MODELLING AND MONITORING OF MEDICAL TIME SERIES

BY

KERRY GORDON, B.SC.

Thesis Submitted to the University of Nottingham for the Degree of Doctor of Philosophy

May 1986

Contents

CHAPTER FIVE: BIVARIATE MODELS FOR UNEQUALLY-SPACED DATA

CHAPTER SIX: APPLICATIONS

o0o

Abstract

In this thesis we examine several extensions to the dynamic linear model framework, outlined by Harrison and Stevens (1976), in order to adapt these models for use in the on-line analysis of medical time series that arise from routine clinical settings. The situation with which we are most concerned is that where we are monitoring individual patients and wish to detect abrupt changes in the patient's condition as soon as possible.

A detailed background to the study and application of dynamic linear models is given, and other techniques for time series monitoring are also discussed when appropriate. We present a selection of specific models that we feel may prove to be of practical use in the modelling and monitoring of medical time series, and we illustrate how these models may be utilized in order to distinguish between a variety of alternative changepoint-types. The sensitivity of these models to the specification of prior information is examined in detail.

The medical background to the time series examined requires the development of models and techniques enabling us to analyze generally unequally-spaced time series. We test the performance of the resulting models and techniques using simulated data. We then attempt to build a framework for bivariate time series modelling, allowing, once more, for the possibility of unequally-spaced data. In particular, we suggest mechanisms whereby causality and feedback may be introduced into such models.

Finally, we report on several applications of this methodology to actual medical time series arising in various contexts, including kidney and bone-marrow transplantation and foetal heart monitoring.

oOo

Acknowledgements

First and foremost, I should like to express my sincere gratitude to Professor A. F. M. Smith, whose continued guidance, encouragement and motivation played a crucial role throughout the period of study.

I am also extremely grateful to Dr. M. S. Knapp and the staff, past and present, of the Unit of Medical Information Technology, Department of Obstetrics and Gynaecology, University Hospital, for providing me with a greater insight into the medical aspects of this study and for their general enthusiasm and support.

Thanks are due, too, to Mrs. Maya Fletcher for her patient but speedy typing of the original manuscript.

The funding for this research was initially provided by the Nottingham and Notts Kidney Fund, and then continued by the National Westminster Bank Nottingham University Research Fund.

To Mum and Dad

CHAPTER ONE

General Introduction

1.1 INTRODUCTION

Rapid growth in information technology has made it possible for more people to be able to store large quantities of numerical information on a computer. In particular it has become possible to store long and dense sequences of time-related data, thus presenting a growing opportunity for statistical time series analysis.

The motivation for this thesis has been provided by medical time series and associated problems of 'real-time' monitoring, where the timing of events of clinical interest is often the critical ingredient. All the statistical modelling in this thesis has, therefore, been based upon the time-domain approach to time series analysis rather than the frequency-domain approach.

It should also be borne in mind that the ideas presented in this thesis have been derived from an attempt to approach problems encountered in routine clinical settings. In particular, the time series of interest consist of measurements on 'human patients', who have presented themselves naturally in the course of clinical practice. Groups of patients, for whom the methods are intended, will therefore tend to be extremely heterogeneous in nature (unlike those from a clinical trial, where controlled conditions often result in homogeneous groups of subjects).

The methodology which we shall adopt, and extend, is related to that of Harrison and Stevens (1976) who set out a Bayesian approach to time series analysis using, as its foundation, the concept of a dynamic linear model (DLM). In order to deal with 'discontinuous' behaviour, multi-process modelling was also incorporated into the DLM framework, resulting in a time-series 'tracking' procedure that would rapidly adjust to sudden changes in the underlying model.

This approach has been shown to be useful not only in the forecasting situation (Harrison and Stevens 1975) but also in the monitoring situation (Smith and West 1983), where the swift detection of sudden changes in pattern is the feature of interest.

We take the view that when dealing with an on-line problem (that is, one where decisions concerning intervention have to be made on a sequential basis), the ability to interpret <u>new</u> information in the light of <u>previous</u> information (i.e. to adopt a Bayesian approach) is essential, especially when it is <u>known</u> that previous interventions have actually taken place.

There are a number of problems, however, which must be overcome if this type of analysis is to be implemented in a clinical setting. Appropriate DLM's must be created for the different series encountered in medical monitoring contexts, and allowance must be made for the possibility of missing or, more generally, unequally-spaced responses, arising from the irregular collection of data which is common to a routine clinical environment. The DLM framework has, therefore, been extended in this thesis in order to handle sets of data arising from periods of medical supervision.

1.2 OUTLINE OF THESIS

In Chapter 2, we discuss the use of dynamic linear models with particular reference to the recursive aspect of the procedures (Kalman 1960, Lindley and Smith 1972, Young 1974, Priestly 1980). The work of Harrison and Stevens (1976) is reviewed, and some extensions to their ideas are discussed. In particular, the use of a conjugate prior distribution (DeGroot 1970, Aitchison and Dunsmore 1975), for the case of unknown observation variance (as proposed by West 1982, Smith and West 1983), is discussed in some detail, and the specific choice of prior parameter estimates is examined. Reference is also made to the work of Godolphin and Harrison (1975) in which attention is drawn to some equivalences between DLM's and the more conventional time series models of Box and Jenkins (1970). The DLM algorithms are described in full for a selection of potentially useful models, with special reference to the medical setting.

In Chapter 3, we review the idea of multi-process models and the assumptions upon which these are based (Kullback and Leibler 1951, Harrison and Stevens 1971,1976). The use of this type of modelling for monitoring time series for changepoints is examined (Harrison and Stevens 1975, Smith and West 1983), and some multi-state representations are presented pictorially for a selection of models. Performance and sensitivity are discussed in relation to the ability to detect changepoints.

In Chapter 4, we extend the ideas of the previous chapters to incorporate the possibility of unequally-spaced responses in time (Clinger and Van Ness 1976, Jones 1980). A general principle is proposed which leads to individual formulations for a selection of models, with particular reference to variance updating. Performance and sensitivity are investigated using simulated data.

In Chapter 5, we investigate the formulation of bivariate time-series models (Tiao and Box 1981). Problems of interpretation are discussed, including the interaction between individual variables (Newbold 1979). We introduce the idea of Markovian state-dependence and we re-examine the relevant probability distributions in the presence of generally unequally-spaced bivariate data. The performance, and sensitivity, of two specific bivariate models, is investigated using simulated data.

Finally, in Chapter 6, we present a set of applications of our methodology to actual medical time series. The setting of kidney transplantation (Smith and Cook 1980, West 1982, Smith and West 1983) is the starting point for these illustrations, and the ideas presented in previous chapters are demonstrated both for this case and for other applications of interest.

1.3 NOTATION

Much of the notation used in this thesis follows directly from that employed by Harrison and Stevens (1976). The symbol '$\sim$' indicates a vector (or matrix) quantity so that, for instance,

$\underset{\sim}{y}_t$ denotes a vector whereas y_t denotes a scalar.

The symbol M is used to denote a 'state' (with respect to the multi-state structure) so that, for instance, $M^{(j)}$ would mean that state j obtains. For the situation where movements between states are subject to Markovian behaviour, two superscripts may be attached to the state symbol M. For instance, ${}^{(i)}M_t^{(j)}$ would mean that state j obtains at time t given that state i obtained at time t - 1.

The time index is attached to various parameters by way of a subscript:

(a) For the case of equally-spaced observations, the letter 't' is used, e.g. $\underset{\sim}{y}_t$, $\underset{\sim}{\theta}_t$, ε_t, etc., so that t not only indexes the number of observations but also corresponds to the actual time (in whatever units are chosen for the application).

(b) For the case of unequally-spaced observations, the letter 'k' is used as the time index, e.g. y_k, $\underset{\sim}{\theta}_k$, ε_k, etc. In this case, k will not necessarily correspond to the actual time at which the kth response was measured, which is instead denoted by T_k.

Other notation is either fairly standard, for example $N(\mu,\sigma^2)$ indicates the normal distribution with mean μ and variance σ^2, $\hat{\underset{\sim}{\phi}}$ the expected value of $\underset{\sim}{\phi}$, etc. or has been defined explicitly as necessary.

oOo

CHAPTER TWO

DYNAMIC LINEAR MODELS

2.1 INTRODUCTION TO DYNAMIC LINEAR MODELS

2.1.1 INTRODUCTORY REMARKS

When we are monitoring a medical time series based on a biochemical or physical indicator (or set of indicators), we are often concerned with detecting changes in pattern, particularly if the transition is from a stable or improving condition to one of deterioration. Much of our attention in this thesis is devoted to the development of certain modelling and inference 'tools' which will allow us to provide a formal framework for this kind of monitoring situation.

In order to determine whether or not there have been deviations from a 'normal' pattern of behaviour, we need to know, of course, what the 'normal' pattern is considered to be. This may appear to be a trivial statement but, in fact, much of the literature of model identification (see, for example, Akaike 1974) is concerned with precisely this problem.

In the medical context we need models of patterns of behaviour which are realistic enough to capture the nature of the series encountered, but simple enough for the non-mathematically-minded clinician to be able to understand them. This means that the model parameters will need to have simple, meaningful interpretations (such as 'level' or 'slope'), preferably bearing some relationship to the underlying mechanisms, be they physiological, pharmacological, pathological or biological. These considerations suggest that state-space representations of time series (see, for example, Anderson and Moore 1979) will be more useful than the conventional functional-form representations proposed by Box and Jenkins (1970).

It will be reasonably straightforward to combine sub-models in a meaningful way in order to create a global model, since the initial models and their parameters have meaningful interpretations in the first place. This facility is essential in medical/biological modelling where, due to complexity, the overall model is often split into so-called 'compartments'. It is at the compartment level that modelling usually takes place; these compartments then being joined together to form a global model (Matis and Wehrly 1979, Carson, Cobelli and Finkelstein 1983).

Another fundamental consideration is that if we use the recursive state-space time series formulation, we can allow for a change in the underlying model between recursions. This has considerable advantages, both for monitoring (where an unexpected change in model may have occurred) and intervention (where a deliberate attempt to change some model parameters may have occurred).

The above paragraphs outline the dynamic linear modelling approach (Harrison and Stevens 1976); the next section discusses some of the existing work related to the use of DLM's.

2.1.2 BACKGROUND TO MODELLING TECHNIQUES

From the point of view of mathematical or statistical modelling, it is often best to work with as simple a model as is consistent with the problem under study. It is not surprising, therefore, that a considerable amount of attention has been given to the application of linear models to statistical data handling. For a comprehensive guide to the use of linear models for classical statistical analysis, see Searle (1971), and for a more recent viewpoint, summarizing the concepts of, so-called, generalized linear models (which avoid classical assumptions of normality) see McCullagh and Nelder (1983), based on the original theoretical work of Nelder and Wedderburn (1972).

In the field of time series analysis, there have been two main branches of study; the time-domain approach and the frequency-domain approach. In this thesis, we shall concentrate on the former and, in particular, on the use of recursive algorithms for time series analysis (see, for example: Young 1974). These are very appealing for the problems we have in mind in that not only do they provide a means for on-line tracking of time-related sequences of observations, but they are also closely related to Bayesian methodology.

Lindley and Smith (1972) presented an hierarchical Bayesian approach to linear modelling problems, an approach which was generalized by Harrison and Stevens (1976) to the time-dependent framework of dynamic linear modelling. Although this latter work was originally geared towards producing practical forecasts (particularly in business and economics) the recursive and flexible nature of the algorithms enabled them to be extended, also, to the monitoring situation, as shown by Smith and West (1983).

The Dynamic Linear Model (DLM)

$$\underset{\sim}{y}_t = \underset{\sim}{H}_t \underset{\sim}{\theta}_t + \underset{\sim}{\varepsilon}_t \tag{2.1}$$

$$\underset{\sim}{\theta}_t = \underset{\sim}{G} \underset{\sim}{\theta}_{t-1} + \underset{\sim}{\omega}_t, \tag{2.2}$$

where

$\underset{\sim}{y}_t$ = vector of observations made at time t
$\underset{\sim}{\theta}_t$ = vector of system parameters at time t
$\underset{\sim}{H}_t$ = known regression matrix at time t
$\underset{\sim}{G}$ = known transition matrix
$\underset{\sim}{\varepsilon}_t, \underset{\sim}{\omega}_t$ = zero-mean, random vectors at time t.

(2.3)

For the basic DLM we make the following assumptions about the quantities outlined in (2.3):

(i) $\underset{\sim}{\varepsilon}_t$ is independent of $\underset{\sim}{\varepsilon}_s$, $\forall\, s \neq t$
(ii) $\underset{\sim}{\varepsilon}_t, \underset{\sim}{\omega}_t$ are independent of $\underset{\sim}{\theta}_{t-1}$, $\forall$ t given the past observations $y_1, \ldots, y_{t-1}$ (denoted by $\underset{\sim}{D}_{t-1}$)
(iii) $\underset{\sim}{\varepsilon}_t$ is independent of $\underset{\sim}{\omega}_t$, $\forall$ t
(iv) $\underset{\sim}{\varepsilon}_t \sim N(\underset{\sim}{0}, \underset{\sim}{E}_t)$; $\underset{\sim}{\omega}_t \sim N(\underset{\sim}{0}, \underset{\sim}{W}_t)$.

(2.4)

The DLM described above has two components: (2.1) is the observation equation, describing the measuring process which relates the current system parameters to the resulting observations; whereas (2.2) is the <u>system</u> equation, describing the process by which the system parameters evolve in time. Clearly, a further assumption here is that the system parameter evolution is Markovian in nature. We note in passing that the case where $\underset{\sim}{\theta}_t = \underset{\sim}{\theta}_{t-1} = \ldots = \underset{\sim}{\theta}$ and $\underset{\sim}{F}_t = \underset{\sim}{F}_{t-1} = \ldots = \underset{\sim}{F}$ are time-independent reduces to the classical linear model formulation.

There are a number of fairly obvious extensions of the above model. For instance, the transition (or system) matrix, $\underset{\sim}{G}$, could be allowed to be time-dependent rather than fixed, resulting in a transition matrix, $\underset{\sim}{G}_t$, which can change from time-point to time-point, further extending the model's flexibility. This extension to the DLM structure will become necessary for the developments we shall describe in Chapter 4.

In recent years, there have been a number of papers describing various extensions of the standard DLM framework. West (1981,1982) discusses robustification of the algorithm by dropping Assumption (iv) above; West et al.(1984) go on to describe non-linear versions of such models, thus formulating a framework for dynamic <u>general</u> linear models. We shall take the view that the normality assumptions are probably adequate for the medical problems studied in this thesis. However, the techniques discussed by West could certainly be combined with the methods described in this thesis, should a situation arise which warrants this kind of approach.

If the assumptions of (2.4) <u>are</u> maintained, and if it is further assumed that, initially,

$$\underset{\sim}{\theta}_o \sim N(\underset{\sim}{m}_o, \underset{\sim}{C}_o) \tag{2.5}$$

and that at time t - 1 the distribution of $\underset{\sim}{\theta}_{t-1}$, given the data, $\underset{\sim}{D}_{t-1}$, up to that point, is described by

$$(\underset{\sim}{\theta}_{t-1} | \underset{\sim}{D}_{t-1}) \sim N(\underset{\sim}{m}_{t-1}, \underset{\sim}{C}_{t-1}) \tag{2.6}$$

then, at time t, the distribution of $\underset{\sim}{\theta}_t$ given all the data, $\underset{\sim}{D}_t$, up to and including time t is given by:

$$(\underset{\sim}{\theta}_t | \underset{\sim}{D}_t) \sim N(\underset{\sim}{m}_t, \underset{\sim}{C}_t) \tag{2.7}$$

where the values of $\underset{\sim}{m}_t$ and $\underset{\sim}{C}_t$ are obtained recursively from the Kalman Filter algorithms (Kalman and Bucy 1961).

Following the notation of (2.3), let

$$\left.\begin{aligned} \underset{\sim}{f}_t &= \underset{\sim}{H}_t \underset{\sim}{G} \underset{\sim}{m}_{t-1} \\ \underset{\sim}{e}_t &= \underset{\sim}{y}_t - \underset{\sim}{f}_t \\ \underset{\sim}{P}_t &= \underset{\sim}{G} \underset{\sim}{C}_{t-1} \underset{\sim}{G}^T + \underset{\sim}{W}_t \\ \underset{\sim}{F}_t &= \underset{\sim}{H}_t \underset{\sim}{P}_t \underset{\sim}{H}_t^T + \underset{\sim}{E}_t \\ \text{and} \quad \underset{\sim}{S}_t &= \underset{\sim}{P}_t \underset{\sim}{H}_t^T (\underset{\sim}{F}_t)^{-1}, \end{aligned}\right\} \tag{2.8}$$

then

$$\underset{\sim}{m}_t = \underset{\sim}{G} \underset{\sim}{m}_{t-1} + \underset{\sim}{S}_t \underset{\sim}{e}_t \tag{2.9}$$

$$\underset{\sim}{C}_t = \underset{\sim}{P}_t - \underset{\sim}{S}_t \underset{\sim}{F}_t \underset{\sim}{S}_t^T. \tag{2.10}$$

In order to calculate the quantities involved, one more assumption has been made; that is, it has been assumed that both the observation and system variances are known for each time t.

To overcome the problem of unknown variances, Harrison and Stevens (1976) suggested the use of discrete-valued grids, covering a range likely to include plausible values for the variances. The 'inefficiency' of this approach had been criticized by Stoodley and Mirnia (1979), and a more efficient procedure for on-line variance updating was put forward by West (1982), based on the idea of a joint conjugate prior distribution for the normal distribution with unknown mean and variance (see, for example, DeGroot 1970).

Let us rewrite

$$\left.\begin{array}{llll} & \mathrm{Var}(\underset{\sim}{\varepsilon}_t) = \underset{\sim}{E}_t & \text{as} & c^2 \underset{\sim}{R}\varepsilon \\ \text{and} & & & \\ & \mathrm{Var}(\underset{\sim}{\omega}_t) = \underset{\sim}{W}_t & \text{as} & c^2 \underset{\sim}{R}\omega \end{array}\right\} \quad (2.11)$$

and assume that

$$\underset{\sim}{\theta}_o \sim N(\underset{\sim}{m}_o, c^2 \underset{\sim}{C}_o) \quad (2.12)$$

and

$$(\underset{\sim}{\theta}_{t-1} | \underset{\sim}{D}_{t-1}) \sim N(\underset{\sim}{m}_{t-1}, c^2 \underset{\sim}{C}_{t-1}) \quad (2.13)$$

(replacing (2.5) and (2.6) respectively) where, in general, c^2 is time-dependent.

Then the equations of (2.8) can be rewritten:

$$\left.\begin{array}{l} \underset{\sim}{f}_t = \underset{\sim}{H}_t \underset{\sim}{G} \underset{\sim}{m}_{t-1} \\ \underset{\sim}{e}_t = \underset{\sim}{y}_t - \underset{\sim}{f}_t \\ \underset{\sim}{P}_t = \underset{\sim}{G} \underset{\sim}{C}_{t-1} \underset{\sim}{G}^T + \underset{\sim}{R}\omega \\ \underset{\sim}{F}_t = \underset{\sim}{H}_t \underset{\sim}{P}_t \underset{\sim}{H}_t^T + \underset{\sim}{R}\varepsilon \\ \underset{\sim}{S}_t = \underset{\sim}{P}_t \underset{\sim}{H}_t^T (\underset{\sim}{F}_t)^{-1} . \end{array}\right\} \quad (2.14)$$

If c^2 is known, the recursion described by (2.14) is identical to that of (2.8). However, when c^2 is unknown West (1982) proposes the following procedure.

Let $\lambda^{-1} = c^2$. We replace (2.13) by

$$(\underset{\sim}{\theta}_{t-1}|\underset{\sim}{D}_{t-1},\lambda) \sim N(\underset{\sim}{m}_{t-1},\lambda^{-1}\underset{\sim}{C}_{t-1}) \tag{2.15}$$

and

$$(\lambda|\underset{\sim}{D}_{t-1}) \sim G(\tfrac{1}{2}n_{t-1},\tfrac{1}{2}r_{t-1}) \tag{2.16}$$

where $U \sim G(\alpha,\beta)$ means that U has a gamma distribution defined by

$$p(u) = \frac{\beta^{\alpha}u^{\alpha-1}e^{-\beta u}}{\Gamma(\alpha)}, \qquad u > 0$$

where

$$\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1}e^{-u}du, \qquad \alpha > 0$$

so that

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \qquad \alpha > 1. \tag{2.17}$$

We also assume that, initially,

$$\lambda \sim G(\tfrac{1}{2}n_o,\tfrac{1}{2}r_o). \tag{2.18}$$

Standard Bayesian analysis (see, for example, DeGroot 1970) shows that

$$(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda) \sim N(\underset{\sim}{m}_t,\lambda^{-1}\underset{\sim}{C}_t) \tag{2.19}$$

and

$$(\lambda|\underset{\sim}{D}_t) \sim G(\tfrac{1}{2}n_t,\tfrac{1}{2}r_t) \tag{2.20}$$

where $\underset{\sim}{m}_t$ and $\underset{\sim}{C}_t$ are defined by (2.9), (2.10) and (2.14), and where

$$n_t = n_{t-1} + 1 \qquad (2.21)$$

$$r_t = r_{t-1} + \underset{\sim}{e}_t^T(\underset{\sim}{F}_t)^{-1}\underset{\sim}{e}_t \qquad (2.22)$$

with $\underset{\sim}{e}_t$ and $\underset{\sim}{F}_t$ as defined by (2.14).

This procedure has been demonstrated (Smith and West 1983, Trimble et al. 1983) to be an improvement on the method of Harrison and Stevens (1976).

Specifying Prior Variance Information

The specification of initial system variance, defined as $\lambda^{-1}\underset{\sim}{R}\omega$ from (2.11), is very much open to debate. An alternative method for expressing the information described by $\underset{\sim}{R}\omega$ is discussed later. In terms of specifying λ, however, one must choose suitable starting values for n_o and r_o in (2.18).

Let us assume, for example, that the variance law is constant (i.e. that c^2 does not depend upon the process mean), and that we have some prior estimate for the value of c^2. It seems sensible to equate the <u>expected value</u>, arising from the proposed gamma distribution, to this estimate.

Theorem 2.1.2

If $\lambda \sim G(\alpha,\beta)$, i.e.

$$p(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}, \qquad \lambda > 0 \qquad (2.23)$$

then

$$E(\lambda^{-1}) = \frac{\beta}{\alpha - 1}, \qquad \alpha > 1 \qquad (2.24)$$

and

$$\text{Var}(\lambda^{-1}) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}, \qquad \alpha > 2. \qquad (2.25)$$

Proof

By straightforward manipulation.

Corollary 2.1.2

For the distribution defined by (2.18) we have

$$E(\lambda^{-1}) = \frac{(\tfrac{1}{2}r_o)}{(\tfrac{1}{2}n_o - 1)} = \frac{r_o}{n_o - 2} \qquad (n_o > 2) \tag{2.26}$$

and

$$\mathrm{Var}(\lambda^{-1}) = \frac{(\tfrac{1}{2}r_o)^2}{(\tfrac{1}{2}n_o - 1)^2(\tfrac{1}{2}n_o - 2)}$$

$$= \frac{2r_o^2}{(n_o - 2)^2(n_o - 4)} \qquad (n_o > 4). \tag{2.27}$$

Since we require non-negative values for $\mathrm{Var}(\lambda^{-1})$ the choice of n_o is restricted to the range $n_o \geq 4$, with an infinite variance resulting from the equality.

NOTE: If we have a fixed estimate of c^2 we can tune our uncertainty in this estimate by adjusting n_o without altering the ratio $r_o/(n_o - 2)$; our uncertainty is decreased as n_o is increased, e.g.

(i) $r_o = 0.3$, $n_o = 5$

$$E(\lambda^{-1}) = \frac{r_o}{n_o - 2} = 0.1$$

$$\mathrm{Var}(\lambda^{-1}) = \frac{2r_o^2}{(n_o - 2)^2(n_o - 4)} = 0.02$$

(ii) $r_o = 4.8$, $n_o = 50$

$$E(\lambda^{-1}) = 0.1$$

$$\mathrm{Var}(\lambda^{-1}) \approx 0.0004.$$

Alternative approaches to the estimation of c^2 have been suggested (Ameen and Harrison 1982, Cantarelis and Johnston 1983) though Ameen and Harrison (1982) acknowledged that the approach taken by West is 'operationally elegant and is properly Bayesian'.

Discounting

A procedure has been proposed by Harrison and his colleagues (Ameen and Harrison, 1982,1983,1984, Harrison and Johnston 1983, Johnston and Harrison 1983) which side-steps the problem of specifying a covariance matrix, $\underset{\sim}{R}_\omega$, strictly in terms of variances and covariances. This method involves the use of, so-called, discount parameters, and is based upon simple ideas of exponential smoothing (see, for example, Brown 1963).

The effect of $\underset{\sim}{R}_\omega$ on the updating recursion appears in (2.14), via

$$\mathrm{Var}(\underset{\sim}{\theta}_t \mid \underset{\sim}{D}_{t-1}) = \underset{\sim}{P}_t = \underset{\sim}{G}\underset{\sim}{C}_{t-1}\underset{\sim}{G}^T + \underset{\sim}{R}_\omega. \tag{2.28}$$

In other words, R_ω increases the system variance. The discounting concept is to replace the additive $\underset{\sim}{R}_\omega$ by a multiplicative quantity, $\underset{\sim}{\Lambda}_t$, so that (2.28) becomes

$$\mathrm{Var}(\underset{\sim}{\theta}_t \mid \underset{\sim}{D}_{t-1}) = \underset{\sim}{P}_t = \underset{\sim}{\Lambda}_t^{\frac{1}{2}}\underset{\sim}{G}\underset{\sim}{C}_{t-1}\underset{\sim}{G}^T\underset{\sim}{\Lambda}_t^{\frac{1}{2}}, \tag{2.29}$$

where $\underset{\sim}{\Lambda}_t$ is a diagonal matrix of positive discount factors, $1/\lambda_{it}$, $i = 1, \ldots, n$, $\lambda_{it} \leq 1$ (where n is the number of parameters in the model).

In this way the effect of $\underset{\sim}{\Lambda}_t$ will also be to increase the system variance, by multiplicatively discounting information.

Notice that different discount factors may be introduced for individual parameters, unlike much of the earlier work in this area, where the choice of a single discount factor for all the parameters had been adopted (see, for example, Sorenson and Sacks 1971).

The rationale behind the adoption of this formulation is threefold. It is asserted that:

(i) most 'practitioners' find it difficult to specify the elements of $\underset{\sim}{R}\omega$, and that 'many modellers have a natural feel for discounting' (Ameen and Harrison 1982);

(ii) the discount vector (matrix) is invariant to scale;

(iii) the limiting form of forecast function is identical to that obtained using the $\underset{\sim}{R}\omega$ formulation as long as $\underset{\sim}{\Lambda}_t$ is expressed in a particular way.

We are mainly interested in the monitoring of medical time series, and therefore have no real concern with forecast functions; limiting forms of these functions are of even less interest, since most of the series we consider are not only finite, but are also reasonably short.

It can also be seen that, using the formulation described by (2.11) to (2.22), the choice of $\underset{\sim}{R}\omega$ is also scale invariant, since measurement of scale is diverted into the estimation of c^2.

In fact, for multiprocess models, the choice of non-zero elements of $\underset{\sim}{R}\omega$ has a simple interpretation. The size of the element depicts: the number of times greater than the variance, c^2, the magnitude of a change is likely to be. For instance, if a level change could plausibly be of a magnitude of up to 20 times that of the normal variation, then the element of R_ω corresponding to level change would be 20. With discount matrices, however, the elements have no comparable interpretation with respect to changepoints.

For these reasons, coupled with the fact that normal discount models are only 'coherent' for series with no missing observations (Ameen and Harrison 1982), whereas we will be investigating the case of unequally-spaced measurements, it was felt that the original formulation of the DLM, involving the R_ω matrix, was more applicable to the problems we shall be considering than the discount factor approach. The former procedure has therefore been adopted in this thesis.

Let us now turn our attention to a number of specific models which fit into the general DLM framework. Since some of these relate closely to the ARIMA class of models, we note first the work of Godolphin and Harrison (1975) in which theoretical equivalences between dynamic linear models and ARIMA models (Box and Jenkins 1970) are derived. It was shown that certain dynamic linear models can be reparareterized to form ARIMA models, though extra restrictions will then be imposed upon the model parameters. For instance, it was reported by Godolphin and Harrison (1975) that the linear growth model can be rewritten in the form of an IMA(0,2,2) model. This fact was well known, but it was shown that for the equivalence to hold true, further restrictions on the moving average parameter space were necessary in addition to the stationary and invertibility conditions.

However, in the work that follows these ideas are bypassed, in that it is typically assumed from the outset that any model structure is associated with the system equations (2.2), and that the observation equation will be of the form

$$y_t = \mu_t + \varepsilon_t, \tag{2.30}$$

i.e. what is observed is merely the level of a series along with observational error. Therefore, if $\theta_t^T = [\mu_t \theta_{1t} \; \ldots \; \theta_{n-1t}]$ then, in DLM terms, H_t will be assumed to be of the form

$$H_t = H = [1 \quad 0 \quad \ldots \quad 0], \quad \forall t \tag{2.31}$$

[NB: The sinusoidal models of Sections 2.3.3 and 2.3.4 are the exceptions to this rule.]

This particular formulation is chosen because it is felt that many 'indicators' of clinical conditions are of a nature whereby their measurement involves a machine-reading (corresponding to, say, a concentration of the indicator in fluid). In other words, the observer, be it human or computer, will 'see' the machine-reading and record this value, thus introducing potential observation errors.

The next section deals with the iteration required between timepoints; in the following section, we describe a selection of useful models.

2.2 RECURSIVE ESTIMATION

Although the general framework of a dynamic linear model incorporates the possibility of multiple observations (see Chapter 5), we restrict ourselves at present to the case of a univariate observation, y_t, for simplicity of exposition.

We have seen in the previous section that if

$$p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_{t-1},\lambda) \sim N(\underset{\sim}{m}_{t-1},\lambda^{-1}\underset{\sim}{C}_{t-1})$$

and

$$p(\lambda|\underset{\sim}{D}_{t-1}) \sim G(\tfrac{1}{2}n_{t-1},\tfrac{1}{2}r_{t-1})$$

then

$$(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda) \sim N(\underset{\sim}{m}_t,\lambda^{-1}\underset{\sim}{C}_t)$$

and

$$(\lambda|\underset{\sim}{D}_t) \sim G(\tfrac{1}{2}n_t,\tfrac{1}{2}r_t)$$

with $\underset{\sim}{m}_t$ and $\underset{\sim}{C}_t$ defined by (2.9), (2.10) and (2.14), and n_t and r_t defined by (2.21), (2.22) and (2.14). It can also be seen that:

$$p(y_t|\underset{\sim}{D}_{t-1},\lambda) \sim N(f_t,\lambda^{-1}F_t) \tag{2.32}$$

where f_t and F_t are defined in (2.14). So,

$$p(y_t|\underset{\sim}{D}_{t-1}) = \int_{\lambda=0}^{\infty} p(y_t|\underset{\sim}{D}_{t-1},\lambda)p(\lambda|\underset{\sim}{D}_{t-1})d\lambda$$

$$= \int_{\lambda=0}^{\infty} (2\pi\lambda^{-1}F_t)^{-\frac{1}{2}}\exp[-\frac{(y_t - f_t)^2}{2\lambda^{-1}F_t}].\frac{(\frac{1}{2}r_{t-1})^{(\frac{1}{2}n_{t-1})}\lambda^{(\frac{1}{2}n_{t-1}-1)}}{\Gamma(\frac{1}{2}n_{t-1})}$$

$$\times \exp[-(\tfrac{1}{2}r_{t-1})\lambda]d\lambda$$

$$\propto F_t^{-\frac{1}{2}}\frac{r_{t-1}^{(\frac{1}{2}n_{t-1})}}{\Gamma(\frac{1}{2}n_{t-1})}\int_0^{\infty}\lambda^{\frac{1}{2}(n_{t-1}+1)-1}\exp[-\frac{\lambda}{2}(r_{t-1}+\frac{(y_t - f_t)^2}{F_t})]d\lambda$$

$$= F_t^{-\frac{1}{2}}\frac{r_{t-1}^{(\frac{1}{2}n_{t-1})}}{\Gamma(\frac{1}{2}n_{t-1})}.\frac{\Gamma(\frac{1}{2}(n_{t-1}+1))}{[\frac{1}{2}(r_{t-1}+\frac{(y_t - f_t)^2}{F_t})]^{\frac{1}{2}(n_{t-1}+1)}}$$

(using the gamma distribution)

$$\propto F_t^{-\frac{1}{2}}r_{t-1}^{(\frac{1}{2}n_{t-1})}r_t^{-\frac{1}{2}n_t}$$ (using the definitions of (2.21) and (2.22))

i.e. $$p(y_t|\underset{\sim}{D}_{t-1}) \propto F_t^{-\frac{1}{2}} r_{t-1}^{\frac{1}{2}n_{t-1}} r_t^{-\frac{1}{2}n_t}, \qquad (2.33)$$

which has the form of a t-density (see, for example, Aitchison and Dunsmore 1975).

This is the predictive density which will be useful not only for providing predictions but also in the derivation of multi-process probabilities (see Section 3.2).

Nuisance Parameters

When the model depends upon either one or more nuisance parameters, $\underset{\sim}{\phi}$, we adopt the following procedure.

First, we assign a probability distribution to $\underset{\sim}{\phi}$ over a sensibly defined range using a discrete-valued grid. For instance, if the system model is a first order autoregressive process and the autoregressive parameter is considered as a nuisance parameter, then a suitable range would be (-1,1) and some kind of distribution would be specified over this range; for example, a flat distribution (e.g. uniform) might express prior ignorance as to the magnitude of $\underset{\sim}{\phi}$.

In this case, we have:

$$p(\underset{\sim}{\theta}_{t-1}|\underset{\sim}{D}_{t-1},\lambda,\underset{\sim}{\phi}) \sim N(\underset{\sim}{m}_{t-1},\lambda^{-1}\underset{\sim}{C}_{t-1}) \qquad (2.34)$$

and

$$p(\lambda|\underset{\sim}{D}_{t-1},\underset{\sim}{\phi}) \sim G(\tfrac{1}{2}n_{t-1},\tfrac{1}{2}r_{t-1}), \qquad (2.35)$$

where, now, $\underset{\sim}{m}_{t-1}$, $\underset{\sim}{C}_{t-1}$ and r_{t-1} are dependent on $\underset{\sim}{\phi}$, i.e. they will typically have different values at each point on the $\underset{\sim}{\phi}$-grid.

Therefore, replacing (2.19) and (2.20), we have:

$$p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda,\underset{\sim}{\phi}) \sim N(\underset{\sim}{m}_t,\lambda^{-1}\underset{\sim}{C}_t) \tag{2.36}$$

and

$$p(\lambda|\underset{\sim}{D}_t,\underset{\sim}{\phi}) \sim G(\tfrac{1}{2}n_t,\tfrac{1}{2}r_t) \tag{2.37}$$

where $\underset{\sim}{m}_t$ and $\underset{\sim}{C}_t$ are defined by (2.9), (2.10) and (2.14); and n_t and r_t are defined by (2.21), (2.22) and (2.14), except that each of the quantities in (2.14) now depends on $\underset{\sim}{\phi}$.

Let the probability distribution of $\underset{\sim}{\phi}$ at time t - 1, over a suitable grid $\underset{\sim}{\Phi}$, be represented by

$$p(\underset{\sim}{\phi}|\underset{\sim}{D}_{t-1}) \sim K_{t-1}(\underset{\sim}{\phi}). \tag{2.38}$$

Then

$$\begin{aligned} p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda) &= \sum_{\underset{\sim}{\Phi}} p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda,\underset{\sim}{\phi})p(\underset{\sim}{\phi}|\underset{\sim}{D}_t,\lambda) \\ &= \sum_{\underset{\sim}{\Phi}} p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda,\underset{\sim}{\phi}).K_t(\underset{\sim}{\phi}), \end{aligned} \tag{2.39}$$

where the first term in the summation is defined by (2.36); this replaces (2.19) when $\underset{\sim}{\phi}$ is present. Also,

$$\begin{aligned} p(\lambda|\underset{\sim}{D}_t) &= \sum_{\underset{\sim}{\Phi}} p(\lambda|\underset{\sim}{D}_t,\underset{\sim}{\phi})p(\underset{\sim}{\phi}|\underset{\sim}{D}_t) \\ &= \sum_{\underset{\sim}{\Phi}} p(\lambda|\underset{\sim}{D}_t,\underset{\sim}{\phi})K_t(\underset{\sim}{\phi}), \end{aligned} \tag{2.40}$$

where the first terms in the summation is defined by (2.37); this replaces (2.20) when $\underset{\sim}{\phi}$ is present. Moreover,

$$\begin{aligned} p(y_t|\underset{\sim}{D}_{t-1}) &= \sum_{\Phi} p(y_t|\underset{\sim}{D}_{t-1},\underset{\sim}{\phi})p(\underset{\sim}{\phi}|\underset{\sim}{D}_{t-1}) \\ &= \sum_{\Phi} p(y_t|\underset{\sim}{D}_{t-1},\underset{\sim}{\phi})K_{t-1}(\underset{\sim}{\phi}), \end{aligned} \tag{2.41}$$

where the first term in the summation is defined by (2.33) for each grid point; this replaces (2.33) when $\underset{\sim}{\phi}$ is present.

Note that (2.39) represents a mixture of normal distributions. We use the minimum Kullback-Leibler divergence criterion (Kullback and Leibler 1951) to approximate (2.39) with:

$$\left.\begin{aligned} \underset{\sim}{m}_t &= \sum_{\underset{\sim}{\Phi}} \underset{\sim}{m}_t(\underset{\sim}{\phi})K_t(\underset{\sim}{\phi}) \\ \text{and} \quad & \\ \underset{\sim}{C}_t &= \sum_{\underset{\sim}{\Phi}}\{\underset{\sim}{C}_t(\underset{\sim}{\phi}) + (\underset{\sim}{m}_t(\underset{\sim}{\phi}) - \underset{\sim}{m}_t)(\underset{\sim}{m}_t(\underset{\sim}{\phi}) - \underset{\sim}{m}_t)^T\}.K_t(\underset{\sim}{\phi}) \end{aligned}\right\} \quad (2.42)$$

Similarly (2.40) is a mixture of gamma distributions which we replace by

$$(r_t)^{-1} = \sum_{\underset{\sim}{\Phi}}(r_t(\underset{\sim}{\phi}))^{-1}K_t(\underset{\sim}{\phi}) \tag{2.43}$$

again derived by minimizing the Kullback-Leibler divergence (see West 1982 for details).

In order to calculate the quantities in (2.42) and (2.43), we need to update the grid weights iteratively, so that

$$K_t(\underset{\sim}{\phi}) = p(\underset{\sim}{\phi}|\underset{\sim}{D}_t) = \frac{p(y_t|\underset{\sim}{D}_{t-1},\underset{\sim}{\phi})p(\underset{\sim}{\phi}|\underset{\sim}{D}_{t-1})}{\sum_{\underset{\sim}{\Phi}} p(y_t|\underset{\sim}{D}_{t-1},\underset{\sim}{\phi})p(\underset{\sim}{\phi}|\underset{\sim}{D}_{t-1})}$$

(using Bayes' theorem), i.e.

$$K_t(\underset{\sim}{\phi}) = \frac{p(y_t|\underset{\sim}{D}_{t-1},\underset{\sim}{\phi})K_{t-1}(\underset{\sim}{\phi})}{\sum_{\underset{\sim}{\Phi}} p(y_t|\underset{\sim}{D}_{t-1},\underset{\sim}{\phi})K_{t-1}(\underset{\sim}{\phi})}, \tag{2.44}$$

where $p(y_t|\underset{\sim}{D}_{t-1},\underset{\sim}{\phi})$ is calculated from (2.33) for each grid-point in $\underset{\sim}{\Phi}$.

2.3 THE MODELS

2.3.1 LINEAR GROWTH

The linear growth model can be written as:

$$y_t = \mu_t + \varepsilon_t \tag{2.45}$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta\mu_t \tag{2.46}$$

$$\beta_t = \beta_{t-1} + \delta\beta_t \tag{2.47}$$

where μ_t is usually interpreted as the system level at time t, and β_t is the incremental growth (i.e. slope) at time t. Furthermore it is assumed that:

(i) $\varepsilon_t \sim N(0,\lambda^{-1} R_\varepsilon)$;

(ii) $\delta\mu_t \sim N(0,\lambda^{-1} R_\mu)$;

(iii) $\delta\beta_t \sim N(0,\lambda^{-1} R_\beta)$,

and that these perturbations are independent of one another. It is important to note that although no suffix t has been attached to the variances, they are assumed to be time-dependent.

Using the DLM representation, we may write:

$$y_t = [1 \quad 0]\begin{pmatrix}\mu_t \\ \beta_t\end{pmatrix} + \varepsilon_t \tag{2.48}$$

$$\begin{pmatrix}\mu_t \\ \beta_t\end{pmatrix} = \begin{pmatrix}1 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1} \\ \beta_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t + \delta\beta_t \\ \delta\beta_t\end{pmatrix} \tag{2.49}$$

i.e.

$$\underset{\sim}{H}_t = \underset{\sim}{H} = [1 \quad 0], \quad \underset{\sim}{\theta}_t = \begin{pmatrix}\mu_t \\ \beta_t\end{pmatrix}, \quad \underset{\sim}{G} = \begin{pmatrix}1 & 1 \\ 0 & 1\end{pmatrix}, \quad \underset{\sim}{\omega}_t = \begin{pmatrix}\delta\mu_t + \delta\beta_t \\ \delta\beta_t\end{pmatrix} \tag{2.50}$$

and

$$\underset{\sim}{R}_\varepsilon = R_\varepsilon, \quad \underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu + R_\beta & R_\beta \\ R_\beta & R_\beta \end{pmatrix} \tag{2.51}$$

Assuming (2.12) and (2.13) to hold, and writing:

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix} m_{t-1} \\ b_{t-1} \end{pmatrix} \tag{2.52}$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix} M_{t-1} & MB_{t-1} \\ MB_{t-1} & B_{t-1} \end{pmatrix} \tag{2.53}$$

we can use the Kalman filter equations of (2.14) to see that:

$$f_t = \underset{\sim}{H}_t \underset{\sim}{G} \underset{\sim}{m}_{t-1} = [1 \quad 0] \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} m_{t-1} \\ b_{t-1} \end{pmatrix} = m_{t-1} + b_{t-1}$$

$$e_t = y_t - f_t = y_t - m_{t-1} - b_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix} = \underset{\sim}{G} \underset{\sim}{C}_{t-1} \underset{\sim}{G}^T + \underset{\sim}{R}_\omega$$

Therefore

$$\begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} M_{t-1} & MB_{t-1} \\ MB_{t-1} & B_{t-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} R_\mu + R_\beta & R_\beta \\ R_\beta & R_\beta \end{pmatrix}$$

$$= \begin{pmatrix} M_{t-1} + B_{t-1} + 2MB_{t-1} + R_\mu + R_\beta & B_{t-1} + MB_{t-1} + R_\beta \\ B_{t-1} + MB_{t-1} + R_\beta & B_{t-1} + R_\beta \end{pmatrix}$$

$$F_t = \underset{\sim}{H}_t \underset{\sim}{P}_t \underset{\sim}{H}_t^T + R_\varepsilon = P_{11t} + R_\varepsilon,$$

$$\underset{\sim}{S}_t = \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} = \underset{\sim}{P}_t \underset{\sim}{H}_t^T (\underset{\sim}{F}_t)^{-1} = \begin{pmatrix} \dfrac{P_{11t}}{P_{11t} + R_\varepsilon} \\ \dfrac{P_{12t}}{P_{11t} + R_\varepsilon} \end{pmatrix}.$$

If $p(\underset{\sim}{\theta}_t | \underset{\sim}{D}_t, \lambda) \sim N(\underset{\sim}{m}_t, \lambda^{-1}\underset{\sim}{C}_t)$ then, from (2.9):

$$\underset{\sim}{m}_t = \begin{pmatrix} m_t \\ b_t \end{pmatrix} = \underset{\sim}{G}\,\underset{\sim}{m}_{t-1} + \underset{\sim}{S}_t e_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} m_{t-1} \\ b_{t-1} \end{pmatrix} + \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} (y_t - m_{t-1} - b_{t-1})$$

$$= \begin{bmatrix} m_{t-1} + b_{t-1} + S_{1t}(y_t - m_{t-1} - b_{t-1}) \\ b_{t-1} + S_{2t}(y_t - m_{t-1} - b_{t-1}) \end{bmatrix}, \tag{2.54}$$

and, from (2.10)

$$\underset{\sim}{C}_t = \underset{\sim}{P}_t - \underset{\sim}{S}_t F_t \underset{\sim}{S}_t^T = \begin{bmatrix} P_{11t} - S_{1t}^2 F_t & P_{12t} - S_{1t}S_{2t}F_t \\ P_{12t} - S_{1t}S_{2t}F_t & P_{22t} - S_{2t}^2 F_t \end{bmatrix}. \tag{2.55}$$

The updating of λ is as defined by (2.16), (2.20) to (2.22).

2.3.2 QUADRATIC GROWTH

Although the linear growth model is very useful in practice, a linear trend will not always fully describe patterns in time series. Therefore higher order polynomials may be more applicable and, in particular, allowance for a quadratic term may be desirable.

The quadratic growth model is a straightforward extension of the linear growth model, described in the previous section:

$$y_t = \mu_t + \varepsilon_t \tag{2.56}$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta\mu_t \tag{2.57}$$

$$\beta_t = \beta_{t-1} + \gamma_t + \delta\beta_t \tag{2.58}$$

$$\gamma_t = \gamma_{t-1} + \delta\gamma_t \tag{2.59}$$

where the parameters and perturbations have a similar interpretation to the linear growth model, with the additional parameter, γ_t, representing the increment in slope at time t. In addition to the assumptions of the previous section, we have a further independent disturbance $\delta\gamma_t \sim N(0, \lambda^{-1} R_\gamma)$.

In DLM form

$$y_t = [1 \quad 0 \quad 0] \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_t \end{pmatrix} + \varepsilon_t \tag{2.60}$$

$$\begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_t + \delta\beta_t + \delta\gamma_t \\ \delta\beta_t + \delta\gamma_t \\ \delta\gamma_t \end{pmatrix}, \tag{2.61}$$

i.e.

$$\underset{\sim}{H}_t = \underset{\sim}{H} = [1 \quad 0 \quad 0], \quad \underset{\sim}{\theta}_t = \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_t \end{pmatrix}, \quad \underset{\sim}{G} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$
$$\underset{\sim}{\omega}_t = \begin{pmatrix} \delta\mu_t + \delta\beta_t + \delta\gamma_t \\ \delta\beta_t + \delta\gamma_t \\ \delta\gamma_t \end{pmatrix} \tag{2.62}$$

and

$$\underset{\sim}{R}_\varepsilon = R_\varepsilon, \quad \underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu + R_\beta + R_\gamma & R_\beta + R_\gamma & R_\gamma \\ R_\beta + R_\gamma & R_\beta + R_\gamma & R_\gamma \\ R_\gamma & R_\gamma & R_\gamma \end{pmatrix} \tag{2.63}$$

If

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix} m_{t-1} \\ b_{t-1} \\ g_{t-1} \end{pmatrix} \tag{2.64}$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix} M_{t-1} & MB_{t-1} & MG_{t-1} \\ MB_{t-1} & B_{t-1} & BG_{t-1} \\ MG_{t-1} & BG_{t-1} & G_{t-1} \end{pmatrix}, \tag{2.65}$$

then, from (2.14), we have:

$$f_t = [1 \quad 0 \quad 0]\begin{pmatrix}1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}m_{t-1}\\ b_{t-1}\\ g_{t-1}\end{pmatrix} = m_{t-1} + b_{t-1} + g_{t-1}$$

$$e_t = y_t - f_t = y_t - m_{t-1} - b_{t-1} - g_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix}P_{11t} & P_{12t} & P_{13t}\\ P_{12t} & P_{22t} & P_{23t}\\ P_{13t} & P_{23t} & P_{33t}\end{pmatrix} = \begin{pmatrix}1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}M_{t-1} & MB_{t-1} & MG_{t-1}\\ MB_{t-1} & B_{t-1} & BG_{t-1}\\ MG_{t-1} & BG_{t-1} & G_{t-1}\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ 1 & 1 & 1\end{pmatrix}$$

$$+ \begin{pmatrix}R_\mu + R_\beta + R_\gamma & R_\beta + R_\gamma & R_\gamma\\ R_\beta + R_\gamma & R_\beta + R_\gamma & R_\gamma\\ R_\gamma & R_\gamma & R_\gamma\end{pmatrix}$$

i.e.

$$P_{11t} = M_{t-1} + B_{t-1} + G_{t-1} + 2(MB_{t-1} + MG_{t-1} + BG_{t-1}) + R_\mu + R_\beta + R_\gamma$$

$$P_{12t} = B_{t-1} + G_{t-1} + MB_{t-1} + MG_{t-1} + 2BG_{t-1} + R_\beta + R_\gamma$$

$$P_{13t} = G_{t-1} + MG_{t-1} + BG_{t-1} + R_\gamma$$

$$P_{22t} = B_{t-1} + G_{t-1} + 2BG_{t-1} + R_\beta + R_\gamma$$

$$P_{23t} = G_{t-1} + BG_{t-1} + R_\gamma$$

$$P_{33t} = G_{t-1} + R_\gamma$$

$$F_t = [1 \quad 0 \quad 0]\begin{pmatrix}P_{11t} & P_{12t} & P_{13t}\\ P_{12t} & P_{22t} & P_{23t}\\ P_{13t} & P_{23t} & P_{33t}\end{pmatrix}\begin{pmatrix}1\\ 0\\ 0\end{pmatrix} + R_\varepsilon = P_{11t} + R_\varepsilon .$$

$$\underset{\sim}{S}_t = \begin{pmatrix}S_{1t}\\ S_{2t}\\ S_{3t}\end{pmatrix} = \begin{pmatrix}P_{11t} & P_{12t} & P_{13t}\\ P_{12t} & P_{22t} & P_{23t}\\ P_{13t} & P_{23t} & P_{33t}\end{pmatrix}\begin{pmatrix}1\\ 0\\ 0\end{pmatrix}\frac{1}{P_{11t} + R_\varepsilon} = \begin{bmatrix}\dfrac{P_{11t}}{P_{11t} + R_\varepsilon}\\ \dfrac{P_{12t}}{P_{11t} + R_\varepsilon}\\ \dfrac{P_{13t}}{P_{11t} + R_\varepsilon}\end{bmatrix}$$

Then

$$\underset{\sim}{m}_t = \begin{pmatrix} m_t \\ b_t \\ g_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} m_{t-1} \\ b_{t-1} \\ g_{t-1} \end{pmatrix} + \begin{pmatrix} S_{1t} \\ S_{2t} \\ S_{3t} \end{pmatrix} (y_t - m_{t-1} - b_{t-1} - g_{t-1})$$

$$= \begin{pmatrix} m_{t-1} + b_{t-1} + g_{t-1} + S_{1t}(y_t - m_{t-1} - b_{t-1} - g_{t-1}) \\ b_{t-1} + g_{t-1} + S_{2t}(y_t - m_{t-1} - b_{t-1} - g_{t-1}) \\ g_{t-1} + S_{3t}(y_t - m_{t-1} - b_{t-1} - g_{t-1}) \end{pmatrix} \qquad (2.66)$$

and

$$\underset{\sim}{C}_t = \begin{pmatrix} M_t & MB_t & MG_t \\ MB_t & B_t & BG_t \\ MG_t & BG_t & G_t \end{pmatrix}$$

$$= \begin{pmatrix} P_{11t} - S_{1t}^2 F_t & P_{12t} - S_{1t}S_{2t}F_t & P_{13t} - S_{1t}S_{3t}F_t \\ P_{12t} - S_{1t}S_{2t}F_t & P_{22t} - S_{2t}^2 F_t & P_{23t} - S_{2t}S_{3t}F_t \\ P_{13t} - S_{1t}S_{3t}F_t & P_{23t} - S_{2t}S_{3t}F_t & P_{33t} - S_{3t}^2 F_t \end{pmatrix} \qquad (2.67)$$

Again, λ updating is defined by (2.16) and (2.20) to (2.22).

2.3.3 A SINUSOIDAL MODEL

Many medical time series exhibit rhythmic behaviour, possibly due to the existence of, so-called, body-clocks which synchronize various bodily functions (More-Ede, Sulzman and Fuller 1982). Cyclic patterns in the human often have a periodicity of about twenty-four hours, reflecting the light/dark, activity/rest phases of daily life (Minors and Waterhouse 1981), although some rhythms, such as respiratory patterns (see, for example, Hrushesky 1984), will clearly have a shorter periodicity and others, such as menstrual cycles (see, for example, Rebar and Yen 1979), a much longer periodicity.

For this reason cyclical models have considerable application to medical time series. In this section, we consider the simplest such model, the sinusoidal waveform. Ideally, we would like to be able to estimate the characteristics of the waveform independently of one another, and so the following configuration is adopted:

$$y_t = \mu_t + c_t\alpha_t + \varepsilon_t \tag{2.68}$$

$$\mu_t = \mu_{t-1} + \delta\mu_t \tag{2.69}$$

$$\alpha_t = \alpha_{t-1} + \delta\alpha_t \tag{2.70}$$

where

$$\left.\begin{aligned} c_t &= \cos(2\pi\omega t + \phi) \\ \omega &= \text{the rhythm frequency} \\ \phi &= \text{the phase} \\ \varepsilon_t &\sim N(0,\lambda^{-1}R_\varepsilon) \\ \delta\mu_t &\sim N(0,\lambda^{-1}R_\mu) \\ \delta\alpha_t &\sim N(0,\lambda^{-1}R_\alpha) \end{aligned}\right\} \tag{2.71}$$

and where μ_t can be interpreted as the <u>level</u> of the series at time t, with α_t representing the rhythm <u>amplitude</u> at time t.

We will often be able to assume that ω is fixed and known (for instance, when a period of 24 hours is specified), whereas it is unlikely that we will be able to stipulate an accurate value for ϕ. Therefore, we adopt the procedure described in Section 2.2, and use a discrete-valued grid to recursively update our beliefs about ϕ. In the absence of further information the range adopted for this purpose is $[0,2\pi)$, representing all possible values for ϕ.

It is, of course, possible to include ω in this updating procedure, by way of a two-dimensional grid. In this case the range for ω will depend upon some prior knowledge of the supposed periodicity and the frequency of data sampling. For example, for hourly measurements of a debatable daily rhythm, a suitably wide range for ω might be $[1/96,1)$, having a periodicity between 1 and 96 hours. Probabilities are then assigned to (ω,ϕ) co-ordinates in such a way that the marginals for ω and ϕ are easily computed (see Appendix A2.2).

Writing (2.68) to (2.70) in DLM form, we have:

$$y_t = [1 \quad c_t]\begin{pmatrix}\mu_t \\ \alpha_t\end{pmatrix} + \varepsilon_t \tag{2.72}$$

$$\begin{pmatrix}\mu_t \\ \alpha_t\end{pmatrix} = \begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1} \\ \alpha_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t \\ \delta\alpha_t\end{pmatrix} \tag{2.73}$$

i.e.

$$\left.\begin{array}{l} \underset{\sim}{H}_t = [1 \quad c_t] = [1 \quad \cos(2\pi\omega t + \phi)] \\ \underset{\sim}{\theta}_t = \begin{pmatrix}\mu_t \\ \alpha_t\end{pmatrix}, \quad \underset{\sim}{G} = \begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}, \quad \underset{\sim}{\omega}_t = \begin{pmatrix}\delta\mu_t \\ \delta\alpha_t\end{pmatrix} \end{array}\right\} \tag{2.74}$$

NOTES:

(i) The form of $\underset{\sim}{H}_t$ is different from that of (2.31), since we wish to keep the estimation of rhythm amplitude independent of the rhythm level;

(ii) although $\underset{\sim}{H}_t$ is not constant, it is known for all times t, for each (ω,ϕ) pairing (or for each value of ϕ, if ω is assumed to be fixed).

Also,

$$\underset{\sim}{R}_\varepsilon = R_\varepsilon, \qquad \underset{\sim}{R}_\omega = \begin{pmatrix}R_\mu & 0 \\ 0 & R_\alpha\end{pmatrix}. \tag{2.75}$$

If

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix} m_{t-1} \\ a_{t-1} \end{pmatrix} \tag{2.76}$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix} M_{t-1} & MA_{t-1} \\ MA_{t-1} & A_{t-1} \end{pmatrix} \tag{2.77}$$

then:

$$f_t = [1 \quad c_t] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} m_{t-1} \\ a_{t-1} \end{pmatrix} = m_{t-1} + c_t a_{t-1}$$

$$e_t = y_t - f_t = y_t - m_{t-1} - c_t a_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} M_{t-1} & MA_{t-1} \\ MA_{t-1} & A_{t-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} R_\mu & 0 \\ 0 & R_\alpha \end{pmatrix}$$

$$= \begin{pmatrix} M_{t-1} + R_\mu & MA_{t-1} \\ MA_{t-1} & A_{t-1} + R_\alpha \end{pmatrix}$$

$$F_t = [1 \quad c_t] \begin{pmatrix} M_{t-1} + R_\mu & MA_{t-1} \\ MA_{t-1} & A_{t-1} + R_\alpha \end{pmatrix} \begin{pmatrix} 1 \\ c_t \end{pmatrix} + R_\varepsilon$$

$$= M_{t-1} + 2c_t MA_{t-1} + c_t^2 A_{t-1} + R_\mu + c_t^2 R_\alpha + R_\varepsilon$$

$$\underset{\sim}{S}_t = \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} = \begin{pmatrix} M_{t-1} + R_\mu & MA_{t-1} \\ MA_{t-1} & A_{t-1} + R_\alpha \end{pmatrix} \begin{pmatrix} 1 \\ c_t \end{pmatrix} \frac{1}{F_t}$$

$$= \begin{pmatrix} \dfrac{M_{t-1} + c_t MA_{t-1} + R_\mu}{F_t} \\ \dfrac{MA_{t-1} + c_t A_{t-1} + c_t R_\alpha}{F_t} \end{pmatrix}$$

So,

$$\underset{\sim}{m}_t = \begin{pmatrix} m_t \\ a_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} m_{t-1} \\ a_{t-1} \end{pmatrix} + \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix}(y_t - m_{t-1} + c_t a_{t-1})$$

$$= \begin{pmatrix} m_{t-1} + S_{1t}(y_t - m_{t-1} - c_t a_{t-1}) \\ a_{t-1} + S_{2t}(y_t - m_{t-1} - c_t a_{t-1}) \end{pmatrix} \tag{2.78}$$

and

$$\underset{\sim}{C}_t = \begin{pmatrix} M_t & MA_t \\ MA_t & A_t \end{pmatrix} = \begin{pmatrix} M_{t-1} + R_\mu - S_{1t}^2 F_t & MA_{t-1} - S_{1t}S_{2t}F_t \\ MA_{t-1} - S_{1t}S_{2t}F_t & A_{t-1} + R_\alpha - S_{2t}^2 F_t \end{pmatrix} \tag{2.79}$$

Because $\underset{\sim}{\phi}$ is present, as either ϕ or (ω,ϕ), we use (2.35), (2.37), (2.40), (2.21) and (2.22) to update λ, and (2.38) and (2.44) to update $\underset{\sim}{\phi}$.

