becomes very complicated for $h > 6$. The error of a prediction $y(t+h|t)$ is defined as

$$\varepsilon(t+h|t) = y(t+h) - y(t+h|t) \tag{64}$$

and the mean sum of squares of the errors of predictions for the period $221 + h$ to 256 is

$$\hat{\sigma}_\varepsilon^2(h) = \frac{1}{36-h} \sum_{t=221}^{256-h} \varepsilon^2(t+h|t) \tag{65}$$

$\hat{\sigma}_\varepsilon^2(j)$ obtained using the linear and polynomial AR models for $h = 1, \ldots, 10$ and the bilinear model for $h = 1, \ldots, 6$ are given in the Table, where $\hat{\sigma}_\varepsilon^2$ is the variance of the residuals from 10 to 221. $\hat{\sigma}_\varepsilon^2(h)$, $h = 1, \ldots, 5$, for the models (60)–(62) are very close to those given by Gabr and Subba Rao (1981).

It is seen from the Table that the polynomial model (63) not only fits the first 221 observations better (smaller $\hat{\sigma}_\varepsilon^2$) but also gives better predictive accuracy except in the case $h = 3$ when compared to the linear models (60) and (61). The model (63) also seems to be better than the bilinear model (62) when it is used to predict several steps ahead.

The performance of long-term predictions using (63) is particularly remarkable. To gain some insight into this, let us consider the unforced response of the general model (1); that is,

$$y_0(t) = f(y_0(t-1), \ldots, y_0(t-n_y), 0, \ldots, 0) \tag{66}$$

Model	Full AR (60)	Subset AR (61)	Subset bilinear (62)	Subset polynomial (63)
$\hat{\sigma}_\varepsilon^2$	198·45	202·28	123·74	121·54
$\hat{\sigma}_\varepsilon^2(1)$	190·87	216·09	123·79	177·69
$\hat{\sigma}_\varepsilon^2(2)$	414·75	429·97	337·60	393·85
$\hat{\sigma}_\varepsilon^2(3)$	652·07	676·22	569·69	786·89
$\hat{\sigma}_\varepsilon^2(4)$	725·71	734·97	658·59	571·94
$\hat{\sigma}_\varepsilon^2(5)$	770·84	775·86	718·06	539·62
$\hat{\sigma}_\varepsilon^2(6)$	786·41	797·50	727·96	554·64
$\hat{\sigma}_\varepsilon^2(7)$	789·01	814·67		545·96
$\hat{\sigma}_\varepsilon^2(8)$	827·79	860·76		457·59
$\hat{\sigma}_\varepsilon^2(9)$	862·06	899·37		396·24
$\hat{\sigma}_\varepsilon^2(10)$	895·58	936·34		447·19

Values of $\hat{\sigma}_\varepsilon^2(h)$ (annual sunspot numbers).

It has long been noticed that the record of sunspot numbers reveals an intriguing cyclical phenomenon of an approximate 11 year period (see Fig. 2). It is therefore reasonable to assume that the underlying process of this time series when unforced should possess some kind of cyclical behaviour such as a limit cycle. The unforced responses for all four models with initial condition $y_0(t) = y(t)$, $t = 1, \ldots, 9$, are shown in Fig. 3. As expected, the responses of the linear and bilinear models die out. The polynomial AR model (63), however, produces a sustained oscillation. The model (63)

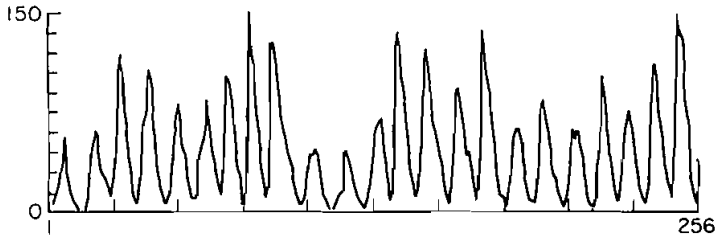
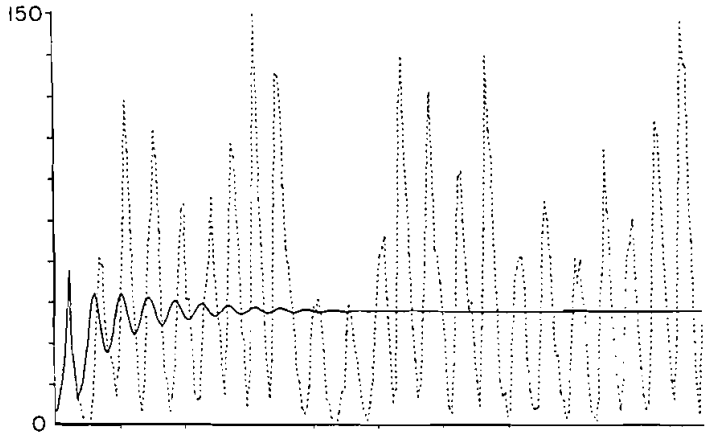


Figure 2. Annual sunspot numbers for the years 1700–1955.