2.3.4 A SINUSOIDAL MODEL WITH LINEAR GROWTH

The sinusoidal model of the previous section can be extended to include a term for slope, resulting in a superposition of the sinusoidal model with the linear growth model of Section 2.3.1.

In DLM form:

$$y_t = [1 \quad 0 \quad c_t]\begin{pmatrix} \mu_t \\ \beta_t \\ \alpha_t \end{pmatrix} + \varepsilon_t \tag{2.80}$$

$$\begin{pmatrix} \mu_t \\ \beta_t \\ \alpha_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \alpha_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_t + \delta\beta_t \\ \delta\beta_t \\ \delta\alpha_t \end{pmatrix}, \tag{2.81}$$

i.e.

$$\left.\begin{aligned} &\underset{\sim}{H}_t = [1 \quad 0 \quad c_t] = [1 \quad 0 \quad \cos(2\pi\omega t + \phi)] \\ &\underset{\sim}{\theta}_t = \begin{pmatrix} \mu_t \\ \beta_t \\ \alpha_t \end{pmatrix}, \quad \underset{\sim}{G} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \underset{\sim}{\omega}_t = \begin{pmatrix} \delta\mu_t + \delta\beta_t \\ \delta\beta_t \\ \delta\alpha_t \end{pmatrix}, \end{aligned}\right\} \quad (2.82)$$

$$\underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu + R_\beta & R_\beta & 0 \\ R_\beta & R_\beta & 0 \\ 0 & 0 & R_\alpha \end{pmatrix}, \quad \underset{\sim}{R}_\varepsilon = R_\varepsilon \quad (2.83)$$

If

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix} m_{t-1} \\ b_{t-1} \\ a_{t-1} \end{pmatrix} \quad (2.84)$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix} M_{t-1} & MB_{t-1} & MA_{t-1} \\ MB_{t-1} & B_{t-1} & BA_{t-1} \\ MA_{t-1} & BA_{t-1} & A_{t-1} \end{pmatrix} \quad (2.85)$$

then

$$f_t = m_{t-1} + b_{t-1} + c_t a_{t-1}$$
$$e_t = y_t - m_{t-1} - b_{t-1} - c_t a_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix} P_{11t} & P_{12t} & P_{13t} \\ P_{12t} & P_{22t} & P_{23t} \\ P_{13t} & P_{23t} & P_{33t} \end{pmatrix}$$

$$= \begin{pmatrix} M_{t-1} + B_{t-1} + 2MB_{t-1} + R_\mu + R_\beta & B_{t-1} + MB_{t-1} + R_\beta & MA_{t-1} + BA_{t-1} \\ B_{t-1} + MB_{t-1} + R_\beta & B_{t-1} + R_\beta & BA_{t-1} \\ MA_{t-1} + BA_{t-1} & BA_{t-1} & A_{t-1} + R_\alpha \end{pmatrix}$$

$$F_t = P_{11t} + 2c_t P_{13t} + c_t^2 P_{33t} + R_\varepsilon$$

$$\underset{\sim}{S}_t = \begin{pmatrix} S_{1t} \\ S_{2t} \\ S_{3t} \end{pmatrix} = \begin{pmatrix} \dfrac{P_{11t} + c_t P_{13t}}{F_t} \\ \dfrac{P_{12t} + c_t P_{23t}}{F_t} \\ \dfrac{P_{13t} + c_t P_{33t}}{F_t} \end{pmatrix}$$

Then

$$\begin{pmatrix} m_t \\ b_t \\ a_t \end{pmatrix} = \begin{pmatrix} m_{t-1} + b_{t-1} + S_{1t}(y_t - m_{t-1} - b_{t-1} - c_t a_{t-1}) \\ b_{t-1} + S_{2t}(y_t - m_{t-1} - b_{t-1} - c_t a_{t-1}) \\ a_{t-1} + S_{3t}(y_t - m_{t-1} - b_{t-1} - c_t a_{t-1}) \end{pmatrix} \qquad (2.86)$$

and

$$\begin{pmatrix} M_t & MB_t & MA_t \\ MB_t & B_t & BA_t \\ MA_t & BA_t & A_t \end{pmatrix} = \begin{pmatrix} P_{11t} - S_{1t}^2 F_t & P_{12t} - S_{1t}S_{2t}F_t & P_{13t} - S_{1t}S_{3t}F_t \\ P_{12t} - S_{1t}S_{2t}F_t & P_{22t} - S_{2t}^2 F_t & P_{23t} - S_{2t}S_{3t}F_t \\ P_{13t} - S_{1t}S_{3t}F_t & P_{23t} - S_{2t}S_{3t}F_t & P_{33t} - S_{3t}^2 F_t \end{pmatrix}$$

(2.87)

$\underset{\sim}{\phi}$ and λ are updated as for the basic sinusoidal model of the previous section. This model may be useful when, for instance, the day-to-day variations consist of a linear trend along with a within-day rhythm.

2.3.5 ARMA MODELS

Patterns in time series are not always obvious, but intuitively one might imagine that 'today's' observation is somehow related to the recent history of measurements; the value of a medical variable is very unlikely to double (say) overnight. In these situations a low-order ARMA model (Box and Jenkins 1970) may provide an adequate description of the structure in the data.

Again, it is pointed out that, in this thesis, the structure is assumed to reside in the system equations. Godolphin and Harrison (1975) have demonstrated that this leads to a different structure for the observations themselves, but we are not concerned with the observation structure here (see Chapter 4 for remarks about model identification using series of observations). Therefore, a grid for an ARMA parameter is specified over the appropriate stationarity region of the parameter space, without the additional restrictions suggested by Godolphin and Harrison (1975).

2.3.5.1: AR(1). The first-order autoregressive model is as follows:

$$y_t = \mu_t + \varepsilon_t \tag{2.88}$$

$$\mu_t - \nu_t = \phi(\mu_{t-1} - \nu_{t-1}) + \delta\mu_t \tag{2.89}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t \tag{2.90}$$

where μ_t can be interpreted as the true (error-free) recording at time t, with ν_t representing the level of the series at time t; ϕ is the autoregressive parameter which, for this model, is considered as a nuisance parameter. Also, we assume that: $\varepsilon_t \sim N(0,\lambda^{-1}R_\varepsilon)$; $\delta\mu_t \sim N(0,\lambda^{-1}R_\mu)$; $\delta\nu_t \sim N(0,\lambda^{-1}R_\nu)$.

The grid method is used to update ϕ, using the 'natural' range (-1,1) corresponding to the range which results in stationarity.

NOTE: (2.90) allows the level to fluctuate in any case, but should we wish other forms of non-stationarity we could incorporate an extra component into the model in a similar way to the inclusion of a slope component in the sinusoidal model (see previous section), thereby restricting ϕ to the stationary range even for non-stationary time series.

In DLM form we have:

$$y_t = [1 \quad 0]\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} + \varepsilon_t \tag{2.91}$$

$$\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} = \begin{pmatrix}\phi & 1-\phi \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1} \\ \nu_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t + \delta\nu_t \\ \delta\nu_t\end{pmatrix} \tag{2.92}$$

i.e.

$$\left.\begin{aligned} &\underset{\sim}{H}_t = \underset{\sim}{H} = [1 \quad 0], \quad \underset{\sim}{\theta}_t = \begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix}, \quad \underset{\sim}{G} = \begin{pmatrix}\phi & 1-\phi \\ 0 & 1\end{pmatrix}, \\ &\underset{\sim}{\omega}_t = \begin{pmatrix}\delta\mu_t + \delta\nu_t \\ \delta\nu_t\end{pmatrix} \end{aligned}\right\} \tag{2.93}$$

and

$$\underset{\sim}{R}_\varepsilon = R_\varepsilon, \quad \underset{\sim}{R}_\omega = \begin{pmatrix}R_\mu + R_\nu & R_\nu \\ R_\nu & R_\nu\end{pmatrix}. \tag{2.94}$$

If

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix}m_{t-1} \\ v_{t-1}\end{pmatrix} \tag{2.95}$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix}M_{t-1} & MV_{t-1} \\ MV_{t-1} & V_{t-1}\end{pmatrix}, \tag{2.96}$$

then

$$f_t = [1 \quad 0]\begin{pmatrix}\phi & 1-\phi \\ 0 & 1\end{pmatrix}\begin{pmatrix}m_{t-1} \\ v_{t-1}\end{pmatrix} = \phi m_{t-1} + (1-\phi)v_{t-1}$$

$$e_t = y_t - \phi m_{t-1} - (1-\phi)v_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix} = \begin{pmatrix} \phi & 1 - \phi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} M_{t-1} & MV_{t-1} \\ MV_{t-1} & V_{t-1} \end{pmatrix} \begin{pmatrix} \phi & 0 \\ 1 - \phi & 1 \end{pmatrix}$$

$$+ \begin{pmatrix} R_\mu + R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix}$$

that is, $\begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix}$ is equal to

$$\begin{pmatrix} \phi^2 M_{t-1} + 2\phi(1 - \phi)MV_{t-1} + (1 - \phi)^2 V_{t-1} + R_\mu + R_\nu & (1 - \phi)V_t + \phi MV_{t-1} + R_\nu \\ (1 - \phi)V_{t-1} + \phi MV_{t-1} + R_\nu & V_{t-1} + R_\nu \end{pmatrix}$$

$$F_t = [1 \quad 0] \begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + R_\varepsilon = P_{11t} + R_\varepsilon$$

$$\underset{\sim}{S}_t = \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} = \begin{pmatrix} \dfrac{P_{11t}}{P_{11t} + R_\varepsilon} \\ \\ \dfrac{P_{12t}}{P_{11t} + R_\varepsilon} \end{pmatrix}$$

So,

$$\underset{\sim}{m}_t = \begin{pmatrix} m_t \\ v_t \end{pmatrix} = \begin{pmatrix} \phi & 1 - \phi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} m_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} e_t$$

$$= \begin{pmatrix} \phi m_{t-1} + (1 - \phi)\, v_{t-1} + S_{1t}(y_t - \phi m_{t-1} - (1 - \phi)\, v_{t-1}) \\ v_{t-1} + S_{2t}(y_t - \phi m_{t-1} - (1 - \phi)\, v_{t-1}) \end{pmatrix} \tag{2.97}$$

$$\underset{\sim}{C}_t = \begin{pmatrix} M_t & MV_t \\ MV_t & V_t \end{pmatrix} = \begin{pmatrix} P_{11t} - S_{1t}^2 F_t & P_{12t} - S_{1t} S_{2t} F_t \\ P_{12t} - S_{1t} S_{2t} F_t & P_{22t} - S_{2t}^2 F_t \end{pmatrix} \tag{2.98}$$

λ and ϕ are updated as in Section 2.3.3 (see Appendix A2.1 for initial setting of the ϕ grid).

2.3.5.2: MA(1). The following configuration is proposed for the first-order moving average model:

$$y_t = \mu_t + \varepsilon_t \tag{2.99}$$

$$\mu_t = \nu_t + \delta\mu_t - \eta\delta\mu_{t-1} \tag{2.100}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t \tag{2.101}$$

where μ_t is the (error-free) time series reading at time t, ν_t is the level of the series at time t and η is the moving average parameter. Also, $\varepsilon_t \sim N(0,\lambda^{-1}R_\varepsilon)$, $\delta\mu_t \sim N(0,\lambda^{-1}R_\mu)$ and $\delta\nu_t \sim N(0,\lambda^{-1}R_\nu)$. Once more, a suitable range for a grid for η is (-1,1), corresponding to the conditions for stationarity.

In DLM form:

$$y_t = [1 \quad 0]\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} + \varepsilon_t \tag{2.102}$$

$$\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} = \begin{pmatrix}0 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1} \\ \nu_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t - \eta\delta\mu_{t-1} + \delta\nu_t \\ \delta\nu_t\end{pmatrix} \tag{2.103}$$

i.e.

$$\left.\begin{aligned} &\underset{\sim}{H}_t = \hat{H} = [1 \quad 0], \qquad \underset{\sim}{\theta}_t = \begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix}, \qquad \underset{\sim}{G} = \begin{pmatrix}0 & 1 \\ 0 & 1\end{pmatrix}, \\ &\underset{\sim}{\omega}_t = \begin{pmatrix}\delta\mu_t - \eta\delta\mu_{t-1} + \delta\nu_t \\ \delta\nu_t\end{pmatrix}\end{aligned}\right\} \tag{2.104}$$

and

$$\underset{\sim}{R}_{\varepsilon} = R_{\varepsilon}, \qquad \underset{\sim}{R}_{\omega} = \begin{pmatrix} R_{\mu} + \eta^2 R_{\mu} + R_{\nu} & R_{\nu} \\ R_{\nu} & R_{\nu} \end{pmatrix} \tag{2.105}$$

NOTE: With $\phi = 0$ in the AR(1) model of Section 2.3.5.1, the $\underset{\sim}{G}$ matrix is identical to the above, since the moving average structure is wholly contained in the $\underset{\sim}{\omega}_t$ vector.

If

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix} m_{t-1} \\ v_{t-1} \end{pmatrix} \tag{2.106}$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix} M_{t-1} & MV_{t-1} \\ MV_{t-1} & V_{t-1} \end{pmatrix} \tag{2.107}$$

then

$$f_t = [1 \quad 0]\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} m_{t-1} \\ v_{t-1} \end{pmatrix} = v_{t-1}$$

$$e_t = y_t - v_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix} = \begin{pmatrix} V_{t-1} + R_{\mu} + \eta^2 R_{\mu} + R_{\nu} & V_{t-1} + R_{\nu} \\ V_{t-1} + R_{\nu} & V_{t-1} + R_{\nu} \end{pmatrix}$$

$$F_t = P_{11t} + R_{\varepsilon}$$

$$\underset{\sim}{S}_t = \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} = \begin{pmatrix} \dfrac{P_{11t}}{P_{11t} + R_{\varepsilon}} \\ \dfrac{P_{12t}}{P_{11t} + R_{\varepsilon}} \end{pmatrix}$$

Then

$$\underset{\sim}{m}_t = \begin{pmatrix} m_t \\ v_t \end{pmatrix} = \begin{pmatrix} v_{t-1} + S_{1t}(y_t - v_{t-1}) \\ v_{t-1} + S_{2t}(y_t - v_{t-1}) \end{pmatrix} \tag{2.108}$$

and

$$\underset{\sim}{C}_t = \begin{pmatrix} M_t & MV_t \\ MV_t & V_t \end{pmatrix} = \begin{pmatrix} P_{11t} - S_{1t}^2 F_t & P_{12t} - S_{1t} S_{2t} F_t \\ P_{12t} - S_{1t} S_{2t} F_t & P_{22t} - S_{2t}^2 F_t \end{pmatrix} \tag{2.109}$$

λ and η are updated as for Section 2.3.3.

2.3.5.3: AR(2). As a straightforward extension of the AR(1) model, the AR(2) model is as follows:

$$y_t = \mu_t + \varepsilon_t \tag{2.110}$$

$$\mu_t - \nu_t = \phi_1(\mu_{t-1} - \nu_{t-1}) + \phi_2(\mu_{t-2} - \nu_{t-2}) + \delta\mu_t \tag{2.111}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t, \tag{2.112}$$

where μ_t is the (error-free) recording at time t, ν_t is the level at time t, and ϕ_1, ϕ_2 are the autoregressive parameters, with: $\varepsilon_t \sim N(0,\lambda^{-1}R_\varepsilon)$, $\delta\mu_t \sim N(0,\lambda^{-1}R_\mu)$, $\delta\nu_t \sim N(0,\lambda^{-1}R_\nu)$.

The region for the (ϕ_1,ϕ_2) grid is again specified by referring to the stationarity conditions: $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, $|\phi_2| < 1$.

In DLM form we may write:

$$y_t = [1 \quad 0 \quad 0] \begin{pmatrix} \mu_t \\ \mu_{t-1} \\ \nu_t \end{pmatrix} + \varepsilon_t \tag{2.113}$$

$$\begin{pmatrix} \mu_t \\ \mu_{t-1} \\ \nu_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & 1 - \phi_1 - \phi_2 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \mu_{t-2} \\ \nu_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_t + \phi_2\delta\nu_{t-1} + \delta\nu_t \\ 0 \\ \delta\nu_t \end{pmatrix} \tag{2.114}$$

i.e.

$$\left.\begin{aligned} &\underset{\sim}{H}_t = \underset{\sim}{H} = [1 \quad 0 \quad 0], \quad \underset{\sim}{\theta}_t = \begin{pmatrix} \mu_t \\ \mu_{t-1} \\ \nu_t \end{pmatrix}, \quad \underset{\sim}{G} = \begin{pmatrix} \phi_1 & \phi_2 & 1 - \phi_1 - \phi_2 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ &\underset{\sim}{\omega}_t = \begin{pmatrix} \delta\mu_t + \phi_2\delta\nu_{t-1} + \delta\nu_t \\ 0 \\ \delta\nu_t \end{pmatrix} \end{aligned}\right\} \quad (2.115)$$

and

$$\underset{\sim}{R}_\varepsilon = R_\varepsilon, \qquad \underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu + \phi_2^2 R_\nu + R_\nu & 0 & R_\nu \\ 0 & 0 & 0 \\ R_\nu & 0 & R_\nu \end{pmatrix} \qquad (2.116)$$

If

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix} m1_{t-1} \\ m0_{t-1} \\ v_{t-1} \end{pmatrix} \qquad (2.117)$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix} M1_{t-1} & MM_{t-1} & MV1_{t-1} \\ MM_{t-1} & MO_{t-1} & MVO_{t-1} \\ MV1_{t-1} & MVO_{t-1} & V_{t-1} \end{pmatrix} \qquad (2.118)$$

then

$$f_t = \phi_1 m1_{t-1} + \phi_2 m0_{t-1} + (1 - \phi_1 - \phi_2)v_{t-1}$$

$$e_t = y_t - \phi_1 m1_{t-1} - \phi_2 m0_{t-1} - (1 - \phi_1 - \phi_2)v_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix} P_{11t} & P_{12t} & P_{13t} \\ P_{12t} & P_{22t} & P_{23t} \\ P_{13t} & P_{23t} & P_{33t} \end{pmatrix}$$

where

$$\begin{aligned} P_{11t} = {} & \phi_1^2 M1_{t-1} + 2\phi_1\phi_2 MM_{t-1} + 2\phi_1(1 - \phi_1 - \phi_2)MV1_{t-1} \\ & + \phi_2^2 MO_{t-1} + 2\phi_2(1 - \phi_1 - \phi_2)MVO_{t-1} + (1 - \phi_1 - \phi_2)^2 V_{t-1} \\ & + R_\mu + R_\nu(1 + \phi_2^2) \end{aligned}$$

$$P_{12t} = \phi_1 M1_{t-1} + \phi_2 MM_{t-1} + (1 - \phi_1 - \phi_2)MV1_{t-1}$$

$$P_{13t} = \phi_1 MV1_{t-1} + \phi_2 MVO_{t-1} + (1 - \phi_1 - \phi_2)V_{t-1} + R_\nu$$

$$P_{22t} = M1_{t-1}$$

$$P_{23t} = MV1_{t-1}$$

$$P_{33t} = V_{t-1} + R_\nu$$

$$F_t = P_{11t} + R_\varepsilon$$

$$\underset{\sim}{S}_t = \begin{pmatrix} S_{1t} \\ S_{2t} \\ S_{3t} \end{pmatrix} = \begin{pmatrix} P_{11t}/F_t \\ P_{12t}/F_t \\ P_{13t}/F_t \end{pmatrix}$$

So,

$$\underset{\sim}{m}_t = \begin{pmatrix} m1_t \\ m0_t \\ v_t \end{pmatrix} = \begin{pmatrix} \phi_1 m1_{t-1} + \phi_2 m0_{t-1} + (1 - \phi_1 - \phi_2)v_{t-1} + S_{1t}e_t \\ m1_{t-1} + S_{2t}e_t \\ v_{t-1} + S_{3t}e_t \end{pmatrix} \tag{2.119}$$

and

$$\underset{\sim}{C}_t = \begin{pmatrix} M1_t & MM_t & MV1_t \\ MM_t & MO_t & MVO_t \\ MV1_t & MVO_t & V_t \end{pmatrix}$$

$$= \begin{pmatrix} P_{11t} - S_{1t}^2 F_t & P_{12t} - S_{1t}S_{2t}F_t & P_{13t} - S_{1t}S_{3t}F_t \\ P_{12t} - S_{1t}S_{2t}F_t & P_{22t} - S_{2t}^2 F_t & P_{23t} - S_{2t}S_{3t}F_t \\ P_{13t} - S_{1t}S_{3t}F_t & P_{23t} - S_{2t}S_{3t}F_t & P_{33t} - S_{3t}^2 F_t \end{pmatrix} \tag{2.120}$$

λ and $\underset{\sim}{\phi}$ are updated as for Section 2.3.3; see Appendix A2.2 for the initial setting of the $\underset{\sim}{\phi}$ grid.

2.3.5.4: MA(2). The MA(2) model is an extension of the MA(1) model of Section 2.3.5.2:

$$y_t = \mu_t + \varepsilon_t \tag{2.121}$$

$$\mu_t = \nu_t + \delta\mu_t - \eta_1\delta\mu_{t-1} - \eta_2\delta\mu_{t-2} \tag{2.122}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t \tag{2.123}$$

where μ_t is the (error-free) recording at time t, ν_t is the level at time t and η_1,η_2 are the moving average parameters, updated by a grid satisfying the stationarity conditions $\eta_1 + \eta_2 < 1$, $\eta_2 - \eta_1 < 1$, $|\eta_2| < 1$. Also, $\varepsilon_t \sim N(0,\lambda^{-1}R_\varepsilon)$, $\delta\mu_t \sim N(0,\lambda^{-1}R_\mu)$ and $\delta\nu_t \sim N(0,\lambda^{-1}R_\nu)$. So, in DLM form, we have:

$$y_t = [1 \quad 0]\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} + \varepsilon_t \tag{2.124}$$

$$\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} = \begin{pmatrix}0 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1} \\ \nu_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t - \eta_1\delta\mu_{t-1} - \eta_2\delta\mu_{t-2} + \delta\nu_t \\ \delta\nu_t\end{pmatrix} \tag{2.125}$$

i.e.

$$\left.\begin{aligned} &\underset{\sim}{H}_t = \underset{\sim}{H} = [1 \quad 0], \quad \underset{\sim}{\theta}_t = \begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix}, \quad \underset{\sim}{G} = \begin{pmatrix}0 & 1 \\ 0 & 1\end{pmatrix}, \\ &\underset{\sim}{\omega}_t = \begin{pmatrix}\delta\mu_t - \eta_1\delta\mu_{t-1} - \eta_2\delta\mu_{t-2} + \delta\nu_t \\ \delta\nu_t\end{pmatrix} \end{aligned}\right\} \tag{2.126}$$

and

$$\underset{\sim}{R}_\varepsilon = R_\varepsilon, \quad \underset{\sim}{R}_\omega = \begin{pmatrix}R_\mu + \eta_1^2 R_\mu + \eta_2^2 R_\mu + R_\nu & R_\nu \\ R_\nu & R_\nu\end{pmatrix} \tag{2.127}$$

If

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix}m_{t-1} \\ v_{t-1}\end{pmatrix} \tag{2.128}$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix}M_{t-1} & MV_{t-1} \\ MV_{t-1} & V_{t-1}\end{pmatrix} \tag{2.129}$$

then,

$$f_t = v_{t-1}$$

$$e_t = y_t - v_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix} = \begin{pmatrix} V_{t-1} + R_\mu + \eta_1^2 R_\mu + \eta_2^2 R_\mu + R_\nu & V_{t-1} + R_\nu \\ V_{t-1} + R_\nu & V_{t-1} + R_\nu \end{pmatrix}$$

$$F_t = P_{11t} + R_\varepsilon$$

$$\underset{\sim}{S}_t = \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} = \begin{pmatrix} \dfrac{P_{11t}}{P_{11t} + R_\varepsilon} \\ \dfrac{P_{12t}}{P_{11t} + R_\varepsilon} \end{pmatrix}$$

So

$$\underset{\sim}{m}_t = \begin{pmatrix} m_t \\ v_t \end{pmatrix} = \begin{pmatrix} v_{t-1} + S_{1t}(y_t - v_{t-1}) \\ v_{t-1} + S_{2t}(y_t - v_{t-1}) \end{pmatrix} \tag{2.130}$$

and

$$\underset{\sim}{C}_t = \begin{pmatrix} M_t & MV_t \\ MV_t & V_t \end{pmatrix} = \begin{pmatrix} P_{11t} - S_{1t}^2 F_t & P_{12t} - S_{1t}S_{2t}F_t \\ P_{12t} - S_{1t}S_{2t}F_t & P_{22t} - S_{2t}^2 F_t \end{pmatrix} \tag{2.131}$$

λ and (η_1, η_2) are updated as for Section 2.3.3.

2.3.5.5: ARMA(1,1). The models of Sections 2.3.5.1 and 2.3.5.2 can be combined to produce the ARMA(1,1) model:

$$y_t = \mu_t + \varepsilon_t \tag{2.132}$$

$$\mu_t - \nu_t = \phi(\mu_{t-1} - \nu_{t-1}) + \delta\mu_t - \eta\delta\mu_{t-1} \tag{2.133}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t, \tag{2.134}$$

where μ_t is the (error-free) recording at time t, ν_t is the level at time t, ϕ is the autoregressive parameter and η is the moving average parameter. Also, $\varepsilon_t \sim N(0, \lambda^{-1}R_\varepsilon)$, $\delta\mu_t \sim N(0, \lambda^{-1}R_\mu)$ and $\delta\nu_t \sim N(0, \lambda^{-1}R_\nu)$.

Continuing to adopt stationarity conditions in order to form the (ϕ,η) grid, we have:

$$|\phi| < 1 \quad \text{and} \quad |\eta| < 1.$$

In DLM form,

$$y_t = [1 \quad 0]\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} + \varepsilon_t \tag{2.135}$$

$$\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} = \begin{pmatrix}\phi & 1-\phi \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1} \\ \nu_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t - \eta\delta\mu_{t-1} + \delta\nu_t \\ \delta\nu_t\end{pmatrix} \tag{2.136}$$

i.e.

$$\left.\begin{aligned} &\underset{\sim}{H}_t = \underset{\sim}{H} = [1 \quad 0], \quad \underset{\sim}{\theta}_t = \begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix}, \quad \underset{\sim}{G} = \begin{pmatrix}\phi & 1-\phi \\ 0 & 1\end{pmatrix}, \\ &\underset{\sim}{\omega}_t = \begin{pmatrix}\delta\mu_t - \eta\delta\mu_{t-1} + \delta\nu_t \\ \delta\nu_t\end{pmatrix}\end{aligned}\right\} \tag{2.137}$$

and

$$\underset{\sim}{R}_\varepsilon = R_\varepsilon, \quad \underset{\sim}{R}_\omega = \begin{pmatrix}R_\mu + \eta^2 R_\mu + R_\nu & R_\nu \\ R_\nu & R_\nu\end{pmatrix} \tag{2.138}$$

If

$$\underset{\sim}{m}_{t-1} = \begin{pmatrix}m_{t-1} \\ v_{t-1}\end{pmatrix} \tag{2.139}$$

and

$$\underset{\sim}{C}_{t-1} = \begin{pmatrix}M_{t-1} & MV_{t-1} \\ MV_{t-1} & V_{t-1}\end{pmatrix} \tag{2.140}$$

then

$$f_t = \phi m_{t-1} + (1-\phi)v_{t-1}$$

$$e_t = y_t - \phi m_{t-1} - (1-\phi)v_{t-1}$$

$$\underset{\sim}{P}_t = \begin{pmatrix} P_{11t} & P_{12t} \\ P_{12t} & P_{22t} \end{pmatrix}$$

$$= \begin{pmatrix} \phi^2 M_{t-1} + 2\phi(1-\phi)MV_{t-1} + (1-\phi)^2 V_{t-1} + R_\mu + \eta^2 R_\mu + R_\nu & \phi MV_{t-1} + (1-\phi)V_{t-1} + R_\nu \\ \phi MV_{t-1} + (1-\phi)V_{t-1} + R_\nu & V_{t-1} + R_\nu \end{pmatrix}$$

$$F_t = P_{11t} + R_\varepsilon$$

$$\underset{\sim}{S}_t = \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} = \begin{pmatrix} \dfrac{P_{11t}}{P_{11t} + R_\varepsilon} \\ \dfrac{P_{12t}}{P_{11t} + R_\varepsilon} \end{pmatrix}$$

Then

$$\underset{\sim}{m}_t = \begin{pmatrix} m_t \\ v_t \end{pmatrix} = \begin{pmatrix} \phi m_{t-1} + (1-\phi)\, v_{t-1} + S_{1t} e_t \\ v_{t-1} + S_{2t} e_t \end{pmatrix} \tag{2.141}$$

and

$$\underset{\sim}{C}_t = \begin{pmatrix} M_t & MV_t \\ MV_t & V_t \end{pmatrix} = \begin{pmatrix} P_{11t} - S_{1t}^2 F_t & P_{12t} - S_{1t} S_{2t} F_t \\ P_{12t} - S_{1t} S_{2t} F_t & P_{22t} - S_{2t}^2 F_t \end{pmatrix} \tag{2.142}$$

λ and (θ,η) are updated as for Section 2.3.3. (See Appendix A2.2 for the initial grid setting.)

o0o

APPENDIX TWO

INITIAL GRID SETTINGS

A2.1 ONE-DIMENSIONAL GRIDS

For those models where $\underset{\sim}{\phi}$ is present with dimension one (e.g. η for the MA(1) model) we must specify a range for the value of ϕ (e.g. (-1,1) for the MA(1) model). Once this range has been specified, we must decide how many nodes are required on the grid. This choice is influenced by two factors, namely the desire to keep computation to a minimum (obviously, the more nodes in the grid the greater the number of calculations involved) and the desire to obtain satisfactory accuracy (the greater the number of nodes the more precise the parameter estimates available).

Let NN be the number of nodes chosen. Then, for a non-informative prior, we could use a uniform distribution, so that:

$$p(\phi = \phi^{(n)}) = \frac{1}{NN}, \quad \forall n = 1, \ldots, NN.$$

In other situations we may have some prior idea of the location of ϕ within the specified range. Let the range for ϕ be (ϕ_L, ϕ_U) and let the prior best estimate for ϕ be ϕ_M located at the Mth node (assuming ϕ_L is the first node and ϕ_U is the NNth). Then a simple way of incorporating this knowledge, without placing too much confidence in ϕ_M, is to set up the following triangular distribution:

$$p(\phi = \phi^{(i)}) = \frac{2}{(NN - 1)} \frac{(i - 1)}{(M - 1)}, \quad i = 1, \ldots, M$$

$$= \frac{2}{(NN - 1)} \frac{(NN - i)}{(NN - M)}, \quad i = M + 1, \ldots, NN.$$

A2.2 TWO-DIMENSIONAL GRIDS

(a) Square Grids. In those models where $\underset{\sim}{\phi}$ has two dimensions it will be a straighforward matter to set up a uniform distribution when the restrictions on $\underset{\sim}{\phi}$ are purely modular, e.g. $|\phi_1| < a$, $|\phi_2| < b$. In this case a square grid results.

If the number of nodes in the direction of ϕ_1 is N1, and in the direction of ϕ_2 is N2 (so that NN = N1 x N2), then:

$$p(\phi_1 = \phi_1^{(n1)} \ \underline{\text{and}} \ \ \phi_2 = \phi_2^{(n2)}) = \frac{1}{NN}, \quad \forall n1 = 1, \ldots, N1; \quad \forall n2 = 1, \ldots, N2$$

with

$$p(\phi_1 = \phi_1^{(n1)}) = \frac{1}{N1}, \quad \forall n1 = 1, \ldots, N1$$

and

$$p(\phi_2 = \phi_2^{(n2)}) = \frac{1}{N2}, \quad \forall n2 = 1, \ldots, N2$$

so that a uniform distribution results for both ϕ_1 and ϕ_2.

(b) Triangular Grids. For the AR(2) and MA(2) models extra restrictions on $\underset{\sim}{\phi}$ result in a triangular grid. In this case a uniform distribution in two dimensions results in a triangular distribution along each of the individual axes.

Clearly, for each of these situations, if we have precise prior knowledge about the location of $\underset{\sim}{\phi}$ it can be incorporated very easily, e.g. if it is known that $\phi_1 = 0.8$ (at node nj, say) and $\phi_2 = -0.6$ (at node nk), then we could set

$$p(\phi_1 = \phi_1^{(n1)} \text{ and } \phi_2 = \phi_2^{(n2)}) = 1, \text{ if } n1 = nj \text{ and } n2 = nk$$

$$= 0, \text{ otherwise.}$$

oOo

CHAPTER THREE

MODELLING DISCONTINUITIES

3.1 INTRODUCTION TO MULTI-PROCESS MODELS

3.1.1 INTRODUCTORY REMARKS

Harrison and Stevens (1976) showed that by a simple extension of the dynamic linear model framework, described in the previous chapter, we can also model time series discontinuities. The idea of multi-process models was introduced, whereby it was considered that any of a number of models could obtain at a particular timepoint. Specifically, they define two classes: one where a unique model obtains at <u>all</u> timepoints, the model being a choice from a discrete set of alternative models; and one where the model chosen need not be unique for all timepoints, but, in fact, can change from one recursion to another.

For the remainder of this thesis we restrict our attention to the latter case of multi-process models, since they facilitate the handling of changepoints in a time series as well as providing the means for changepoint detection (Smith and West 1983, Trimble et al. 1983).

Multi-process modelling permits the entire model to change between time points; however, the accommodation of innovations is managed by manipulation of the observation and system variances, R_ε and R_ω, without altering the underlying model structure, defined by the system matrix, G. This involves the setting up of a, so-called, multistate structure for the variances.

Let J be the number of models (states) which could obtain at any time t. Then the multistate structure is described by $\{M_t^{(j)}, p_o^{(j)};\ j = 1, \ldots, J\}$, where: $M_t^{(j)}$ means that model j obtains at time t and $p_o^{(j)}$ is the probability that model j obtains.

It has been assumed here that the probability of a particular model obtaining at time t is independent of the previous time series history, i.e. it does not depend upon the model which obtained at the previous timepoint; this assumption is relaxed in Chapter 5.

Before we discuss the mechanics of time series monitoring by means of the multistate dynamic linear model, we shall briefly review previous literature related to the monitoring of time-related observations and to the problem of changepoint determination.

3.1.2 BACKGROUND TO MONITORING TECHNIQUES

Time series arise in a great many fields of application and so it is not surprising that developments in the area of time series <u>monitoring</u> have originated in a variety of contexts.

Early contributions sprung from the field of control engineering where one of the most influential contributions was that of

Kalman (1960,1961), who developed general results for filtering and prediction problems with the aim of achieving optimum control (resulting in the, so-called, Kalman Filter). At around the same time, results on optimal control policies, involving optimal parameter estimates, were published by Joseph and Tou (1961). Thereafter followed a succession of papers in related areas, including: control theory (Filippov 1962, Markus and Lee 1962, Zadeh and Desoer 1963, McGill 1965), navigation, guidance and missile tracking (Schmidt 1966, Sorenson 1966). The Kalman Filter and associated methodology mainly appeared in the control theory literature, although some statisticians also made early contributions (see, for example, Box and Jenkins 1962).

Alongside the techniques for optimal control, came methods for forecasting from observed time series, mainly arising from the work of economists (see, for example, Winters 1960, Brown 1963). This work continued throughout the following two decades (e.g. Kirby 1967, Crane and Crotty 1967, Wheelwright and Makridakis 1973) and, again, statisticians played an active part (e.g. Cox 1961, Box and Jenkins 1970) along with operational researchers (e.g. Trigg 1964, Trigg and Leach 1967). The (1967) paper of Trigg and Leach deals with subject matter which is much closer to the kind of situation in which we are interested: that of automatic responses to out-of-control signals.

In their paper, Trigg and Leach treat the traditional forecast system smoothing constant (see, for example, Brown 1963) as if it were a parameter, and control its value automatically by equating it to the error-tracking signal (see, for example, Montgomery and Johnson 1976 for an account of tracking signals).

When the system is 'out-of-control' (i.e. the error tracking signal is 'large') the smoothing constant is increased, thereby increasing the weighting given to more recent data, in order to facilitate rapid adjustment to the change in time series pattern. This method of tracking, and associated techniques, usually comes under the heading of adaptive control, and represents a way of adjusting to changes in time series rather than detecting these changes.

Other statistical approaches have also been made to model changepoint phenomena in time series. In (1965) Box and Tiao investigated level changes in time series represented by ARIMA models, and extended some of these ideas in their (1975) paper. It was assumed, retrospectively, that the timepoint at which the single possible level change had occurred was known, but that the magnitude of the change was unknown and, therefore, needed to be estimated. This amounts to the assumption that a known intervention has occurred, resulting in a possible change of unknown size. This assumption was relaxed by Smith (1977) who employed similar models, but without specifying the timepoint at which a change may have occurred. Instead, a prior distribution for the timepoint of change was selected and then updated, using Bayes theorem, into the corresponding posterior. The mode, say, of this posterior distribution could then be taken as the timepoint at which a level change was most likely to have happened. Clearly, once a level change has been located in a series (if one exists at all), the series can be split into two at the changepoint and the process repeated on the two 'halves', in order to provide a pragmatic procedure for finding further level changes.

Of course, a change in process level is not the only time series innovation in which we are interested. Other literature of interest has focussed on detecting changes in the slope of a straight line, from the work of Bacon and Watts (1971) through to the paper by Smith and Cook (1980); or has been concerned with detecting changes in regression models (see, for example, Ferreira 1975, Brown, Durbin and Evans 1975) or with the problem of detecting changes in linear models in general (see, for example, Chin Choy and Broemeling 1980, Holbert 1982, Booth and Smith 1982).

There have been many attempts to develop techniques for the accommodation of outliers (see, for example, Dixon 1953, Box and Tiao 1968, Abraham and Box 1978). Although relatively few of these have been concerned with outliers in time-related sequences of observations, the work of Fox (1972) and of Abraham and Box (1979) is relevant. A corner of the literature has also touched upon the problem of detecting a change in the variance of a time series model (Wichern, Miller and Hsu 1976, Hsu 1979, Diaz 1982).

Much of the work referenced thus far has been concerned with the retrospective identification of time series discontinuities. We now turn our attention to procedures for the prospective, or 'on-line', detection of changepoints.

Cumulative Sum Techniques. One of the best known methods for the prospective detection of a shift in the level of a time series is the cumulative sum (CUSUM) technique, dating from the work of Page (1954), and described in detail by van Dobben de Bruyn (1968).

The basic idea of the CUSUM is very simple. Let $y_1, y_2, \ldots, y_i, \ldots$ be an observed time series which has a 'target' mean of T (often derived as the mean of a control sample). Then, rather than simply plotting the values of y_t against time t, the CUSUM plots the values of Σ_t against time, where

$$\Sigma_i = (y_1 - T) + (y_2 - T) + \ldots + (y_i - T). \tag{3.1}$$

If the time series is stable (i.e. is not subject to a shift in mean) the values of y_t will fluctuate around the target value T, and so the values of Σ_t will fluctuate around zero. However, a change in the mean level of the observation series will result in a slope change away from zero in the CUSUM sequence. The magnitude of the slope is equal to the level change in the observation series.

As far as process control is concerned, a 'mask' is usually adopted such that if the Σ_t series drifts beyond one of the boundaries, the process is deemed to be out-of-control (see, for example, Edwards 1980 for the mechanics of this procedure). However, this highlights an important drawback to the CUSUM procedure in the medical monitoring context, namely that, unless one is willing to tolerate a large number of false alarms, the delay between the timepoint at which the process becomes unstable and the timepoint at which the CUSUM mask boundary has been crossed may be too great. Hinkley (1971) has produced some theoretical results concerning inferences about the timepoint of change using the CUSUM technique and further results have been obtained by Johnson and Bagshaw (1974), Bagshaw and Johnson (1977) and Ezzet (1985).

Another major problem with the CUSUM method is that it is designed, primarily, to capture changes in the level of a time series only; it is not designed to distinguish between several types of discontinuity.

However, in (1979) Stoodley and Mirnia adopted a backward CUSUM scheme (as in Harrison and Davies 1964), in conjunction with a modified form of the linear growth model, in order to detect transients, changes in level and changes in the slope of a time series. However, their choice of limits for the CUSUM component and their definitions of changepoint type were somewhat arbitrary and, as they pointed out, lacking in 'theoretical foundation'. It is, therefore, very difficult to extend these methods so that they apply to other time series models than linear growth without retaining the ad-hoc threshold definitions.

In fact Ameen and Harrison (1983) did extend the methods used by Stoodley and Mirnia (1979), so that a backward CUSUM scheme could be used to monitor the forecasts of a time series represented by a general dynamic linear model. A specific dynamic linear model was chosen, which described the 'stable state' of the series (in the terminology of Section 3.1), and this model was used to produce one-step-ahead forecasts and their corresponding forecast errors. The errors were then monitored by the backward CUSUM method, as in Harrison and Davies (1964). If a change is signalled by the CUSUM scheme then a set of multi-process models is applied to the ensuing data (i.e. a multistate structure is employed), until such time as the posterior probability of a return to the stable state is greater than a particular threshold, at which point the single stable-state model, with updated parameters, is re-introduced, and the CUSUM scheme reset. In this approach, the type of changepoint can be determined by the multistate component of the system, whereas the timing of change is determined by the CUSUM component.

This method of monitoring and forecasting is, in fact, a modification of the earlier work of Harrison and Stevens (1976), in which they suggested the imposition of a multistate structure onto the dynamic linear model framework, discussed in detail in Chapter 2 of this thesis. However, if the concept of discounting is introduced (Ameen and Harrison 1982) as a replacement for the updating of the covariance matrix of the normal dynamic linear model (see Section 2.1.2), this complicates the use of multi-process models for modelling discontinuities since, as mentioned earlier, the discount parameters do not have a simple interpretation with regards to changepoints. We require, therefore, the introduction of another monitoring technique alongside the dynamic linear model structure when the discounting principle is used; Ameen and Harrison (1983) adopt the CUSUM scheme and claim that this monitoring system is more efficient than the original multi-process system, since the phases where a single model is applied to the data (i.e. in periods of stability) result in fewer calculations. This is, of course, true although the decrease in computation may not be great, since the competing models (which are introduced when the CUSUM component registers an out-of-control signal) need to be constantly updated, albeit marginally, even when the system has reverted to its stable state. It can also be seen that although the methods of Ameen and Harrison (1983) avoid the ad-hoc definitions of changepoint-type used by Stoodley and Mirnia (1979), they do rely on the introduction of an arbitrary threshold in order to determine when the return to stability has occurred. They also rely on the usual CUSUM 'cut-off' rules for the initial detection of instability, therefore possibly resulting in an inadmissable delay in the detection of the onset of instability.

Let us now turn our attention to the addition of a multi-state structure to the normal dynamic linear model framework, as described by Harrison and Stevens (1976). By retaining the error-covariance-matrix formulation, as opposed to the discount-matrix formulation, we may easily incorporate models for the detection of the onset of instability without having to call on alternative monitoring schemes.

Multistate Modelling. Define the multistate structure as in Section 3.1.1. Then the dynamic linear model given by (2.1) and (2.2) can be extended to:

$$\underset{\sim}{y}_t = \underset{\sim}{H}_t \underset{\sim}{\theta}_t + \underset{\sim}{\varepsilon}_t^{(j)} \qquad (3.2)$$

$$\underset{\sim}{\theta}_t = \underset{\sim}{G}\underset{\sim}{\theta}_{t-1} + \underset{\sim}{\omega}_t^{(j)} \qquad (3.3)$$

for j = 1, ..., J, where

$$\underset{\sim}{E}_t^{(j)} = \mathrm{var}(\underset{\sim}{\varepsilon}_t^{(j)}) = \lambda^{-1}\underset{\sim}{R}_\varepsilon^{(j)} \qquad (3.4)$$

$$\underset{\sim}{W}_t^{(j)} = \mathrm{var}(\underset{\sim}{\omega}_t^{(j)}) = \lambda^{-1}\underset{\sim}{R}_\omega^{(j)} \qquad (3.5)$$

so that (3.2) and (3.3) represent J possible models, differing only through the elements of $\underset{\sim}{R}_\varepsilon$ and $\underset{\sim}{R}_\omega$. It will be shown in Section 3.3 how particular choices of $\underset{\sim}{R}_\varepsilon$ and $\underset{\sim}{R}_\omega$ can result in a variety of changepoint models.

As well as the assumptions given by (2.4), it will also be assumed here (as noted earlier) that:

$$p(M_t^{(j)}|H) = p_o^{(j)}, \quad j = 1, \ldots, J, \ \forall t \tag{3.6}$$

where H is the process history <u>prior</u> to time t. We may now rewrite the Kalman Filter recursion, given in (2.12) to (2.22), as follows (it being understood throughout that i,j run through the range 1, ..., J):

$$\underset{\sim}{\theta}_o \sim N(\underset{\sim}{m}_o, \lambda^{-1}\underset{\sim}{C}_o) \tag{3.7}$$

Equation (2.15) is replaced by:

$$p(\underset{\sim}{\theta}_{t-1}|\underset{\sim}{D}_{t-1},\lambda,M_{t-1}^{(i)}) \sim N(\underset{\sim}{m}_{t-1}^{(i)},\lambda^{-1}\underset{\sim}{C}_{t-1}^{(i)}) \tag{3.8}$$

Equation (2.16) is replaced by:

$$p(\lambda|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)}) \sim G(\tfrac{1}{2}n_{t-1},\tfrac{1}{2}r_{t-1}^{(i)}) \tag{3.9}$$

where, again, $U \sim G(\alpha,\beta)$ means that U has a gamma distribution. We assume that the initialization of (3.9) is given by (2.18).

Upon receipt of $\underset{\sim}{y}_t$, we can update (3.8) and (3.9) to give:

$$p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda,M_{t-1}^{(i)},M_t^{(j)}) \sim N(\underset{\sim}{m}_t^{(ij)},\lambda^{-1}\underset{\sim}{C}_t^{(ij)}) \tag{3.10}$$

and

$$p(\lambda|\underset{\sim}{D}_t,M_{t-1}^{(i)},M_t^{(j)}) \sim G(\tfrac{1}{2}n_t,\tfrac{1}{2}r_t^{(ij)}) \tag{3.11}$$

where $\underset{\sim}{m}_t^{(ij)}$ and $\underset{\sim}{C}_t^{(ij)}$ are given by the Kalman Filter recursions:

$$\underset{\sim}{m}_t^{(ij)} = \underset{\sim}{G}\underset{\sim}{m}_{t-1}^{(i)} + \underset{\sim}{S}_t^{(ij)}\underset{\sim}{e}_t^{(i)} \tag{3.12}$$

$$\underset{\sim}{C}_t^{(ij)} = \underset{\sim}{P}_t^{(ij)} - \underset{\sim}{S}_t^{(ij)}\underset{\sim}{F}_t^{(ij)}(\underset{\sim}{S}_t^{(ij)})^T \tag{3.13}$$

and where, by standard Bayesian conjugate analysis, n_t and $r_t^{(ij)}$ are given by:

$$n_t = n_{t-1} + 1 \tag{3.14}$$

$$r_t^{(ij)} = r_{t-1}^{(i)} + (\underset{\sim}{e}_t^{(i)})^T(\underset{\sim}{F}_t^{(ij)})^{-1}\underset{\sim}{e}_t^{(i)} \tag{3.15}$$

with

$$\left.\begin{aligned}
\underset{\sim}{f}_t^{(i)} &= \underset{\sim}{H}_t\underset{\sim}{G}\underset{\sim}{m}_{t-1}^{(i)} \\
\underset{\sim}{e}_t^{(i)} &= \underset{\sim}{y}_t - \underset{\sim}{f}_t^{(i)} \\
\underset{\sim}{P}_t^{(ij)} &= \underset{\sim}{G}\underset{\sim}{C}_{t-1}^{(i)}\underset{\sim}{G}^T + \underset{\sim}{R}_\omega^{(j)} \\
\underset{\sim}{F}_t^{(ij)} &= \underset{\sim}{H}_t\underset{\sim}{P}_t^{(ij)}\underset{\sim}{H}_t^T + \underset{\sim}{R}_\varepsilon^{(j)} \\
\underset{\sim}{S}_t^{(ij)} &= \underset{\sim}{P}_t^{(ij)}\underset{\sim}{H}_t^T(\underset{\sim}{F}_t^{(ij)})^{-1}
\end{aligned}\right\} \tag{3.16}$$

It should be clear that whereas (3.8) and (3.9) describe J models, (3.10) and (3.11) describe J^2 models. In other words, each iteration produces a J-fold increase in the number of models under consideration and, plainly, this will soon explode even when the number of states, J, is small: e.g. when J = 2 there would be over 1000 models by the time t = 10!

We will therefore have to approximate the forms of (3.10) and (3.11), so that they resemble (3.8) and (3.9), in order to avoid this problem. The next section describes a pragmatically successful algorithm for the general class of multistate dynamic linear models.

3.2 RECURSIVE ESTIMATION

As in Chapter 2, we shall restrict ourselves, for ease of exposition, to a univariate observation y_t.

From Equations (3.2), (3.3) and (3.16) we have:

$$p(y_t|\underset{\sim}{D}_{t-1},\lambda,M_{t-1}^{(i)},M_t^{(j)}) \sim N(f_t^{(i)},\lambda^{-1}F_t^{(ij)}) \tag{3.17}$$

so that

$$p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)}) = \int_0^\infty p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)},\lambda)p(\lambda|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)})d\lambda$$

$$= \int_0^\infty p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)},\lambda)p(\lambda|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)})d\lambda$$

and, following the arguments of Section 2.2, we can use (3.17) and (3.9) to show that:

$$Z_t^{(ij)} = p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)}) \propto (F_t^{(ij)})^{-\frac{1}{2}}(r_{t-1}^{(i)})^{\frac{1}{2}n_{t-1}}(r_t^{(ij)})^{-\frac{1}{2}n_t} \tag{3.18}$$

If we let

$$p_t^{(j)} = p(M_t^{(j)}|\underset{\sim}{D}_t) \tag{3.19}$$

and

$$p_t^{(ij)} = p(M_{t-1}^{(i)},M_t^{(j)}|\underset{\sim}{D}_t) \tag{3.20}$$

Then ., using (3.19),

$$p(y_t|\underset{\sim}{D}_{t-1}) = \sum_{i=1}^{J} p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)})p(M_{t-1}^{(i)}|\underset{\sim}{D}_{t-1})$$

$$= \sum_{i=1}^{J}[\sum_{j=1}^{J} p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)})p(M_t^{(j)}|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)})]p_{t-1}^{(i)}$$

Using (3.6) we see that

$$p(y_t|\underset{\sim}{D}_{t-1}) = \sum_{i=1}^{J} \sum_{j=1}^{J} Z_t^{(ij)} p_o^{(j)} p_{t-1}^{(i)} \tag{3.21}$$

From Equation (3.20),

$$p_t^{(ij)} = p(M_{t-1}^{(i)}, M_t^{(j)} | \underset{\sim}{D}_t)$$

$$= p(y_t|\underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}, M_t^{(j)}) p(M_{t-1}^{(i)}, M_t^{(j)} | \underset{\sim}{D}_{t-1}) / p(y_t|\underset{\sim}{D}_{t-1})$$

(using Bayes theorem)

$$= Z_t^{(ij)} p(M_t^{(j)} | \underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}) p(M_{t-1}^{(i)} | \underset{\sim}{D}_{t-1}) / p(y_t|\underset{\sim}{D}_{t-1})$$

so that

$$p_t^{(ij)} = Z_t^{(ij)} p_o^{(j)} p_{t-1}^{(i)} / p(y_t|\underset{\sim}{D}_{t-1}) \tag{3.22}$$

(using Equations (3.6) and (3.19)), where $Z_t^{(ij)}$ is given by (3.18) and where the denominator is given by (3.21).

Completing the Recursion

1. Collapsing Procedures: In order to close the recursion, we need to approximate (3.10) and (3.11) so that they take the forms of (3.8) and (3.9) respectively. The normal approximation used is that proposed by Harrison and Stevens (1976); West (1982) pointed out that the 'collapsing' mixture employed was, in fact, that which minimized the well-known Kullback-Leibler divergence, and the gamma approximation we use is also that which minimizes the Kullback-Leibler divergence (see West (1982) for details). We therefore make the following assumptions:

$$\left.\begin{array}{l}\text{We approximate the } J^2 \text{ normal distributions:}\\ N(\underset{\sim}{m}_t^{(ij)}, \lambda^{-1}\underset{\sim}{C}_t^{(ij)}),\ i = 1, \ldots, J,\ j = 1, \ldots, J\\ \text{by the } J \text{ mixtures: } N(\underset{\sim}{m}_t^{(j)}, \lambda^{-1}\underset{\sim}{C}_t^{(j)}),\ j = 1, \ldots, J\end{array}\right\} \quad (3.23)$$

where

$$\underset{\sim}{m}_t^{(j)} = \sum_{i=1}^{J} (p_t^{(ij)}/p_t^{(j)})\underset{\sim}{m}_t^{(ij)} \quad (3.24)$$

and

$$\underset{\sim}{C}_t^{(j)} = \sum_{i=1}^{J} (p_t^{(ij)}/p_t^{(j)})\{\underset{\sim}{C}_t^{(ij)} + (\underset{\sim}{m}_t^{(ij)} - \underset{\sim}{m}_t^{(j)}) \times (\underset{\sim}{m}_t^{(ij)} - \underset{\sim}{m}_t^{(j)})^T\} \quad (3.25)$$

We approximate the J^2 gamma distributions: $G(\frac{1}{2}n_t, \frac{1}{2}r_t^{(ij)})$, $i = 1, \ldots, J$, $j = 1, \ldots, J$ by the J mixtures:

$$G(\tfrac{1}{2}n_t, \tfrac{1}{2}r_t^{(j)}),\ j = 1, \ldots, J, \quad (3.26)$$

where

$$(r_t^{(j)})^{-1} = \sum_{i=1}^{J} (p_t^{(ij)}/p_t^{(j)})(r_t^{(ij)})^{-1} \quad (3.27)$$

2. *Updating Procedures:* We have now shown how to update the parameter estimates and covariances from (3.8) and (3.9) to (3.23) and (3.26) having received the latest observation, y_t. To complete the process we must specify the form of $p_t^{(j)}$.

Now

$$p_t^{(j)} = p(M_t^{(j)} | \underset{\sim}{D}_t)$$
$$= p(y_t | \underset{\sim}{D}_{t-1}, M_t^{(j)}) p(M_t^{(j)} | \underset{\sim}{D}_{t-1}) / p(y_t | \underset{\sim}{D}_{t-1})$$

(using Bayes theorem)

$$= \frac{\left[\sum_{i=1}^{J} p(y_t | \underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}, M_t^{(j)}) p(M_{t-1}^{(i)} | \underset{\sim}{D}_{t-1}, M_t^{(j)})\right] . p_o^{(j)}}{p(y_t | \underset{\sim}{D}_{t-1})}$$

(using Equation (3.6)), so that

$$p_t^{(j)} = \sum_{i=1}^{J} Z_t^{(ij)} p_{t-1}^{(i)} p_o^{(j)} / p(y_t | \underset{\sim}{D}_{t-1}) \tag{3.28}$$

where $Z_t^{(ij)}$ is given by (3.18) and where the denominator is given by (3.21).

State Probabilities

As a by-product of the parameter-updating process, we have calculated the quantities $p_t^{(j)}$, $j = 1, \ldots, J$. It can easily be seen that $p_t^{(j)} = p(M_t^{(j)} | \underset{\sim}{D}_t)$ denotes the probability that state j obtains at time t, given all the data up to and including time t. So that, for instance, if state j represents the change in level model (see Section 3.3) $p_t^{(j)}$ is the probability of a change in level at time t, and therefore can be used to indicate the timing of the changepoint. However, when some change in pattern occurs at time t, it may not be readily apparent which particular one of several alternative types of change of pattern has obtained until further information is available. It may be essential, therefore, to be able to update our beliefs about the state obtaining at time t having received observations $y_{t+1}, y_{t+2}, \ldots,$ in addition to those up to time t.

Let

$$O_t^{(i)} = p(M_{t-1}^{(i)} | \underset{\sim}{D}_t) \tag{3.29}$$

and

$$T_t^{(h)} = p(M_{t-2}^{(h)} | \underset{\sim}{D}_t) \tag{3.30}$$

so that $O_t^{(i)}$ denotes the <u>one-step-back</u> probability of state i obtaining at time t - 1, and $T_t^{(h)}$ denotes the two-steps-back probability of state h obtaining at time t - 2. For example, if h is the level change model, $T_t^{(h)}$ denotes the probability of a level change at time t - 2 given all the data up to and including time t - 2 and the additional observations y_{t-1} and y_t.

We use Bayes theorem in order to calculate the quantities in (3.29) and (3.30), so that:

$$O_t^{(i)} = p(M_{t-1}^{(i)} | \underset{\sim}{D}_t)$$

$$= p(y_t | \underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}) p(M_{t-1}^{(i)} | \underset{\sim}{D}_{t-1}) / p(y_t | \underset{\sim}{D}_{t-1})$$

$$= \sum_{j=1}^{J} Z_t^{(ij)} p_o^{(j)} p_{t-1}^{(i)} / p(y_t | \underset{\sim}{D}_{t-1}) \qquad (3.31)$$

with $Z_t^{(ij)}$ specified by (3.18) and the denominator by (3.21).