(*a*)

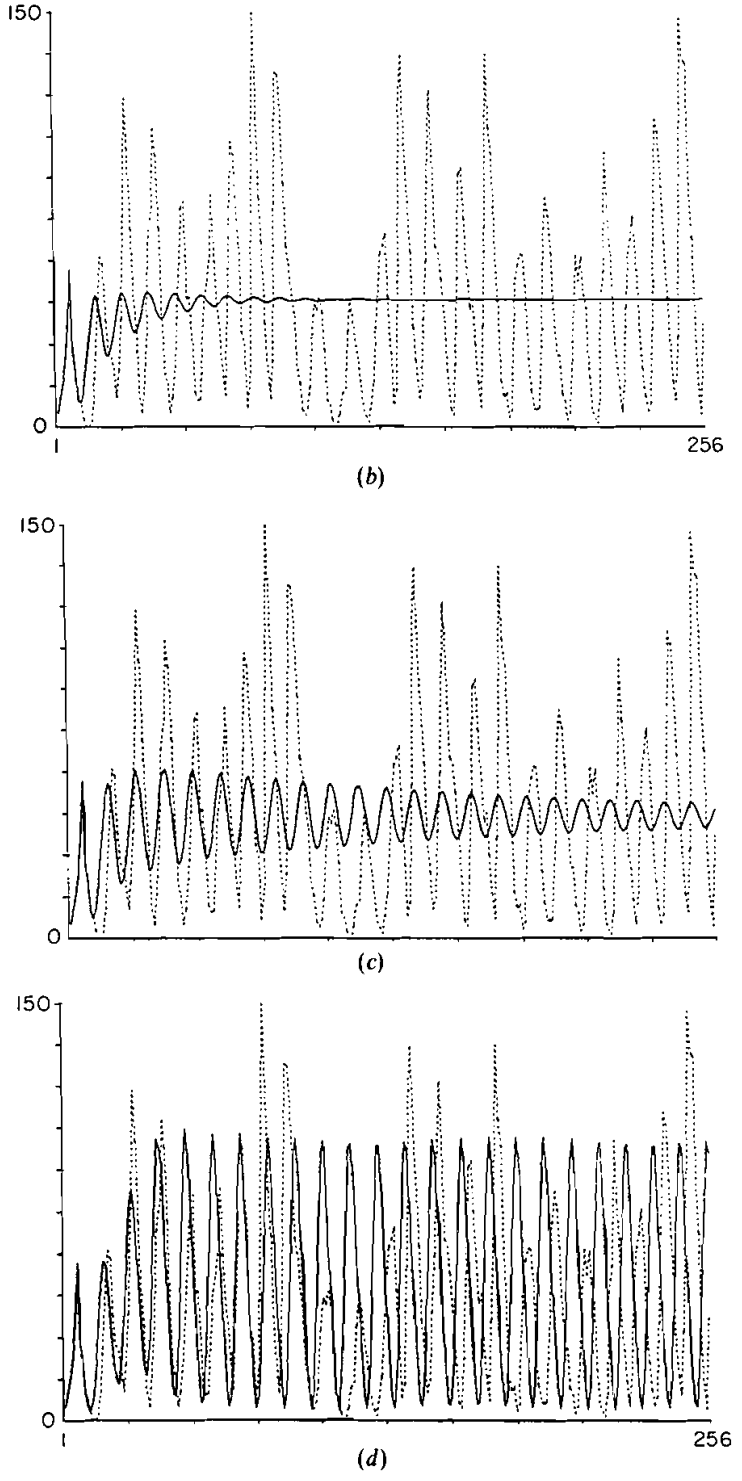


Figure 3. Unforced response: (*a*) full AR model; (*b*) subset AR model; (*c*) subset bilinear model; (*d*) subset polynomial AR model. ——, model unforced response; – – – –, observations.

may therefore be a better approximation to the underlying process in this aspect and this may also explain the better performance of long-term predictions using this model.

6. Conclusions

A unified approach to non-linear time-series modelling has been introduced. Several non-linear time-series models have been studied under the framework of a general representation known as the NARMA model. The TNARMA time-series model has also been discussed.