Similarly,

$$T_t^{(h)} = p(M_{t-2}^{(h)} | \underset{\sim}{D}_t)$$

$$= p(y_t | \underset{\sim}{D}_{t-1}, M_{t-2}^{(h)}) p(M_{t-2}^{(h)} | \underset{\sim}{D}_{t-1}) / p(y_t | \underset{\sim}{D}_{t-1})$$

$$= \sum_{j=1}^{J} p(y_t | \underset{\sim}{D}_{t-1}, M_{t-2}^{(h)}, M_t^{(j)}) p_o^{(j)} . O_{t-1}^{(h)} / p(y_t | \underset{\sim}{D}_{t-1})$$

$$= \frac{\sum_{j=1}^{J} \sum_{i=1}^{J} Z_t^{(ij)} . p(M_{t-1}^{(i)} | \underset{\sim}{D}_{t-1}, M_{t-2}^{(h)}, M_t^{(j)}) . p_o^{(j)} . O_{t-1}^{(h)}}{p(y_t | \underset{\sim}{D}_{t-1})}$$

$$= \frac{\sum_{j=1}^{J} \sum_{i=1}^{J} Z_t^{(ij)} [Z_{t-1}^{(hi)} p_o^{(i)} / \sum_{i=1}^{J} Z_{t-1}^{(hi)} p_o^{(i)}] . p_o^{(j)} . O_{t-1}^{(h)}}{p(y_t | \underset{\sim}{D}_{t-1})} \qquad (3.32)$$

where $Z_t^{(ij)}$ is specified by (3.18), $Z_{t-1}^{(hi)}$ has been calculated at time t - 1 from (3.18), $O_{t-1}^{(h)}$ is given by (3.31) and the denominator is given by (3.21).

Summary of Iteration

(i) Using (3.6), (3.8) and (3.9) as the starting point, calculate (3.12), (3.13), (3.14) and (3.15) via the quantities in (3.16);

(ii) Use (3.14), (3.15) and (3.16) to calculate (3.18);

(iii) Use (3.18) to calculate (3.21);

(iv) Use (3.18) and (3.21) to calculate (3.22), (3.28), (3.31) and (3.32);

(v) Use (3.22) and (3.28) along with (3.12), (3.13), (3.14) and (3.15) to calculate (3.24), (3.25) and (3.27);

(vi) Use (3.6), (3.23) and (3.26) as the starting point for the next iteration.

Nuisance Parameters

If the model depends upon one or more nuisance parameters, $\underset{\sim}{\phi}$, we adopt the procedure outlined in Chapter 2 and specify a probability distribution for $\underset{\sim}{\phi}$ over a suitably chosen discrete grid (see Appendix 2 for initial conditions).

We replace (3.8) and (3.9) by:

$$p(\underset{\sim}{\theta}_{t-1}|\underset{\sim}{D}_{t-1},\lambda,M_{t-1}^{(i)},\underset{\sim}{\phi}) \sim N(\underset{\sim}{m}_{t-1}^{(i)},\lambda^{-1}\underset{\sim}{C}_{t-1}^{(i)}) \tag{3.33}$$

and

$$p(\lambda|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},\underset{\sim}{\phi}) \sim G(\tfrac{1}{2}n_{t-1},\tfrac{1}{2}r_{t-1}^{(i)}) \tag{3.34}$$

where $\underset{\sim}{m}_{t-1}^{(i)}$, $\underset{\sim}{C}_{t-1}^{(i)}$ and $r_{t-1}^{(i)}$ now all depend on $\underset{\sim}{\phi}$, so that if there are NN nodes in the $\underset{\sim}{\phi}$ grid, $\underset{\sim}{\Phi}$, (3.33) and (3.34) represent NN possible Normal-gamma distributions.

It can be readily seen that:

$$p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda,M_{t-1}^{(i)},M_t^{(j)},\underset{\sim}{\phi}) \sim N(\underset{\sim}{m}_t^{(ij)},\lambda^{-1}\underset{\sim}{C}_t^{(ij)}) \tag{3.35}$$

and

$$p(\lambda|\underset{\sim}{D}_t,M_{t-1}^{(i)},M_t^{(j)},\underset{\sim}{\phi}) \sim G(\tfrac{1}{2}n_t,\tfrac{1}{2}r_t^{(ij)}) \tag{3.36}$$

where $\underset{\sim}{m}_t^{(ij)}$, $\underset{\sim}{C}_t^{(ij)}$, n_t and $r_t^{(ij)}$ are defined by (3.12), (3.13), (3.14), (3.15) and (3.16) for <u>each</u> node in Φ.

Let

$$p(\underset{\sim}{\phi}|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)}) = K_{t-1}^{(i)}(\underset{\sim}{\phi}) \tag{3.37}$$

and

$$p(\underset{\sim}{\phi}|\underset{\sim}{D}_t,M_{t-1}^{(i)},M_t^{(j)}) = K_t^{(ij)}(\underset{\sim}{\phi}) \tag{3.38}$$

Then, we replace (3.10) by:

$$p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda,M_{t-1}^{(i)},M_t^{(j)}) = \sum_{\underset{\sim}{\Phi}} p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda,M_{t-1}^{(i)},M_t^{(j)},\underset{\sim}{\phi})p(\underset{\sim}{\phi}|\underset{\sim}{D}_t,\lambda,M_{t-1}^{(i)},M_t^{(j)})$$

$$= \sum_{\underset{\sim}{\Phi}} p(\underset{\sim}{\theta}_t|\underset{\sim}{D}_t,\lambda,M_{t-1}^{(i)},M_t^{(j)},\underset{\sim}{\phi})K_t^{(ij)}(\underset{\sim}{\phi}) \tag{3.39}$$

where the first term in the summation is given by (3.35). In practice, (3.39) is a mixture of normals and we approximate by an $N(\underset{\sim}{m}_t^{(ij)},\lambda^{-1}\underset{\sim}{C}_t^{(ij)})$ distribution, where

$$\underset{\sim}{m}_t^{(ij)} = \sum_{\underset{\sim}{\Phi}} \underset{\sim}{m}_t^{(ij)}(\underset{\sim}{\phi}).K_t^{(ij)}(\underset{\sim}{\phi}) \tag{3.40}$$

and

$$\underset{\sim}{C}_t^{(ij)} = \sum_{\underset{\sim}{\Phi}}[\underset{\sim}{C}_t^{(ij)}(\underset{\sim}{\phi}) + (\underset{\sim}{m}_t^{(ij)}(\underset{\sim}{\phi}) - \underset{\sim}{m}_t^{(ij)})(\underset{\sim}{m}_t^{(ij)}(\underset{\sim}{\phi}) - \underset{\sim}{m}_t^{(ij)})^T].K_t^{(ij)}(\underset{\sim}{\phi}) \tag{3.41}$$

Similarly, we replace (3.11) by:

$$p(\lambda|\underset{\sim}{D}_t,M_{t-1}^{(i)},M_t^{(j)}) = \sum_{\underset{\sim}{\phi}} p(\lambda|\underset{\sim}{D}_t,M_{t-1}^{(i)},M_t^{(j)},\underset{\sim}{\phi})K_t^{(ij)}(\underset{\sim}{\phi}) \tag{3.42}$$

where the first term in the summation is given by (3.36). Since (3.42) is a mixture of gamma distributions we approximate it by

$$(r_t^{(ij)})^{-1} = \sum_{\underset{\sim}{\phi}} (r_t^{(ij)}(\underset{\sim}{\phi}))^{-1}K_t^{(ij)}(\underset{\sim}{\phi}). \tag{3.43}$$

We now replace (3.18) by:

$$\begin{aligned} Z_t^{(ij)}(\underset{\sim}{\phi}) &= p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)},\underset{\sim}{\phi}) \\ &\propto (F_t^{(ij)})^{-\frac{1}{2}}(r_{t-1}^{(i)})^{\frac{1}{2}n_{t-1}}(r_t^{(ij)})^{-\frac{1}{2}n_t} \end{aligned} \tag{3.44}$$

where $F_t^{(ij)}$, $r_{t-1}^{(i)}$ and $r_t^{(ij)}$ now depend on $\underset{\sim}{\phi}$.

In order to calculate (3.40), (3.41) and (3.43) we need to derive $K_t^{(ij)}(\underset{\sim}{\phi})$, where, using Bayes theorem,

$$\begin{aligned} K_t^{(ij)}(\underset{\sim}{\phi}) &= p(\underset{\sim}{\phi}|\underset{\sim}{D}_t,M_{t-1}^{(i)},M_t^{(j)}) \\ &= \frac{p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)},\underset{\sim}{\phi})p(\underset{\sim}{\phi}|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)})}{\sum_{\underset{\sim}{\phi}} p(y_t|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)},\underset{\sim}{\phi})p(\underset{\sim}{\phi}|\underset{\sim}{D}_{t-1},M_{t-1}^{(i)},M_t^{(j)})} \\ &= \frac{Z_t^{(ij)}(\underset{\sim}{\phi})K_{t-1}^{(i)}(\underset{\sim}{\phi})}{\sum_{\phi} Z_t^{(ij)}(\underset{\sim}{\phi})K_{t-1}^{(i)}(\underset{\sim}{\phi})} \end{aligned} \tag{3.45}$$

where $Z_t^{(ij)}(\underset{\sim}{\phi})$ is given by (3.44).

Updating the Probabilities:

1. $$p_t^{(ij)} = p(M_{t-1}^{(i)}, M_t^{(j)} | \underset{\sim}{D}_t)$$

$$= p(y_t | \underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}, M_t^{(j)}) p(M_{t-1}^{(i)}, M_t^{(j)} | \underset{\sim}{D}_{t-1}) / p(y_t | \underset{\sim}{D}_{t-1})$$

$$= \sum_{\underset{\sim}{\phi}} Z_t^{(ij)}(\underset{\sim}{\phi}) K_{t-1}^{(i)}(\underset{\sim}{\phi}) . p_o^{(j)} p_{t-1}^{(i)} / p(y_t | \underset{\sim}{D}_{t-1}) \qquad (3.46)$$

where $Z_t^{(ij)}(\underset{\sim}{\phi})$ is given by (3.44).

2. $$p_t^{(j)} = p(M_t^{(j)} | \underset{\sim}{D}_t)$$

$$= \sum_{i=1}^{J} p(y_t | \underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}, M_t^{(j)}) p_{t-1}^{(i)} p_o^{(j)} / p(y_t | \underset{\sim}{D}_{t-1})$$

(from Equation (3.28))

$$= \sum_{i=1}^{J} \sum_{\underset{\sim}{\phi}} Z_t^{(ij)}(\underset{\sim}{\phi}) K_{t-1}^{(i)}(\underset{\sim}{\phi}) . p_{t-1}^{(i)} p_o^{(j)} / p(y_t | \underset{\sim}{D}_{t-1}) \qquad (3.47)$$

where $Z_t^{(ij)}(\underset{\sim}{\phi})$ is given by (3.44).

3. $$o_t^{(i)} = p(M_{t-1}^{(i)} | \underset{\sim}{D}_t)$$

$$= \sum_{j=1}^{J} p(y_t | \underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}, M_t^{(j)}) p_{t-1}^{(i)} p_o^{(j)} / p(y_t | \underset{\sim}{D}_{t-1})$$

(from Equation (3.31))

$$= \sum_{j=1}^{J} \sum_{\underset{\sim}{\phi}} Z_t^{(ij)}(\underset{\sim}{\phi}) K_{t-1}^{(i)}(\underset{\sim}{\phi}) . p_{t-1}^{(i)} p_o^{(j)} / p(y_t | \underset{\sim}{D}_{t-1}) \qquad (3.48)$$

where $Z_t^{(ij)}(\underset{\sim}{\phi})$ is given by (3.44).

4. $$T_t^{(h)} = p(M_{t-2}^{(h)}|\underset{\sim}{D}_t)$$

$$= \{ \sum_{j=1}^{J} \sum_{i=1}^{J} p(y_t|\underset{\sim}{D}_{t-1}, M_{t-2}^{(h)}, M_{t-1}^{(i)}, M_t^{(j)})$$

$$\times \frac{p(y_{t-1}|\underset{\sim}{D}_{t-2}, M_{t-2}^{(h)}, M_{t-1}^{(i)})p_o^{(i)}}{\sum_{i=1}^{J} p(y_{t-1}|\underset{\sim}{D}_{t-2}, M_{t-2}^{(h)}, M_{t-1}^{(i)})p_o^{(i)}} . p_o^{(j)} 0_{t-1}^{(h)}\}/p(y_t|\underset{\sim}{D}_{t-1})$$

(from Equation (3.32))

$$= \{ \sum_{j=1}^{J} \sum_{i=1}^{J} [\sum_{\underset{\sim}{\phi}} Z_t^{(ij)}(\underset{\sim}{\phi})K_{t-1}^{(hi)}(\underset{\sim}{\phi})]$$

$$\times \frac{[\sum_{\underset{\sim}{\phi}} Z_{t-1}^{(hi)}(\underset{\sim}{\phi})K_{t-2}^{(h)}(\underset{\sim}{\phi})]p_o^{(i)}}{\sum_{i=1}^{J} \sum_{\underset{\sim}{\phi}} Z_{t-1}^{(hi)}(\underset{\sim}{\phi})K_{t-2}^{(h)}(\underset{\sim}{\phi})p_o^{(i)}} . p_o^{(j)} 0_{t-1}^{(h)}\}/p(y_t|\underset{\sim}{D}_{t-1})$$

$$= \{ \sum_{j=1}^{J} \sum_{i=1}^{J} \sum_{\underset{\sim}{\phi}}[Z_t^{(ij)}(\underset{\sim}{\phi}) . \frac{Z_{t-1}^{(hi)}(\underset{\sim}{\phi})K_{t-2}^{(h)}(\underset{\sim}{\phi})}{\sum_{\underset{\sim}{\phi}} Z_{t-1}^{(hi)}(\underset{\sim}{\phi})K_{t-2}^{(h)}(\underset{\sim}{\phi})}]$$

$$\times \frac{[\sum_{\underset{\sim}{\phi}} Z_{t-1}^{(hi)}(\underset{\sim}{\phi})K_{t-2}^{(h)}(\underset{\sim}{\phi})p_o^{(i)}]}{[\sum_{i=1}^{J} \sum_{\underset{\sim}{\phi}} Z_{t-1}^{(hi)}(\underset{\sim}{\phi})K_{t-2}^{(h)}(\underset{\sim}{\phi})p_o^{(i)}]} . p_o^{(j)} 0_{t-1}^{(h)}\}/p(y_t|\underset{\sim}{D}_{t-1}) \qquad (3.49)$$

(using Equation (3.45)), where $Z_t^{(ij)}(\underset{\sim}{\phi})$ is given by (3.44), $Z_{t-1}^{(hi)}(\underset{\sim}{\phi})$ has been calculated at time t - 1 from (3.44), and where $0_{t-1}^{(h)}$ has been calculated at time t - 1 from (3.48).

5. $K_t^{(j)}(\underset{\sim}{\phi}) = p(\underset{\sim}{\phi}|\underset{\sim}{D}_t, M_t^{(j)})$

$$= \sum_{i=1}^{J} p(\underset{\sim}{\phi}|\underset{\sim}{D}_t, M_{t-1}^{(i)}, M_t^{(j)}) p(M_{t-1}^{(i)}|\underset{\sim}{D}_t, M_t^{(j)})$$

$$= \sum_{i=1}^{J} \frac{Z_t^{(ij)}(\underset{\sim}{\phi}) K_{t-1}^{(i)}(\underset{\sim}{\phi})}{p(y_t|\underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}, M_t^{(j)})} \cdot \frac{p(y_t|\underset{\sim}{D}_{t-1}, M_{t-1}^{(i)}, M_t^{(j)}) p_{t-1}^{(i)}}{p(y_t|\underset{\sim}{D}_{t-1}, M_t^{(j)})}$$

(using Bayes theorem)

$$= \frac{\sum_{i=1}^{J} Z_t^{(ij)}(\underset{\sim}{\phi}) K_{t-1}^{(i)}(\underset{\sim}{\phi}) p_{t-1}^{(i)}}{\sum_{i=1}^{J} \sum_{\underset{\sim}{\Phi}} Z_t^{(ij)}(\underset{\sim}{\phi}) K_{t-1}^{(i)}(\underset{\sim}{\phi}) p_{t-1}^{(i)}} \qquad (3.50)$$

where $Z_t^{(ij)}(\underset{\sim}{\phi})$ is given by (3.44).

Summary of Iteration in the Presence of Nuisance Parameters, ϕ

(i) Using (3.6), (3.33), (3.34) and (3.37) as the starting point, calculate (3.12), (3.13), (3.14) and (3.15) via the quantities in (3.16) for <u>each</u> node in $\underset{\sim}{\Phi}$;

(ii) Use (3.14), (3.15) and (3.16) to calculate (3.44) for each node in $\underset{\sim}{\Phi}$;

(iii) Use (3.37) and (3.44) to calculate (3.45), (3.46), (3.47), (3.48), (3.49) and (3.50);

(iv) Use (3.45) along with (3.12), (3.13), (3.14) and (3.15) to calculate (3.40), (3.41) and (3.43);

(v) Use (3.40), (3.41), (3.43), (3.46) and (3.47) to calculate (3.24), (3.25) and (3.27);

(vi) Use (3.6), (3.35) and (3.36) as the starting point for the next iteration.

3.3 THE CHANGEPOINT MODELS

In this section we shall examine the implications of attaching a multistate structure to the models described in Section 2.3. As has been mentioned previously, changepoint models can be introduced through the adjustment of observation and system variances, without having to change the underlying model structure described by $\underset{\sim}{G}$.

3.3.1 LINEAR GROWTH

Consider the linear growth model of Section 2.3.1 with a multistate structure imposed according to (3.2) and (3.3):

$$y_t = \mu_t + \varepsilon_t^{(j)} \tag{3.51}$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta\mu_t^{(j)} \tag{3.52}$$

$$\beta_t = \beta_{t-1} + \delta\beta_t^{(j)} \tag{3.53}$$

with

$$\left.\begin{aligned} \varepsilon_t^{(j)} &\sim N(0,\lambda^{-1}R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} &\sim N(0,\lambda^{-1}R_\mu^{(j)}) \\ \delta\beta_t^{(j)} &\sim N(0,\lambda^{-1}R_\beta^{(j)}) \end{aligned}\right\} \tag{3.54}$$

We have four simple states (i.e. J = 4):

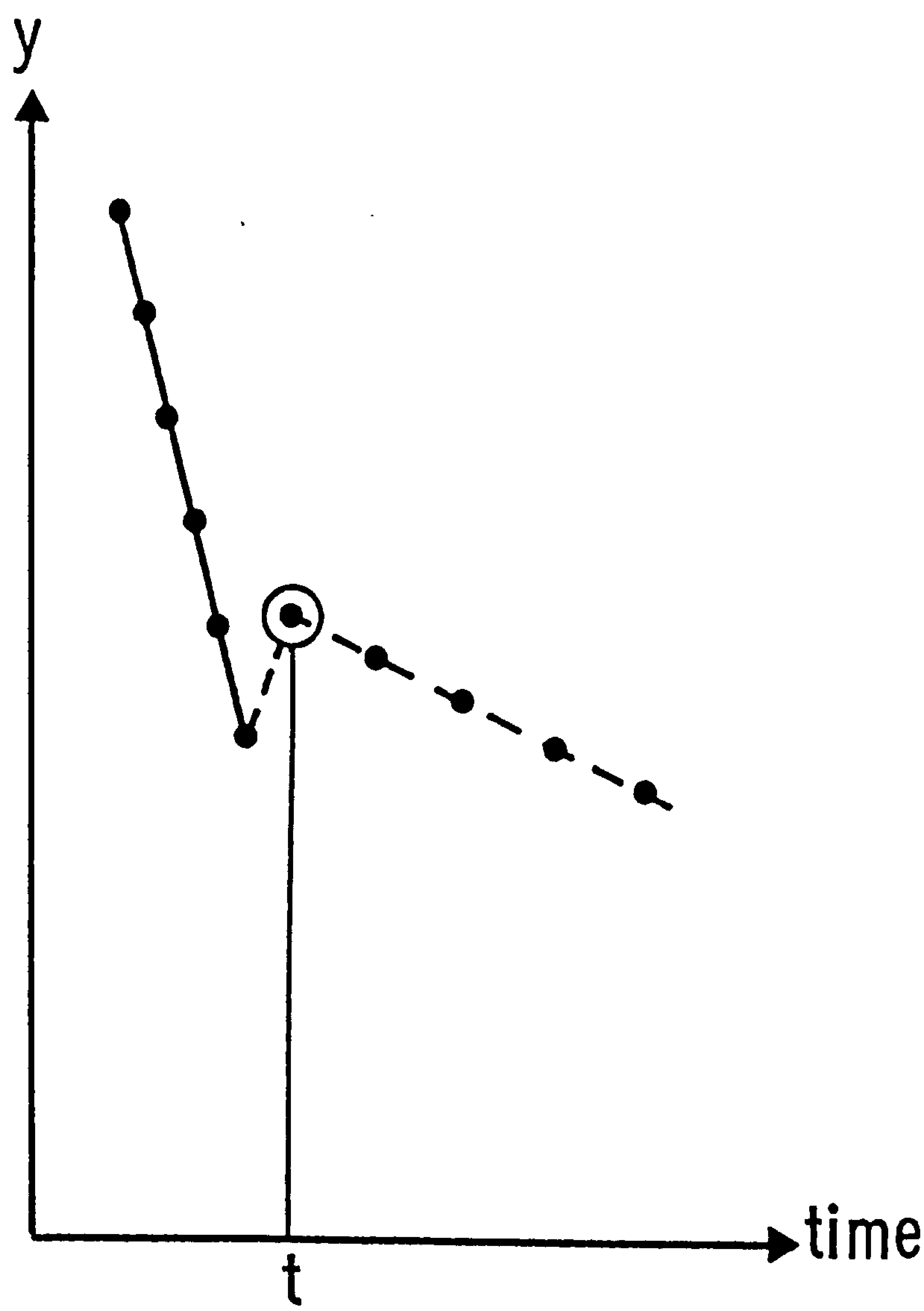
(i) $j = 1$: $R_\varepsilon^{(1)} = 1$, $R_\mu^{(1)} = 0$, $R_\beta^{(1)} = 0$. In this case $\delta\mu_t^{(j)}$ and $\delta\beta_t^{(j)}$ are identically zero, and therefore μ_t and β_t will not be perturbed, which represents the system in the steady state.

(ii) If $\delta\mu_t^{(j)}$ were non-zero only at time t, this would affect μ_t and, since μ_t is related to μ_{t+1} according to (3.52), μ_{t+1}, μ_{t+2} and so on would also be influenced. Therefore, a single non-zero $\delta\mu_t^{(j)}$ results in a change in the level of the measurements, y, beginning at time t. This can be achieved by setting: $R_\varepsilon^{(2)} = 1$, $R_\mu^{(2)}$ = positive, $R_\beta^{(2)} = 0$.

(iii) If $\delta\beta_t^{(j)}$ were non-zero only at time t, this would affect β_t and, since β_t is related to β_{t+1} according to (3.53), β_{t+1}, β_{t+2} and so on would also be influenced. Also, according to (3.52), μ_{t+1} will be influenced by β_{t+1} and μ_t (which is affected by β_t). In other words, a single non-zero $\delta\beta_t^{(j)}$ results in an incremental effect on the level, μ, and therefore produces a change in the slope of the measurements, y, beginning at time t. This can be achieved by setting: $R_\varepsilon^{(3)} = 1$, $R_\mu^{(3)} = 0$, $R_\beta^{(3)}$ = positive.

(iv) If $\varepsilon_t^{(j)}$ were very large only at time t, this would affect y_t according to (3.51) but not future values of y, since y_{t+1} is not directly related to y_t. Therefore, a single large $\varepsilon_t^{(j)}$ results in a transient observation at time t. This can be achieved by setting $R_\varepsilon^{(4)}$ = large positive, $R_\mu^{(4)} = 0$, $R_\beta^{(4)} = 0$.

Clearly, it would be possible to extend the multistate structure by including simple-state combinations in the overall model, e.g. with $R_\varepsilon^{(5)} = 1$, $R_\mu^{(5)}$ = positive, $R_\beta^{(5)}$ = positive we may model the situation where there is a concurrent level change and slope change (see Figure 3.1).



Level Change + Slope Change

$\delta\mu_t > 0$, $\delta\beta_t > 0$

FIGURE 3.1

For the remainder of this thesis, however, we restrict ourselves to a single changepoint type at any one timepoint. See Figure 3.2 for a pictorial display of the linear growth multistate structure where, for clarity of presentation, it is assumed that $\underset{\sim}{C}_t = 0 \ \forall t$ (where

$$\lambda^{-1}\underset{\sim}{C}_t = \text{var}\left(\begin{pmatrix}\mu_t \\ \beta_t\end{pmatrix} \middle| \underset{\sim}{D}_t\right).$$

NOTE: It can be seen from Figure 3.2 that the current observation, y_t (circled in diagram), may not be sufficient to discriminate between changepoint-types, and that it is necessary to receive y_{t+1} in order to be able to attempt to identify a specific change in pattern.

3.3.2 QUADRATIC GROWTH

For the multistate quadratic growth model:

$$y_t = \mu_t + \varepsilon_t^{(j)} \qquad (3.55)$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta\mu_t^{(j)} \qquad (3.56)$$

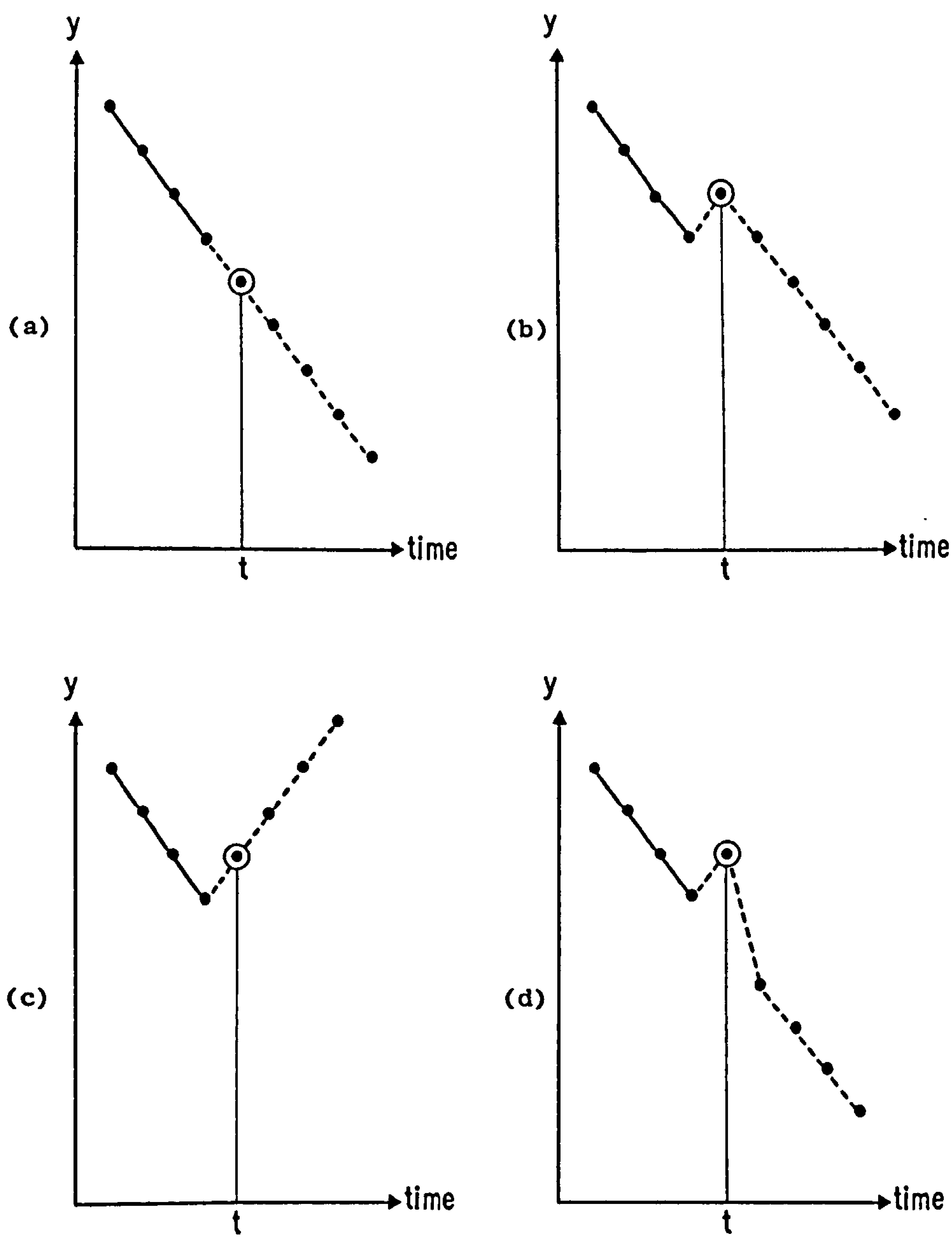
$$\beta_t = \beta_{t-1} + \gamma_t + \delta\beta_t^{(j)} \qquad (3.57)$$

$$\gamma_t = \gamma_{t-1} + \delta\gamma_t^{(j)} \qquad (3.58)$$

where

$$\left.\begin{aligned} \varepsilon_t^{(j)} &\sim N(0,\lambda^{-1} R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} &\sim N(0,\lambda^{-1} R_\mu^{(j)}) \\ \delta\beta_t^{(j)} &\sim N(0,\lambda^{-1} R_\beta^{(j)}) \\ \delta\gamma_t^{(j)} &\sim N(0,\lambda^{-1} R_\gamma^{(j)}) \end{aligned}\right\} \qquad (3.59)$$

the four error terms combine with the steady-state model to produce five simple states, i.e. $J = 5$.



(a) j = 1: steady state; (b) j = 2: level change, $\delta\mu_t > 0$;

(c) j = 3: slope change, $\delta\beta_t > 0$; (d) j = 4: transient, ε_t large

FIGURE 3.2

Pictorially, the multistate structure is shown in Figure 3.3, with $\underset{\sim}{C}_t = 0$ for clarity.

3.3.3 SINUSOIDAL MODEL

$$y_t = \mu_t + c_t\alpha_t + \varepsilon_t^{(j)} \tag{3.60}$$

$$\mu_t = \mu_{t-1} + \delta\mu_t^{(j)} \tag{3.61}$$

$$\alpha_t = \alpha_{t-1} + \delta\alpha_t^{(j)} \tag{3.62}$$

where

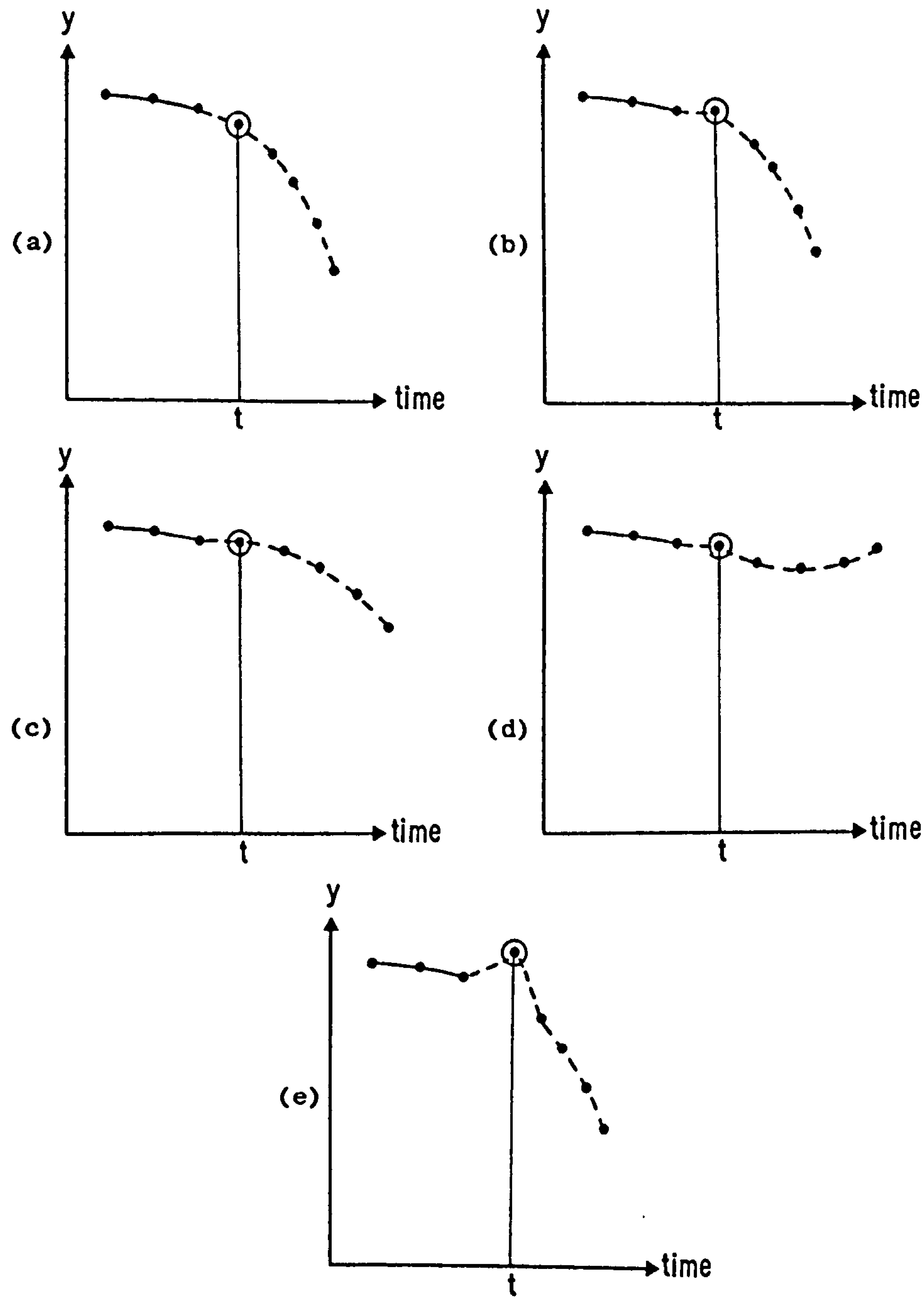
$$c_t = \cos(2\pi\omega t + \phi)$$

and

$$\left.\begin{aligned} \varepsilon_t^{(j)} &\sim N(0,\lambda^{-1}R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} &\sim N(0,\lambda^{-1}R_\mu^{(j)}) \\ \delta\alpha_t^{(j)} &\sim N(0,\lambda^{-1}R_\alpha^{(j)}) \end{aligned}\right\} \tag{3.63}$$

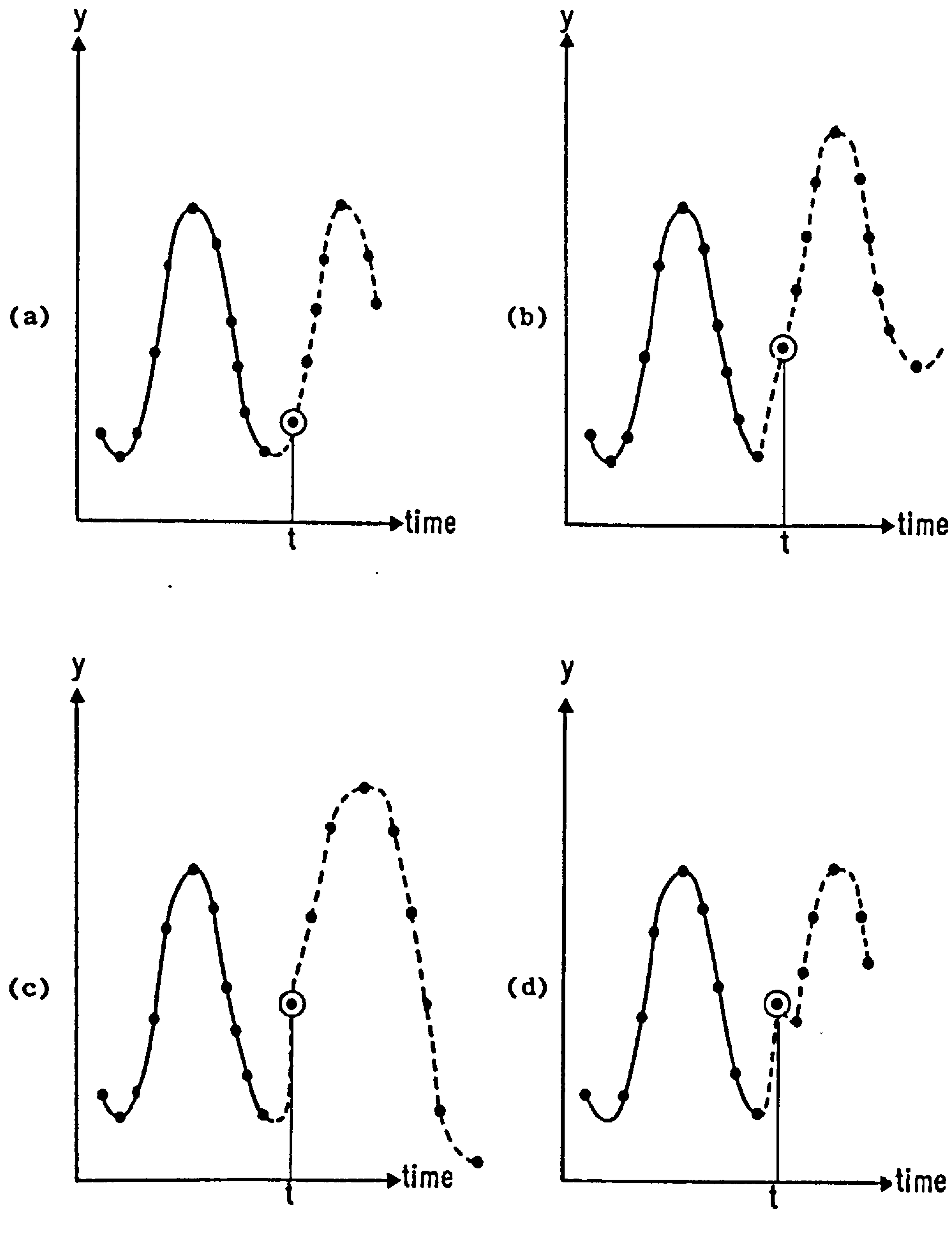
This multistate structure is demonstrated in Figure 3.4, assuming $\underset{\sim}{C}_t = 0$ for clarity.

Note that there are nuisance parameters in this model since $c_t = \cos(2\pi\omega t + \phi)$ with ω (the frequency) and ϕ (the phase) treated as nuisance parameters. The dynamic linear model specified above does not consider sudden changes in ω or ϕ. If this situation were likely to arise we would need to formulate an alternative dynamic linear model in order to handle it. However, for many medical time series ω is fairly rigid, since the rhythmic frequency is likely to be reasonably stable (either a twenty-hour-hour rhythm, seven-day rhythm, a twenty-eight-day cycle, etc.); ϕ is also unlikely to change rapidly, e.g. in rheumatoid arthritis, the time at which patients exhibit the most severe symptoms (usually in the morning) has been shown to be fairly constant across individuals (Kowanko et al. 1982).



(a) j = 1: steady state; (b) j = 2: level change, $\delta\mu_t > 0$;
(c) j = 3: slope change, $\delta\beta_t > 0$; (d) j = 4: quadratic change, $\delta\gamma_t > 0$;
(e) j = 5: transient, ε_t large.

FIGURE 3.3



(a) j = 1: steady state; (b) j = 2: level change, $\delta\mu_t > 0$;

(c) j = 3: amplitude change, $\delta\alpha_t > 0$; (d) j = 4: transient, ε_t large

FIGURE 3.4

3.3.4 SINUSOIDAL MODEL WITH LINEAR GROWTH

$$y_t = \mu_t + c_t\alpha_t + \varepsilon_t^{(j)} \tag{3.64}$$

$$\mu_t = \mu_{t-1} + \beta_t + \delta\mu_t^{(j)} \tag{3.65}$$

$$\beta_t = \beta_{t-1} + \delta\beta_t^{(j)} \tag{3.66}$$

$$\alpha_t = \alpha_{t-1} + \delta\alpha_t^{(j)} \tag{3.67}$$

where

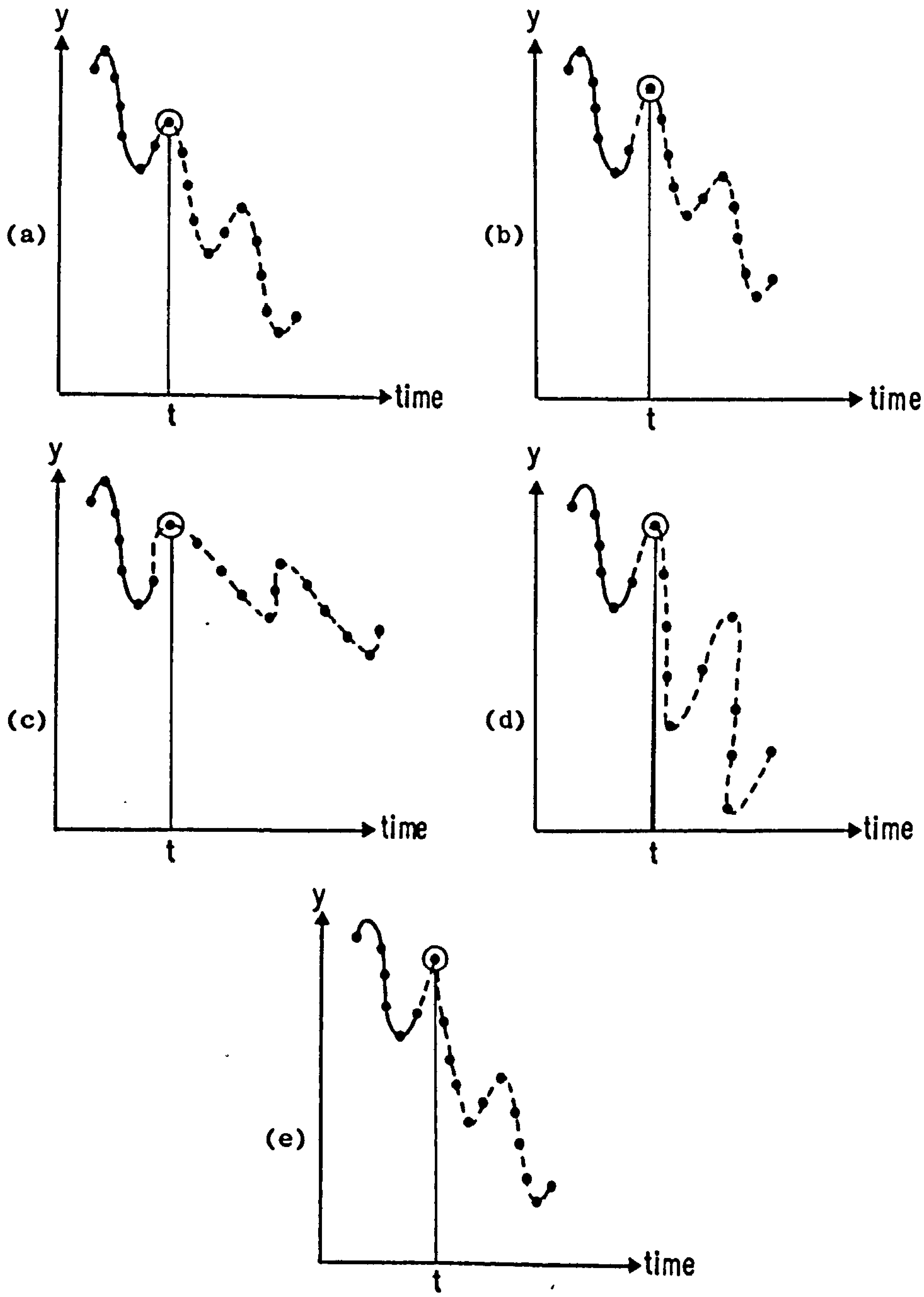
$$\left.\begin{array}{l} \varepsilon_t^{(j)} \sim N(0,\lambda^{-1}R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} \sim N(0,\lambda^{-1}R_\mu^{(j)}) \\ \delta\beta_t^{(j)} \sim N(0,\lambda^{-1}R_\beta^{(j)}) \\ \delta\alpha_t^{(j)} \sim N(0,\lambda^{-1}R_\alpha^{(j)}) \end{array}\right\} \tag{3.68}$$

The four error terms combine with the steady state to produce a five-state model (J = 5) as demonstrated by Figure 3.5 with $\underset{\sim}{C}_t = 0$ for clarity.

3.3.5 ARMA MODELS

For the models presented in Section 2.3.5, the ARMA parameters are treated as nuisance parameters. We shall not therefore be concerned with sudden changes in these parameters, although, of course, the grid method of estimation described in Section 2.2 allows for the accommodation of gradual changes, by recursively updating the distribution of $\underset{\sim}{\phi}$ upon receipt of successive observations.

Our experience has been that for those medical situations thus encountered where an ARMA model is adequate, sudden changes in the ARMA parameters have no apparent physical interpretation. If, however, a situation arises where a change in an ARMA parameter <u>is</u> seen to be important, one could specify an alternative



(a) j = 1: steady state; (b) j = 2: level change, $\delta\mu_t > 0$

(c) j = 3: slope change, $\delta\beta_t' > 0$; (d) j = 4: amplitude change, $\delta\alpha_t > 0$;

(d) j = 5: transient, ε_t large.

FIGURE 3.5

dynamic linear model along the lines of that described by Harrison and Stevens (1976), or one could use the techniques proposed by Ezzet and Smith (1985), based upon the score statistic, which may be more appropriate to this detection problem.

3.3.5.1: AR(1).

$$y_t = \mu_t + \varepsilon_t^{(j)} \tag{3.69}$$

$$\mu_t - \nu_t = \phi(\mu_{t-1} - \nu_{t-1}) + \delta\mu_t^{(j)} \tag{3.70}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t^{(j)} \tag{3.71}$$

where

$$\left.\begin{array}{l} \varepsilon_t^{(j)} \sim N(0,\lambda^{-1}R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} \sim N(0,\lambda^{-1}R_\mu^{(j)}) \\ \delta\nu_t^{(j)} \sim N(0,\lambda^{-1}R_\nu^{(j)}) \end{array}\right\} \tag{3.72}$$

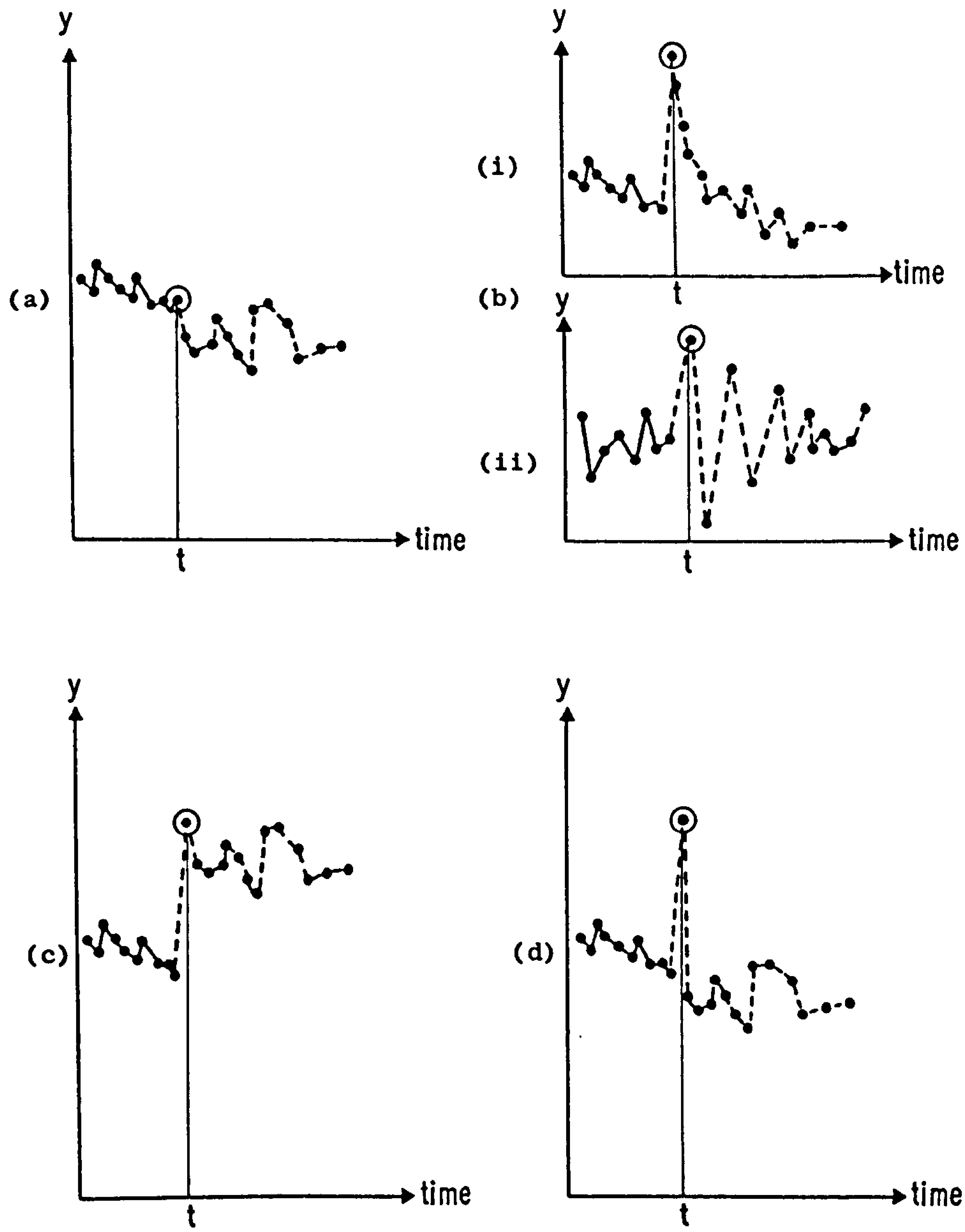
The multistate structure is shown in Figure 3.6. Notice that the changepoint phenomenon arising from a large $\delta\mu_t^{(j)}$ takes on a different appearance depending upon whether the autoregressive parameter, ϕ, is positive or negative; this type of discontinuity is referred to as an 'impulse'. Figures 3.7 to 3.10 demonstrate this changing impulse characteristic for a variety of ϕ values in the range (-1,1), using simulated data with a changepoint (impulse) occurring at time t = 50.

3.3.5.2: MA(1).

$$y_t = \mu_t + \varepsilon_t^{(j)} \tag{3.73}$$

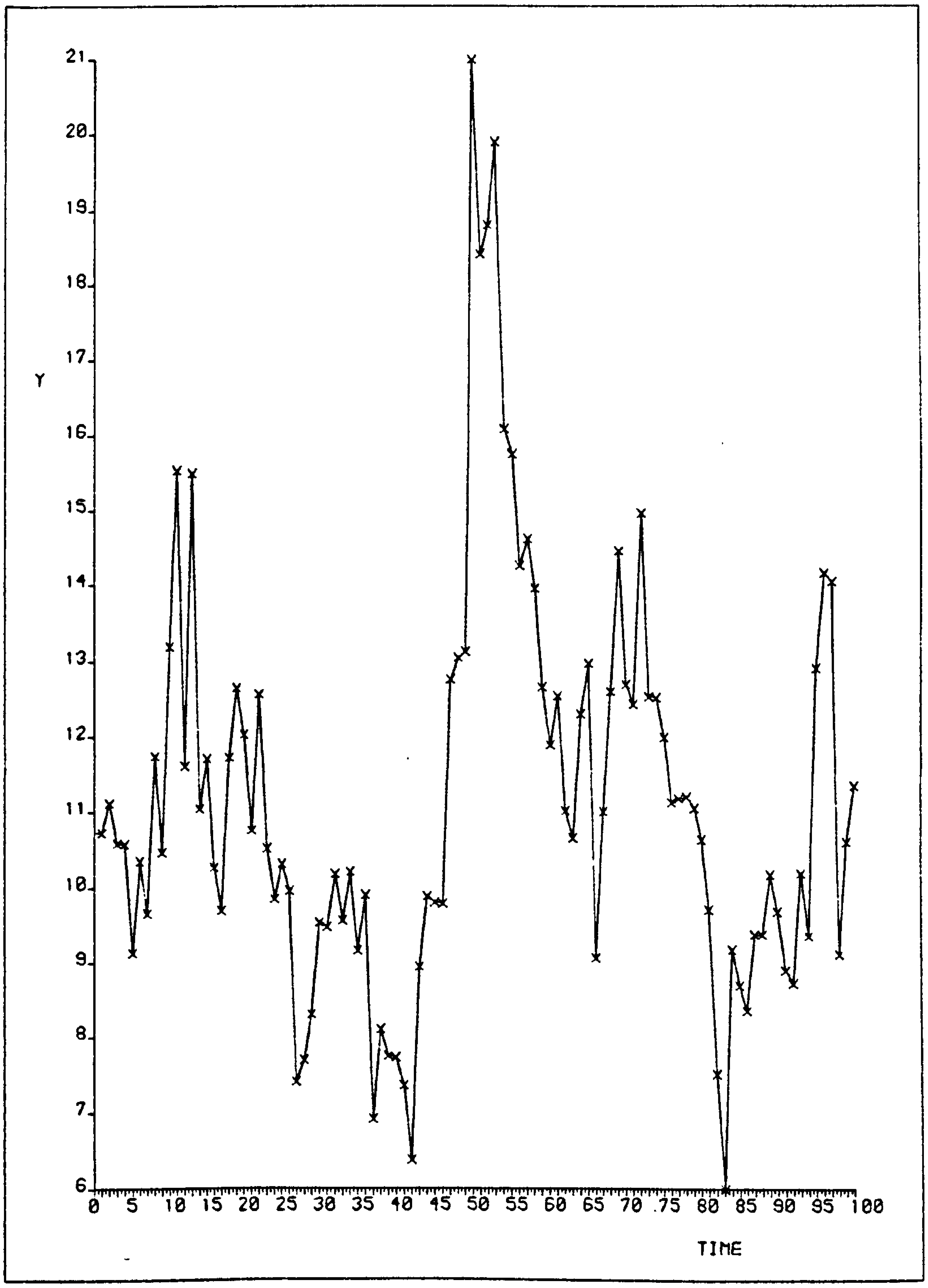
$$\mu_t = \nu_t + \delta\mu_t^{(j)} - \eta\delta\mu_{t-1}^{(i)} \tag{3.74}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t^{(j)} \tag{3.75}$$



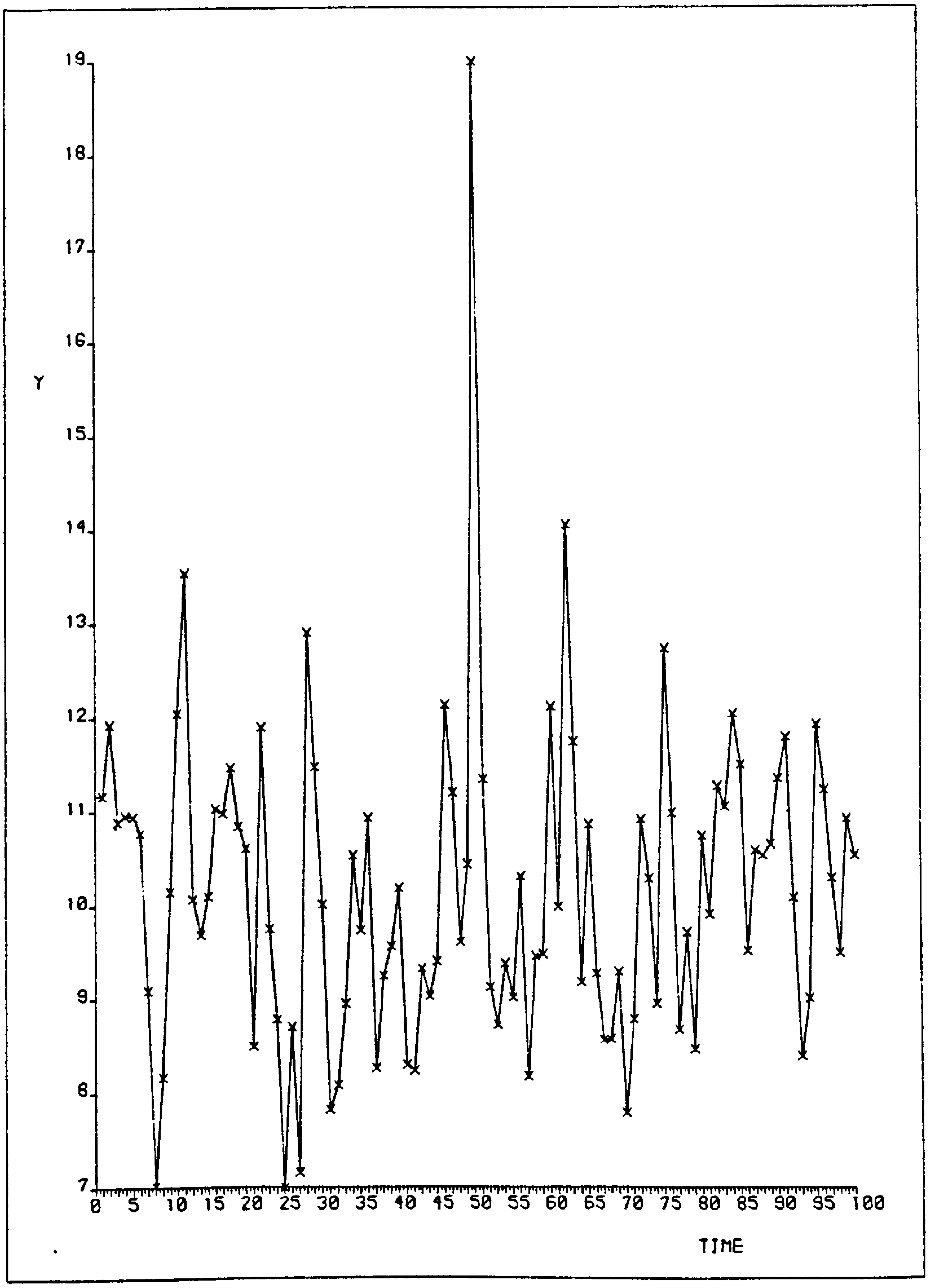
(a) j = 1: steady state; (b) j = 2: impulse, $\delta\mu_t$ large,
(i) $\phi > 0$, (ii) $\phi < 0$
(c) j = 3: level change, $\delta\nu_t > 0$; (d) j = 4: transient, ε_t large.

FIGURE 3.6



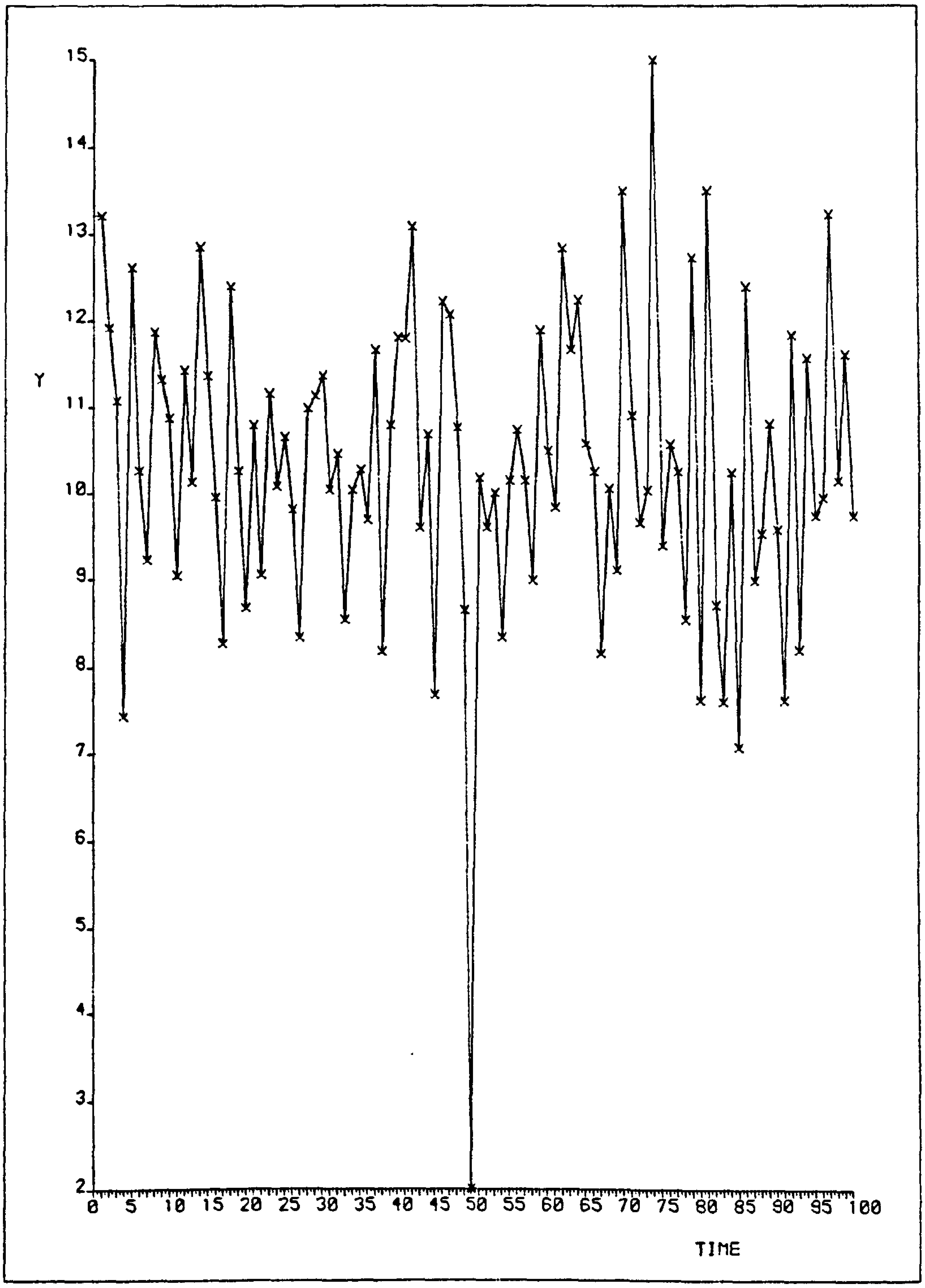
$\phi = 0.9$

FIGURE 3.7



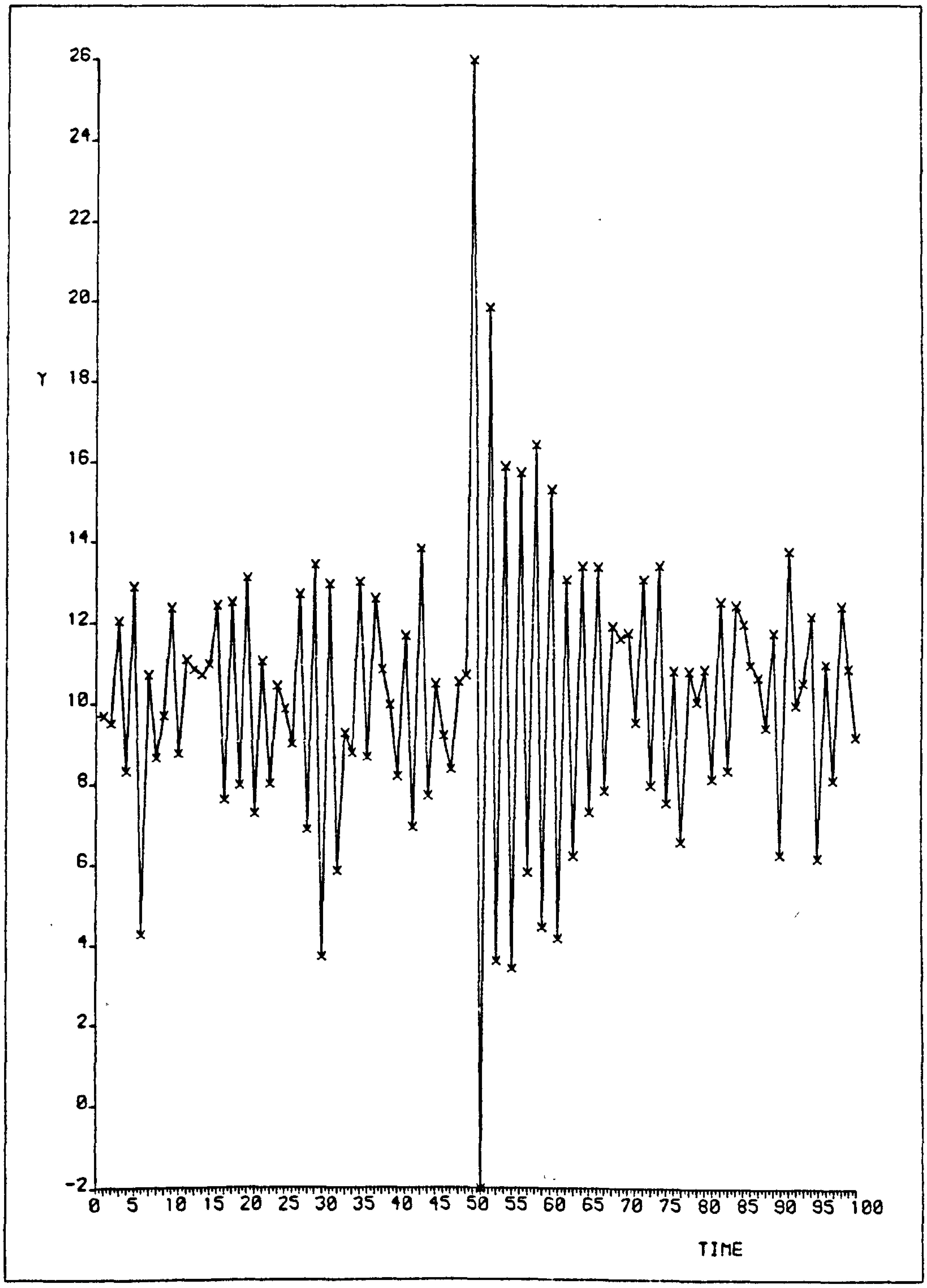
$\phi = 0.2$

FIGURE 3.8



ϕ = -0.2

FIGURE 3.9



$\phi = -0.9$

FIGURE 3.10

where

$$\left.\begin{aligned} \varepsilon_t^{(j)} &\sim N(0,\lambda^{-1}R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} &\sim N(0,\lambda^{-1}R_\mu^{(j)}) \\ \delta\nu_t^{(j)} &\sim N(0,\lambda^{-1}R_\nu^{(j)}) \end{aligned}\right\} \quad (3.76)$$

The multistate structure is given in Figure 3.11 where, once more, the impulse is dependent upon the sign of η. Figures 3.12 to 3.15 demonstrate this dependence, using simulated data with an impulse at time t = 48.

3.3.5.3: AR(2).

$$y_t = \mu_t + \varepsilon_t^{(j)} \quad (3.77)$$

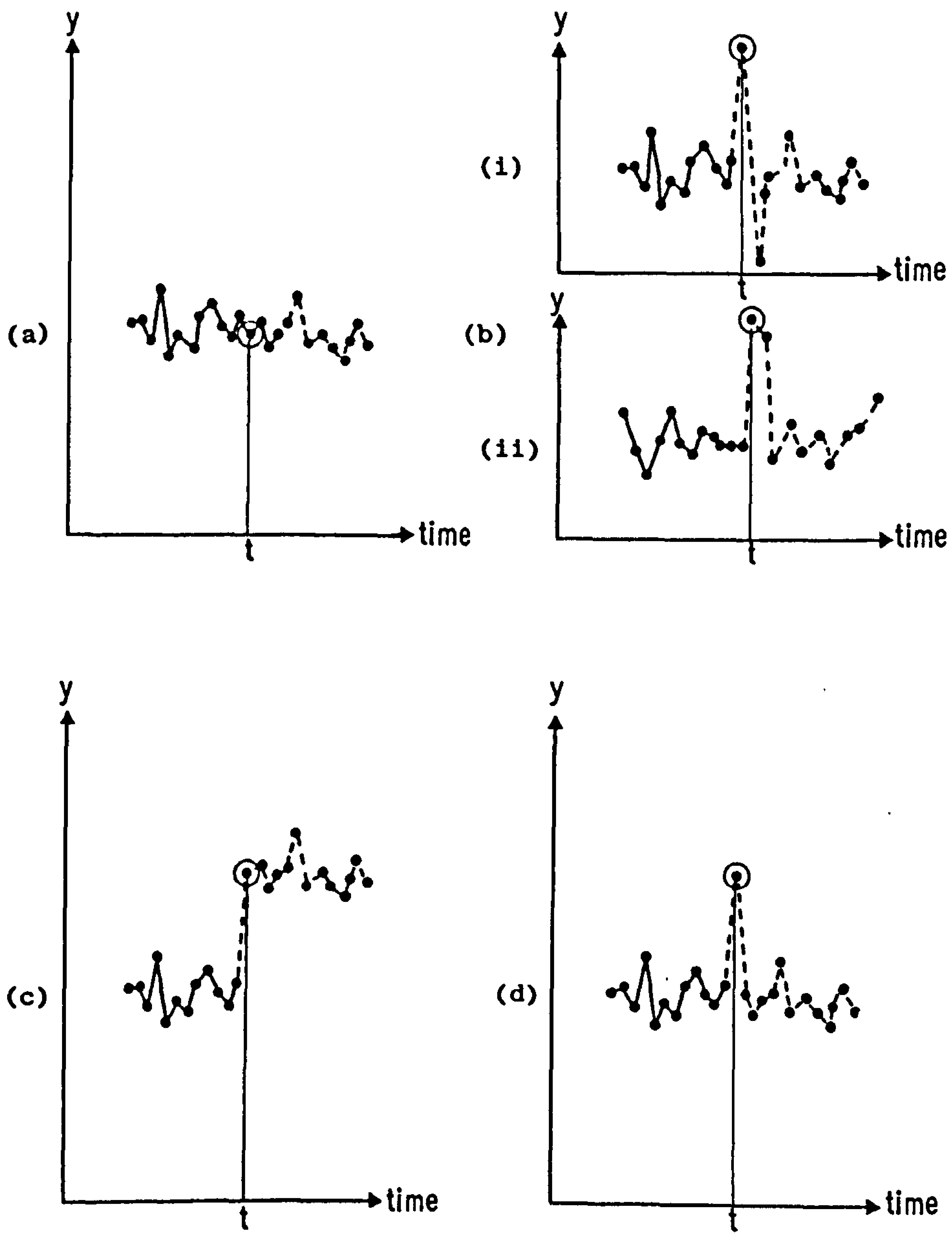
$$\mu_t - \nu_t = \phi_1(\mu_{t-1} - \nu_{t-1}) + \phi_2(\mu_{t-2} - \nu_{t-2}) + \delta\mu_t^{(j)} \quad (3.78)$$

$$\nu_t = \nu_{t-1} + \delta\nu_t^{(j)} \quad (3.79)$$

where

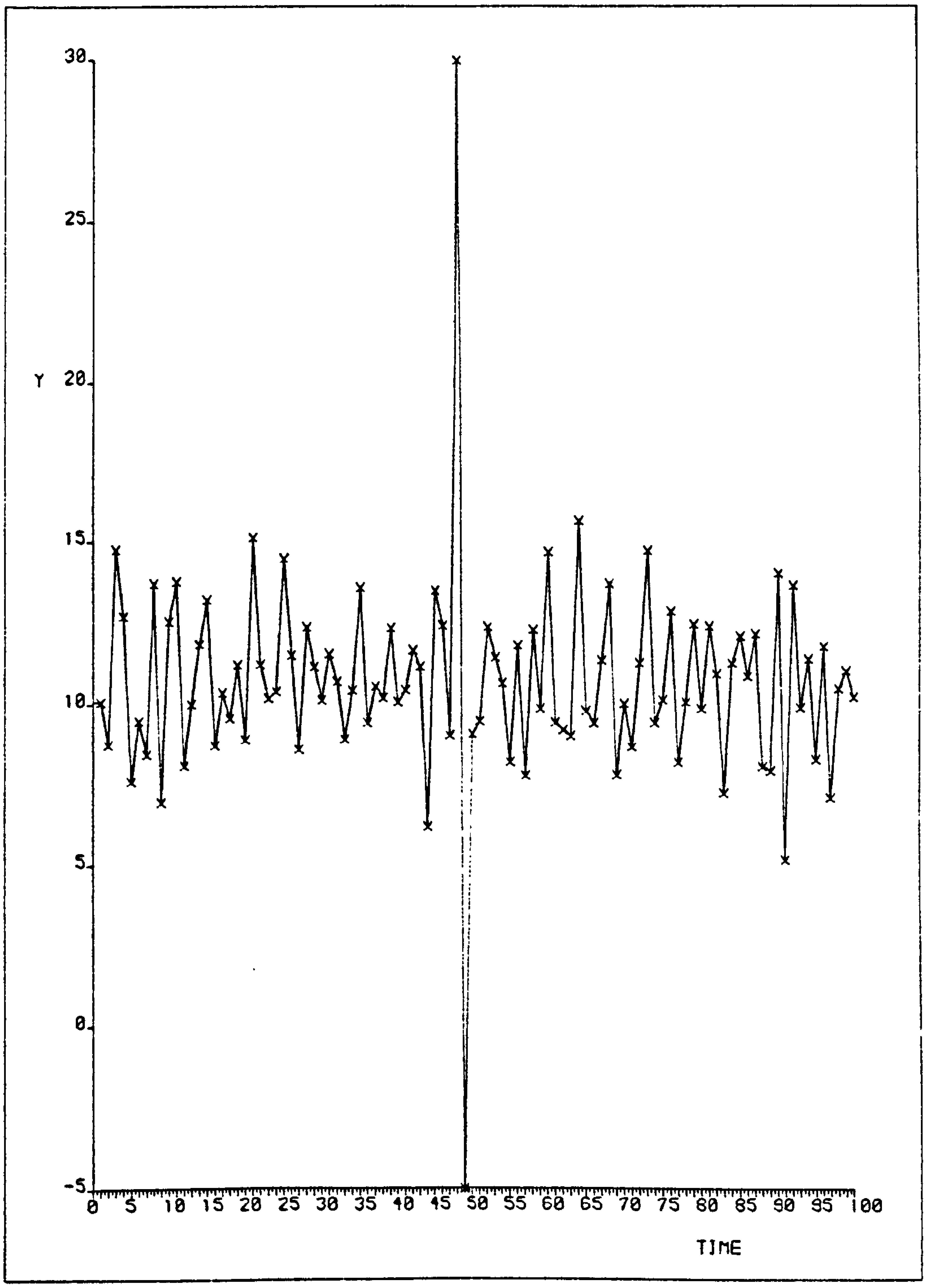
$$\left.\begin{aligned} \varepsilon_t^{(j)} &\sim N(0,\lambda^{-1}R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} &\sim N(0,\lambda^{-1}R_\mu^{(j)}) \\ \delta\nu_t^{(j)} &\sim N(0,\lambda^{-1}R_\nu^{(j)}) \end{aligned}\right\} \quad (3.80)$$

The multistate structure, using typical values of ϕ_1,ϕ_2 is given in Figure 3.16; Figures 3.17 to 3.20 show the effect of changes in ϕ_1 and ϕ_2 on the impulse characteristic (changepoint at time t = 50). Note that for Figures 3.17 and 3.18, $\phi_1^2 + 4\phi_2 < 0$, resulting in a damped sinusoidal autocorrelation function and pseudoperiodic patterns in the time series, whereas for Figures 3.19 and 3.20, $\phi_1^2 + 4\phi_2 > 0$ resulting in an autocorrelation function which is a mixture of damped exponentials (see Box and Jenkins (1970) for the derivation of this property).



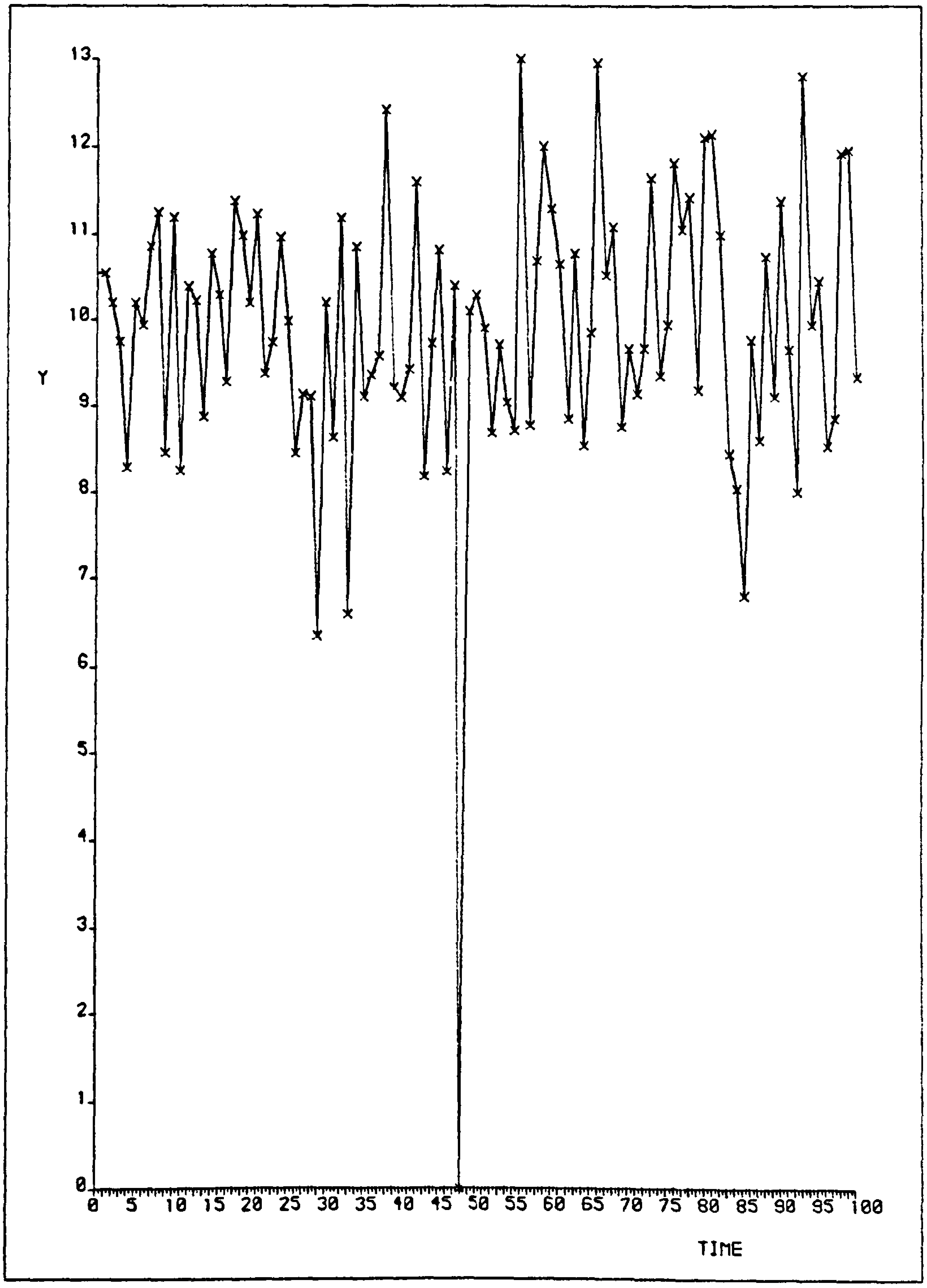
(a) j = 1: steady state; (b) j = 2: impulse, $\delta\mu_t$ large
(i) $\eta > 0$, (ii) $\eta < 0$
(c) j = 3: level change, $\delta\nu_t > 0$; (d) j = 4: transient, ε_t large

FIGURE 3.11



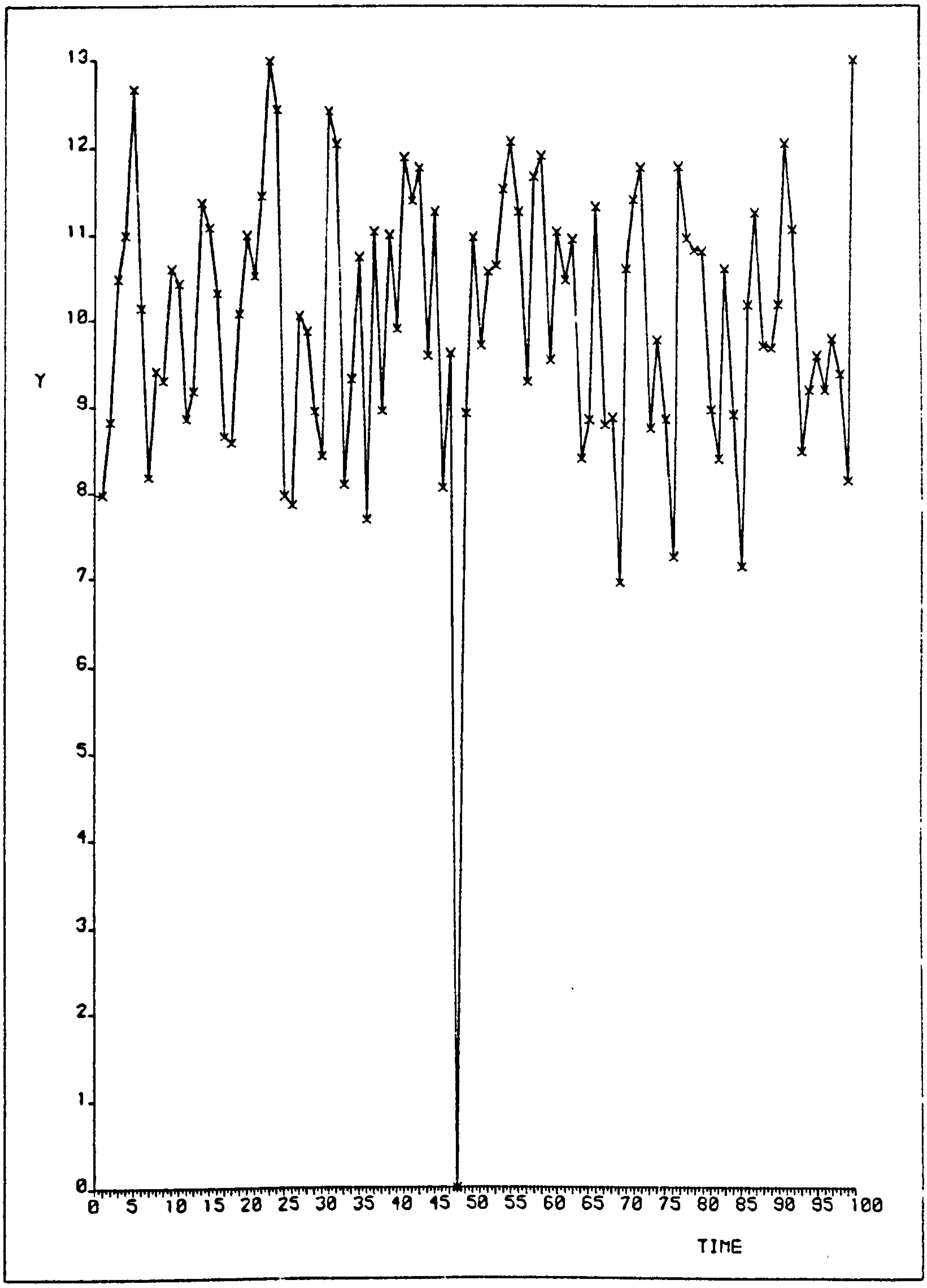
$\eta = 0.9$

FIGURE 3.12



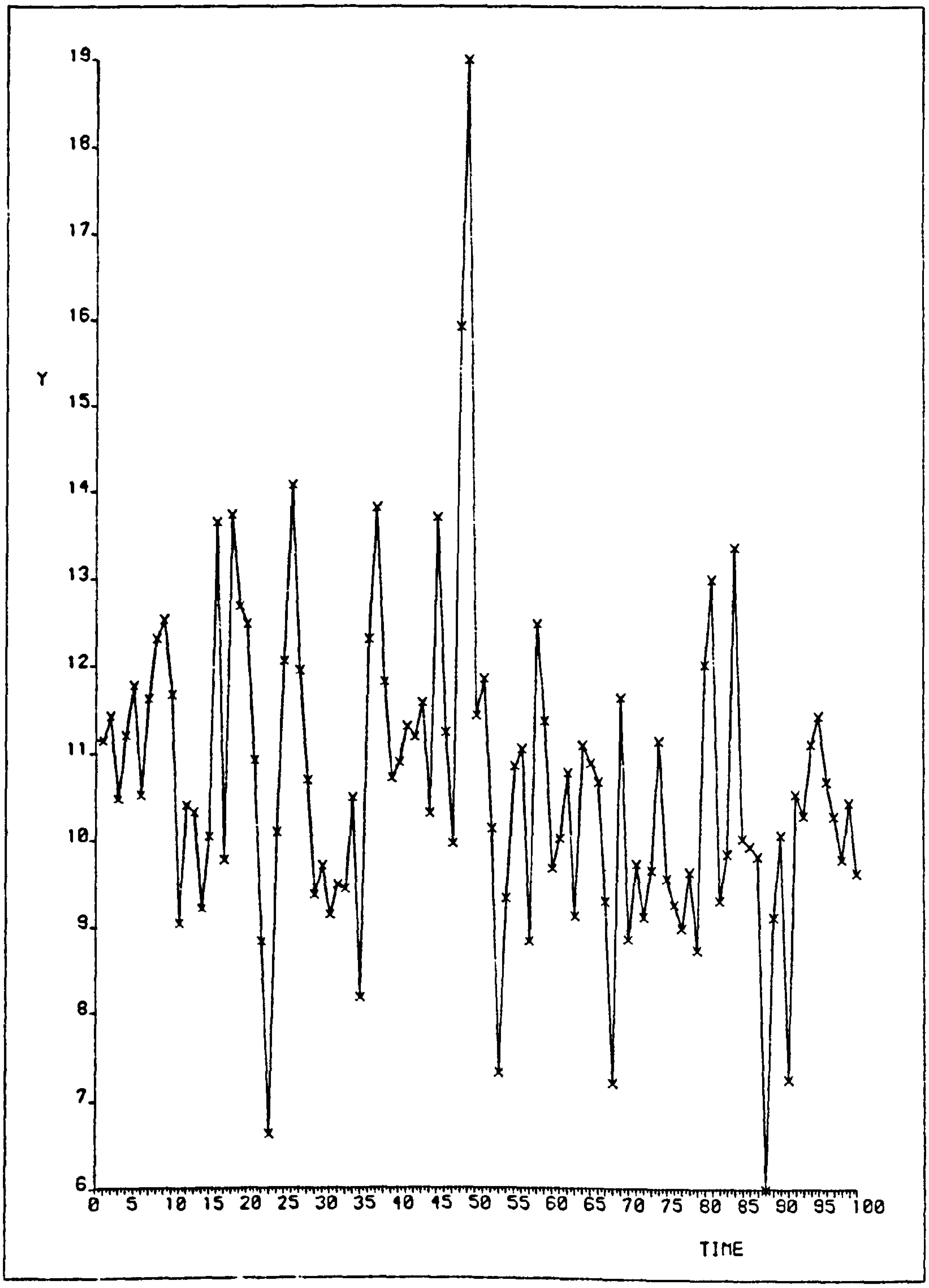
$\eta = 0.2$

FIGURE 3.13



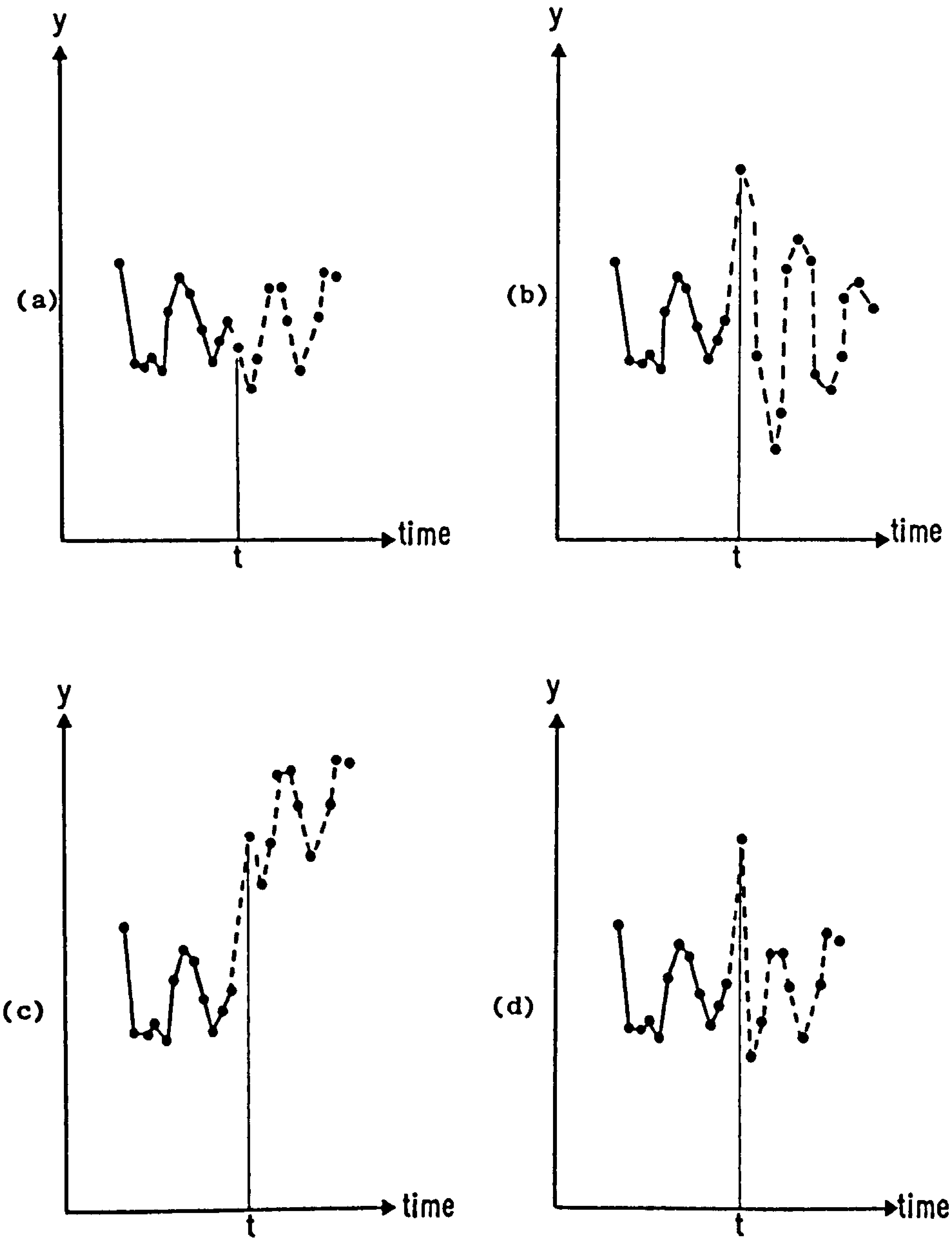
$\eta = -0.2$

FIGURE 3.14



$\eta = -0.9$

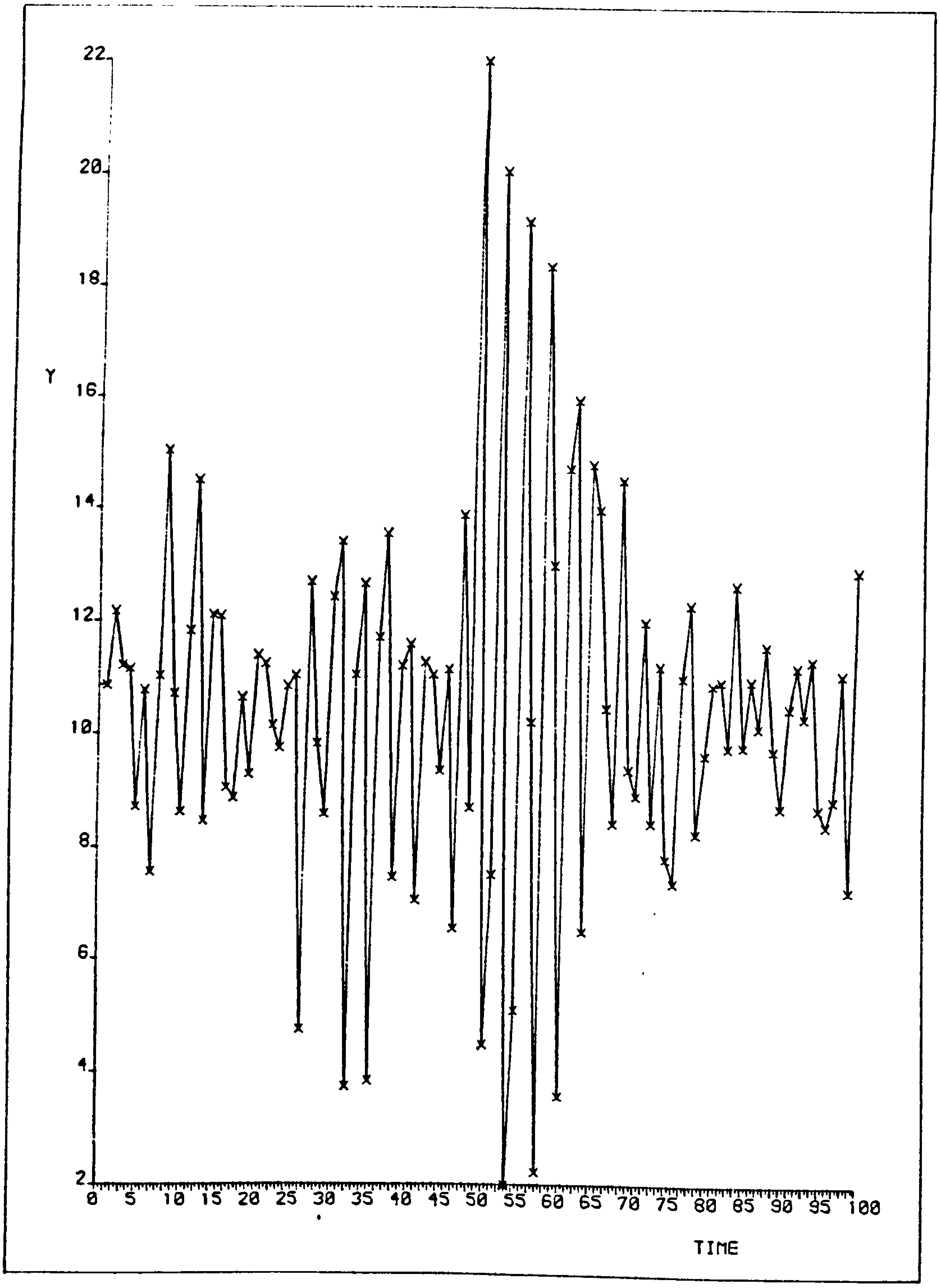
FIGURE 3.15



$\phi_1 = 0.8; \quad \phi_2 = -0.8$

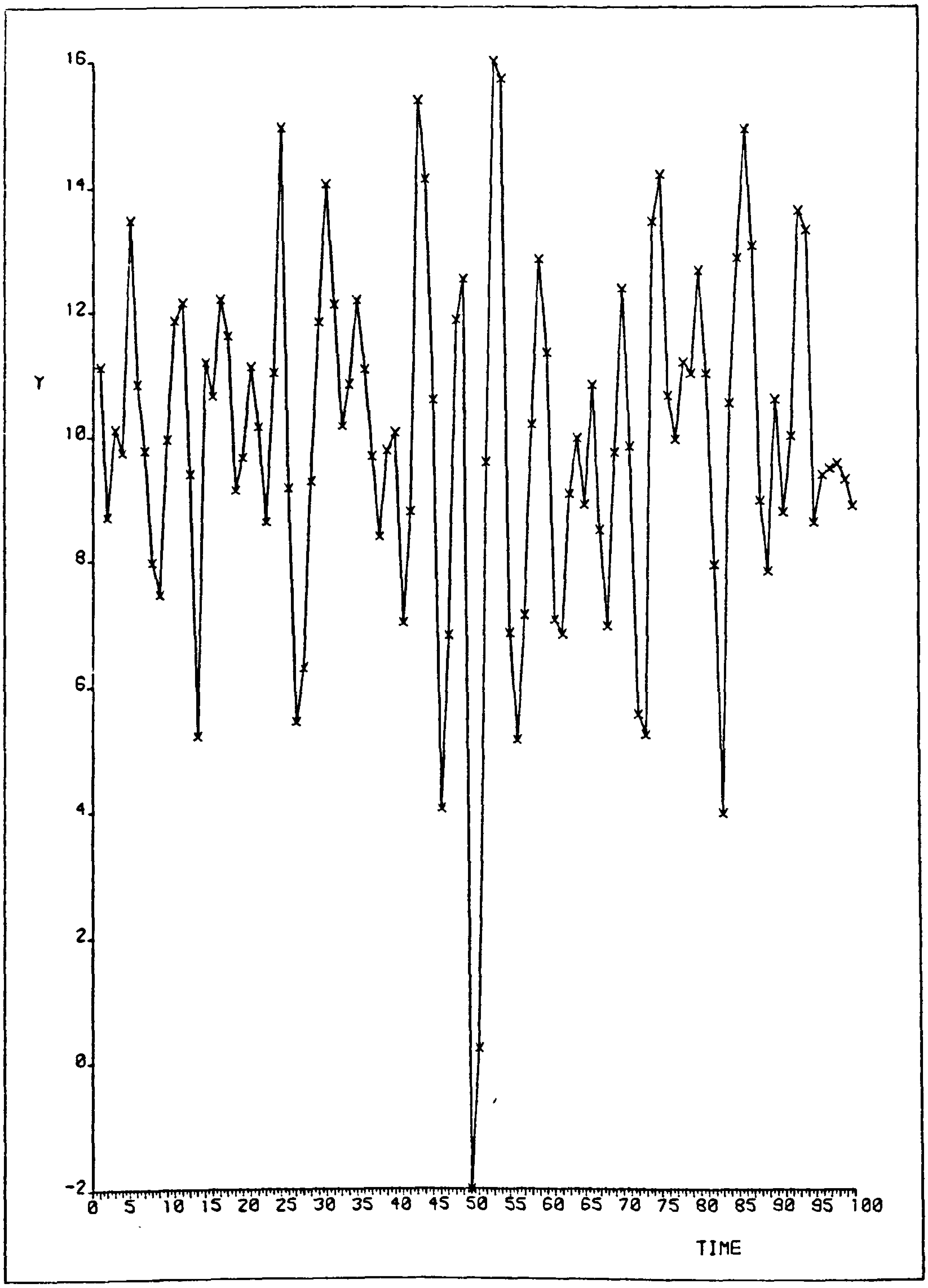
(a) j = 1: steady state; (b) j = 2: impulse, $\delta\mu_t$ large;
(c) j = 3: level change, $\delta\nu_t > 0$; (d) j = 4: transient, ε_t large

FIGURE 3.16



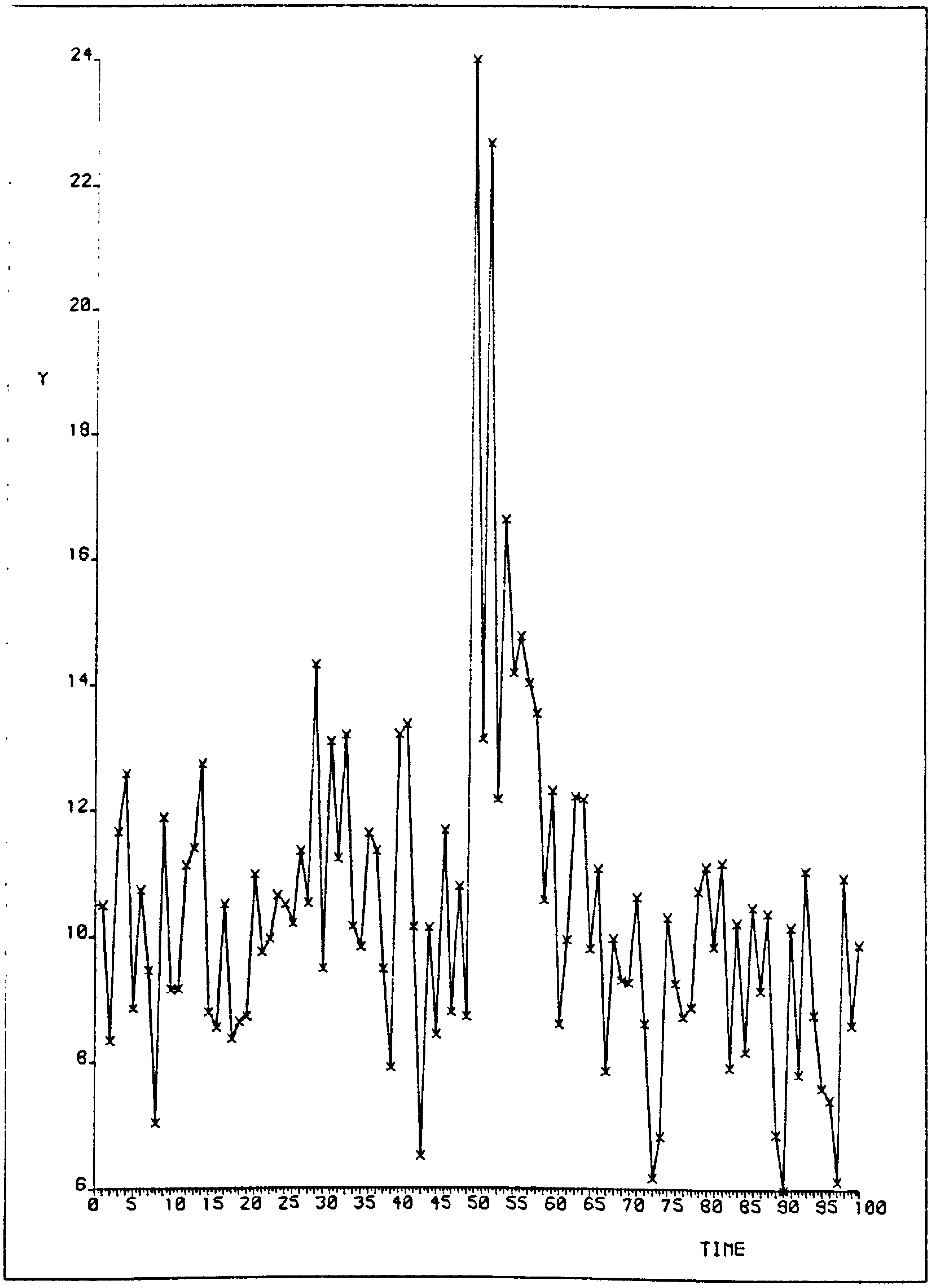
$\phi_1 = -0.8, \ \phi_2 = -0.8$

FIGURE 3.17



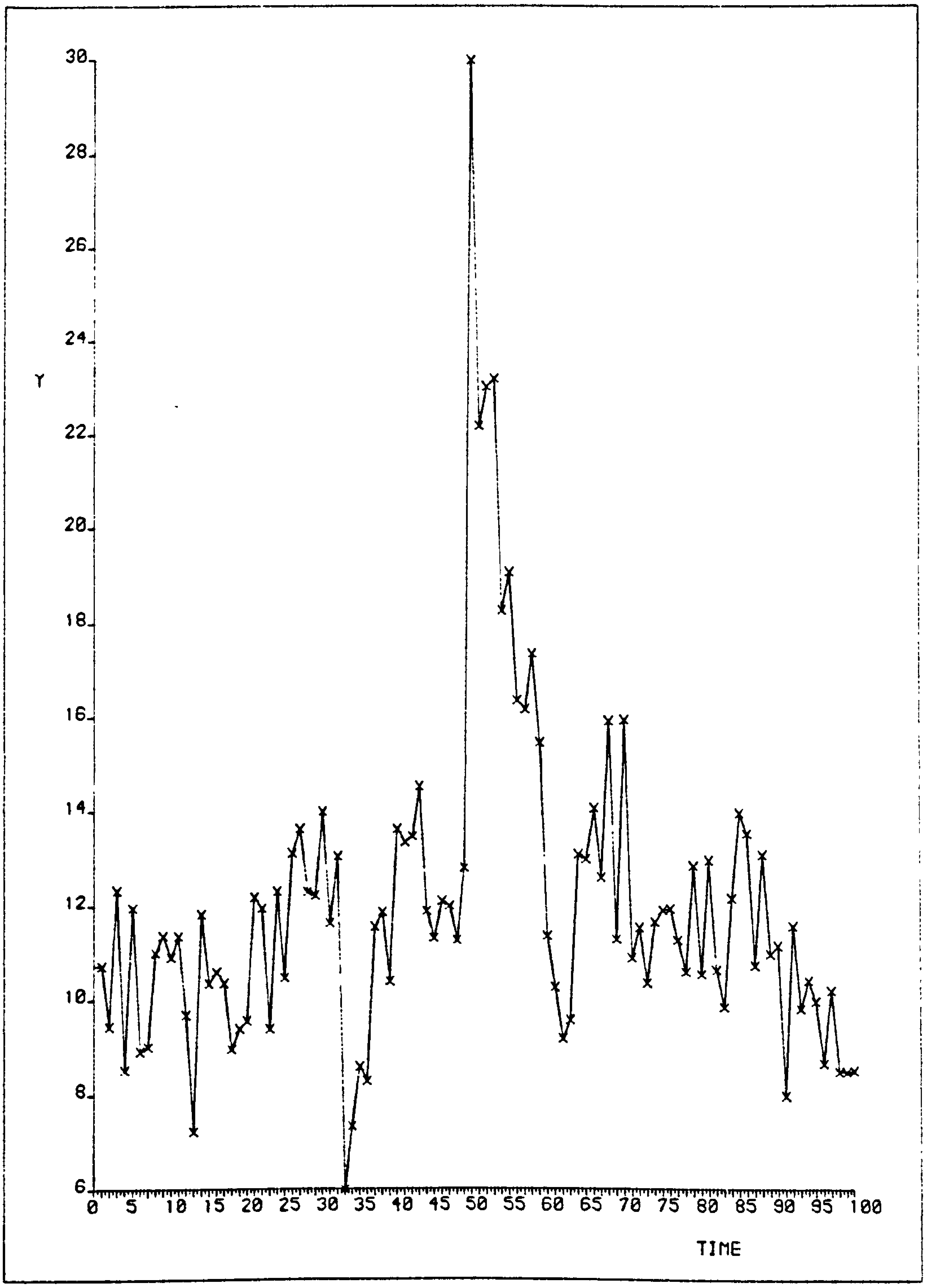
$\phi_1 = 0.8,\ \phi_2 = -0.8$

FIGURE 3.18



$\phi_1 = 0.2, \phi_2 = 0.6$

FIGURE 3.19



$\phi_1 = 0.6, \phi_2 = 0.2$

FIGURE 3.20

3.3.5.4: MA(2).

$$y_t = \mu_t + \varepsilon_t^{(j)} \tag{3.81}$$

$$\mu_t = \nu_t + \delta\mu_t^{(j)} - \eta_1\delta\mu_{t-1}^{(i)} - \eta_2\delta\mu_{t-2}^{(h)} \tag{3.82}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t \tag{3.83}$$

where

$$\left.\begin{array}{l} \varepsilon_t^{(j)} \sim N(0,\lambda^{-1}R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} \sim N(0,\lambda^{-1}R_\mu^{(j)}) \\ \delta\nu_t^{(j)} \sim N(0,\lambda^{-1}R_\nu^{(j)}) \end{array}\right\} \tag{3.84}$$

The multistate structure, using typical values of η_1,η_2, is shown in Figure 3.21; Figures 3.22 to 3.25 demonstrate the effect of changes in η_1 and η_2 on the impulse characteristic (with changepoint at time t = 48).

3.3.5.5: ARMA(1,1).

$$y_t = \mu_t + \varepsilon_t^{(j)} \tag{3.85}$$

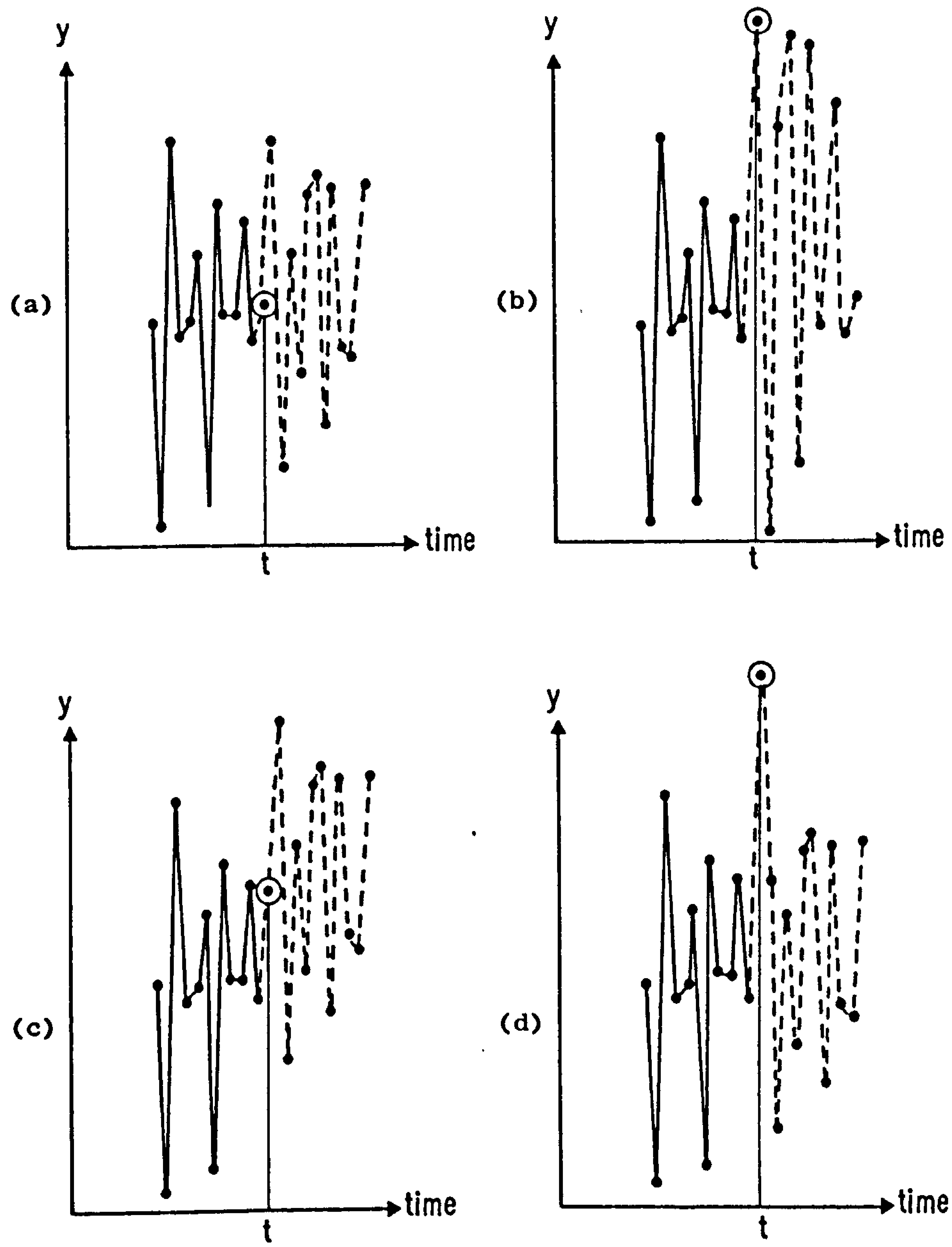
$$\mu_t - \nu_t = \phi(\mu_{t-1} - \nu_{t-1}) + \delta\mu_t^{(j)} - \eta\delta\mu_{t-1}^{(i)} \tag{3.86}$$

$$\nu_t = \nu_{t-1} + \delta\nu_t^{(j)} \tag{3.87}$$

where

$$\left.\begin{array}{l} \varepsilon_t^{(j)} \sim N(0,\lambda^{-1}R_\varepsilon^{(j)}) \\ \delta\mu_t^{(j)} \sim N(0,\lambda^{-1}R_\mu^{(j)}) \\ \delta\nu_t^{(j)} \sim N(0,\lambda^{-1}R_\nu^{(j)}) \end{array}\right\} \tag{3.88}$$

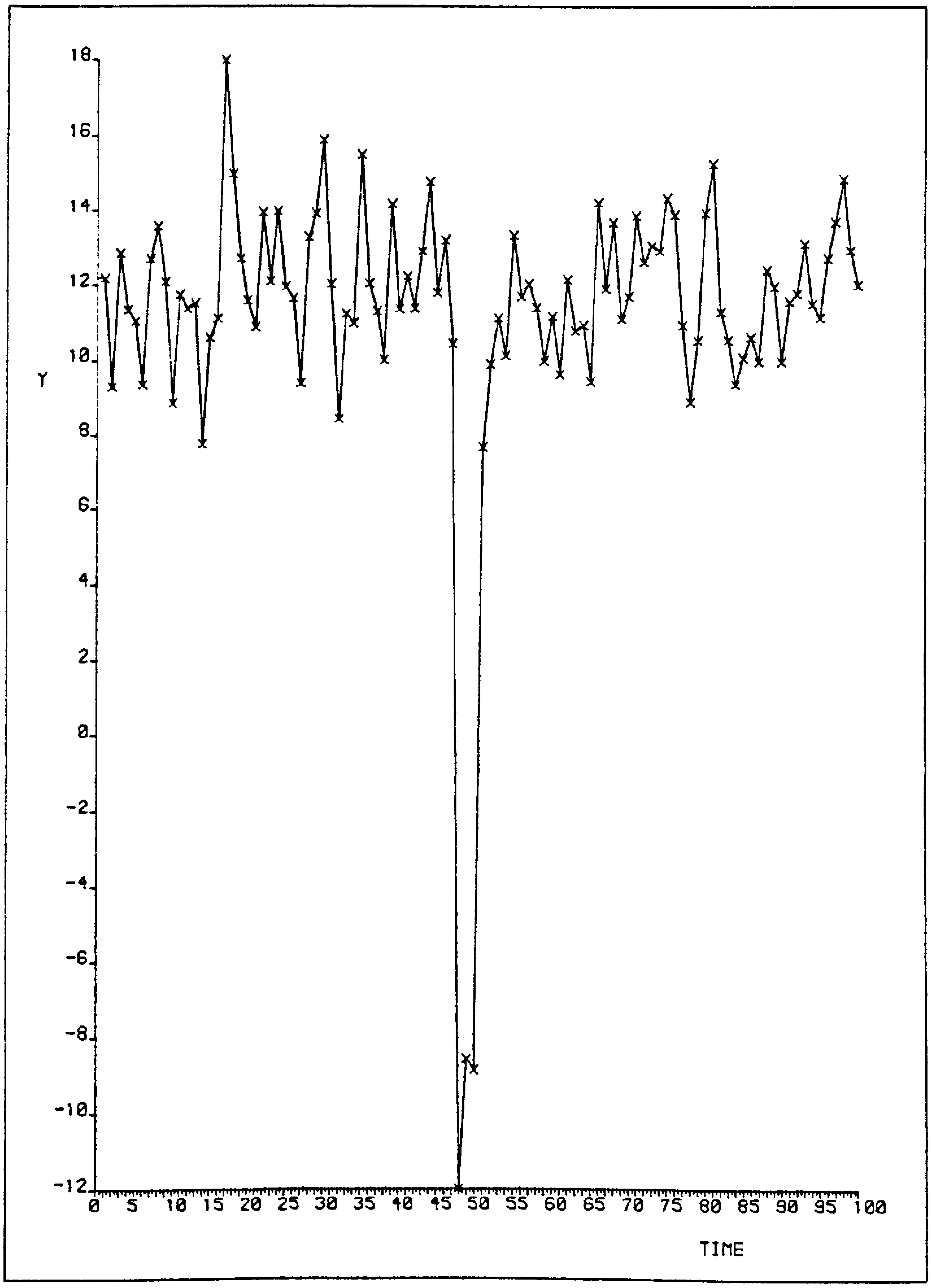
The multistate structure, using typical values of ϕ,η, is shown in Figure 3.26; Figures 3.27 to 3.30 show the effect of changes in ϕ and η on the impulse characteristic (with the change-point at time t = 49).