The general stationarity conditions have been investigated and the novel concepts of global and local stationarity have been introduced. If the stationarity of a model is globally valid on the normed space of all the stationary distributed noise processes then the stationarity is referred to as a *global stationarity*. If, on the other hand, the stationarity is only valid in a region of this space then it is called a *local stationarity*. The general invertibility conditions have been given as a dual case to stationarity, and the new concepts of global and local m-invertibility have been defined. If the invertibility of a model is independent of the statistical properties of the observations then it is known as *global m-invertibility*; otherwise it is called a *local m-invertibility*. The implications of these concepts for polynomial time-series models has been studied and it has been shown that, although Granger and Andersen (1978 a) were correct in criticizing the explosive nature of polynomial models, this model can still be very useful in modelling time series when the underlying process is stable and non-divergent. The time series of annual sunspot numbers has been used to illustrate these proposals.

The derivations of algorithms for selecting subset models for all the models discussed in this paper have been completed and extensive applications of these algorithms to real-life time series are currently under way. The results will be published in due course.

ACKNOWLEDGMENT

This work is supported by the Science and Engineering Research Council (Grant GR/D/30587).

REFERENCES

BILLINGS, S. A., and CHEN, S., 1989, Identification of non-linear rational systems using a prediction-error estimation algorithm. *International Journal Systems Science*, **20**, 467–494.

BOX, G. E. P., and JENKINS, G. M., 1976, *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day).

CHEN, S., and BILLINGS, S. A., 1989 a, Recursive prediction error parameter estimator for non-linear models. *International Journal of Control*, **49**, 569–594; 1989 b, Representations of non-linear systems: the NARMAX model. *Ibid.*, **49**, 1013–1032.

FLIESS, M., and NORMAND-CYROT, D., 1982, On the approximation of nonlinear systems by some simple state-space models. *Proceedings of the 6th IFAC Symposium on Identification and System Parameter Estimation*, Washington D.C., edited by G. A. Bekey (Pergamon Press), pp. 511–514.

GABR, M. M., and SUBBA RAO, T., 1981, The estimation and prediction of subset bilinear time series models with applications. *Journal of Time Series Analysis*, **2**, 155–171.

GRANGER, C. W. J., and ANDERSEN, A. P., 1978 a, *An Introduction to Bilinear Time Series Models* (Göttingen: Vandenhoeck and Ruprecht); 1978 b, On the invertibility of time series models. *Stochastic Processes and their Applications*, **8**, 87–92.

HAGGAN, V., HERAVI, S. M., and PRIESTLEY, M. B., 1984, A study of the application of state-dependent models in non-linear time series analysis. *Journal of Time Series Analysis*, **5**, 69–102.

HALLIN, M., 1980, Invertibility and generalized invertibility of time series models. *Journal of the Royal Statistical Society*, B **42**, 210–212.

JONES, D. A., 1976, Non-linear autoregressive processes. Ph.D. thesis, University of London.

LAI, H. Y., and HSIEH, S.-H., 1988, Threshold modelling of nonlinear dynamic systems. *Proceedings of 6th International Conference on Modal Analysis*, Hyatt Orlando, U.S.A., February, edited by D. J. DeMichele (SEM), pp. 89–95.

LEONTARITIS, I. J., and BILLINGS, S. A., 1985 a, Input–output parametric models for non-linear systems, Part I: Deterministic non-linear systems. *International Journal of Control*, **41**, 303–328; 1985 b, Input–output parametric models for non-linear systems, Part II: Stochastic non-linear systems. *Ibid.*, **41**, 329–344.

LIN, J., HAN, C., ROBERTS, P. D., and WAN, B., 1989, A new approach to stochastic optimizing control of steady state systems using dynamic information. *International Journal of Control*, **50**, 2151–2172.

OZAKI, T., 1981, Non-linear threshold autoregressive models for non-linear random vibrations. *Journal of Applied Probability*, **18**, 443–451; 1985, Non-linear time series models and dynamical systems. *Handbook of Statistics*, 5: *Time Series in the Time Domain*, edited by E. J. Hannan, P. R. Krishnaiah and M. M. Rao (Amsterdam: North-Holland), pp. 25–83.

PRIESTLEY, M. B., 1980, State-dependent models: a general approach to non-linear time series analysis. *Journal of Time Series Analysis*, **1**, 47–71.

QUINN, B. G., 1982, Stationarity and invertibility of simple bilinear models. *Stochastic Processes and their Applications*, **12**, 225–230.

SONTAG, E. D., 1979, *Polynomial Response Maps*. Lecture Notes in Control and Information Sciences, Vol. 13 (Berlin: Springer-Verlag).

SUBBA RAO, T., and GABR, M. M., 1984, *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Lecture Notes in Statistics (New York: Springer-Verlag), Vol. 24.

TJØSTHEIM, D., 1986, Some doubly stochastic time series models. *Journal of Time Series Analysis*, **7**, 51–72.

TONG, H., 1983, *Threshold Models in Non-linear Time Series Analysis*. Lecture Notes in Statistics (New York: Springer-Verlag), Vol. 4.

TONG, H., and LIM, K. S., 1980, Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society*, B **42**, 245–292.