$\eta_1 = 0.8; \quad \eta_2 = -0.8$

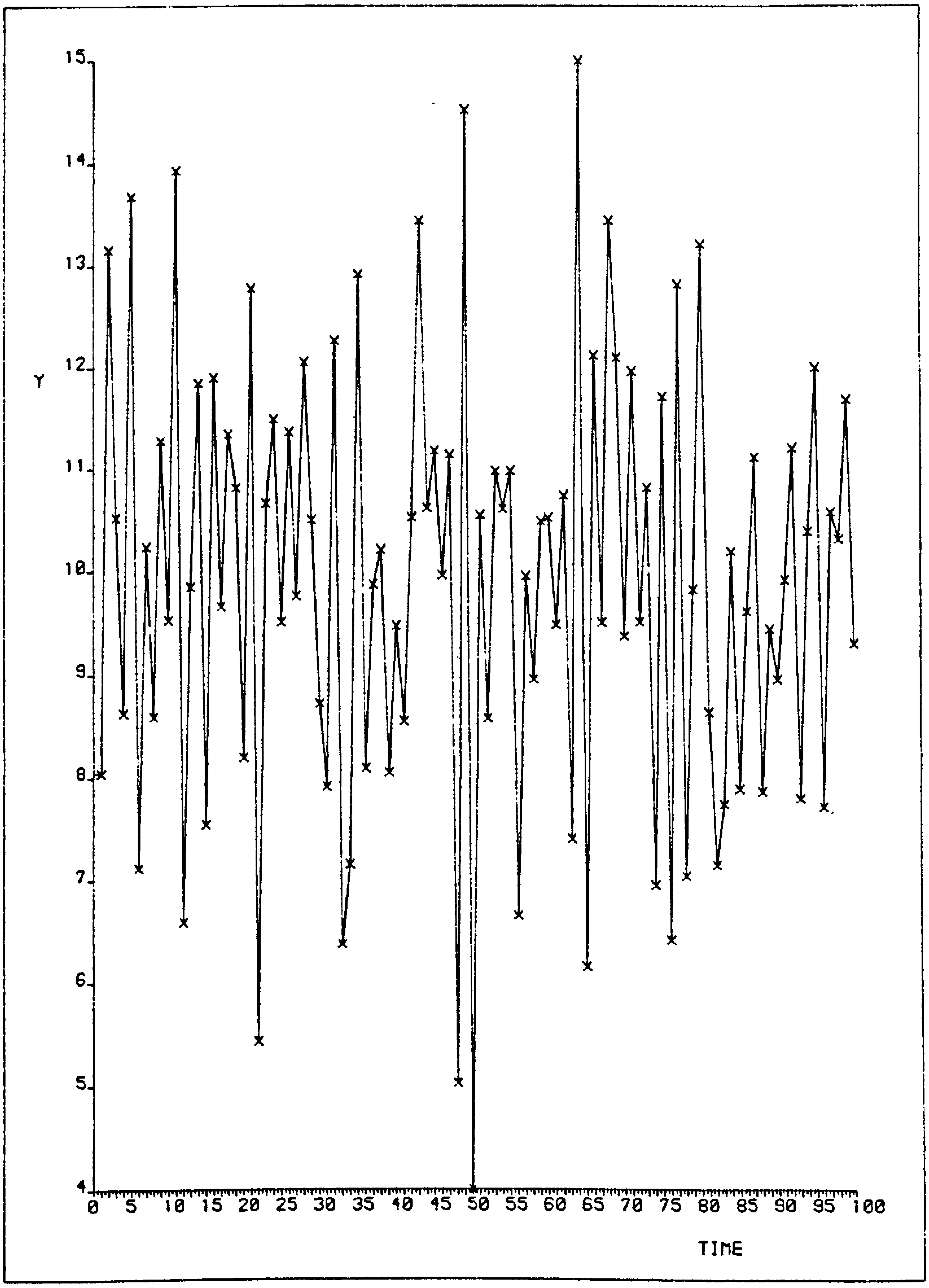
(a) $j = 1$: steady state; (b) $j = 2$: impulse, $\delta\mu_t$ large;
(c) $j = 3$: level change, $\delta\nu_t > 0$; (d) $j = 4$: transient, ε_t large

FIGURE 3.21



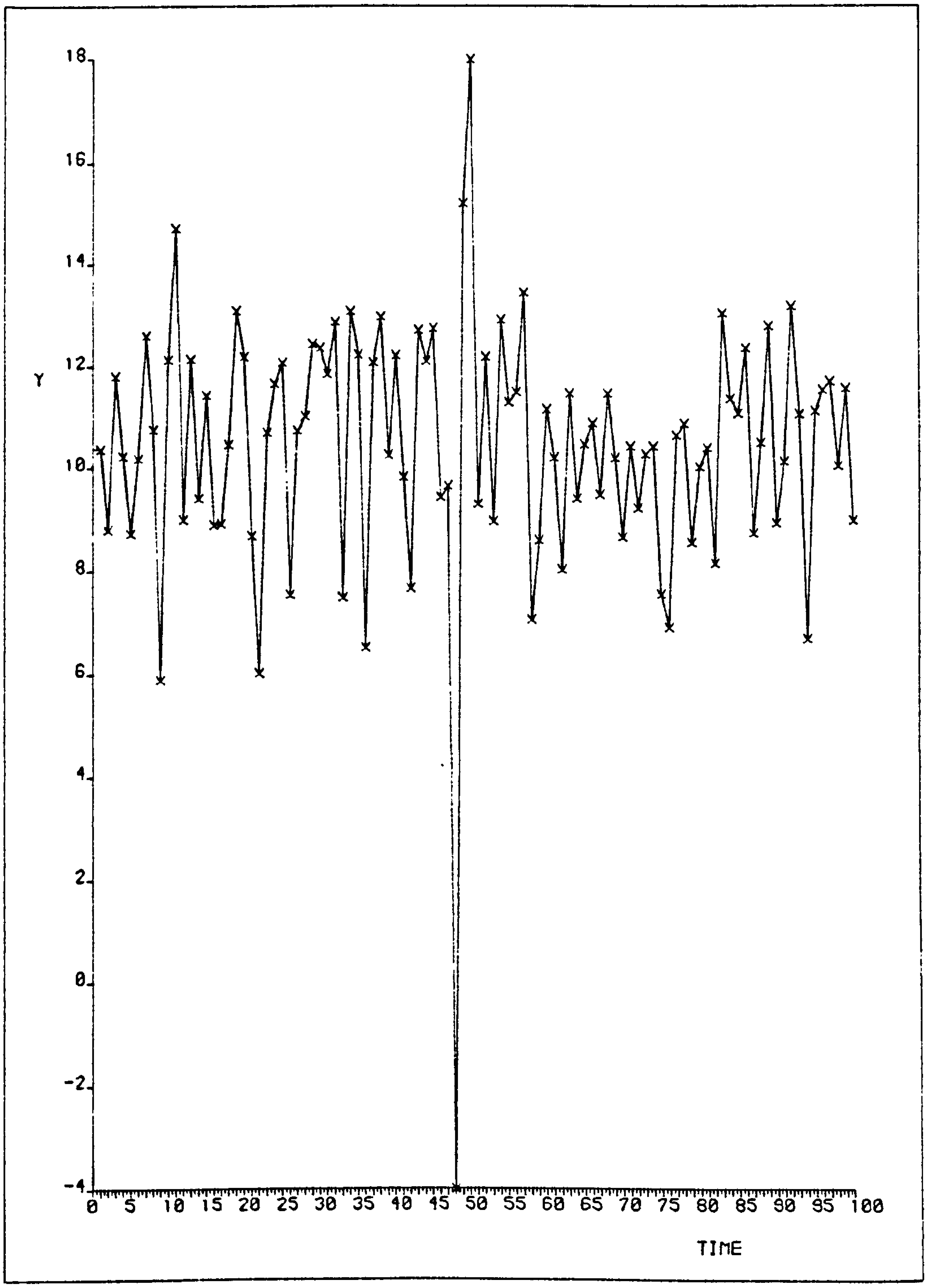
$\phi_1 = -0.8, \phi_2 = -0.8$

FIGURE 3.22



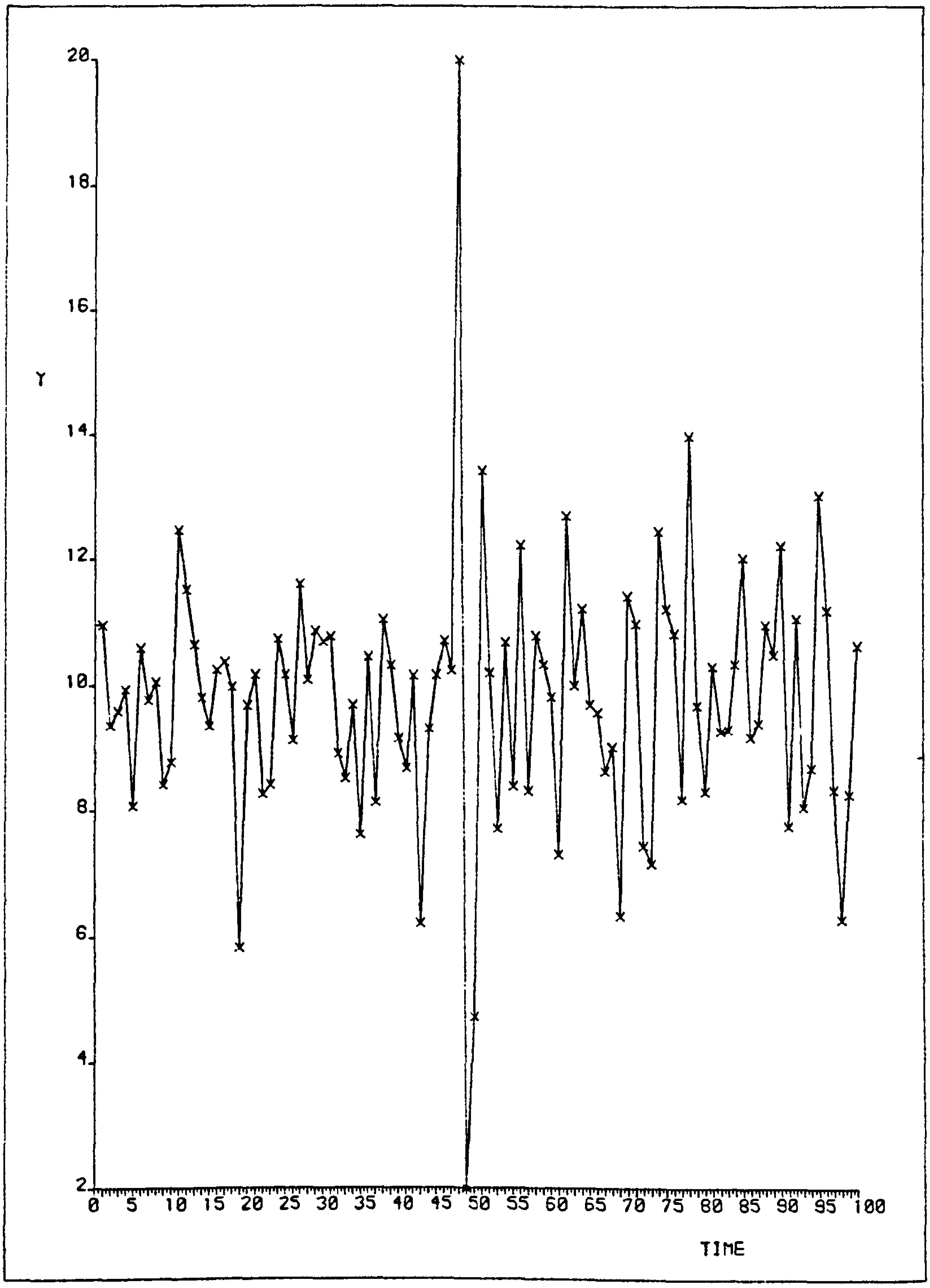
$\eta_1 = 0.8, \eta_2 = -0.8$

FIGURE 3.23



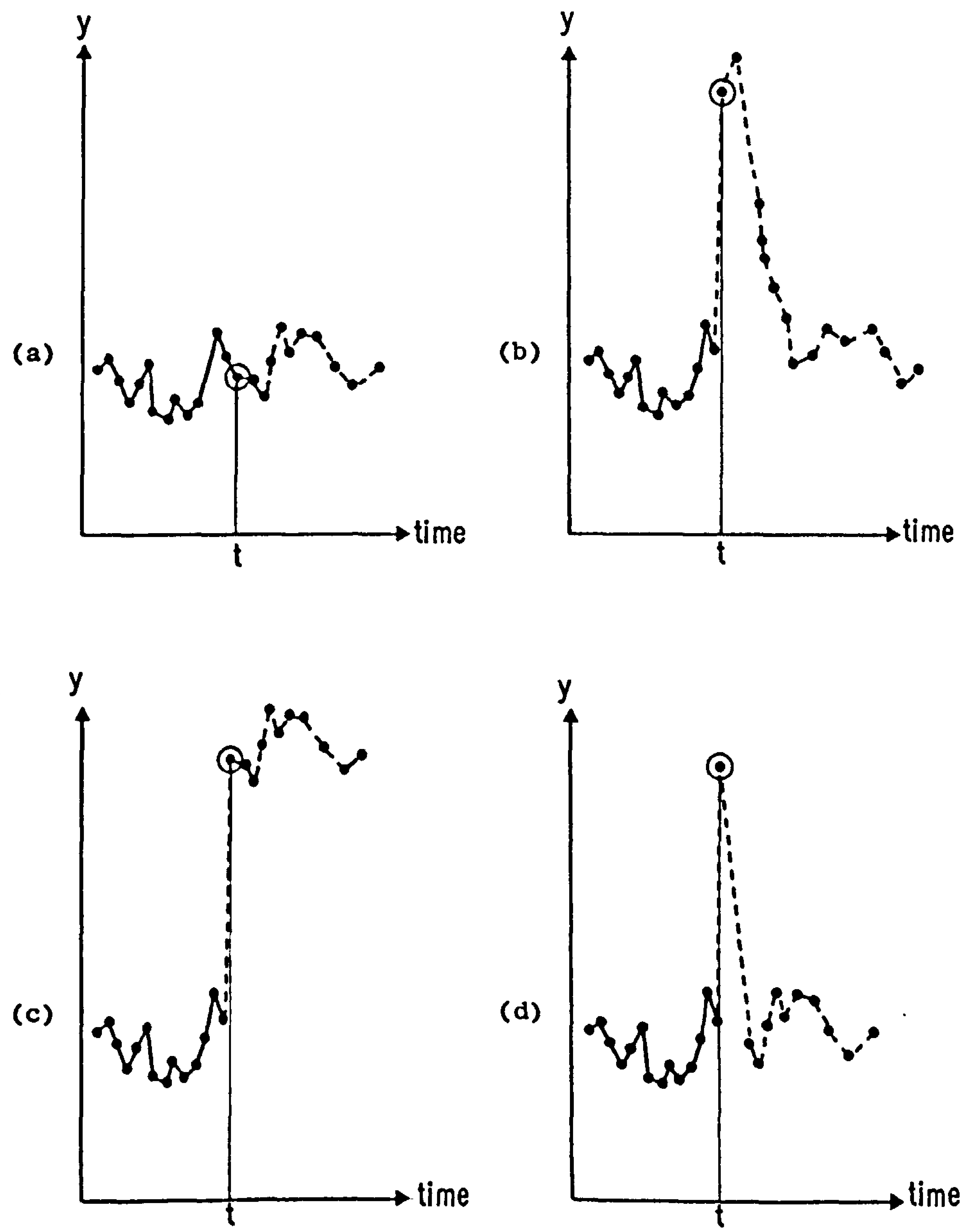
$\eta_1 = 0.2,\ \eta_2 = 0.6$

FIGURE 3.24



$\eta_1 = 0.6, \eta_2 = 0.2$

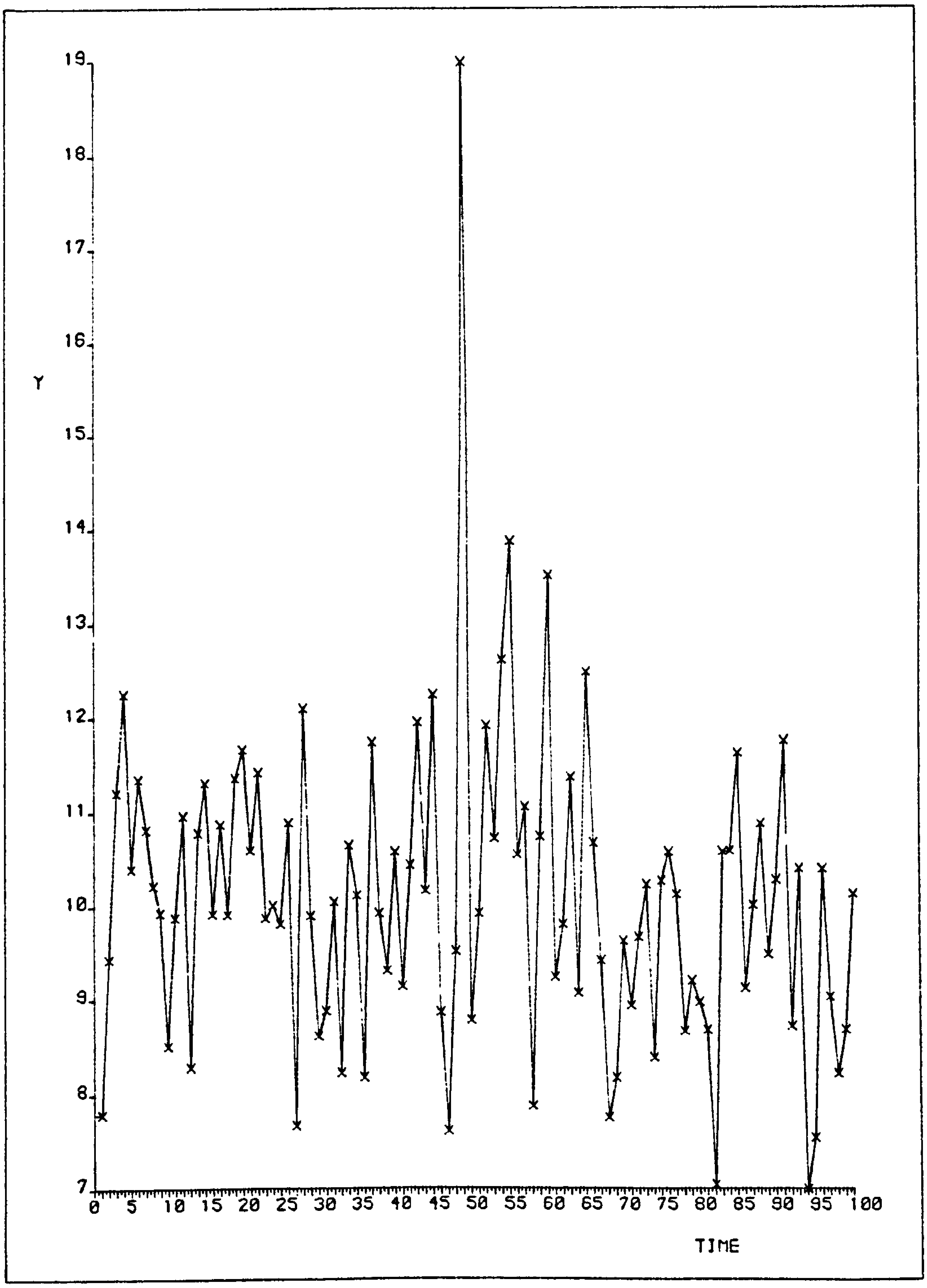
FIGURE 3.25



$\phi = 0.8$; $\eta = -0.8$

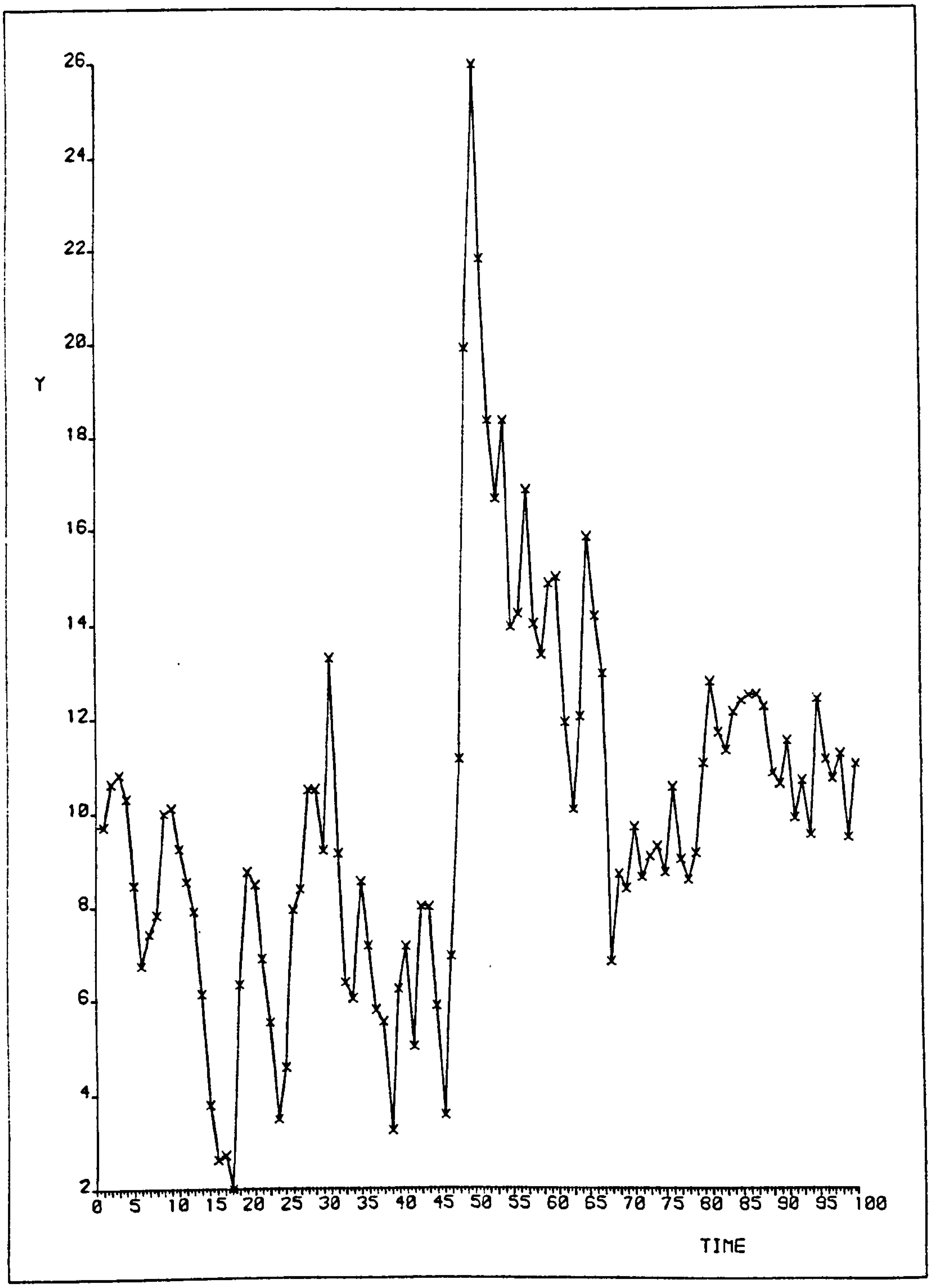
(a) $j = 1$: steady state; (b) $j = 2$: impulse, $\delta\mu_t$ large;
(c) $j = 3$: level change, $\delta\nu_t > 0$; (d) $j = 4$: transient, ε_t large.

FIGURE 3.26



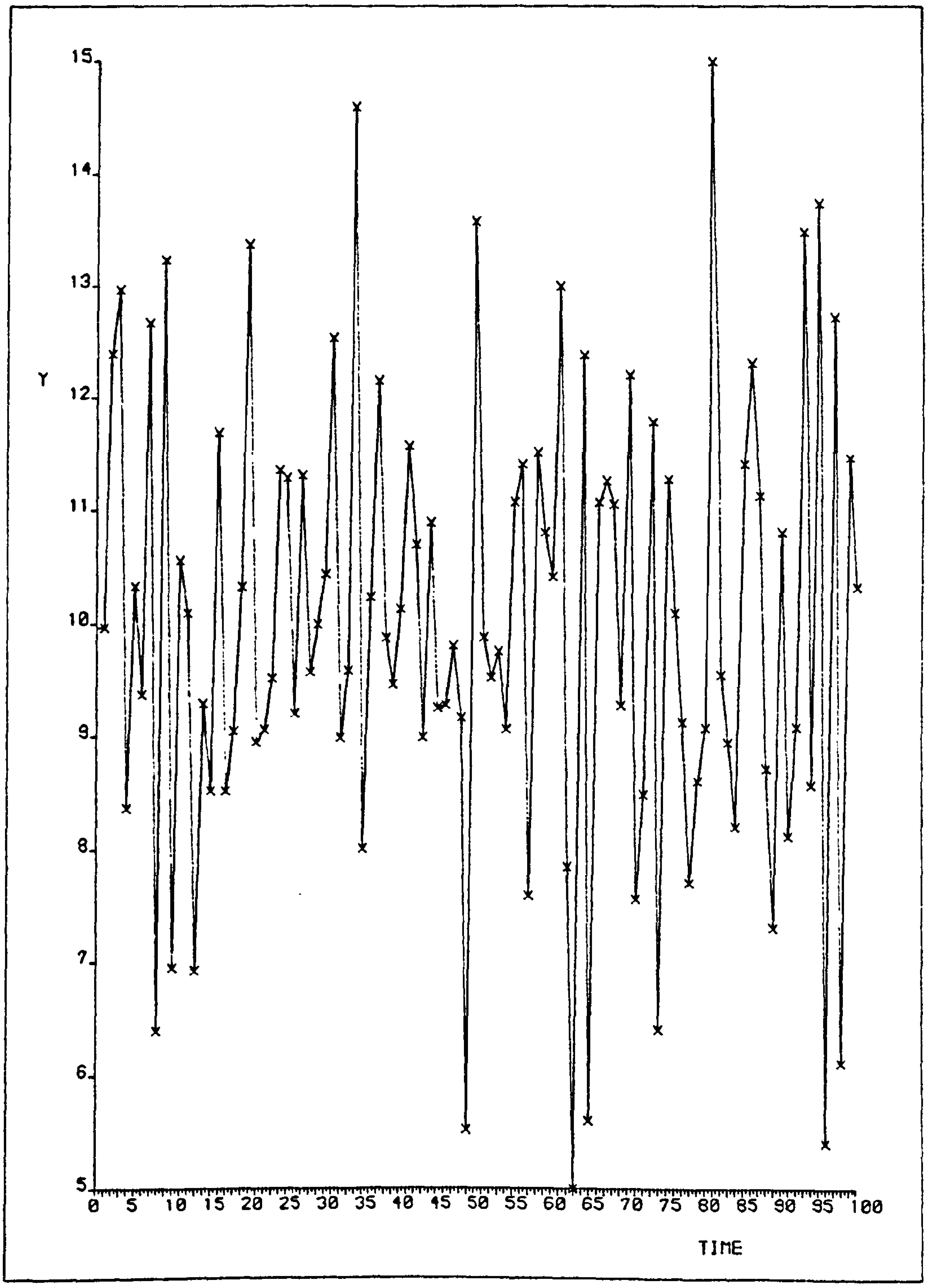
$\phi = -0.8, \eta = -0.8$

FIGURE 3.27



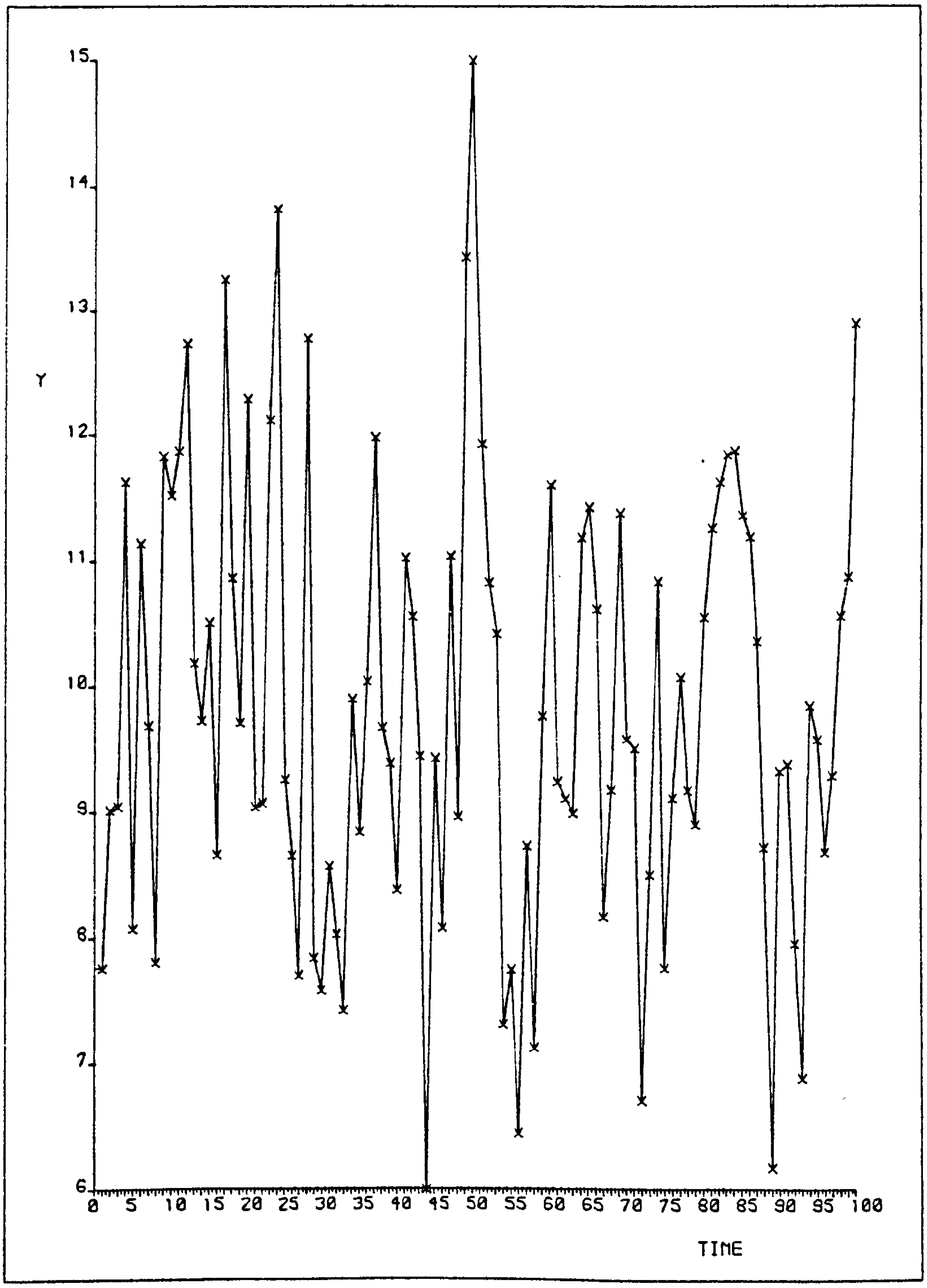
$\phi = 0.8, \eta = -0.8$

FIGURE 3.28



$\phi = 0.2, \eta = 0.6$

FIGURE 3.29



$\phi = 0.6,\ \eta = 0.2$

FIGURE 3.30

3.4 PERFORMANCE AND SENSITIVITY

In this section we shall investigate the 'performance' of the models described in Chapters 2 and 3. To do this, we select three specific models for illustration; namely the linear growth model (of Sections 2.3.1 and 3.3.1), the sinusoidal model (of Sections 2.3.3 and 3.3.3) and the first order autoregressive model (of Sections 2.3.5.1 and 3.3.5.1). In each case performance is assessed on a simulated time series which, in its stable state, exhibits the behaviour associated with the model in question, though several simulated changepoints have been induced in each of the three series (see Appendices 3 for details).

Measures of Performance

Performance is assessed in two areas:

(a) Event detection,

(b) Estimation and forecasting.

In each of these areas we consider two aspects:

(a) Event detection:

(i) magnitude of signal - how certain were we that a changepoint was observed?

(ii) false alarms - how many times did signals occur when no changepoint was present?

We use the one-step-back probabilities of a specific changepoint-type (as defined by Equations (3.31) or (3.48)) as the 'signal' referred to above.

(b) Estimation and Forecasting:

(i) Parameter Estimation - compare estimates of model and nuisance parameters, obtained via the multistate Kalman Filter recursions, with the actual (pre-set) parameter values;

(ii) Observation Estimation, i.e. Forecasts - we use one-step-ahead forecasts, given by Equation (2.14), to assess the local forecasting ability of the models; as well as presenting plots of these forecasts, f_t, superimposed on the corresponding time series, y_t, and plots of the associated one-step-ahead forecast errors, $e_t (= y_t - f_t)$, we also use two quantitative measures:

1. Sum of squares of forecast errors = $SSFE(t) = \sum_{i=1}^{t} e_i^2$ (used by Stoodley and Mirnia 1979), so that a lower SSFE(t) implies better forecasting;

2. Mean absolute deviation = $MAD(t) = \frac{1}{t} \sum_{i=1}^{t} |e_i|$ (used by Ameen and Harrison 1982), so that a lower MAD(t) also implies better forecasting.

Sensitivity

We shall examine how sensitive the models are to small changes in the following parameters:

(a) (n_o, r_o) pairs - keeping the initial variance estimate constant, according to Equation (2.28);

(b) r_o - changing the initial variance estimate;

(c) $R_\varepsilon, \underset{\sim}{R}_\omega, \underset{\sim}{p}_o$ - changing the multistate conditions;

(d) $\underset{\sim}{m}_o$ - small changes to each component in turn;

(e) NN - the number of nodes for the $\underset{\sim}{\phi}$ grid (AR(1) and sinusoidal models only).

Sensitivity will be judged on changes in the performance measures described earlier.

3.4.1 LINEAR GROWTH

3.4.1.1: Initial Setting. For the data set described in Appendix A3.1 the following prior values were employed:

$$\underset{\sim}{m}_o = \begin{pmatrix} 100 \\ 5 \end{pmatrix} ; \quad \underset{\sim}{C}_o = \begin{pmatrix} 10 & 0 \\ 0 & 0.5 \end{pmatrix} ;$$

$$n_o = 5; \quad r_o = 45; \quad (\text{so } E(c^2) = r_o/(n_o - 2) = 15)$$

	j = 1	j = 2	j = 3	j = 4
$p_o^{(j)}$ =	0.85	0.06	0.07	0.02
$R_\varepsilon^{(j)}$ =	1	1	1	30
$R_\mu^{(j)}$ =	0	20	0	0
$R_\beta^{(j)}$ =	0	0	10	0

(a) The event detection techniques summarized in Section 3.2 were applied to the data, and the results are shown in Figure 3.31; the uppermost plot shows the actual data series plotted against time, and the lower three plots show each of the relevant one-step-back probabilities associated with the corresponding observation.

Notice that:

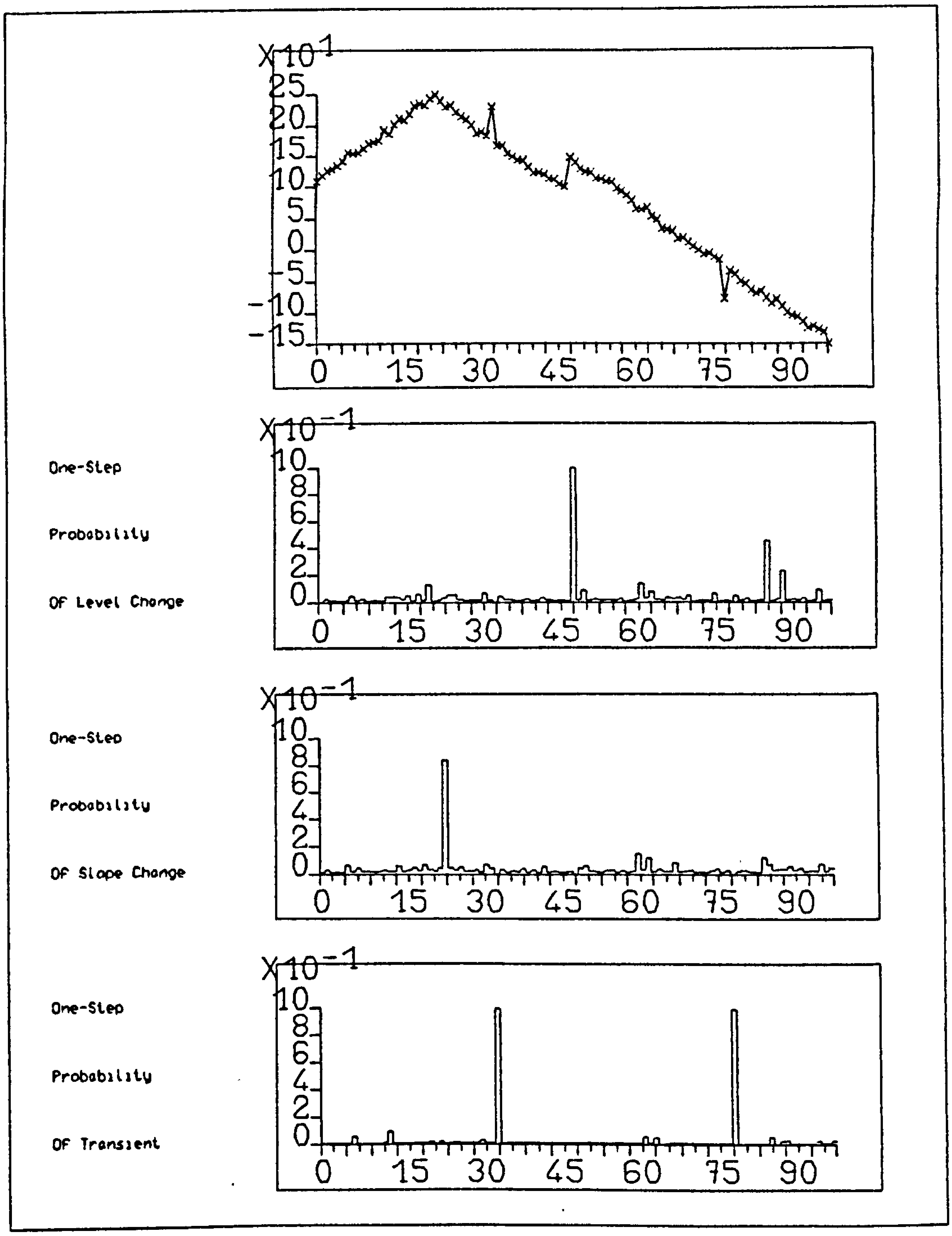

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.31

$$0_{26}^{(3)} = 0.799$$
$$0_{36}^{(4)} = 1.000$$
$$0_{51}^{(2)} = 1.000$$

and $$0_{81}^{(4)} = 1.000$$ (all to three decimal places).

If we use $0_t^{(i)} > 0.2$ $(i = 2,3,4)$, say, as a criterion for a positive signal, then the number of false alarms (false positives) = NFP = 2.

(b) (i) The final estimate of $\underset{\sim}{\theta}$ is $\underset{\sim}{m}_{100} = \begin{pmatrix} -116.9 \\ -7.8 \end{pmatrix}$ (compared to the theoretical values of $\begin{pmatrix} -117.5 \\ -5.0 \end{pmatrix}$).

(ii) The one-step-ahead forecasts (asterisks) are shown along with the raw data in the uppermost plot of Figure 3.32; the lower plot shows the progression of one-step-ahead forecast errors.

In this case:

$$SSFE(100) = SSFE = 13878$$

and $$MAD(100) = MAD = 7.85.$$

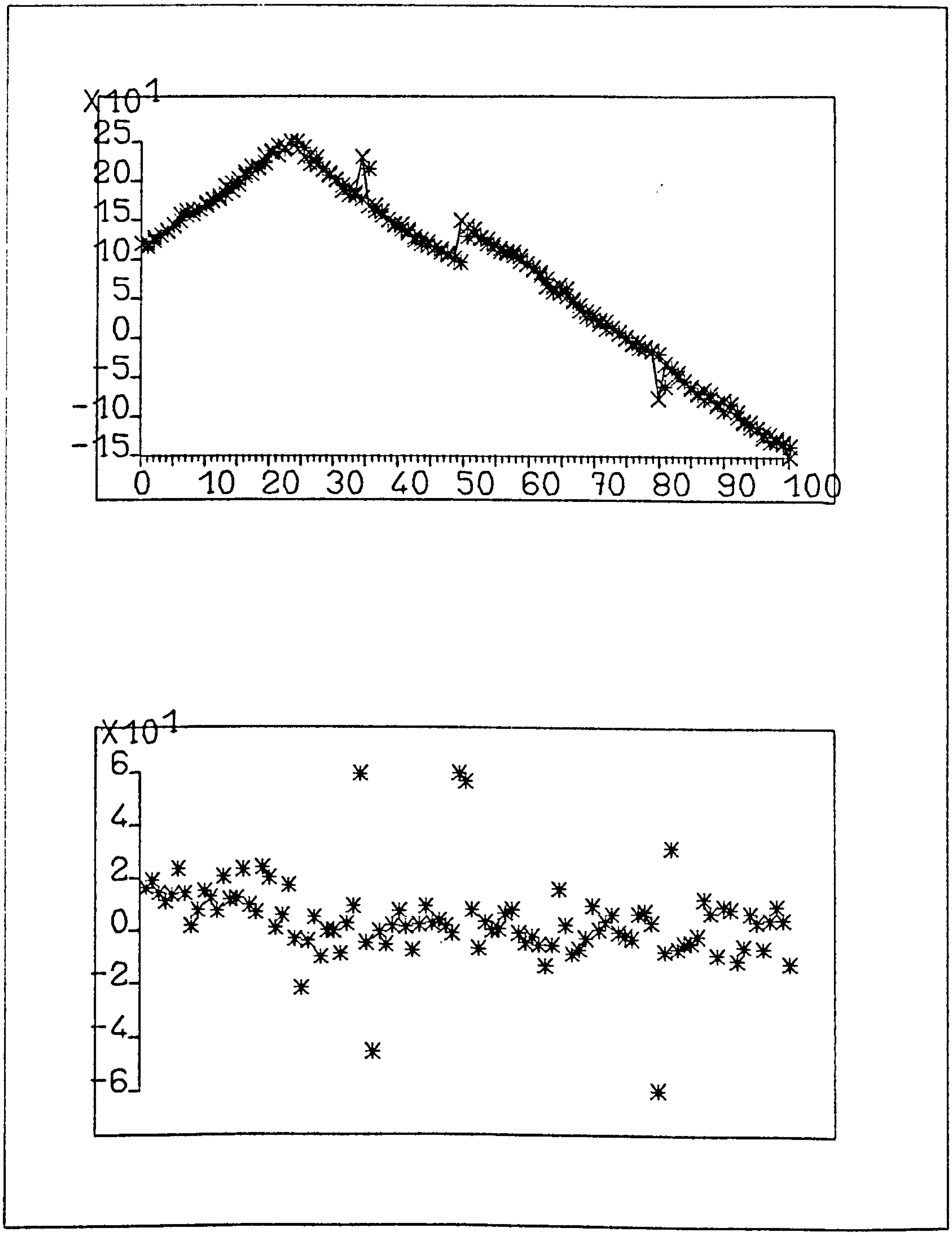
3.4.1.2: Sensitivity Analysis. For each case the remaining parameter settings, initially, are unchanged from those given in the previous section.

(a) (n_o, r_o)

(i) $n_o = 25$; $r_o = 345$

(ii) $n_o = 50$; $r_o = 720$.

The multistate Kalman Filter results along with one-step-ahead forecasts (asterisks) are shown in Figure 3.33 and Figure 3.34 respectively, for (i) and (ii). In this case:



X10^1
25
20
15
10
5
0
-5
-10
-15
0 10 20 30 40 50 60 70 80 90 100
X10^1
6
4
2
0
-2
-4
-6

FIGURE 3.32

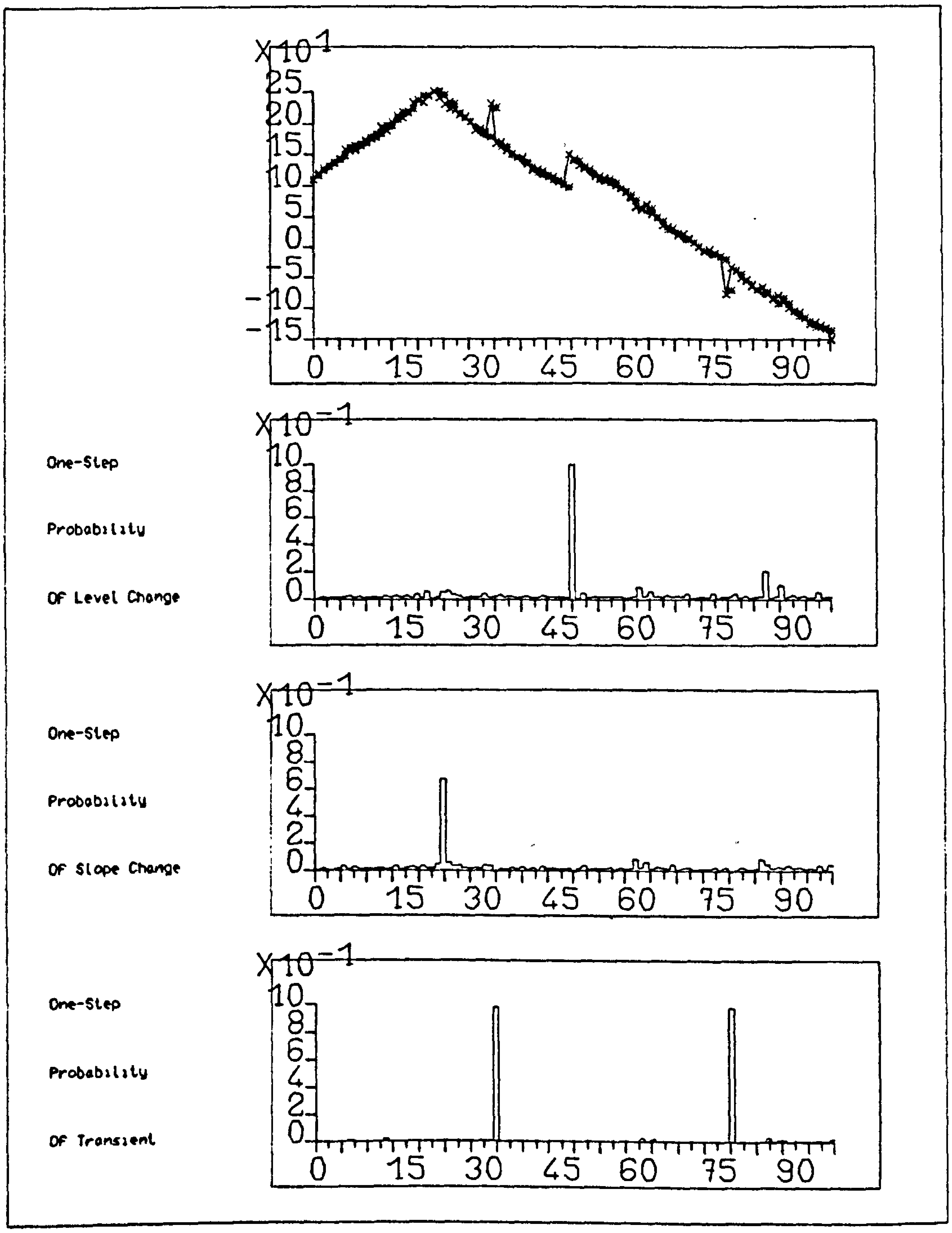

One-Step
Probability
Of Level Change
One-Step
Probability
Of Slope Change
One-Step
Probability
Of Transient

FIGURE 3.33

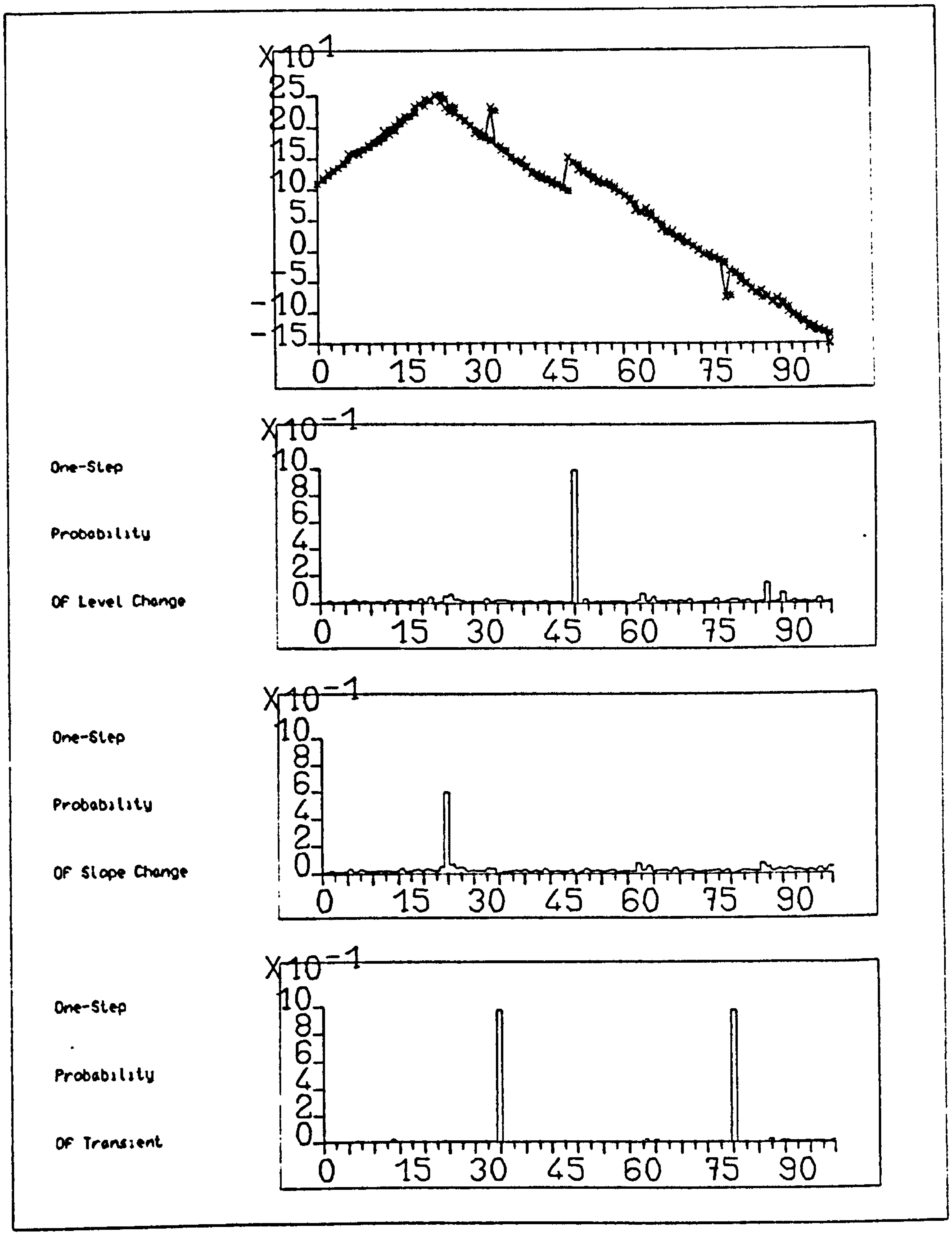

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.34

	(i)	(ii)
$0_{26}^{(3)}$	0.674	0.597
$0_{36}^{(4)}$	0.980	0.976
$0_{51}^{(2)}$	0.996	0.995
$0_{81}^{(4)}$	0.976	0.973
NFP	1	0
$\underset{\sim}{m}_{100}$	$\begin{pmatrix}-116.5\\-7.9\end{pmatrix}$	$\begin{pmatrix}-116.0\\-7.5\end{pmatrix}$
SSFE	15672	16017
MAD	8.11	8.18

(b) r_o

(i) $r_o = 15$ (ii) $r_o = 135$ (see Figures 3.35 and 3.36)

	(i)	(ii)
$0_{26}^{(3)}$	0.955	0.641
$0_{36}^{(4)}$	1.000	0.979
$0_{51}^{(2)}$	1.000	0.996
$0_{81}^{(4)}$	0.999	0.981
NFP	18	1
$\underset{\sim}{m}_{100}$	$\begin{pmatrix}-117.7\\-8.9\end{pmatrix}$	$\begin{pmatrix}-116.6\\-8.0\end{pmatrix}$
SSFE	13982	15690
MAD	8.04	8.11

(c) Multistate Conditions

1. $R_\varepsilon, \underset{\sim}{R}_\omega$

(i) $R_\mu^{(2)} = 60$, $R_\beta^{(3)} = 30$, $R_\varepsilon^{(4)} = 90$

(ii) $R_\mu^{(2)} = 180$, $R_\beta^{(3)} = 90$, $R_\varepsilon^{(4)} = 270$ (see Figures 3.37 and 3.38)

	(i)	(ii)
$0_{26}^{(3)}$	0.838	0.858

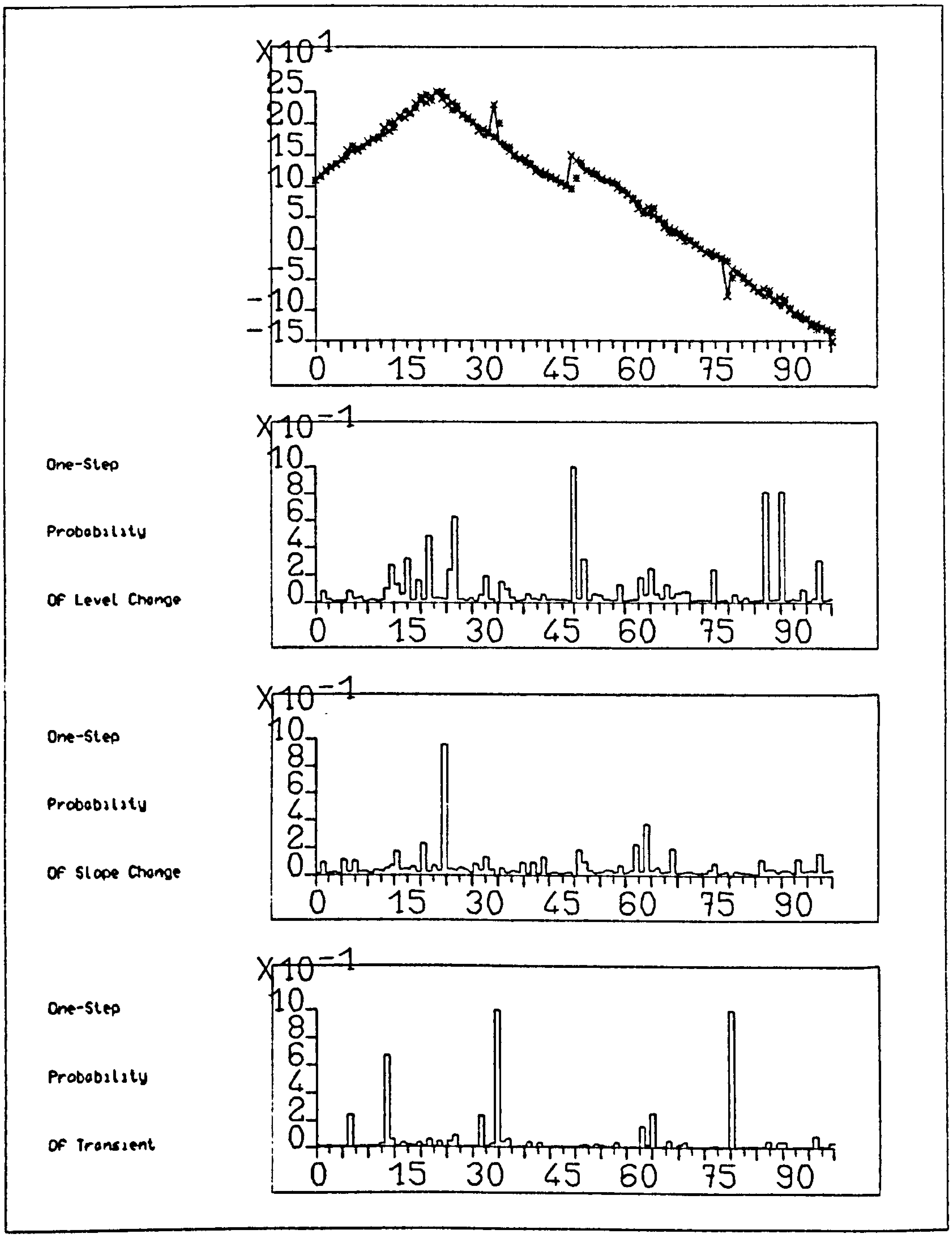
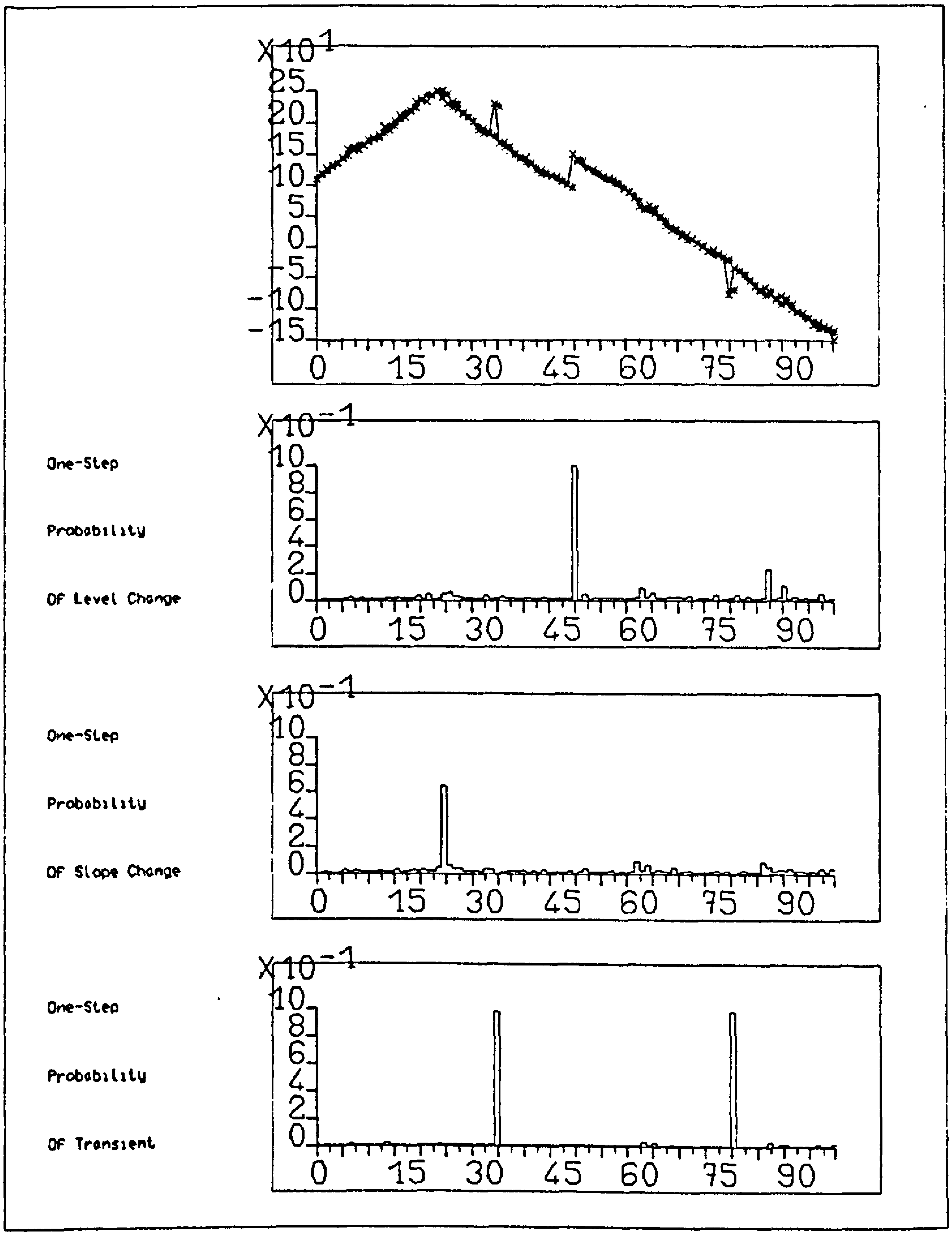

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.35



X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.36

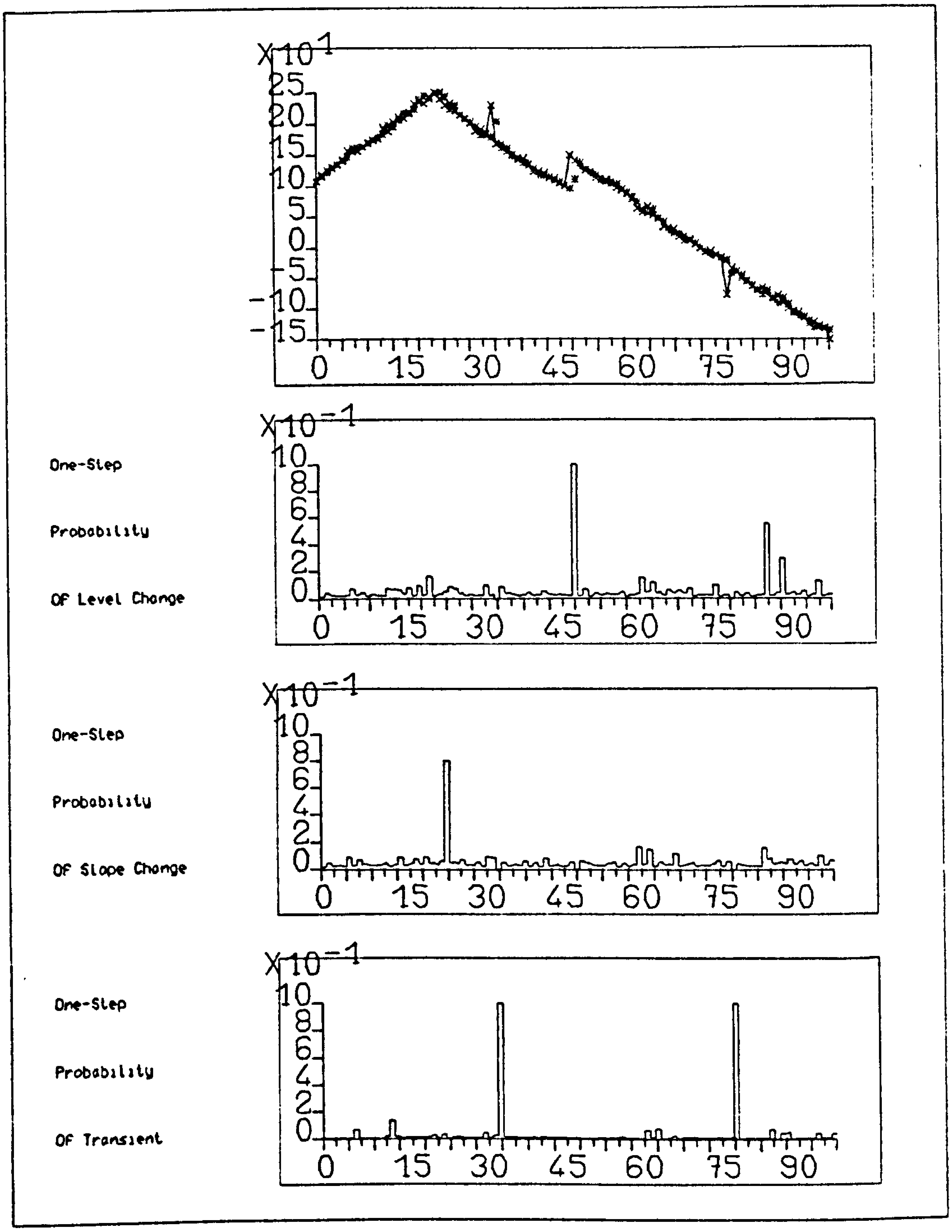
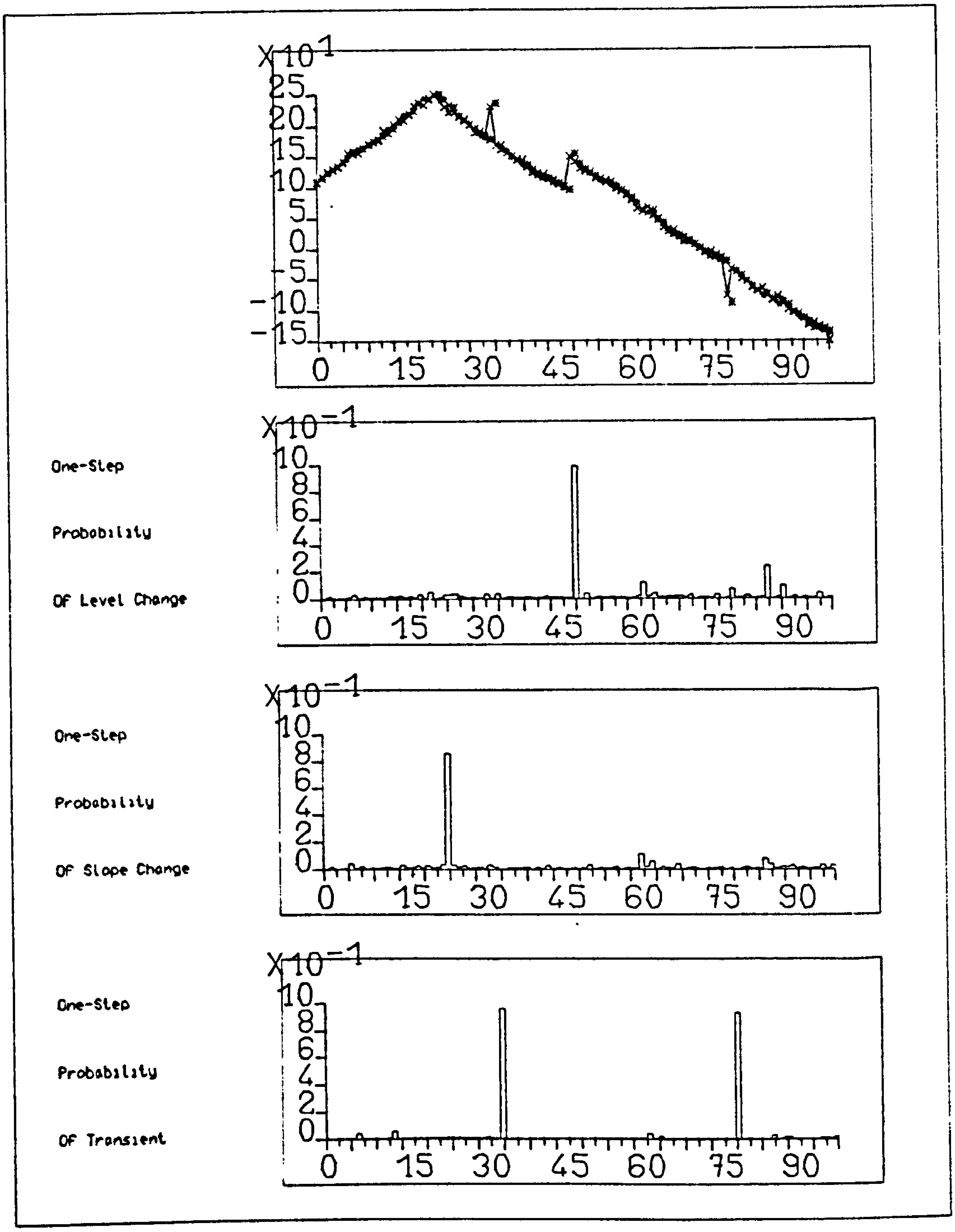

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.37



X 10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
X 10^-1
One-Step
Probability
Of Level Change
10
8
6
4
2
0
0 15 30 45 60 75 90
X 10^-1
One-Step
Probability
Of Slope Change
10
8
6
4
2
0
0 15 30 45 60 75 90
X 10^-1
One-Step
Probability
Of Transient
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.38

$0_{36}^{(4)}$	0.994	0.959
$0_{51}^{(2)}$	0.998	0.991
$0_{81}^{(4)}$	0.991	0.926
NFP	2	1
$\underset{\sim}{m}_{100}$	$\begin{pmatrix}-117.3\\-8.7\end{pmatrix}$	$\begin{pmatrix}-116.7\\-8.1\end{pmatrix}$
SSFE	14724	18200
MAD	7.95	8.46

2. $\underset{\sim}{p}_o$: $p_o^{(1)} = 0.97$, $p_o^{(2)} = 0.01$, $p_o^{(3)} = 0.01$, $p_o^{(4)} = 0.01$

(see Figure 3.39)

$0_{26}^{(3)} = 0.905$

$0_{36}^{(4)} = 0.999$

$0_{51}^{(2)} = 0.998$

$0_{81}^{(4)} = 0.999$

NFP = 0

$\underset{\sim}{m}_{100} = \begin{pmatrix}-113.9\\-5.6\end{pmatrix}$

SSFE = 13609

MAD = 7.64

(d) $\underset{\sim}{m}_o$

1. m_o

(i) $m_o = 33.33$ (ii) $m_o = 300$ (see Figures 3.40 and 3.41)

	(i)	(ii)
$0_{26}^{(3)}$	0.767	0.865
$0_{36}^{(4)}$	0.990	0.994
$0_{51}^{(2)}$	0.998	0.998
$0_{81}^{(4)}$	0.986	0.991
NFP	1	3
$\underset{\sim}{m}_{100}$	$\begin{pmatrix}-117.1\\-8.5\end{pmatrix}$	$\begin{pmatrix}-117.4\\-8.7\end{pmatrix}$
SSFE	20666	53679
MAD	8.70	10.15

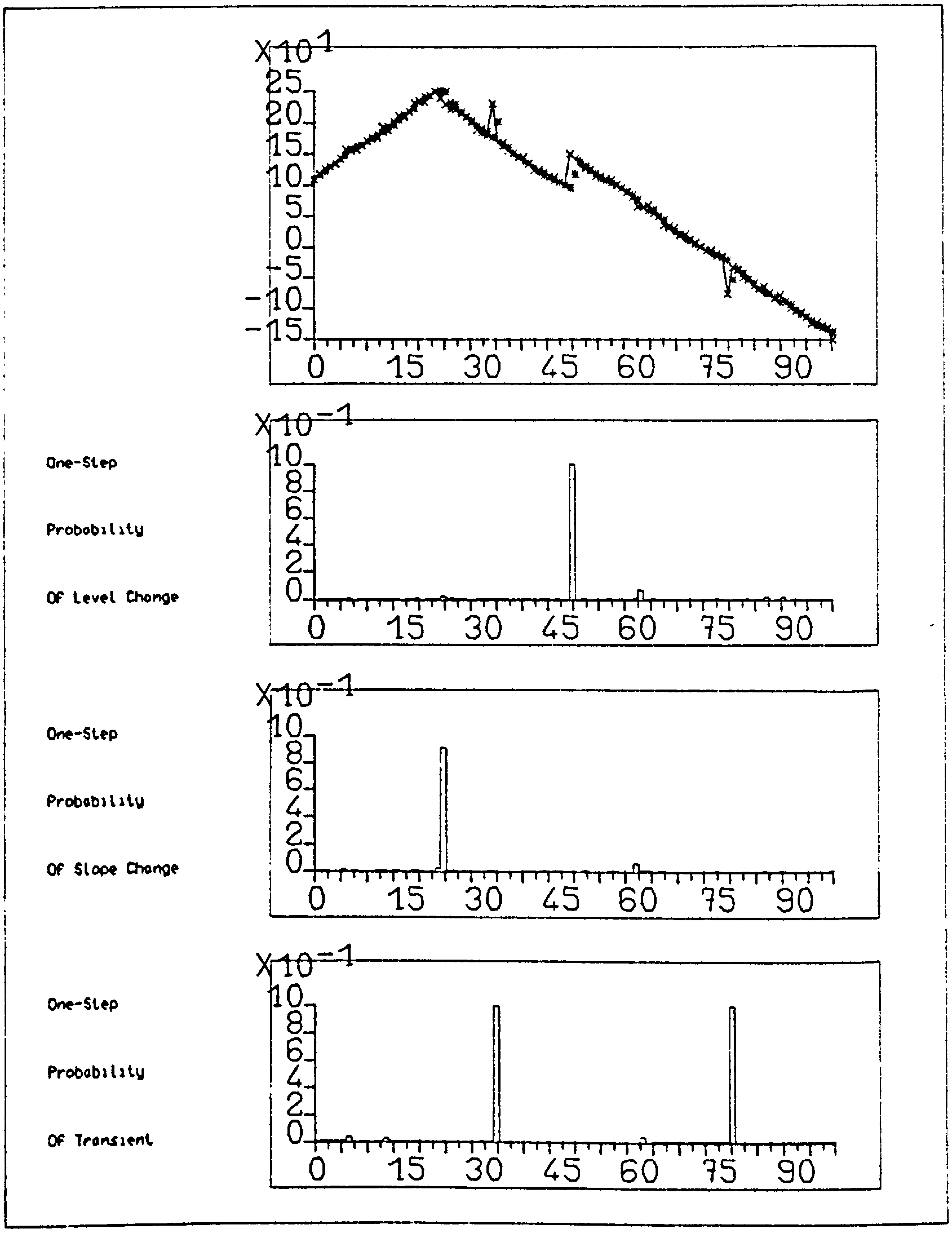

X10^1
One-Step
Probability
Of Level Change
One-Step
Probability
Of Slope Change
One-Step
Probability
Of Transient
X10^-1

FIGURE 3.39

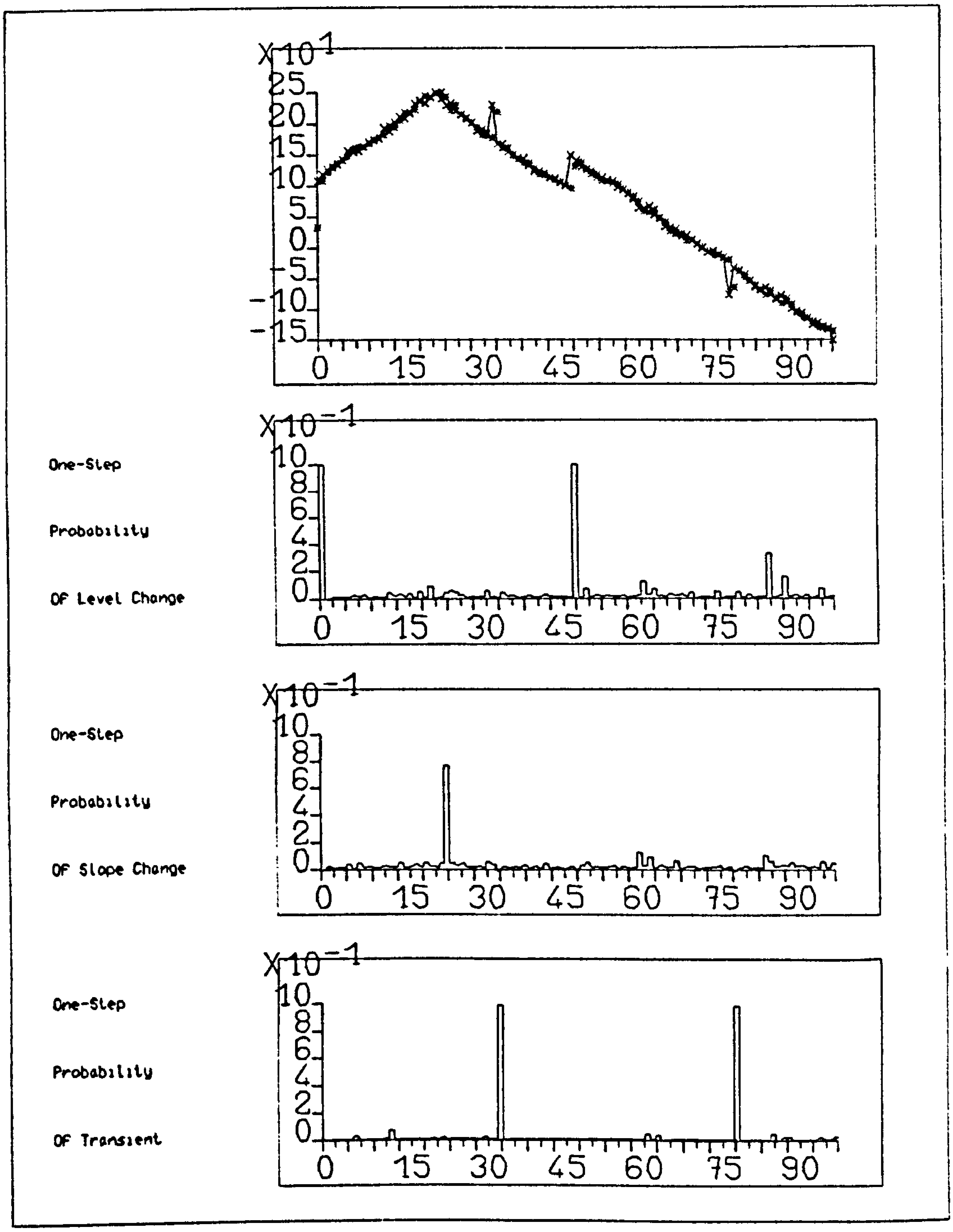

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.40

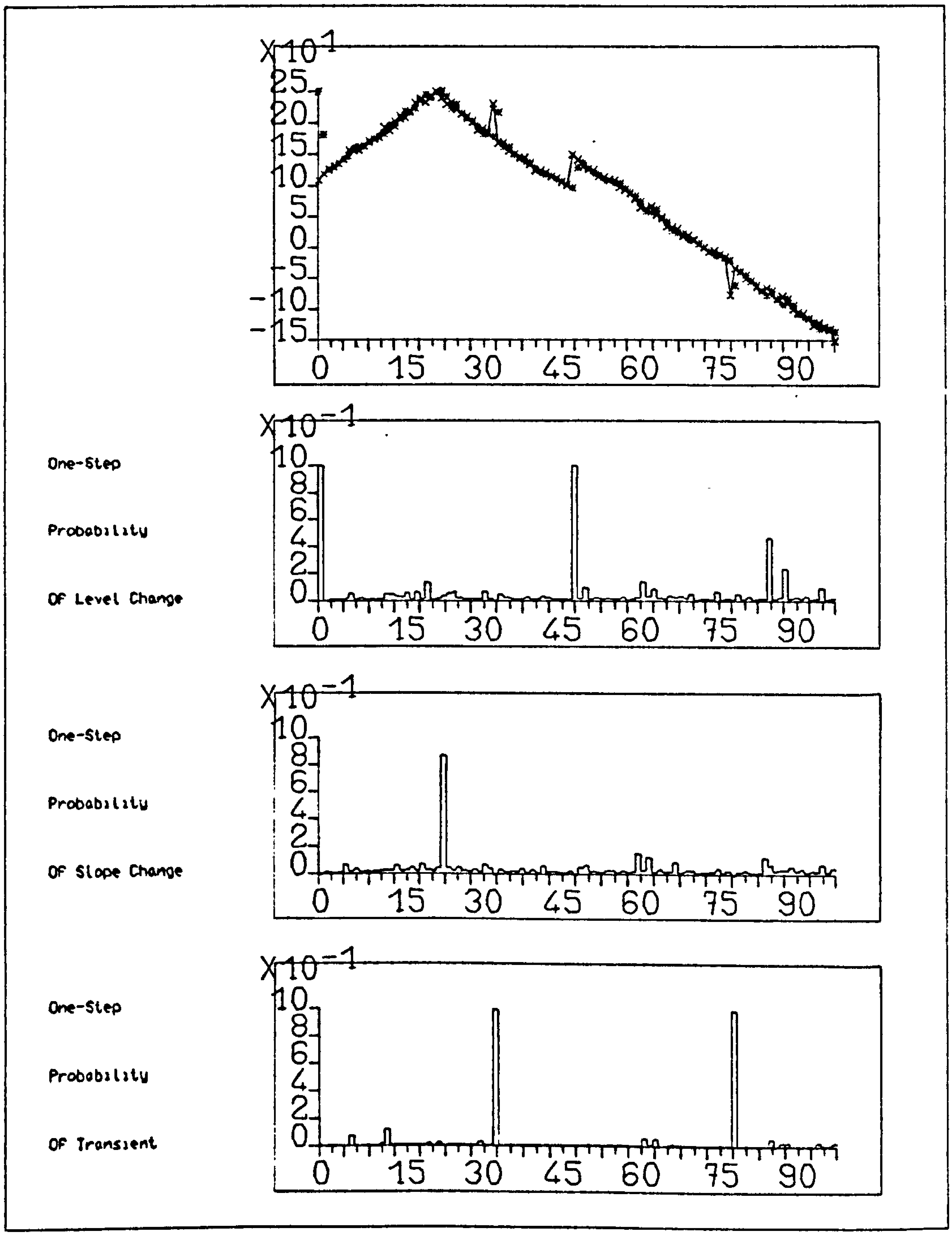

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.41

2. b_o

(i) $b_o = 1.667$ (ii) $b_o = 15$ (iii) $b_o = -5$ (see Figures 3.42, 3.43 and 3.44)

	(i)	(ii)	(iii)
$O_{26}^{(3)}$	0.782	0.771	0.786
$O_{36}^{(4)}$	0.988	0.988	0.989
$O_{51}^{(2)}$	0.997	0.997	0.997
$O_{81}^{(4)}$	0.986	0.986	0.986
NFP	1	2	3
$\underset{\sim}{m}_{100}$	$\begin{pmatrix}-117.0\\-8.4\end{pmatrix}$	$\begin{pmatrix}-116.9\\-8.3\end{pmatrix}$	$\begin{pmatrix}-117.0\\-8.4\end{pmatrix}$
SSFE	15441	15054	15923
MAD	8.15	7.86	8.27

3.4.2 SINUSOIDAL MODEL

3.4.2.1: Initial Setting. For the data set described in Appendix A3.2 the following prior values were employed:

$$\underset{\sim}{m}_o = \begin{pmatrix}100\\30\end{pmatrix}; \quad \underset{\sim}{C}_o = \begin{pmatrix}10 & 0\\0 & 3\end{pmatrix}$$

$$n_o = 5; \quad r_o = 45 \quad (\text{therefore } E(c^2) = 15)$$

with $p_o^{(j)}$, $R_\varepsilon^{(j)}$, $\underset{\sim}{R}_\omega^{(j)}$, j = 1, ..., 4, identical to that given in Section 3.4.1.1.

For the ϕ grid, a range of $(0^0,360^o]$ was used with NN = number of nodes = 36.

Using these values the multistate Kalman Filter results are shown in Figure 3.45, and the corresponding forecasts and forecast errors in Figure 3.46.

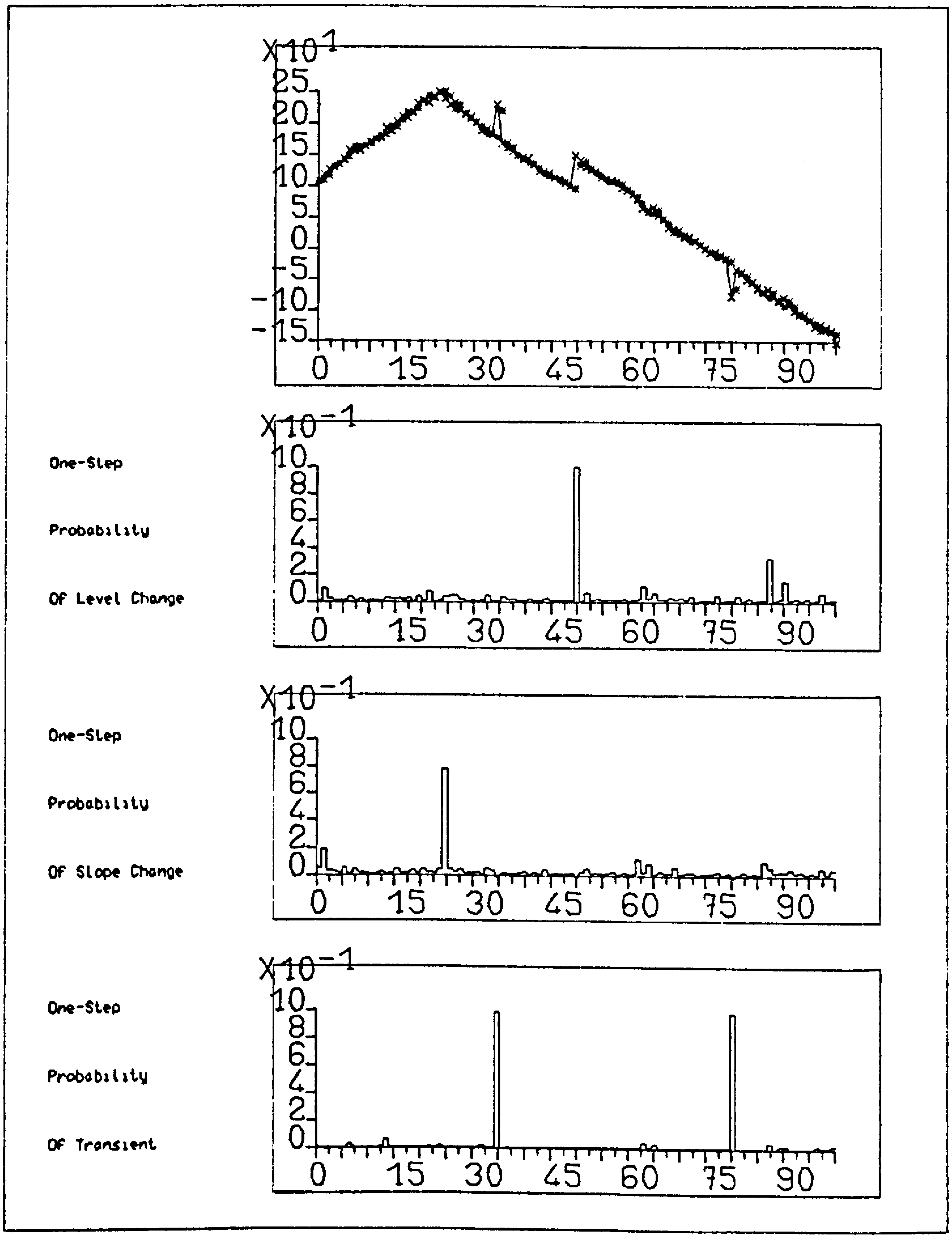

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.42

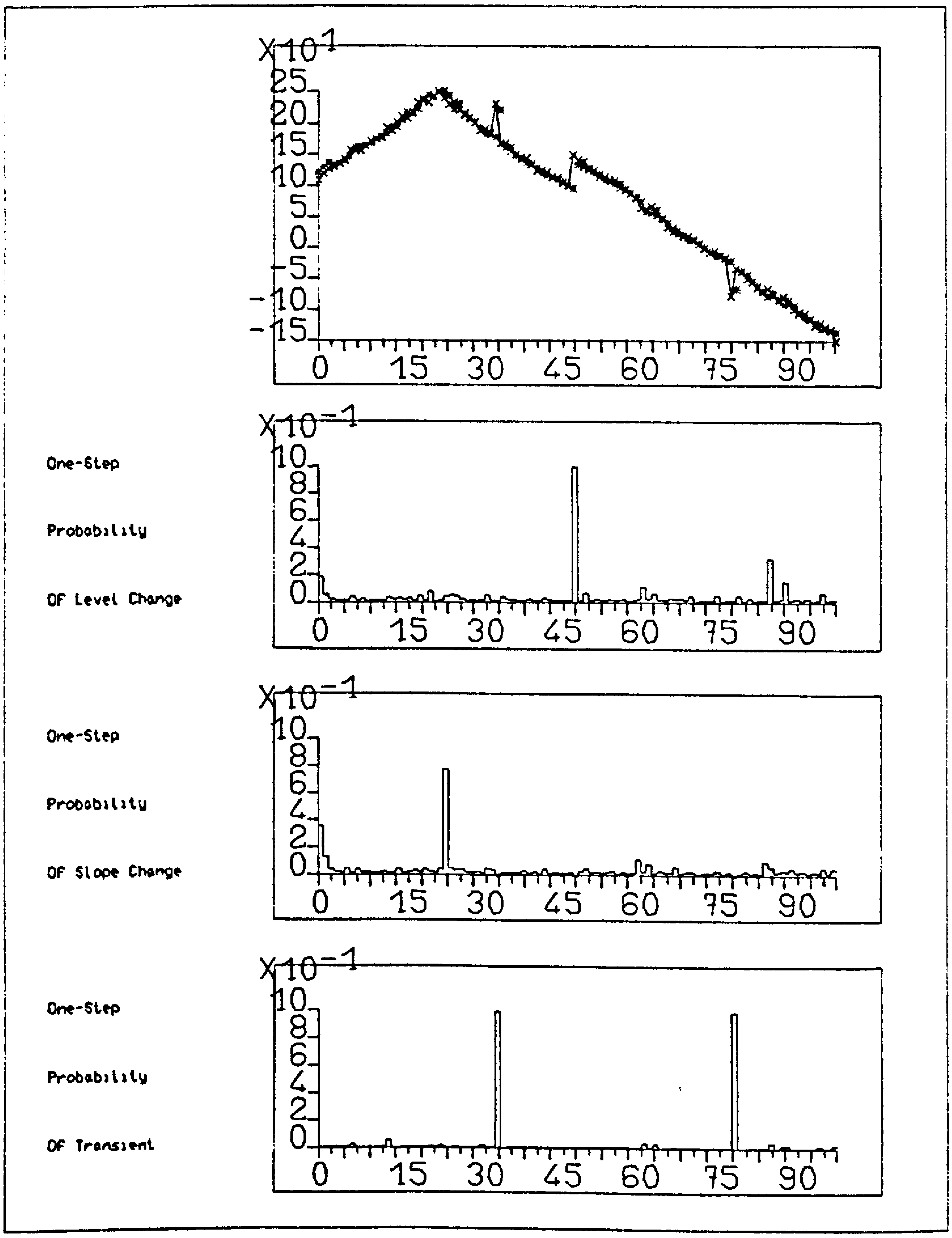

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.43

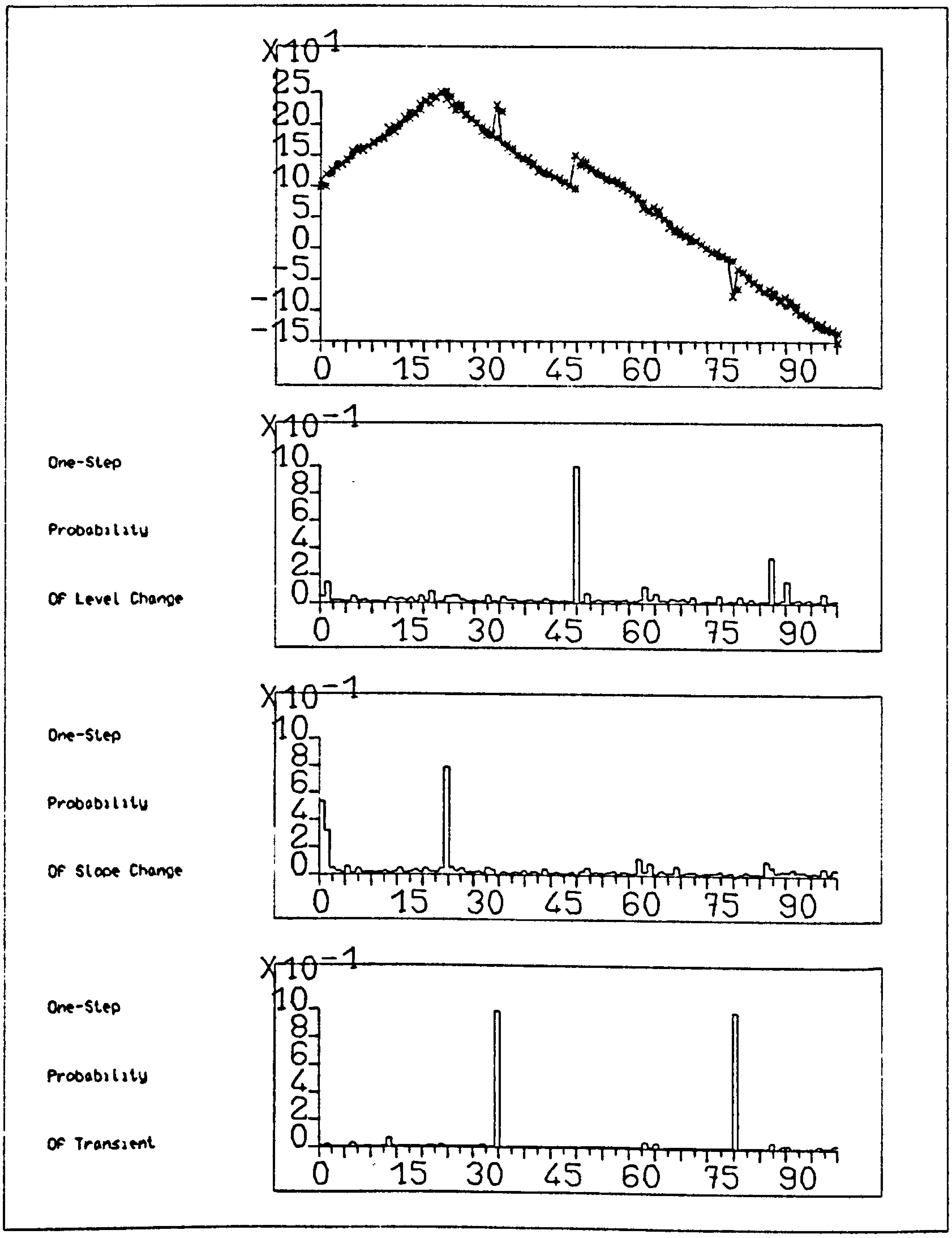

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.44

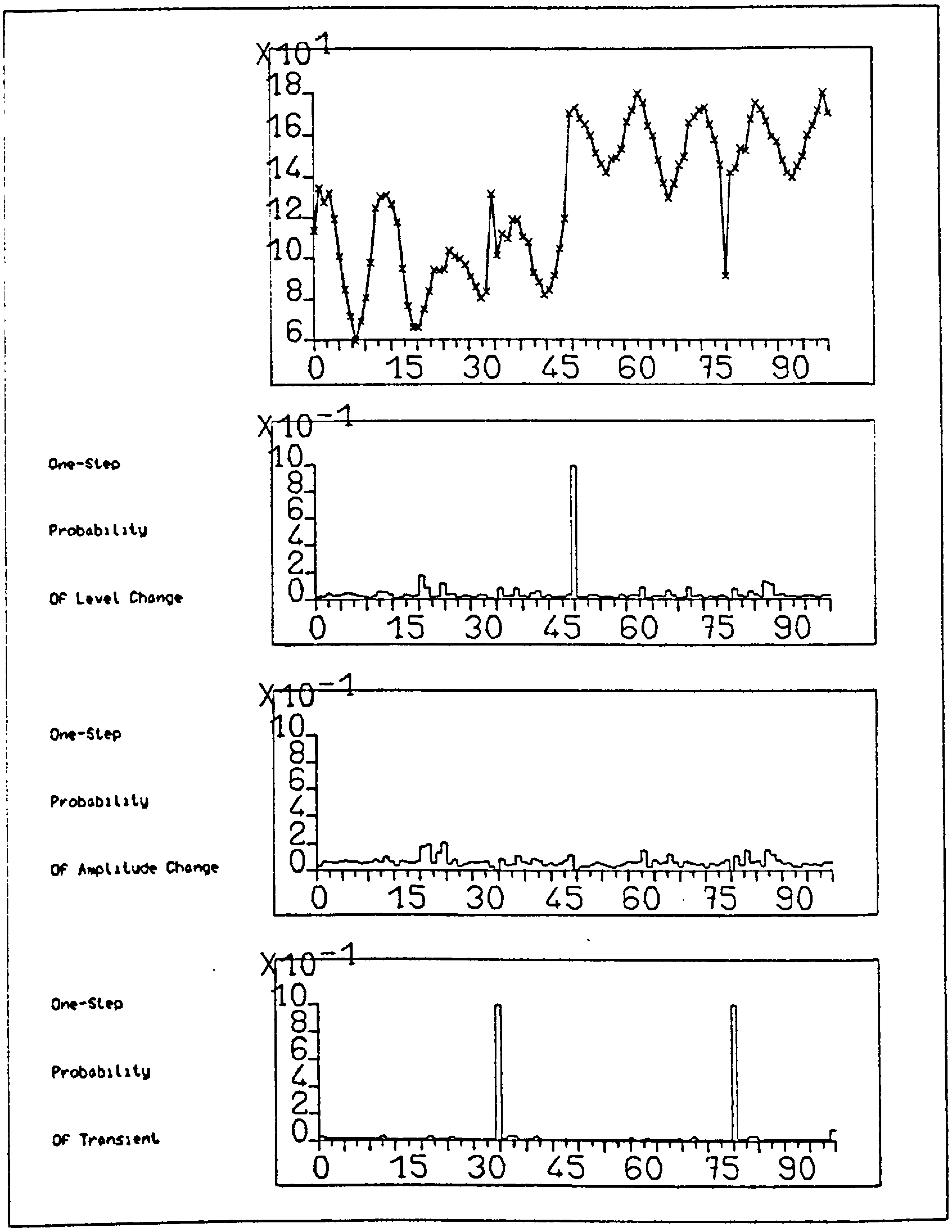

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
X10^-1
10
8
6
4
2
0
One-Step
Probability
Of Level Change
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Amplitude Change
10
8
6
4
2
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Transient
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.45

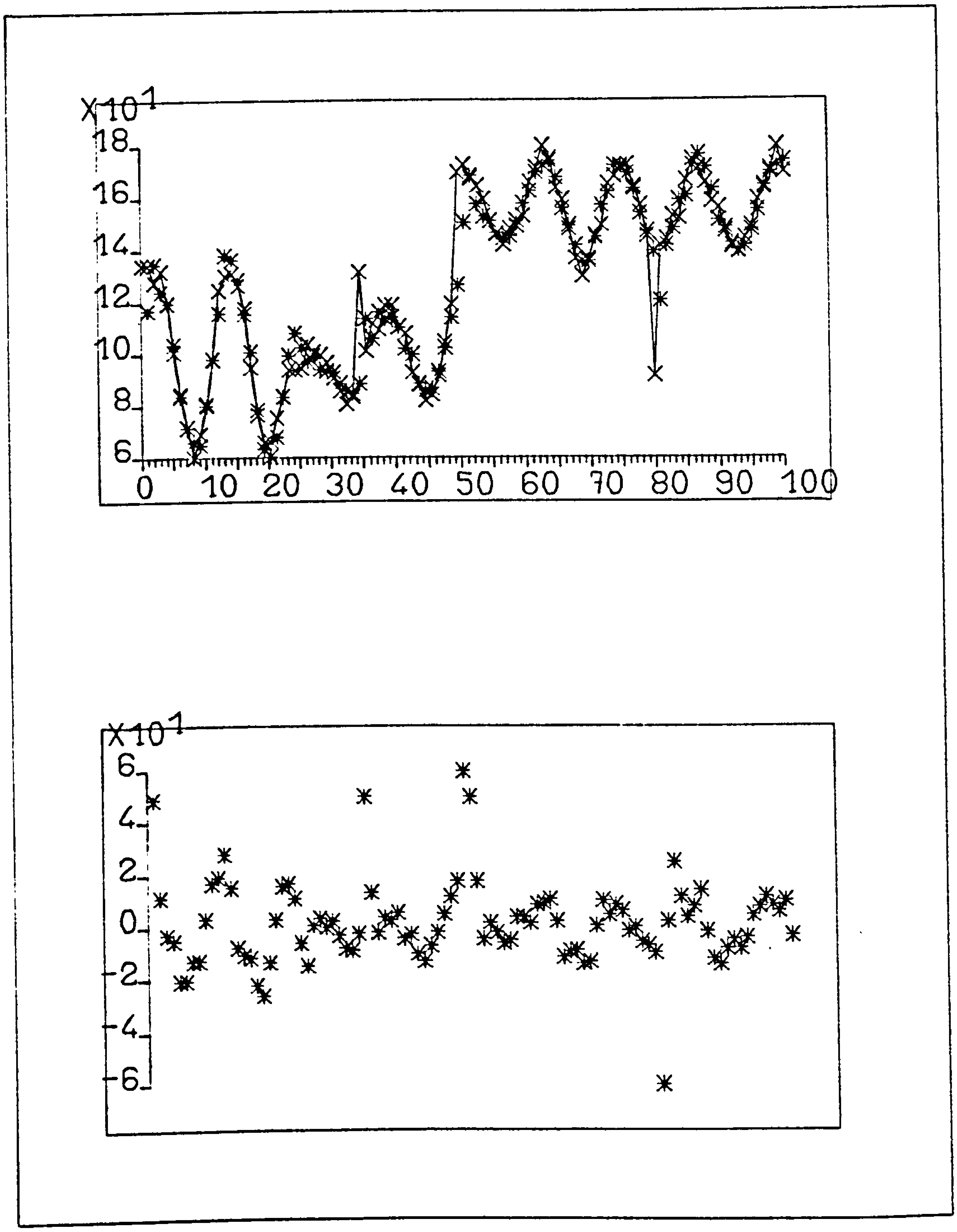


FIGURE 3.46

We have: $0_{26}^{(2)} = 0.213$

$0_{36}^{(4)} = 0.997$

$0_{51}^{(3)} = 0.992$

$0_{81}^{(4)} = 1.000$

$NFP = 0$

$\underset{\sim}{m}_{100} = \begin{pmatrix} 150.0 \\ 15.4 \end{pmatrix}$ $\left(\underset{\sim}{\theta} = \begin{pmatrix} 150.0 \\ 15.0 \end{pmatrix}\right.$, theoretically)

$\hat{\phi}_{100} = 90.0$ ($\phi = 90.0^{\circ}$, theoretically)

SSFE = 17297

MAD = 9.1

3.4.2.2: Sensitivity Analysis.

Multistate Kalman Filter results for (i) to (xii) in Table 3.1 are displayed in Figures 3.47 to 3.58 respectively. Figures 3.59 and 3.60 show the progression of the ϕ grid through the analysis for NN = 36 and for NN = 12, respectively. In these three-dimensional plots the x-axis denotes the ϕ range $(0^{\circ},360^{\circ}]$, the y-axis denotes time (0,100] and the z-axis denotes probability (0.0,1.0).

NOTE: $\omega = 1/12$ has been assumed fixed and known.

3.4.3 AR(1)

3.4.3.1: Initial Setting. For the data set described in Appendix A3.3 the following prior values were employed:

$$\underset{\sim}{m}_o = \begin{pmatrix} 10 \\ 10 \end{pmatrix}; \quad \underset{\sim}{C}_o = \begin{pmatrix} 15 & 0 \\ 0 & 15 \end{pmatrix}$$

$$n_o = 5; \quad r_o = 3.0 \quad (\text{therefore } E(c^2) = 1.0)$$

with $p_o^{(j)}$, $R_{\varepsilon}^{(j)}$, $\underset{\sim}{R}_{\omega}^{(j)}$ $j = 1, \ldots, 4$ identical to the values given in Section 3.4.1.1, and with NN = 21 for the ϕ-grid.

TABLE 3.1

	$0_{26}^{(2)}$	$0_{36}^{(4)}$	$0_{51}^{(3)}$	$0_{81}^{(4)}$	NFP	$\underset{\sim}{m}_{100}$	$\hat{\phi}_{100}$	SSFE	MAD
INITIAL SETTING	0.213	0.997	0.992	1.000	0	$\begin{pmatrix}150.0\\15.4\end{pmatrix}$	90.0	17297	9.1
CHANGES									
(i) $n_o = 25,\ r_o = 345$	0.141	0.989	0.969	1.000	0	$\begin{pmatrix}150.0\\15.5\end{pmatrix}$	90.0	17566	9.2
(ii) $n_o = 50,\ r_o = 720$	0.119	0.984	0.948	0.999	0	$\begin{pmatrix}150.0\\15.5\end{pmatrix}$	90.1	17756	9.3
(iii) $r_o = 15$	0.265	0.999	0.997	1.000	3	$\begin{pmatrix}150.0\\15.3\end{pmatrix}$	90.0	17300	9.1
(iv) $r_o = 135$	0.085	0.985	0.962	1.000	0	$\begin{pmatrix}150.0\\15.4\end{pmatrix}$	90.1	17986	9.3
(v) $R_\mu^{(2)} = 60, R_\alpha^{(3)} = 30, R_\varepsilon^{(4)} = 90$	0.272	0.970	0.950	0.997	0	$\begin{pmatrix}149.9\\15.4\end{pmatrix}$	90.0	17764	9.3
(vi) $R_\mu^{(2)} = 180, R_\alpha^{(3)} = 90, R_\varepsilon^{(4)} = 270$	0.336	0.805	0.848	0.965	1	$\begin{pmatrix}149.9\\15.4\end{pmatrix}$	90.0	18462	9.4
(vii) $p_o^{(1)} = 0.97, p_o^{(i)} = 0.01$ $i = 2,3,4$	0.171	0.998	0.983	1.000	0	$\begin{pmatrix}150.0\\15.2\end{pmatrix}$	90.0	18185	9.5
(viii) $m_o = 33.3$	0.070	1.000	0.997	1.000	9	$\begin{pmatrix}149.9\\15.0\end{pmatrix}$	90.0	42418	12.8
(ix) $m_o = 300.0$	0.095	0.999	0.998	1.000	6	$\begin{pmatrix}149.9\\15.0\end{pmatrix}$	90.0	53063	12.3
(x) $a_o = 10$	0.070	1.000	0.997	1.000	8	$\begin{pmatrix}149.9\\15.1\end{pmatrix}$	90.0	28081	11.7
(xi) $a_o = 90$	0.078	0.942	0.987	1.000	20	$\begin{pmatrix}152.8\\11.8\end{pmatrix}$	108.6	24758	11.2
(xii) NN = 12	0.197	0.996	0.987	1.000	0	$\begin{pmatrix}150.0\\15.4\end{pmatrix}$	90.0	17081	9.0

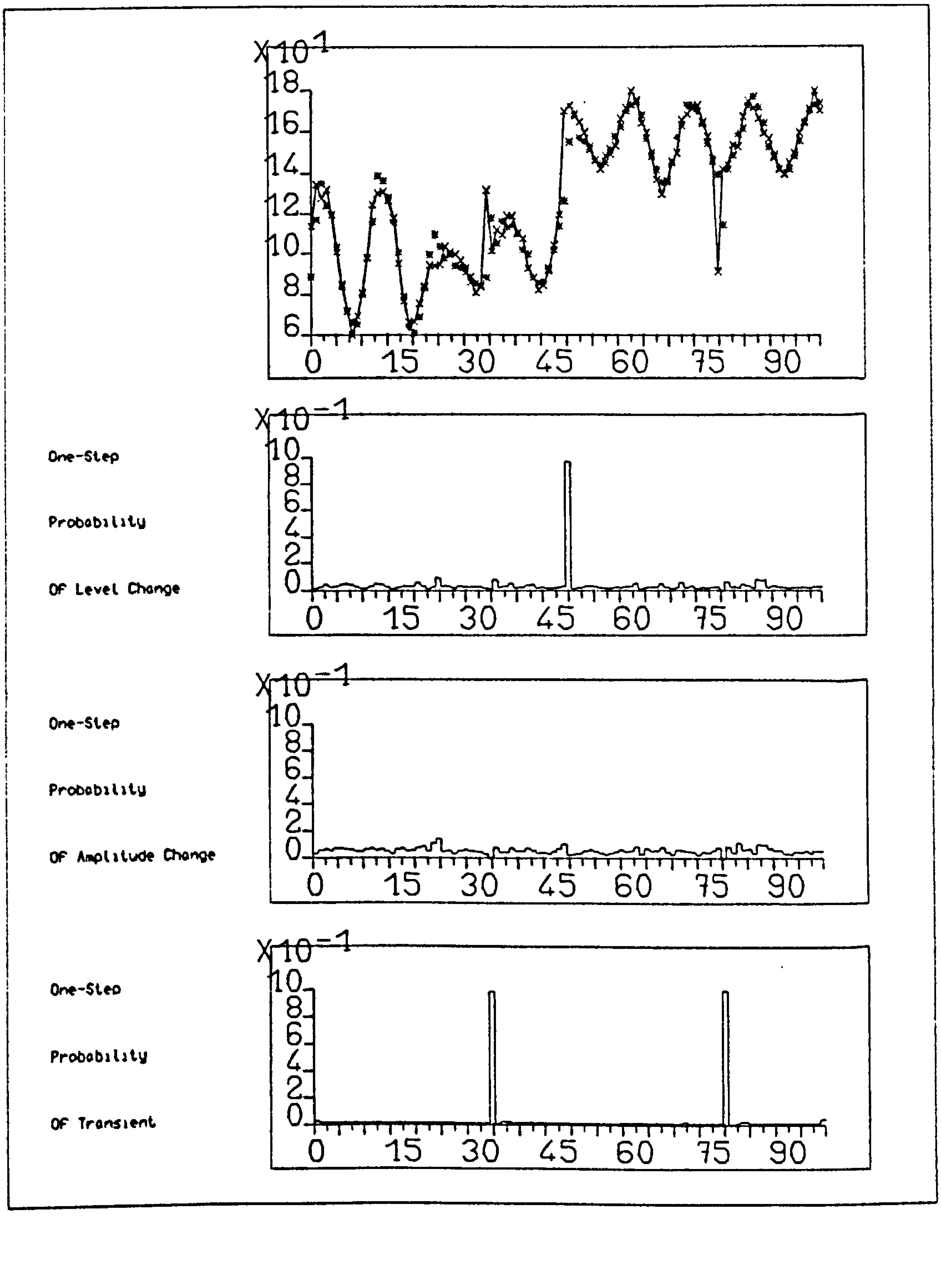

X10^1
18
16
14
12
10
8
6
0
15
30
45
60
75
90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0
15
30
45
60
75
90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0
15
30
45
60
75
90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0
15
30
45
60
75
90

FIGURE 3.47

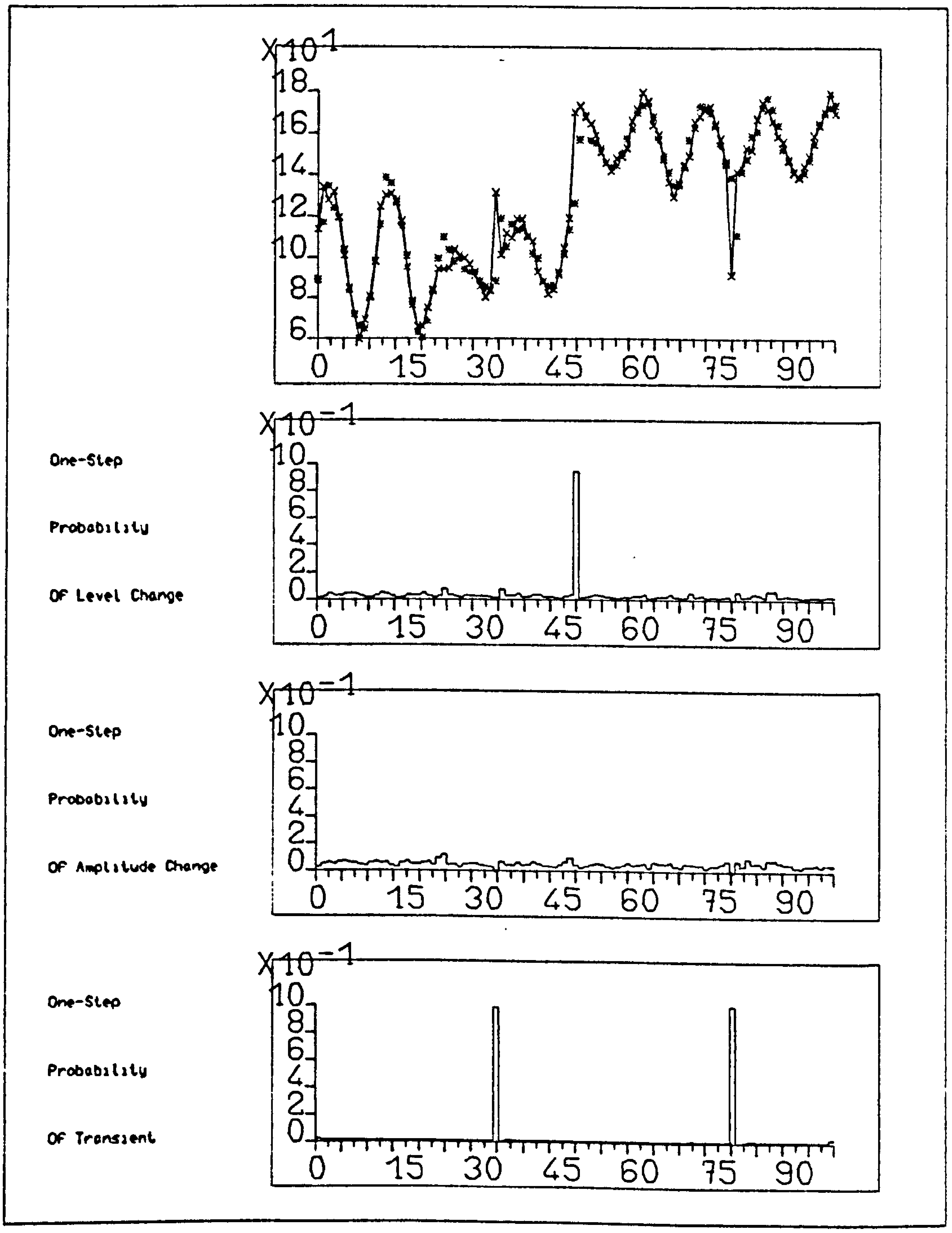

X10^1
18
16
14
12
10
8
6
0
15
30
45
60
75
90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0
15
30
45
60
75
90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0
15
30
45
60
75
90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0
15
30
45
60
75
90

FIGURE 3.48

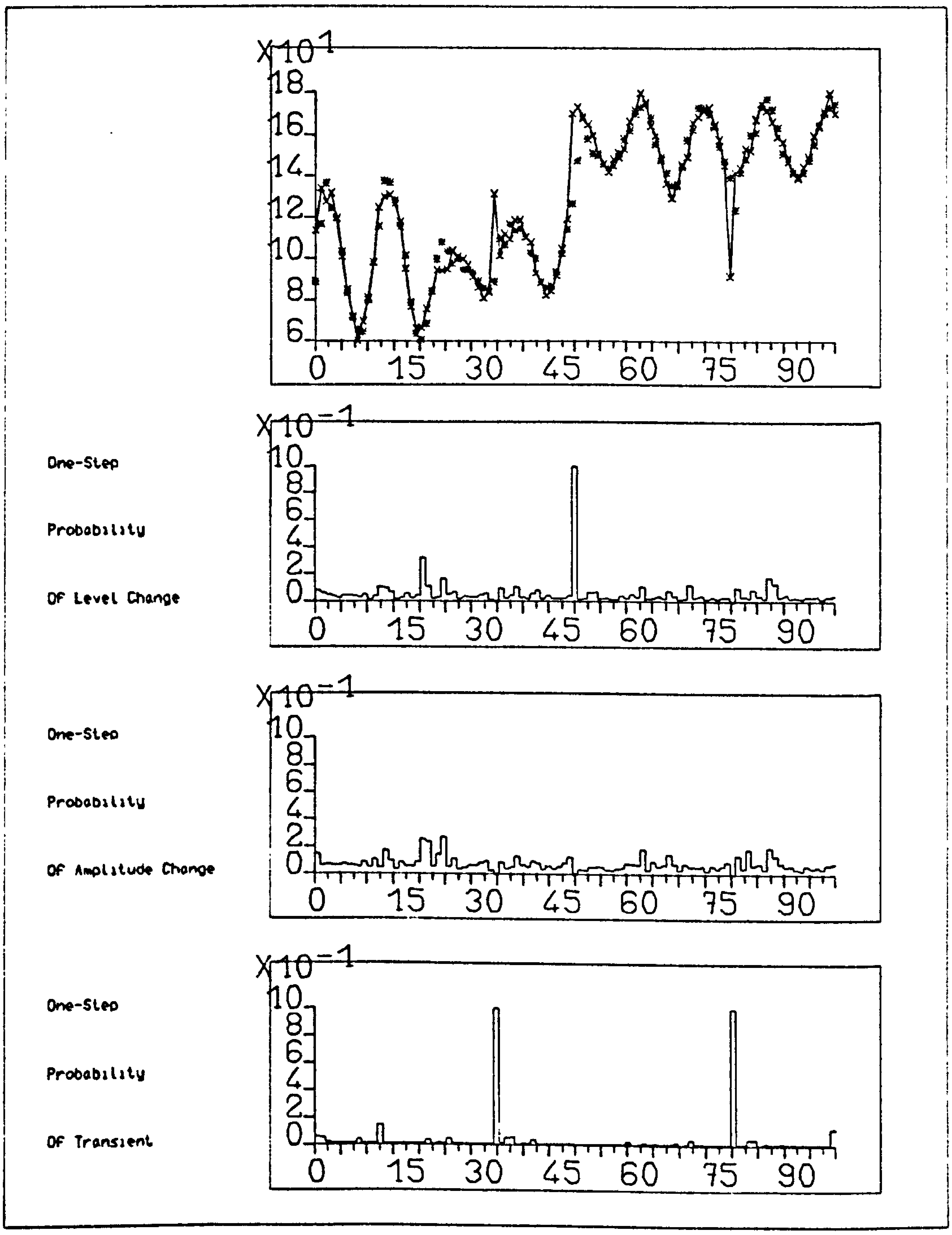

One-Step
Probability
Of Level Change
One-Step
Probability
Of Amplitude Change
One-Step
Probability
Of Transient

FIGURE 3.49

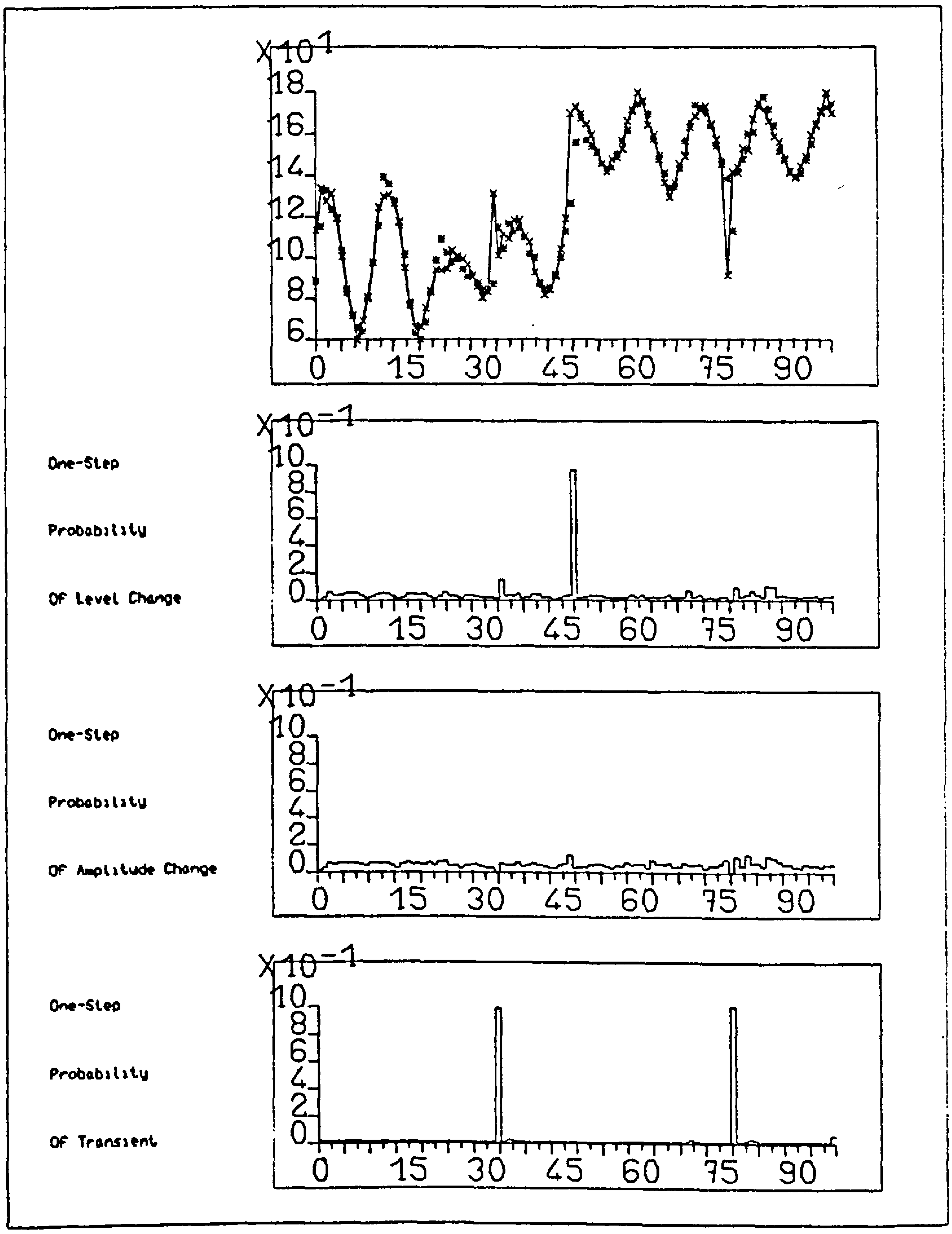
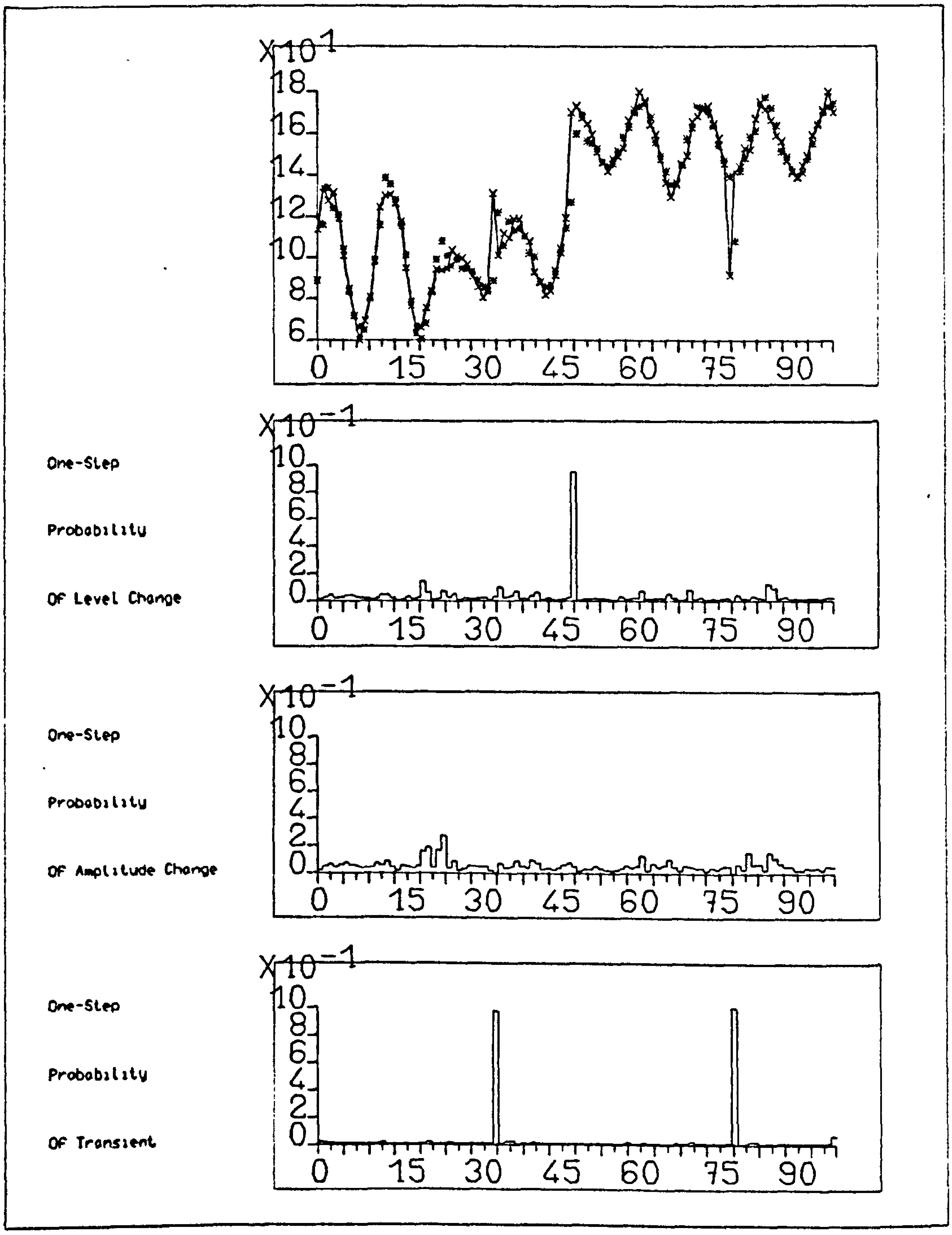

X10^1
18
16
14
12
10
8
6
0
15
30
45
60
75
90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
One-Step
Probability
Of Amplitude Change
One-Step
Probability
Of Transient

FIGURE 3.50



X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.51

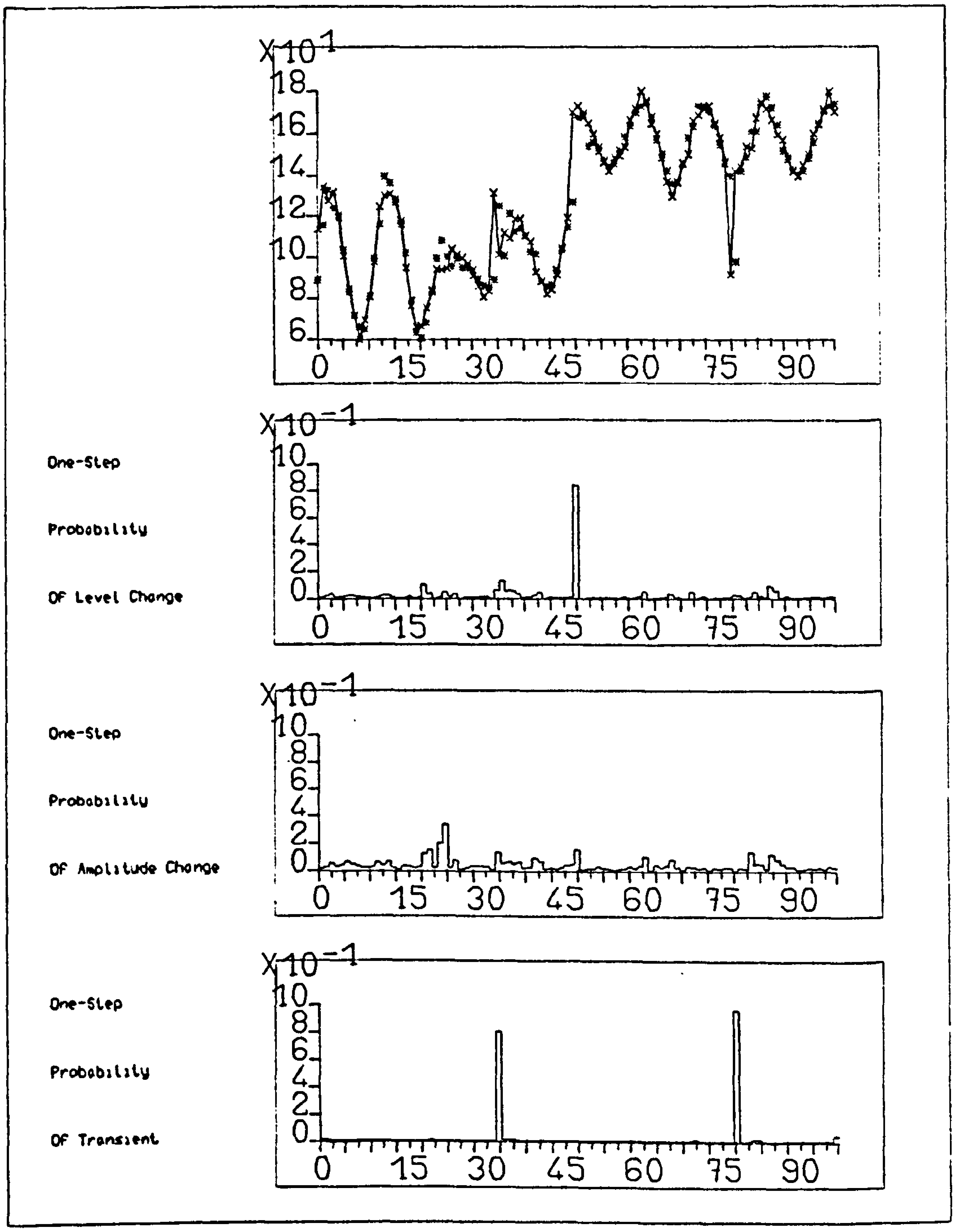

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Level Change
10
8
6
4
2
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Amplitude Change
10
8
6
4
2
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Transient
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.52

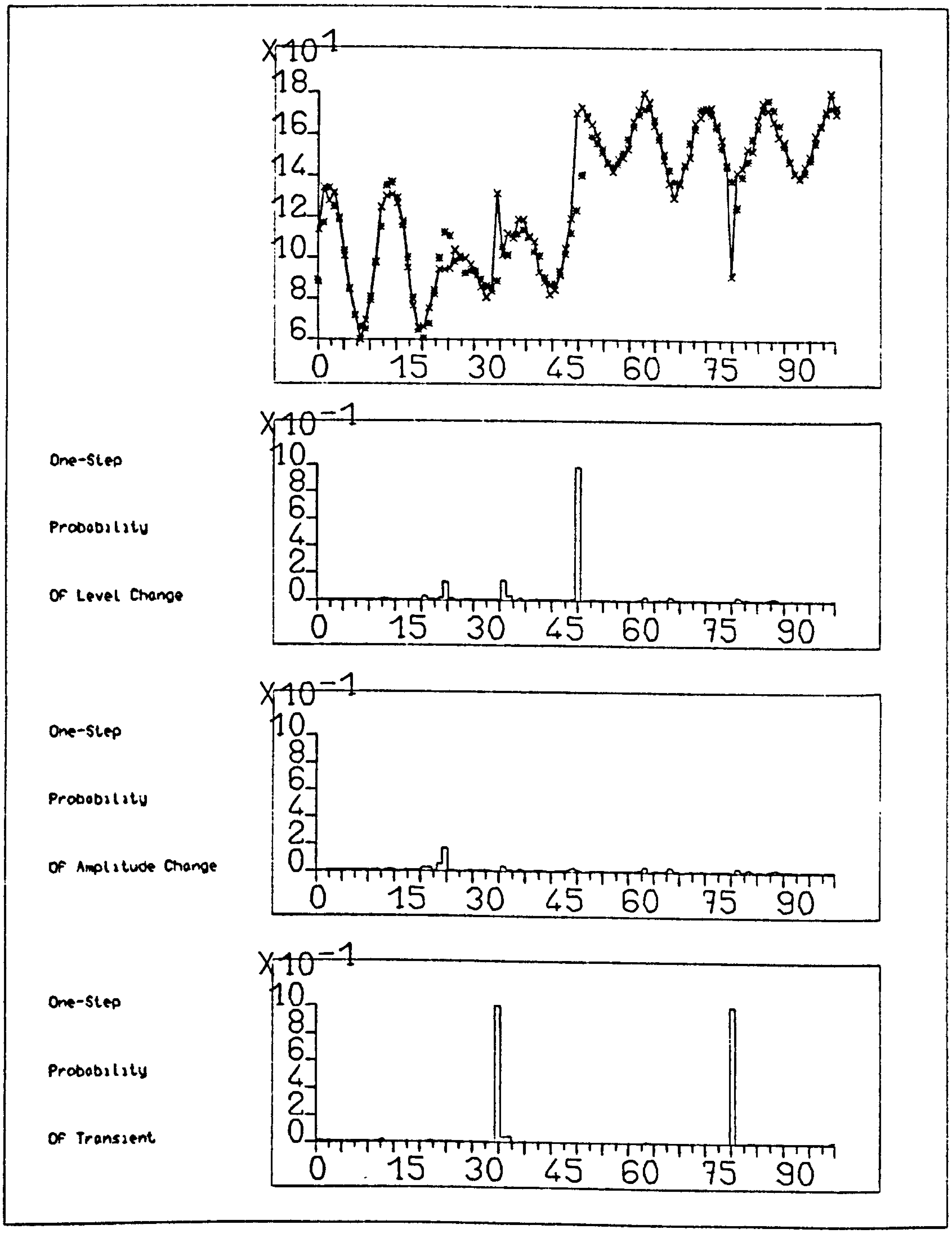
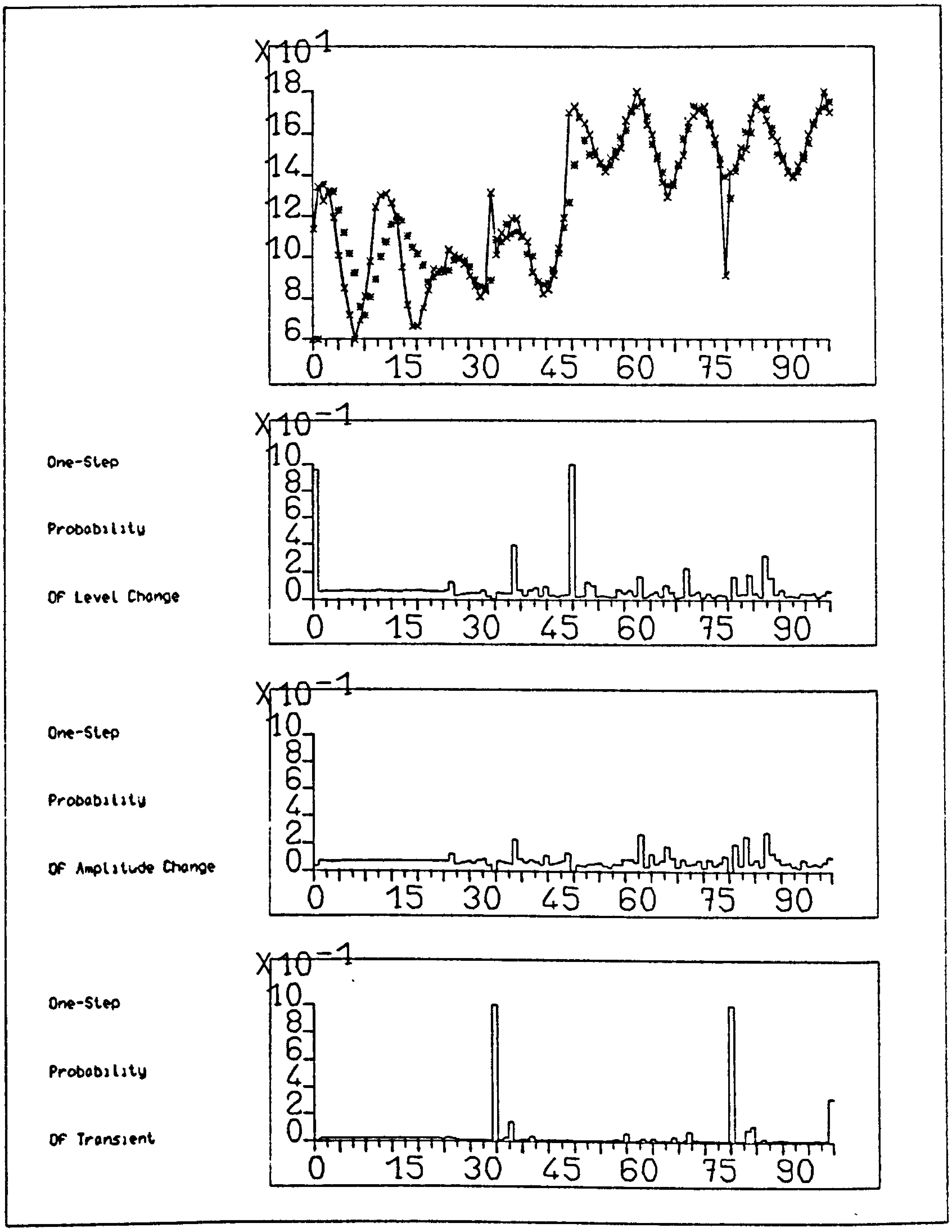

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.53



X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.54

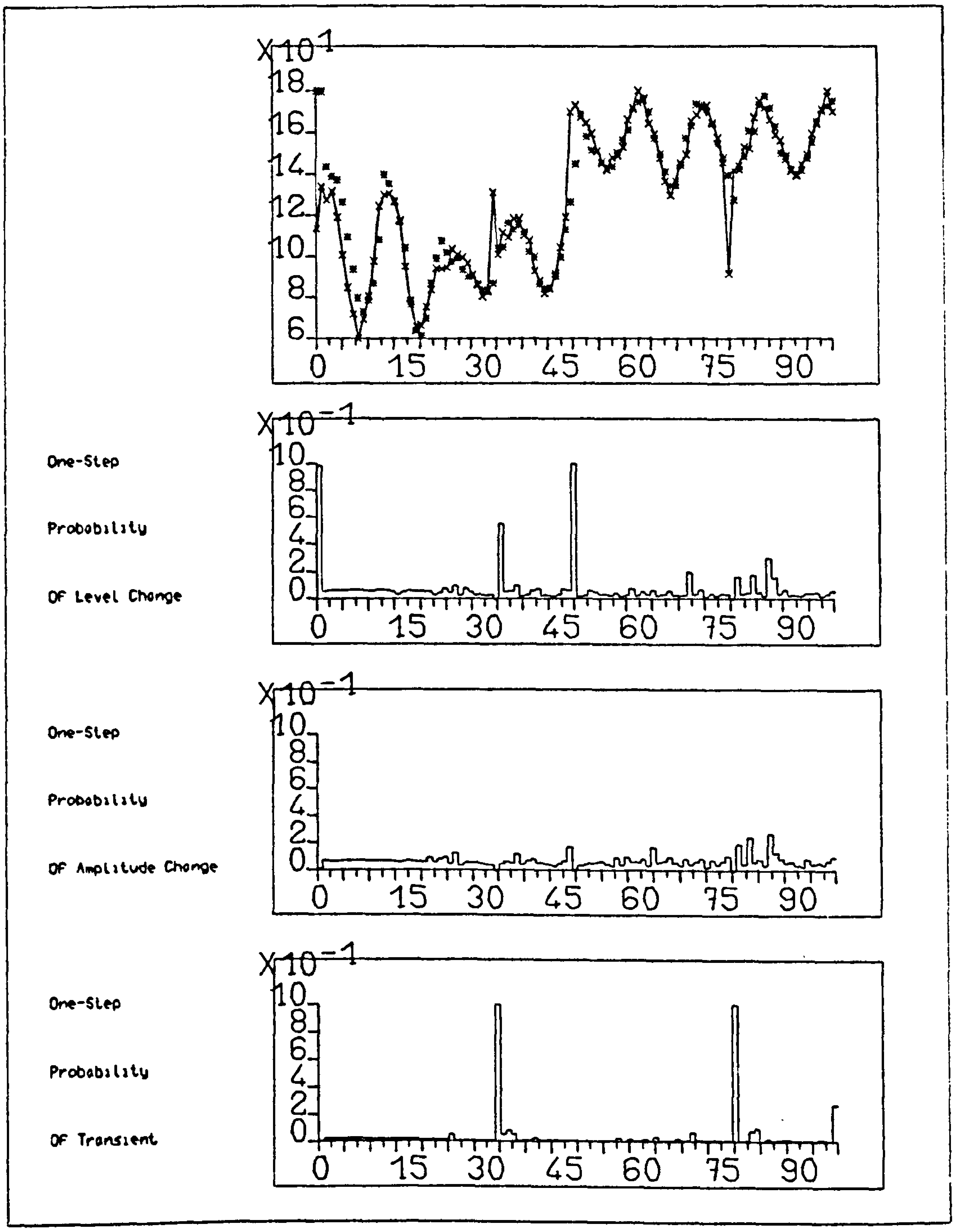

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.55

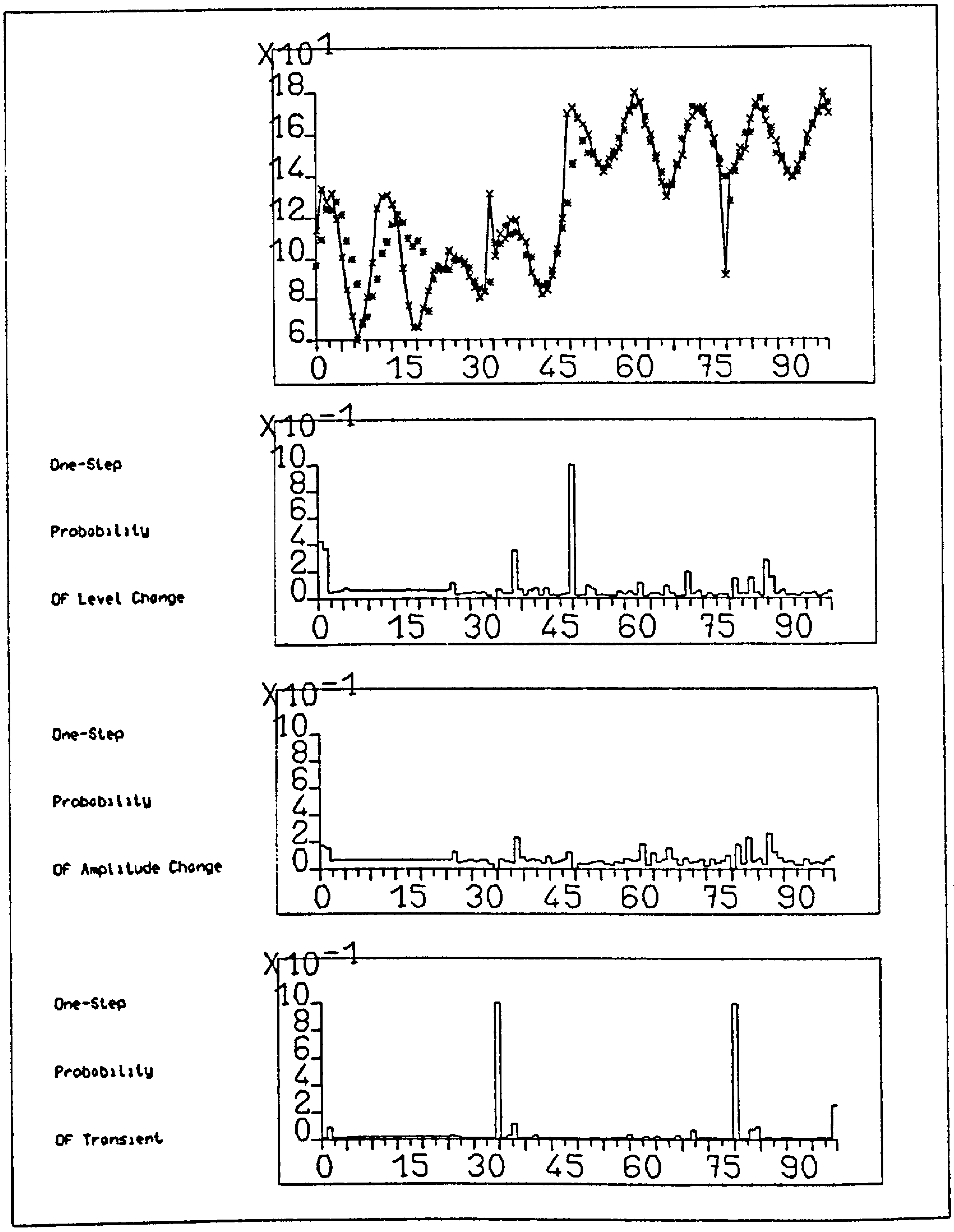

X10^1
One-Step
Probability
Of Level Change
X10^-1
One-Step
Probability
Of Amplitude Change
X10^-1
One-Step
Probability
Of Transient
X10^-1

FIGURE 3.56

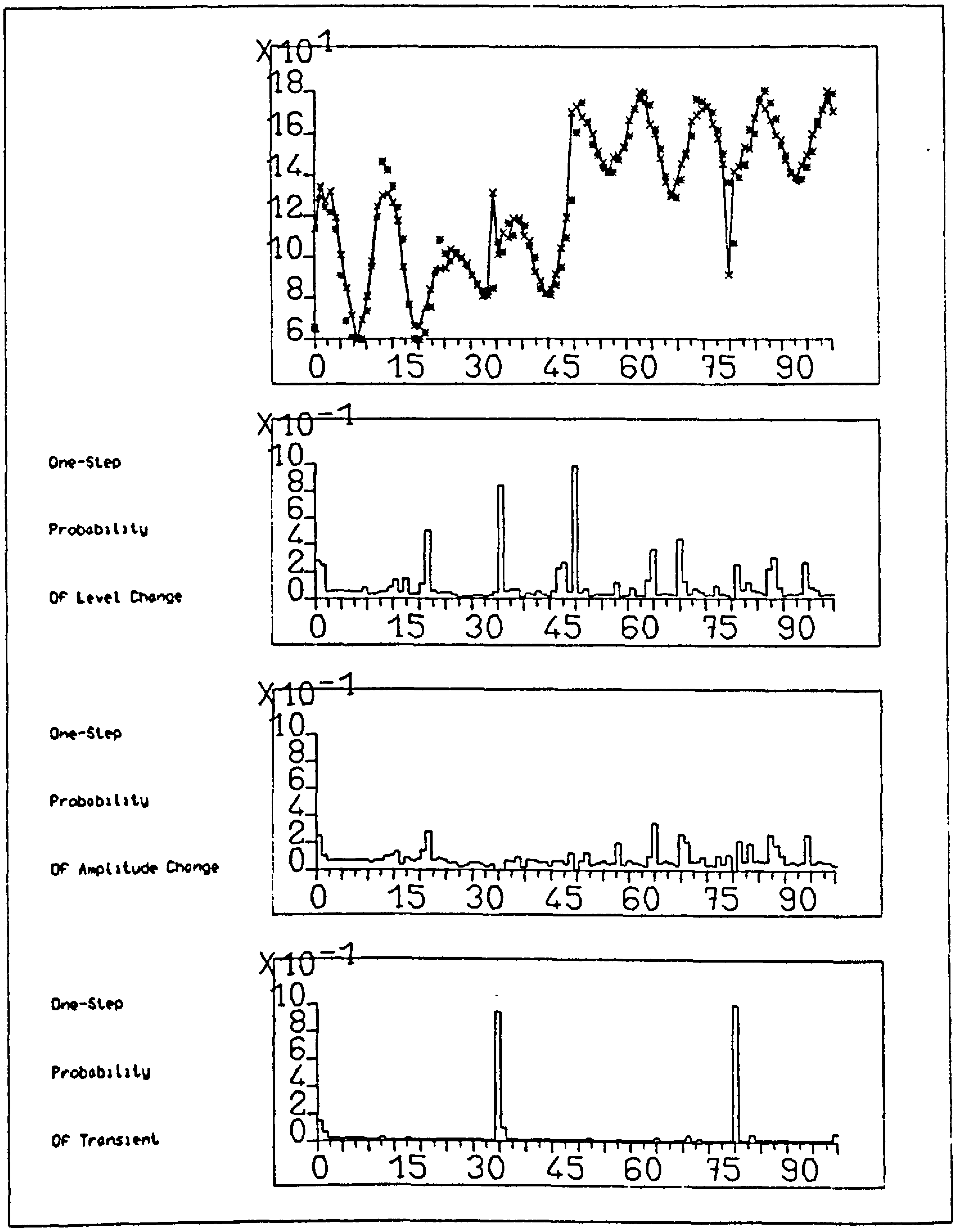

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.57

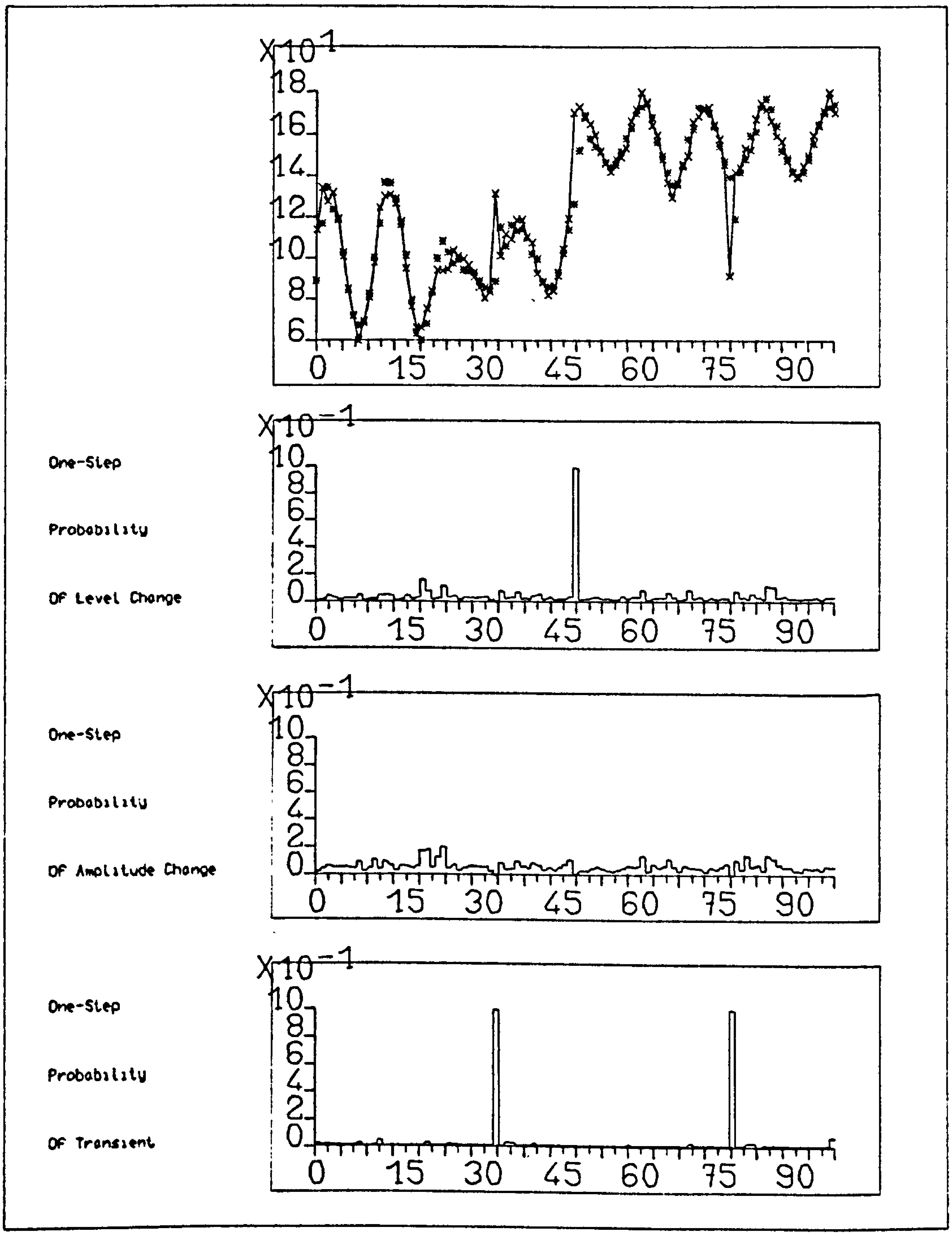

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.58

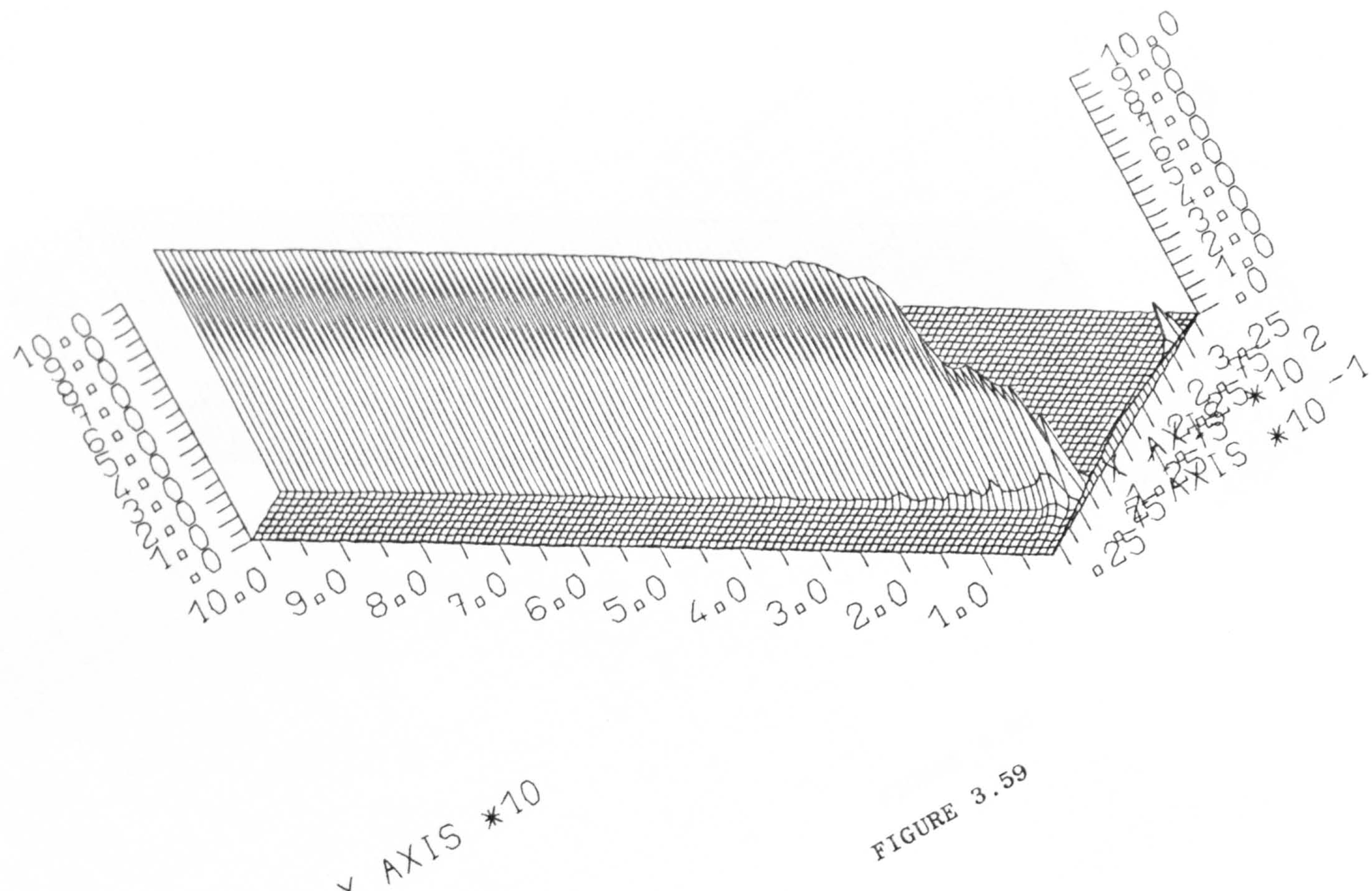


FIGURE 3.59

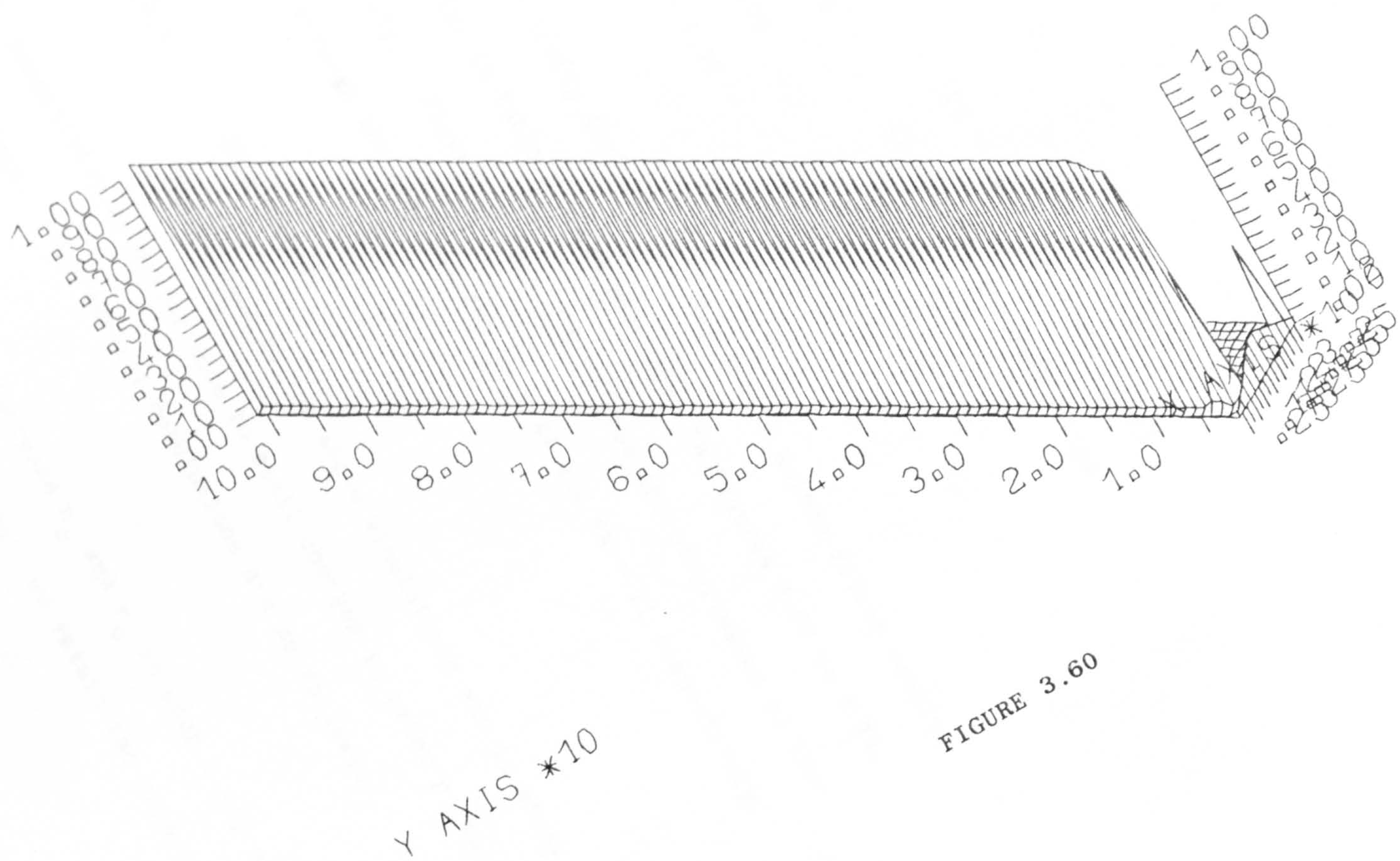


FIGURE 3.60

Using these values the multistate Kalman Filter results are shown in Figure 3.61, and the corresponding forecasts and forecast errors in Figure 3.62.

We have:

$$0_{26}^{(2)} = 0.679$$
$$0_{31}^{(3)} = 0.842$$
$$0_{36}^{(3)} = 0.742$$
$$0_{51}^{(4)} = 0.840$$
$$0_{76}^{(2)} = 0.298$$
$$0_{81}^{(4)} = 0.999$$
$$\text{NFP} = 3$$
$$v_{100} = 18.6 \qquad (\nu = 18.9, \text{ theoretically})$$
$$\hat{\phi}_{100} = 0.72 \qquad (\phi = 0.70, \text{ theoretically})$$
$$\text{SSFE} = 756$$
$$\text{MAD} = 1.77$$

3.4.3.2: Sensitivity Analysis. Multistate Kalman Filter results for (i) to (x) in Table 3.2 are displayed in Figures 3.63 to 3.72, respectively. Figures 3.73 and 3.74 show the progression of the ϕ grid through the analysis for NN = 21 and NN = 11 respectively.

3.4.4 CONCLUSIONS

The purpose of the above simulations was to study the sensitivity of performance to small changes in prior parameter settings. Our general conclusions are as follows:

(a) If we increase n_o and r_o in such a way as to keep the ratio $r_o/(n_o - 2)$ a constant, we retain the same initial estimate

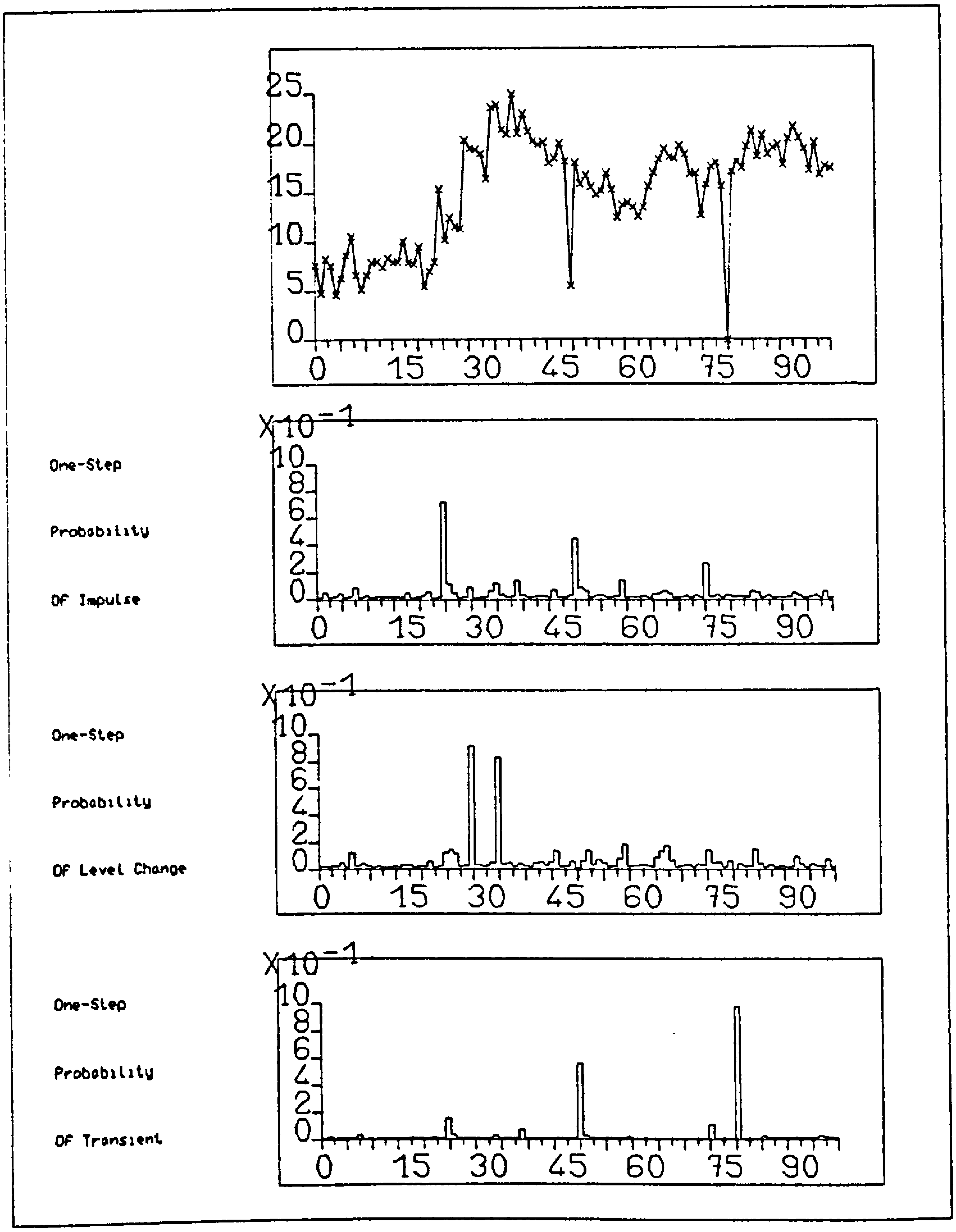

One-Step
Probability
Of Impulse
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient
X10^-1

FIGURE 3.61

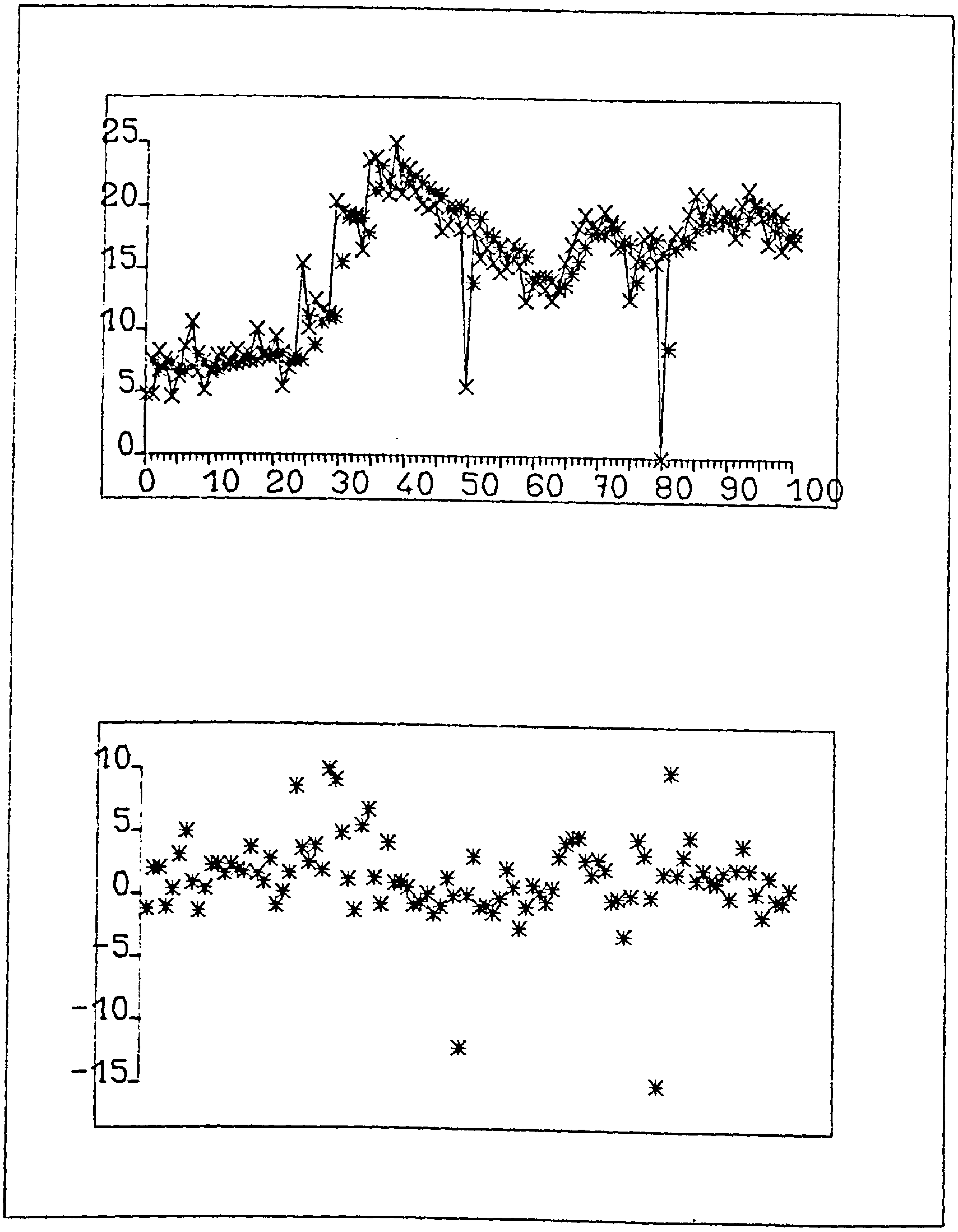
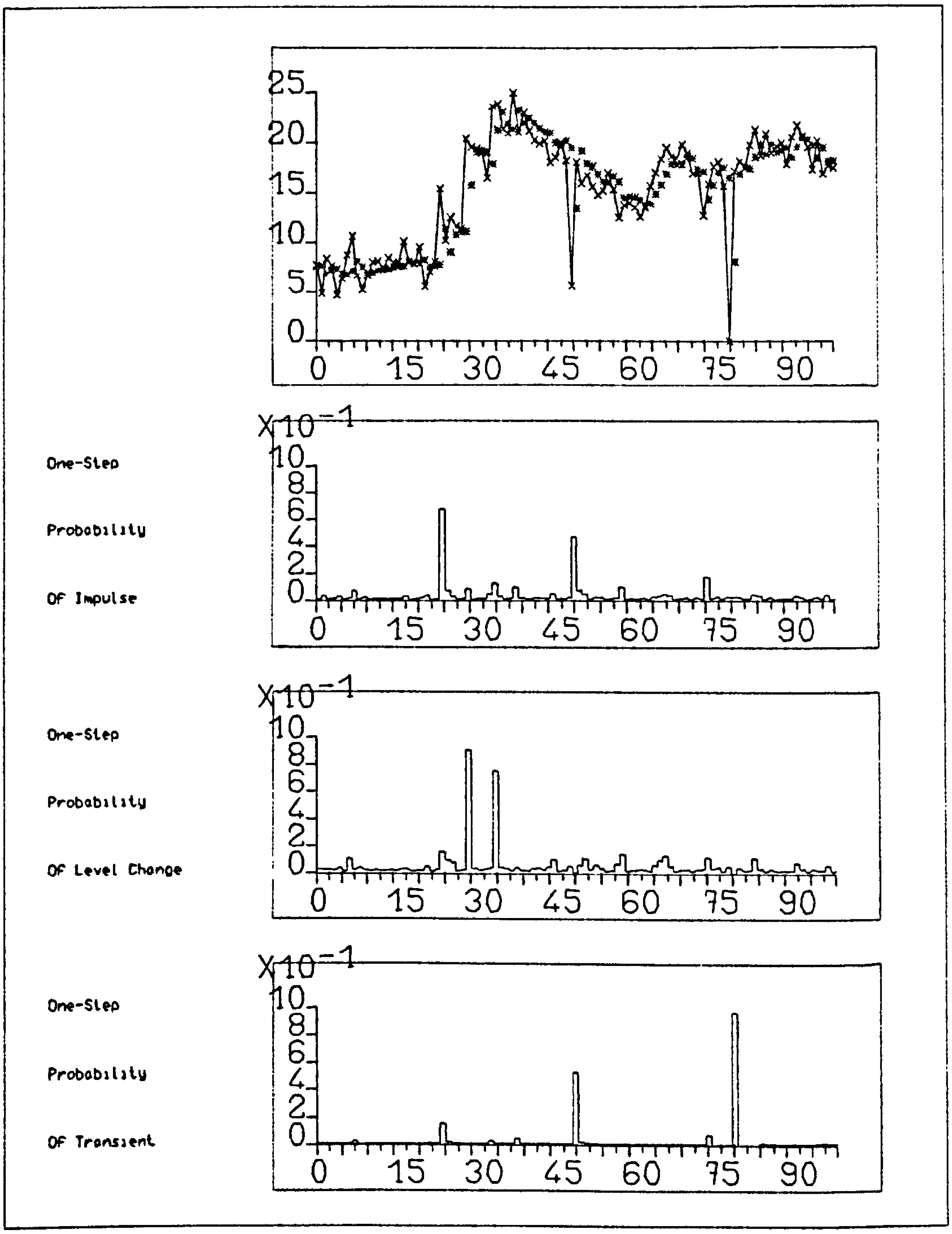


FIGURE 3.62

TABLE 3.2

	$0_{26}^{(2)}$	$0_{31}^{(3)}$	$0_{36}^{(3)}$	$0_{51}^{(4)}$	$0_{76}^{(2)}$	$0_{81}^{(4)}$	NFP	V_{100}	$\hat{\phi}_{100}$	SSFE	MAD
INTIAL SETTING	0.679	0.842	0.742	0.840	0.298	0.999	3	18.6	0.72	756	1.77
CHANGES											
(i) $n_o = 25,\ r_o = 23.0$	0.676	0.905	0.751	0.524	0.172	0.968	1	18.7	0.68	793	1.81
(ii) $n_o = 50,\ r_o = 48.0$	0.650	0.900	0.689	0.507	0.140	0.960	1	18.7	0.68	797	1.81
(iii) $r_o = 1.0$	0.746	0.945	0.929	0.582	0.694	0.955	14	18.6	0.37	739	1.76
(iv) $r_o = 9.0$	0.606	0.891	0.660	0.520	0.147	0.964	1	18.7	0.68	792	1.81
(v) $R_\mu^{(2)} = 60, R_\nu^{(3)} = 30, R_\varepsilon^{(4)} = 90$	0.716	0.911	0.827	0.552	0.259	0.974	1	18.7	0.67	781	1.81
(vi) $R_\mu^{(2)} = 180, R_\nu^{(3)} = 90, R_\varepsilon^{(4)} = 270$	0.723	0.938	0.824	0.423	0.150	0.933	1	18.8	0.69	833	1.84
(vii) $p_o^{(1)} = 0.97,\ p_o^{(i)} = 0.01$, $i = 2,3,4$	0.572	0.899	0.485	0.822	0.034	0.999	1	18.6	0.84	797	1.80
(viii) $v_o = 3.33$	0.722	0.841	0.740	0.761	0.241	0.986	2	18.7	0.73	874	1.88
(ix) $v_o = 30.0$	0.709	0.852	0.740	0.718	0.228	0.986	2	18.6	0.75	1686	2.29
(x) NN = 11	0.716	0.912	0.828	0.546	0.272	0.974	1	18.7	0.65	780	1.81



25
20
15
10
5
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Impulse
10 8 6 4 2 0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Level Change
10 8 6 4 2 0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Transient
10 8 6 4 2 0
0 15 30 45 60 75 90

FIGURE 3.63

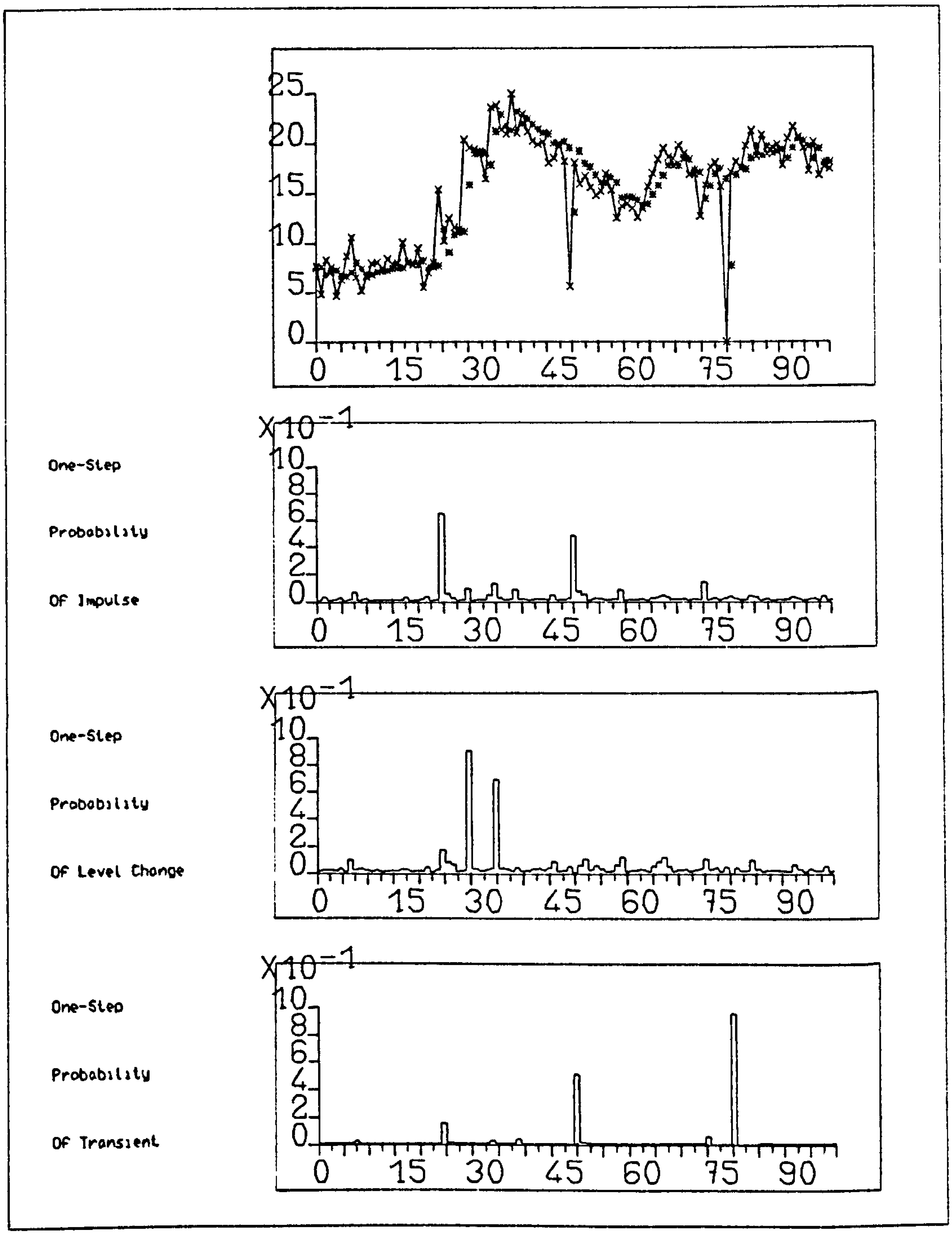
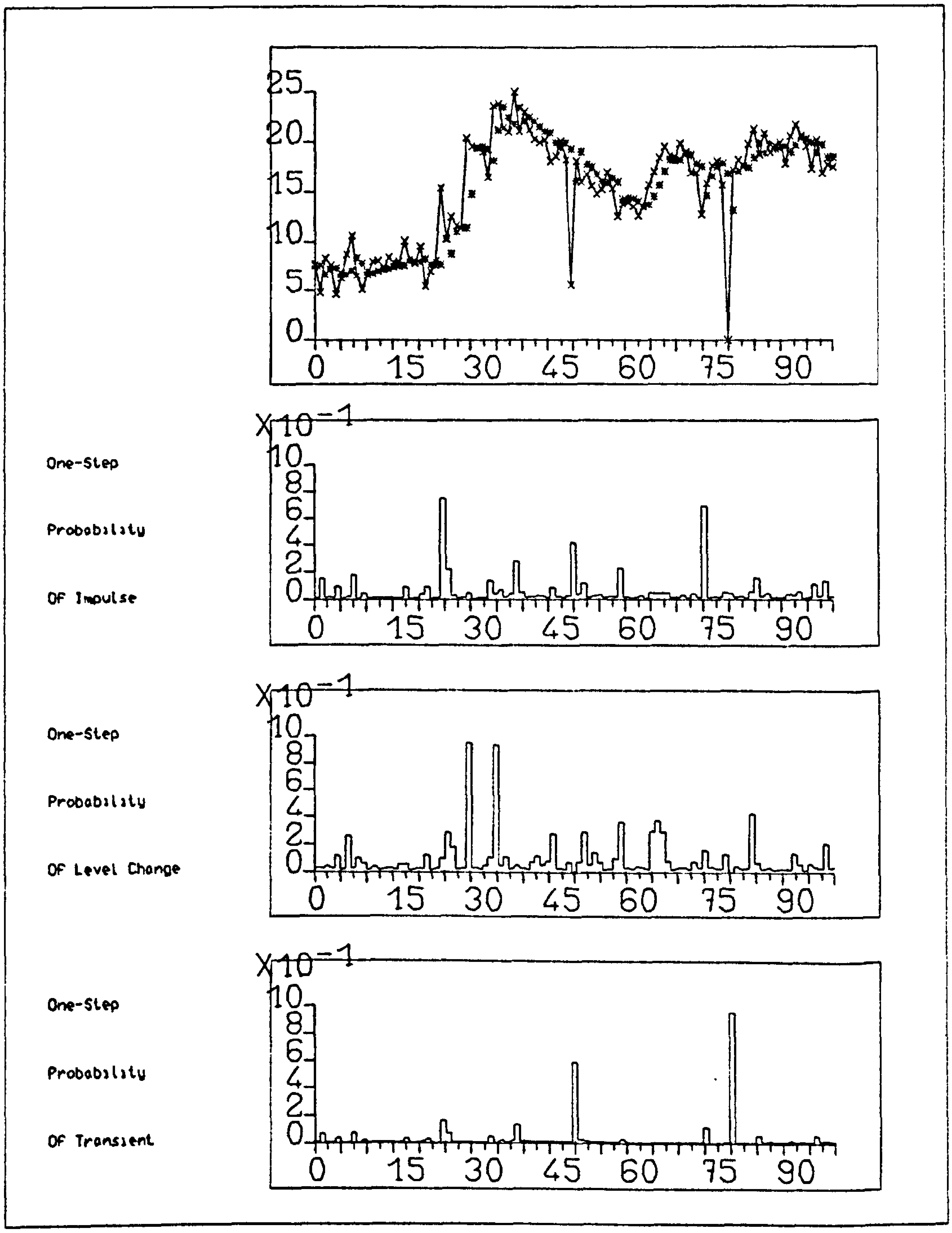

25
20
15
10
5
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Impulse
10
8
6
4
2
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Level Change
10
8
6
4
2
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Transient
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.64



25
20
15
10
5
0
0 15 30 45 60 75 90
One-Step
Probability
Of Impulse
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.65

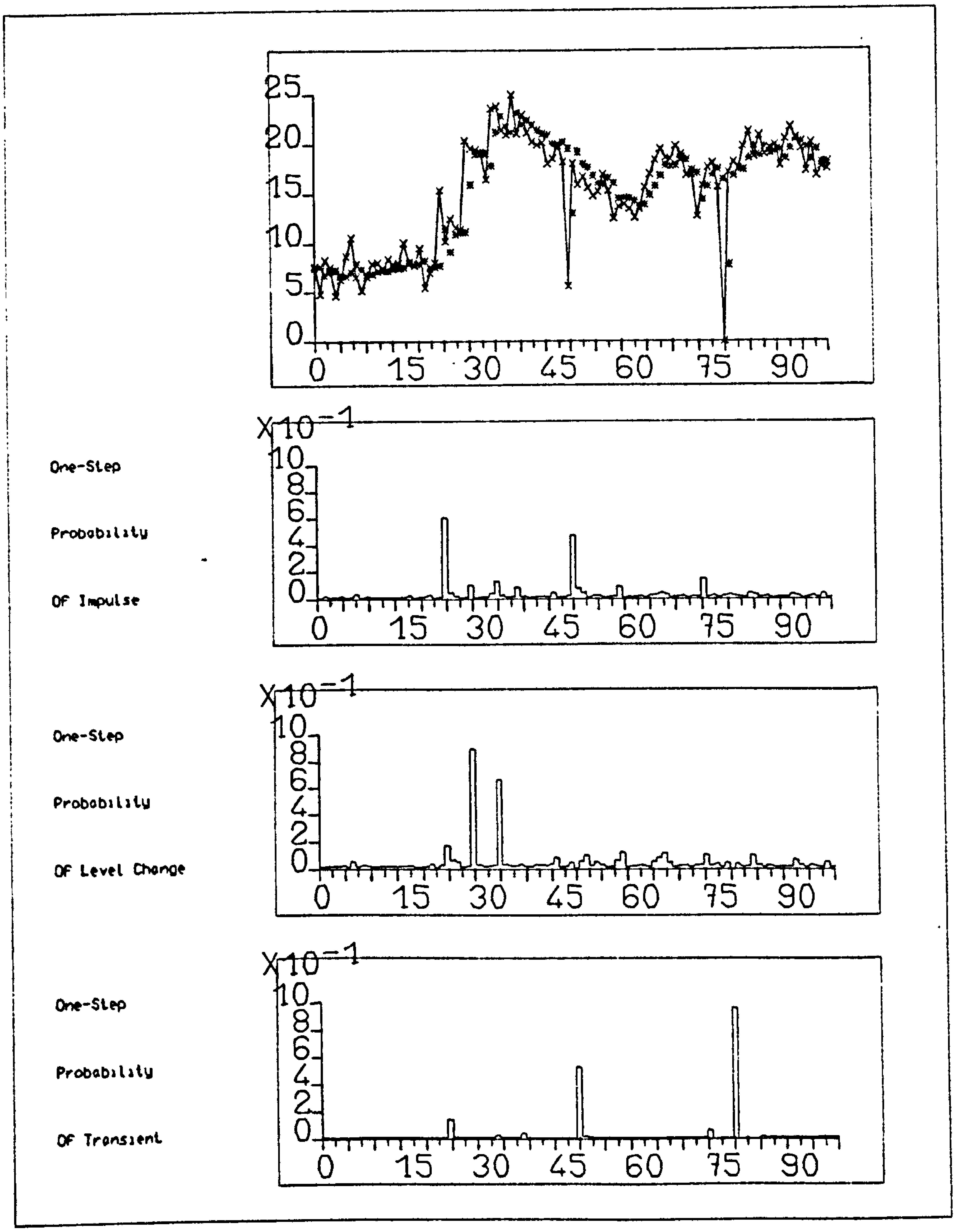

25
20
15
10
5
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Impulse
10
8
6
4
2
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Level Change
10
8
6
4
2
0
0 15 30 45 60 75 90
X10^-1
One-Step
Probability
Of Transient
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.66

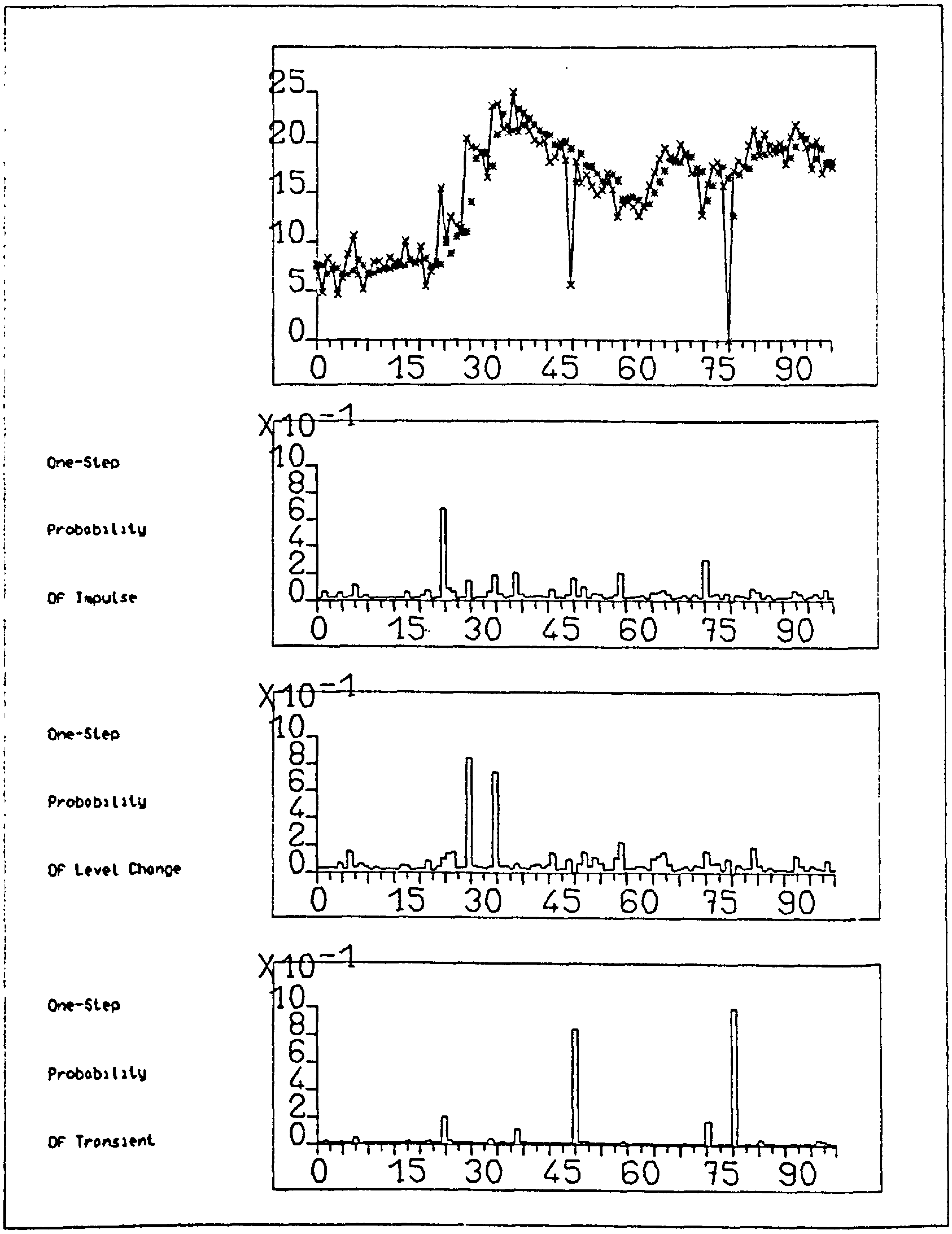

One-Step
Probability
Of Impulse
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient

FIGURE 3.67

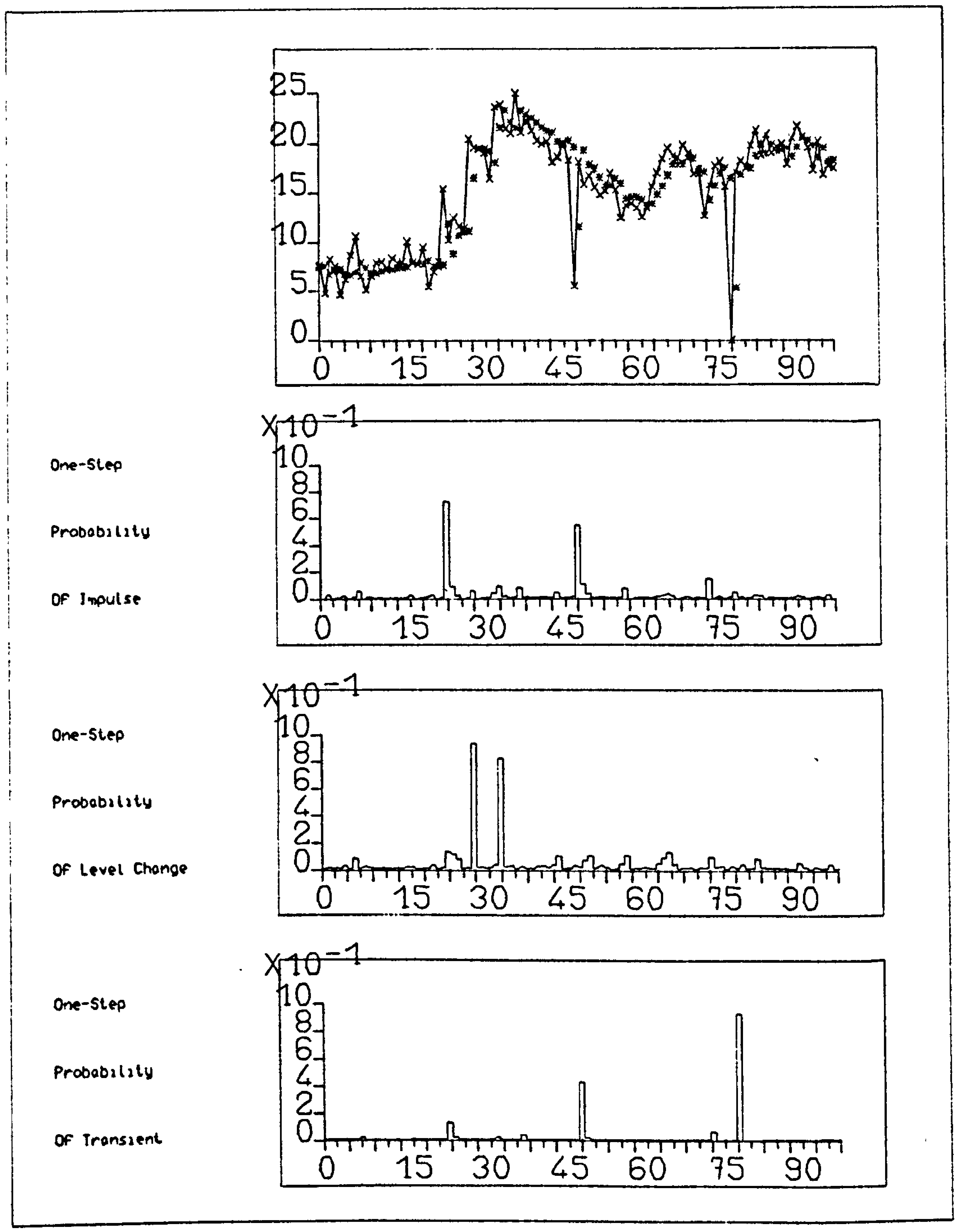
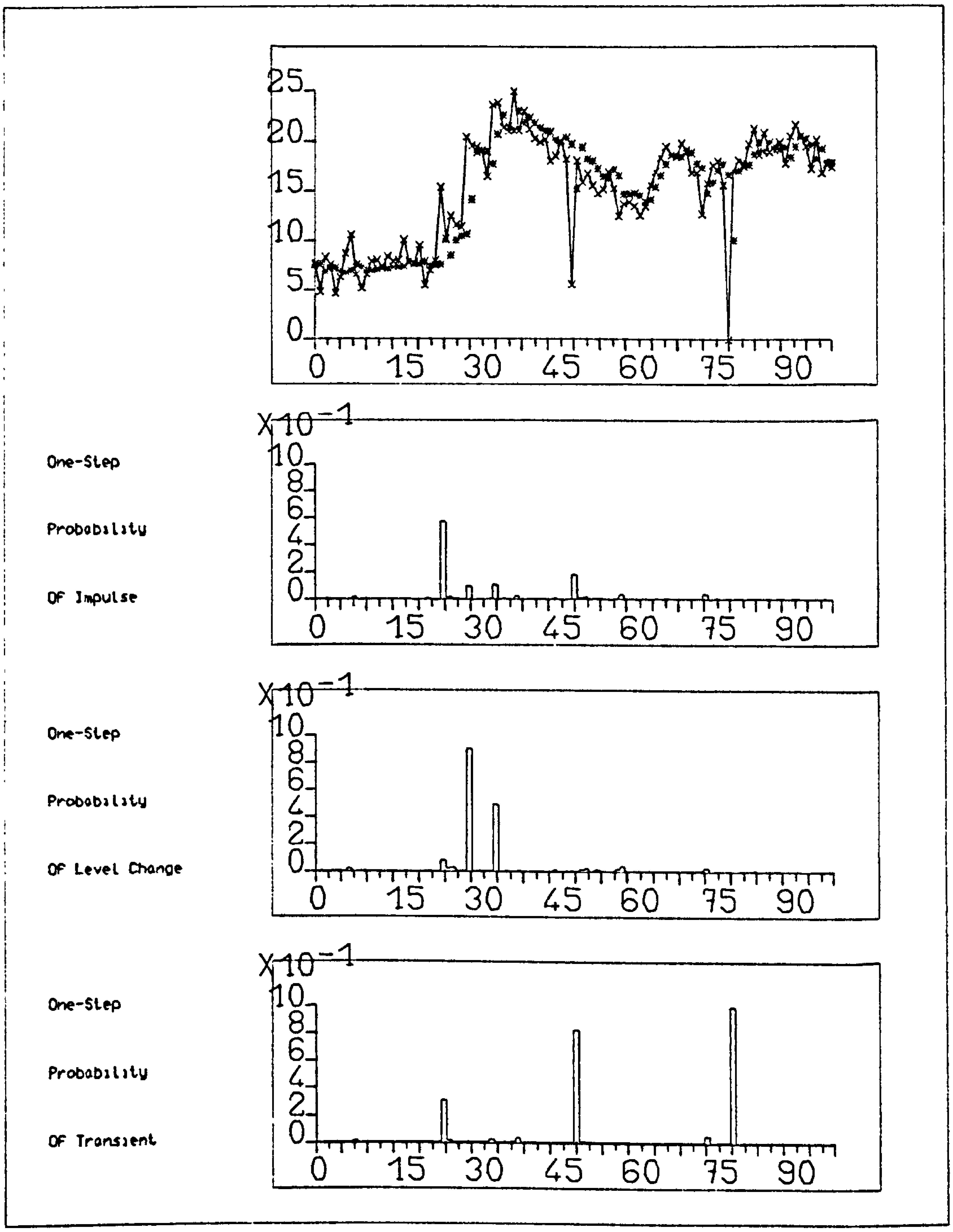
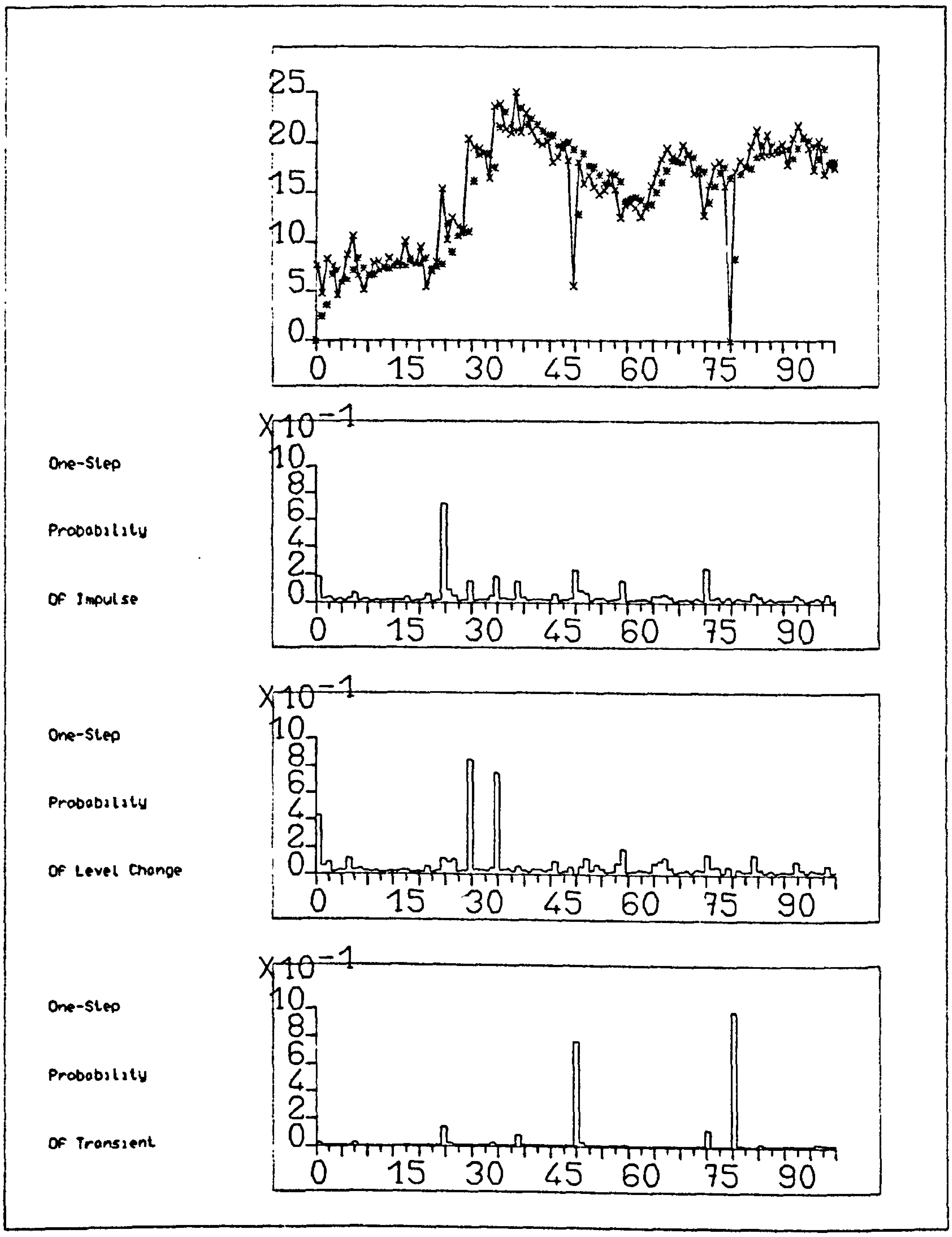

25
20
15
10
5
0
0 15 30 45 60 75 90
One-Step
Probability
Of Impulse
X10⁻¹
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10⁻¹
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10⁻¹
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.68



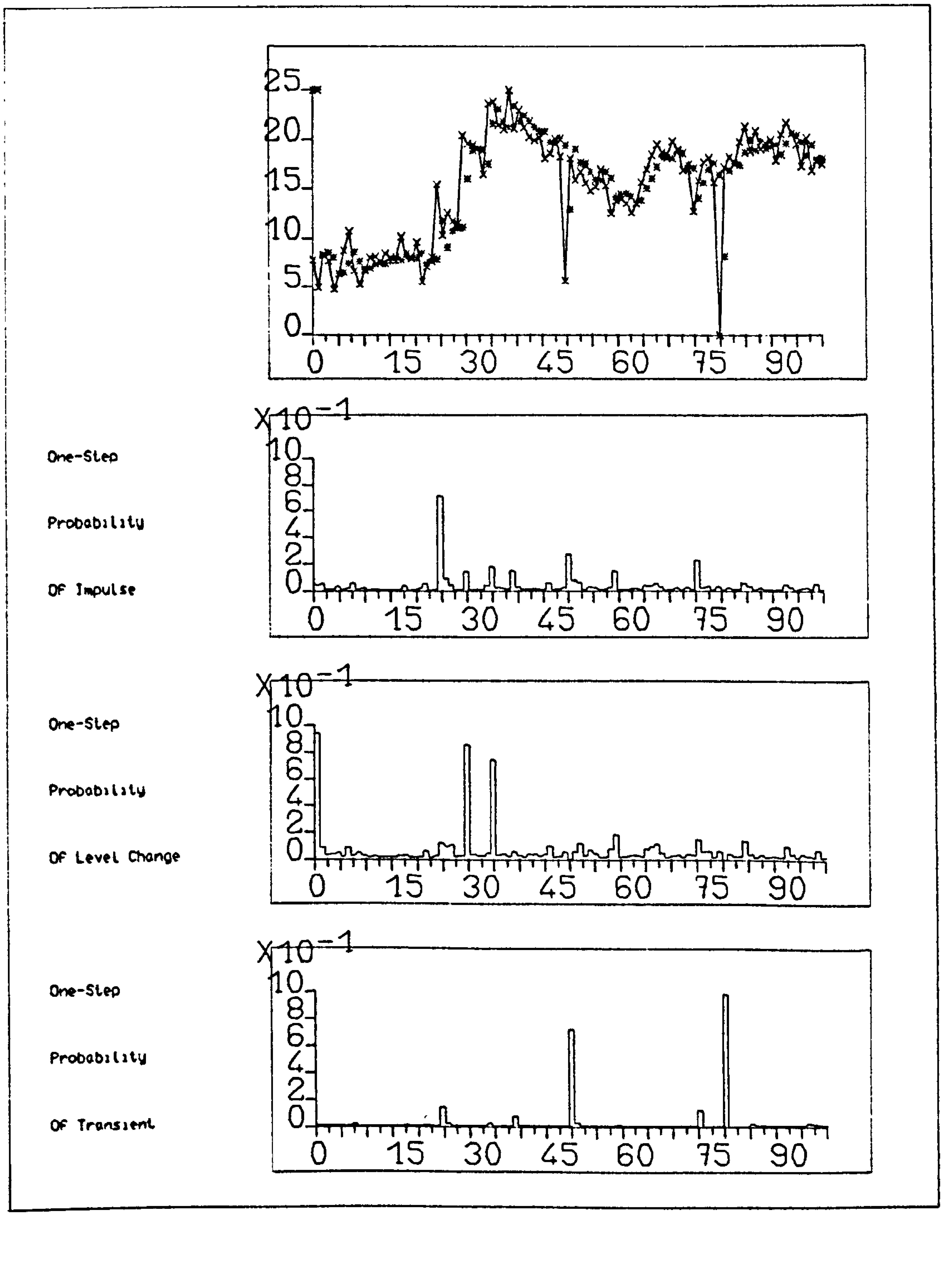
One-Step
Probability
Of Impulse
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient
X10^-1

FIGURE 3.69



One-Step
Probability
Of Impulse
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient

FIGURE 3.70



25
20
15
10
5
0
0 15 30 45 60 75 90
One-Step
Probability
Of Impulse
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 3.71

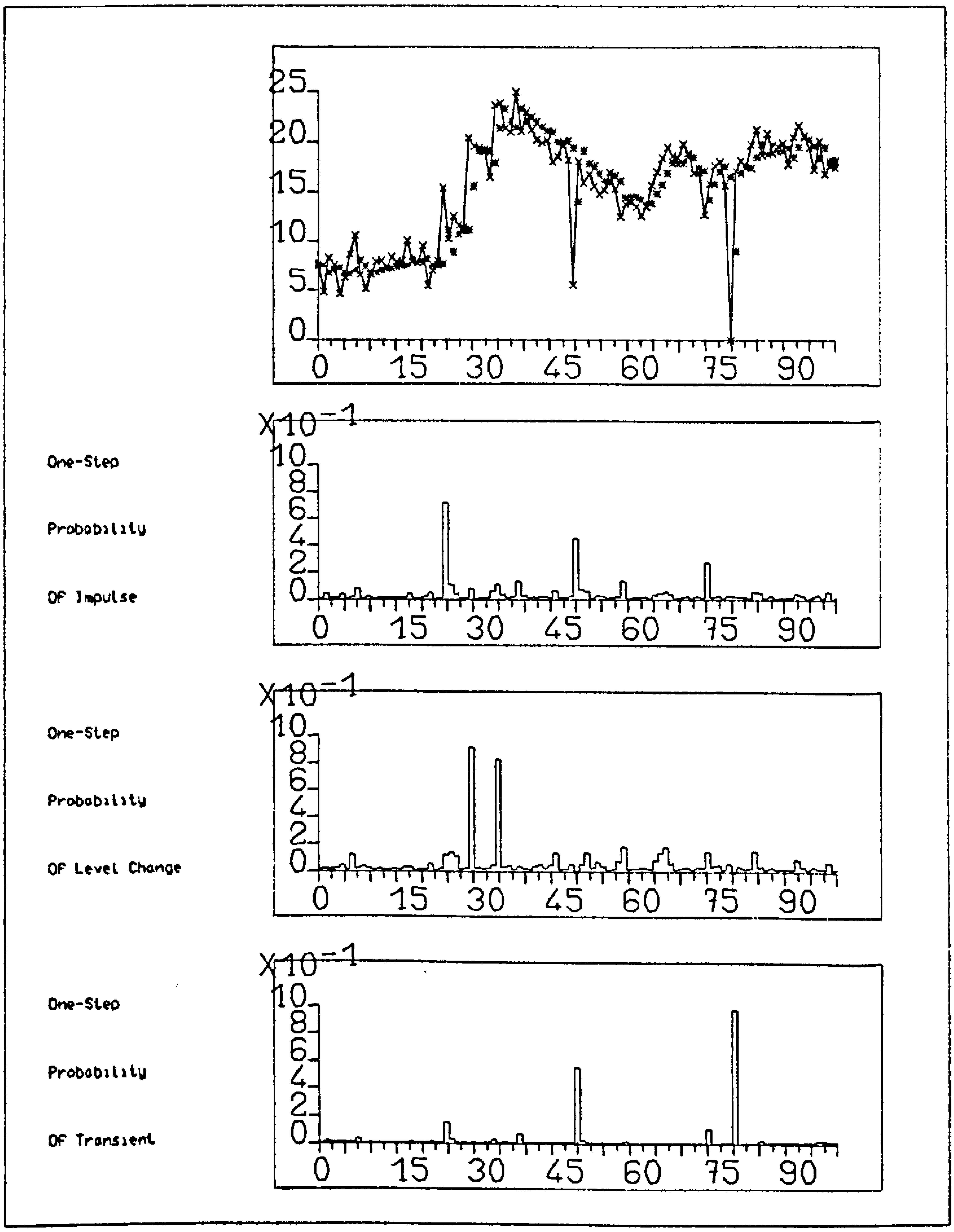
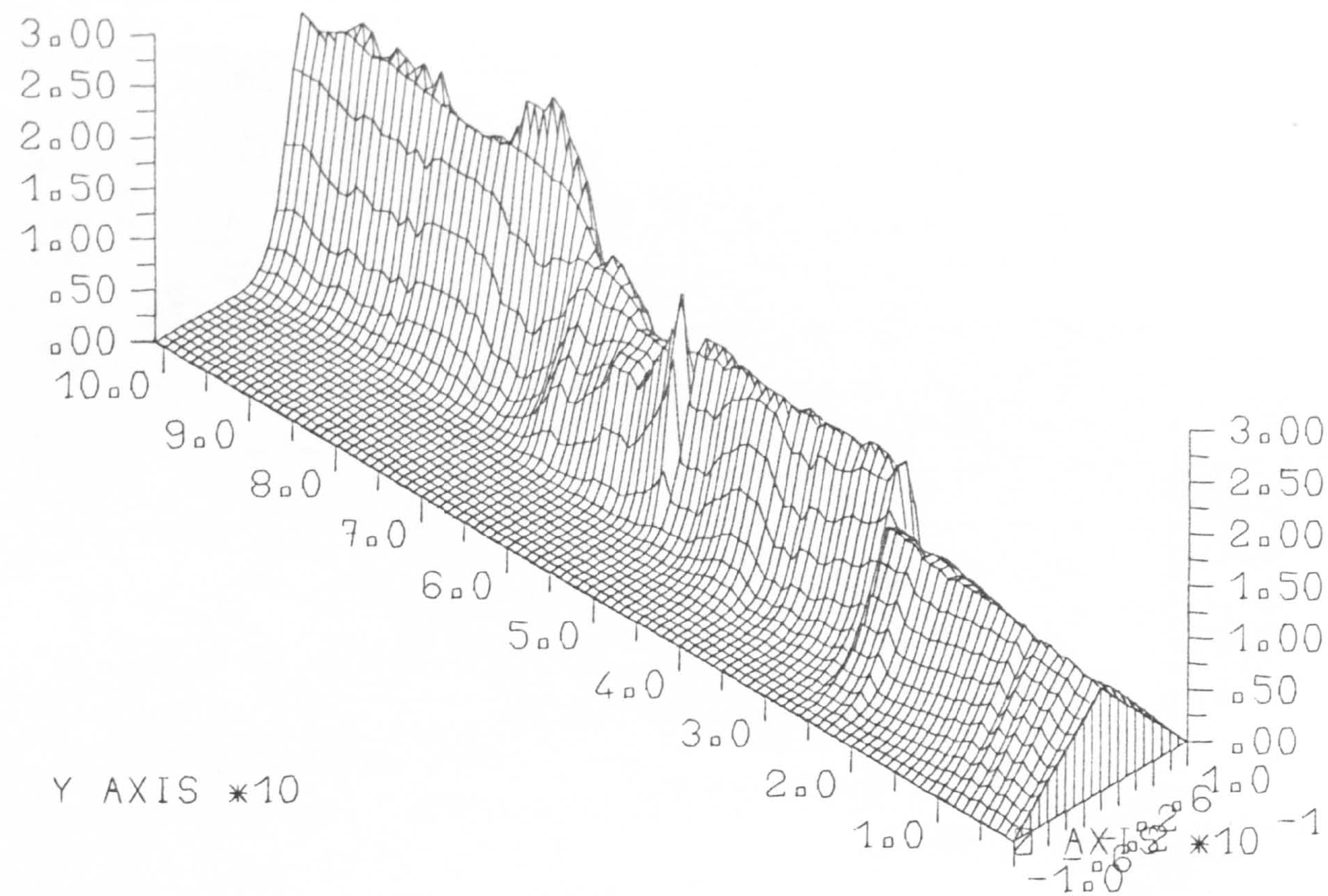

One-Step
Probability
Of Impulse
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient
X10^-1

FIGURE 3.72



3.00
2.50
2.00
1.50
1.00
.50
.00
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
3.00
2.50
2.00
1.50
1.00
.50
.00
*10 −1

FIGURE 3.73

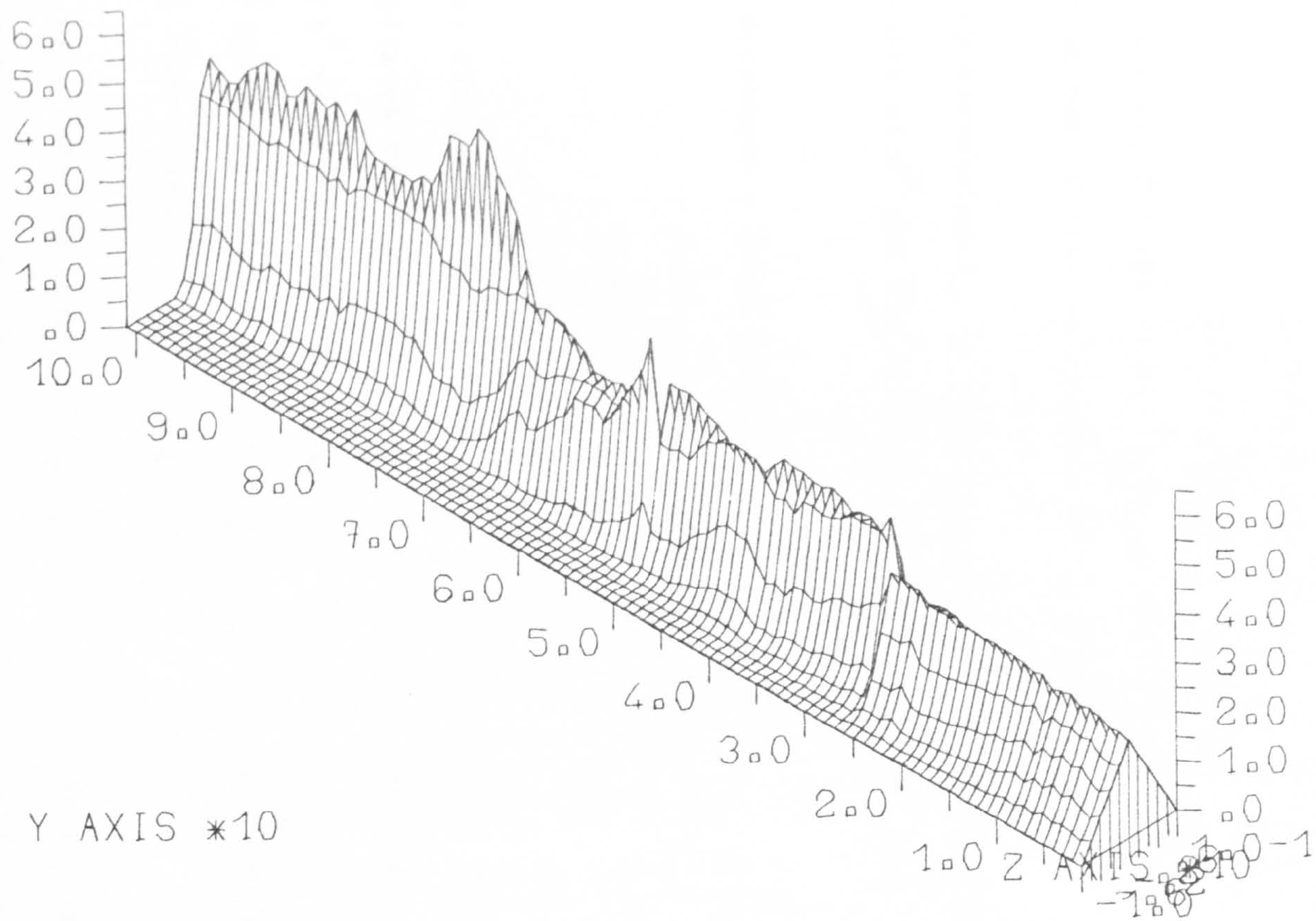


FIGURE 3.74

of $\lambda^{-1} = c^2$, but our <u>variance</u> on that estimate is smaller (see Section 2.1.2), corresponding to greater confidence in our initial estimate of λ^{-1}.

The effects of this increase on the various performance measures are very small. Such changes as we have observed as as follows:

As (n_o, r_o) increases - event probabilities decrease
- number of false positives decreases
- SSFE/MAD increases
- $\underset{\sim}{\theta}, \phi$ estimation: no change.

This indicates that, for the series under study, c^2 was perhaps more changeable than our initial variance on c^2 allowed for.

(b) By decreasing r_o, holding n_o constant, we lower our initial estimate of c^2 and the following changes in performance obtain:

As r_o decreases - event probabilities increase
- number of false positives increases
- SSFE/MAD decreases
- $\underset{\sim}{\theta}, \phi$ estimation: no change.

Most of these changes are small, but the increase in the number of false positives can be large. Clearly, with too small an estimate of c^2 we are likely to signal observations which are <u>outside</u> the stated variance range, though <u>within</u> the <u>correct</u> variance range, thus precipitating false alarms.

(c) By increasing $R_\varepsilon, R_{\underset{\sim}{\omega}}$ we would expect event detection to be more difficult, since the magnitude of a change needs to be greater to reveal its presence. Referring to the results in Sections 3.4.1 to 3.4.3, however, we see that changes in $R_\varepsilon, R_{\underset{\sim}{\omega}}$ have little effect on either the event probabilities or the number of false positives, as the learning procedure for λ adjusts itself to take account of the sizes of these multipliers (see Figures 3.38, 3.52 and 3.68). However, the SSFE and MAD increase noticeably (implying that observation tracking is poorer).

This indicates that the initial choice of $R_\varepsilon, R_{\tilde{\omega}}$ is not critical to the event detection problem, though some care must be taken if we seek to use the models for forecasting.

(d) Making the initial changepoint probabilities small ($p_o^{(i)}$, i = 2,3,4) has a negligible effect on either event detection or forecasting and estimation.

(e) Changes to the components of $\tilde{m}_o$ have little effect on the detection of changes; however the number of false positives may increase, especially in the initial stages while the system learns about the correct 'level' of $\tilde{m}_o$ (see Figures 3.41, 3.55 and 3.71; see Chapter 6 for an example where this property is exploited). It takes the system a little time to re-adjust, as reflected by the inflated values of SSFE and MAD, mainly due to an initial period of poor tracking.

(f) We might expect that the results would be much less accurate if we were to decrease the number of nodes in the ϕ grid, but a change in NN seems to have hardly any effect on any of the performance measures.

N.B. Notice the effect of an impulse on the ϕ-updating procedure (Figures 3.73 and 3.74); the amount of information about ϕ provided by an impulse is considerable and the grid method of estimation reacts quickly.

As an attempt at an overall summary, we might say that:

(i) as far as estimation is concerned, the models are not very sensitive to the starting values of any of the parameters;

(ii) the most sensitive parameter is $\underset{\sim}{m}_0$: a poor initial estimate may result in too many false positives (especially early on) and lead to poor forecasts;

(iii) r_0 is also fairly critical, as the number of false alarms will be great if r_0 is too small;

(iv) the models are not sensitive to the choice of NN;

(v) changes in the specification of R_ε and $\underset{\sim}{R}_\omega$ have more effect on forecasting than on change detection (re. Section 2.1.2).

oOo

APPENDIX THREE

DATA SETS

A3.1 LINEAR GROWTH

Using the starting values of $\underset{\sim}{\theta}_o = \begin{pmatrix}100\\5\end{pmatrix}$,100 observations were generated according to the linear growth model of Section 2.3.1 with the errors, ε_t, simulated from $\varepsilon_t \sim N(0,15)$.

At t = 25 the slope was reversed in sign, so that $\beta_t = -5$, $t \geq 25$.

At t = 50 the level was increased by 50%, so that $\mu_{50} = 3\mu_{49}/2 + \beta_{50}$.

At t = 35, $\frac{1}{2}\mu_{35}$ was added to y_{35} (in addition to ε_{35}) resulting in a transient observation; a second transient was created at t = 80 by subtracting $\frac{1}{2}\mu_{80}$ from y_{80}.

The following data set obtained:

TIME	OBSERVATION	TIME	OBSERVATION
1.00	103.79	51.00	131.51
2.00	112.76	52.00	122.24
3.00	119.21	53.00	118.36
4.00	121.93	54.00	117.32
5.00	125.98	55.00	108.12
6.00	132.89	56.00	107.61
7.00	144.86	57.00	104.21
8.00	143.62	58.00	103.28
9.00	145.05	59.00	93.68
10.00	150.60	60.00	90.26
11.00	157.31	61.00	84.68
12.00	159.62	62.00	77.64
13.00	162.02	63.00	65.47
14.00	175.72	64.00	64.90
15.00	171.36	65.00	67.75
16.00	183.24	66.00	55.49
17.00	190.49	67.00	51.63
18.00	189.18	68.00	39.56

TIME	OBSERVATION	TIME	OBSERVATION
19.00	197.00	69.00	38.23
20.00	208.88	70.00	35.83
21.00	212.22	71.00	25.67
22.00	209.98	72.00	26.92
23.00	218.91	73.00	20.30
24.00	224.29	74.00	15.27
25.00	215.60	75.00	9.59
26.00	207.26	76.00	3.82
27.00	209.75	77.00	5.41
28.00	199.45	78.00	0.17
29.00	193.58	79.00	-3.59
30.00	189.76	80.00	-56.25
31.00	182.29	81.00	-18.70
32.00	171.58	82.00	-23.26
33.00	173.82	83.00	-32.89
34.00	168.86	84.00	-36.84
35.00	207.50	85.00	-45.25
36.00	154.23	86.00	-49.28
37.00	154.43	87.00	-46.39
38.00	143.58	88.00	-55.75
39.00	139.13	89.00	-63.08
40.00	134.74	90.00	-58.03
41.00	133.71	91.00	-67.29
42.00	124.52	92.00	-75.54
43.00	116.52	93.00	-79.86
44.00	117.02	94.00	-81.74
45.00	114.46	95.00	-88.94
46.00	107.64	96.00	-97.98
47.00	106.68	97.00	-95.29
48.00	99.97	98.00	-99.94
49.00	96.19	99.99	-103.46
50.00	138.32	100.00	-119.56

A3.2 SINUSOIDAL MODEL

Using the starting values of $\underset{\sim}{\theta}_o = \begin{pmatrix} 100 \\ 30 \end{pmatrix}$ and $\phi = \pi/2$, 100 observations were generated according to the sinusoidal model of Section 2.3.3 with $\varepsilon_t \sim N(0,15)$.

At t = 25 the amplitude was <u>decreased</u> by 50%, so that $\alpha_t = 15$, $t \geq 25$.

At t = 50 the level was <u>increased</u> by 50%, so that $\mu_t = 150$, $t \geq 50$.

At t = 35, $\frac{1}{2}\mu_{35} = 50$ was added to y_{35} resulting in a transient observation; a second transient was created at t = 80 by subtracting $\frac{1}{2}\mu_{80} = 75$ from y_{80}.

The following data set obtained:

TIME	OBSERVATION	TIME	OBSERVATION
1.00	111.89	51.00	164.21
2.00	130.09	52.00	159.50
3.00	124.26	53.00	156.90
4.00	128.02	54.00	152.32
5.00	116.83	55.00	145.03
6.00	100.29	56.00	140.44
7.00	86.02	57.00	136.86
8.00	74.48	58.00	142.41
9.00	64.00	59.00	143.18
10.00	72.32	60.00	146.70
11.00	82.53	61.00	158.19
12.00	97.73	62.00	163.02
13.00	121.45	63.00	170.49
14.00	126.29	64.00	166.28
15.00	127.04	65.00	156.36
16.00	123.18	66.00	152.22
17.00	115.28	67.00	142.12
18.00	95.18	68.00	132.09
19.00	78.83	69.00	125.88
20.00	69.48	70.00	131.90
21.00	69.74	71.00	140.00
22.00	77.68	72.00	143.51
23.00	85.32	73.00	157.65
24.00	94.46	74.00	160.16
25.00	94.33	75.00	162.86
26.00	94.95	76.00	164.06
27.00	102.85	77.00	156.72
28.00	100.44	78.00	150.35
29.00	99.33	79.00	139.88
30.00	96.73	80.00	92.01
31.00	91.54	81.00	136.67
32.00	87.02	82.00	138.58
33.00	82.49	83.00	146.65
34.00	85.00	84.00	145.93
35.00	127.50	85.00	158.90
36.00	100.75	86.00	165.75
37.00	110.16	87.00	162.84
38.00	108.08	88.00	158.00
39.00	116.23	89.00	151.65
40.00	116.22	90.00	149.36
41.00	108.84	91.00	141.46
42.00	106.49	92.00	136.32
43.00	93.25	93.00	134.27
44.00	89.04	94.00	139.12
45.00	83.77	95.00	143.24
46.00	85.65	96.00	152.02
47.00	92.15	97.00	156.13
48.00	103.56	98.00	162.32
49.00	116.62	99.00	170.26
50.00	161.52	100.00	161.46

A3.3 AR(1)

Using the starting values of $\underset{\sim}{\theta}_o = \begin{pmatrix}10\\10\end{pmatrix}$ and $\phi = 0.7$, 100 observations were generated according to the AR(1) model of Section 2.3.5.1 with $\varepsilon_t \sim N(0,1.0)$.

At t = 30 and t = 35 level changes were simulated by setting $R_{\nu_{30}} = R_{\nu_{35}} = 10$, i.e. $\delta\nu_{30} \sim N(0,10)$ and $\delta\nu_{35} \sim N(0,10)$.

At t = 25 and t = 75 impulses were simulated by setting $R_{\mu_{25}} = R_{\mu_{75}} = 20$, i.e. $\delta\mu_{25} \sim N(0,20)$ and $\delta\mu_{75} \sim N(0,20)$.

At t = 50 and t = 80 transients were simulated by setting $R_{\varepsilon_{50}} = R_{\varepsilon_{80}} = 30$, i.e. $\varepsilon_{50} \sim N(0,30)$ and $\varepsilon_{80} \sim N(0,30)$.

The following data set obtained:

TIME	OBSERVATION	TIME	OBSERVATION
1.00	10.14	51.00	18.43
2.00	7.86	52.00	16.72
3.00	10.65	53.00	17.38
4.00	10.07	54.00	16.46
5.00	7.71	55.00	15.82
6.00	9.03	56.00	16.15
7.00	10.95	57.00	17.57
8.00	12.49	58.00	16.23
9.00	9.28	59.00	13.97
10.00	8.14	60.00	14.99
11.00	9.28	61.00	15.14
12.00	10.35	62.00	14.80
13.00	10.41	63.00	14.04
14.00	9.95	64.00	14.78
15.00	10.70	65.00	16.51
16.00	10.36	66.00	17.57
17.00	10.39	67.00	18.68
18.00	12.06	68.00	19.51
19.00	10.34	69.00	18.87
20.00	10.22	70.00	18.76
21.00	11.61	71.00	19.76
22.00	8.39	72.00	19.09
23.00	9.61	73.00	17.50
24.00	10.38	74.00	17.52
25.00	16.30	75.00	14.15
26.00	12.17	76.00	16.66
27.00	13.99	77.00	18.10
28.00	13.27	78.00	18.40
29.00	13.10	79.00	16.51

TIME	OBSERVATION	TIME	OBSERVATION
30.00	20.26	80.00	4.00
31.00	19.60	81.00	17.69
32.00	19.50	82.00	18.47
33.00	19.18	83.00	17.96
34.00	17.16	84.00	19.70
35.00	22.82	85.00	20.90
36.00	22.99	86.00	18.90
37.00	21.05	87.00	20.56
38.00	20.72	88.00	19.09
39.00	23.92	89.00	19.60
40.00	20.79	90.00	19.80
41.00	22.33	91.00	18.21
42.00	20.90	92.00	20.28
43.00	20.12	93.00	21.20
44.00	19.88	94.00	20.35
45.00	20.07	95.00	19.50
46.00	18.43	96.00	17.77
47.00	18.81	97.00	19.97
48.00	19.98	98.00	17.41
49.00	18.56	99.00	18.09
50.00	8.44	100.00	17.91

oOo

CHAPTER FOUR

UNEQUALLY-SPACED MEASUREMENTS

4.1 INTRODUCTION TO TIME SERIES WITH UNEQUALLY-SPACED DATA

4.1.1 INTRODUCTORY REMARKS

Medical time series obtained during routine clinical supervision have certain special characteristics, which need to be taken into account at the modelling stage if we are to utilize the methods discussed in earlier chapters. In particular, measurements are likely to be received at unevenly-spaced timepoints. There are several reasons why this might be the case, including the closure of laboratories (for instance at weekends., holidays, etc.), occasional sample mishandling, missed appointments, the fact that the patient might leave the 'in-patient' environment (making it more difficult to obtain measurements) and, most important of all, a change in the severity of the patient's condition. Clearly, if the health of a patient deteriorates the frequency of sampling is likely to be increased, since the clinician will wish to keep a closer check on the patient's condition; conversely, the patient may be monitored less frequently if his/her condition appears to be improving. In extreme cases, where the clinician is on the lookout for sudden events which might radically change the course of a 'disease', the interaction between event detection and the underlying between-sample time interval is of fundamental importance, and will be examined in Chapter 5.

Using the terminology of Jones (1985), we might say that the medical time series we are interested in are likely to be truly unequally-spaced rather than equally-spaced with occasional missing observations, although the latter could also arise from sample mishandling, missed appointments, etc. It will be necessary, therefore, to reformulate the models described earlier in order to accommodate the possibility of truly unequally-spaced observations.

A number of factors have influenced our strategy for modelling unequally-spaced time series:

(i) the desire to incorporate changepoint models and hence to retain a state-space (recursive) formulation based on a discrete-time, rather than a continuous-time, representation for unequally-spaced models (see Section 4.2 for a fuller discussion of this point);

(ii) the relative absence of literature relating to irregularly-spaced time series in the time domain;

(iii) the desire to develop a modelling framework that is both conceptually simple and widely applicable to a range of specific models.

NOTATION: Due to the algebraic complexity of the following sections, we make some preliminary remarks on notation.

It is assumed that 'k' will denote the time index for an unequally-spaced time series so that, for instance, y_k denotes the kth observation and not, in general, the observation at time k (unless the first k observations happen to be evenly-spaced).

The time at which the kth observation is made will be denoted by T_k, so that for the case of equally-spaced measurements $T_k = k, \forall k$.

The interval between consecutive observations will be denoted by $d_k = T_k - T_{k-1}$ (initially it is assumed that $T_o = 0$, so that $d_1 = T_1$). Notice that if $d_k > 1$ then no observation is made at time $T_{k-1} + 1$, since $T_k > T_{k-1} + 1$. Notice, further, that $d_k > 0$ $\forall k$, since $T_k > T_{k-1}$ and that if T_k is measured in whole units for each k, $d_k \varepsilon Z^+$.

Further notation will be defined as necessary.

4.1.2 BACKGROUND TO UNEQUALLY-SPACED TIME SERIES MODELLING

The topic of unequally-spaced measurements in time is one which has attracted relatively little attention in the time series literature. One (tentative) explanation for this, perhaps, is that the majority of actual time series applications for which much of the existing theory has been developed have involved regularly spaced data, with very few 'missing' observations. For instance, in data arising from economic sources which inspired many of the early ideas in time series analysis (see, for example, Box and Jenkins 1970, Wheelwright and Makridakis 1973, Montgomery and Johnson 1976) unbroken weekly, monthly or quarterly series were very much the norm. However, in recent years, as the scope of applications has increased, a number of authors have suggested that this is not invariably the case and that, for a variety of reasons, economic and commercial time series may well contain missing observations or be unequally-spaced (Doran 1974, Robinson 1977, Harvey and Pierse 1984). Certainly, this is so, in the strict sense, where 'monthly' data is concerned, owing to the differing lengths of calendar months.

Although, historically, literature on unequally-spaced time series is sparse, it appears that the subject has generated a good deal more interest during the last ten years or so. Among the earliest papers were those of Jones (1962) and Parzen (1963), who both deal with the implementation of spectral analysis in the presence of missing observations and, in particular, periodically missing observations, e.g. daily measurements, but with no measurements obtainable at weekends. This work was extended by Clinger and Van Ness (1976), who develop a general cyclic sampling scheme with observations at:

$$T_1, \ldots, T_k, T_1 + m, \ldots, T_k + m, T_1 + 2m, \ldots, T_k + 2m, \ldots$$

In this case, too, the spectral approach to time series analysis has been adopted.

Spectral approaches were also used by Scheinok (1965) and Bloomfield (1970) although, instead of periodic sampling, these authors assumed that the presence or absence of an observation was governed by a random mechanism, so that missing time-points arise probabilistically.

Other work on missing observations in time series includes that of Sargan and Drettakis (1974) and Dunsmuir and Robinson (1981) on the accommodation of missing values in ARMA models. Sargan and Drettakis (1974) use a state-space approach estimating the missing observations, whereas Dunsmuir and Robinson (1981) apply frequency-domain methodology to the problem of estimating spectral densities in the presence of missing data.

In (1980),Jones presented a method for calculating the exact likelihoods of ARMA models in the presence of missing observations, using a state-space approach involving the Kalman Filter. This work, complemented by the recent advances in maximum likelihood estimation techniques (see, for example, Harvey and Phillips 1979), constitutes a breakthrough in the area of time series analysis with missing data and seems to have stimulated a renewed interest in the whole topic (Ansley and Kohn 1983, Harvey and Pierse 1984). Indeed, it is the basic principle from Jones (1980) that will be utilized in this thesis, as outlined in the following section.

Ideas for modelling time series with generally unequally-spaced observations, rather than purely missing values, stem from the literature on aliassing (see, for example, Shapiro and Silverman 1960, Loynes 1969). There was very little activity in this area up until the late 1970's when Clinger and Van Ness (1976) re-introduced the periodic sampling concept. Once again, a spectral approach was adopted by many of the authors who tackled the problem (Robinson 1977,1980). The general strategy has been to fit a continuous-time model to a discrete set of observations which are generally unequally-spaced. In conjunction with the spectral analysis approach of Robinson (1980), some authors adopted a time domain representation of the continuous modelling of discrete time series (Phadke and Wu 1974, Jones 1981, Kitagawa 1984).

A detailed account of continuous modelling for discrete observations, using a state-space approach, can be found in Jones (1985), in which ARIMA models are the main focus of attention. An application of these techniques can be found in Jones and Tryon (1986).

4.2 GENERAL FRAMEWORK

4.2.1 UNDERLYING ASSUMPTIONS

We note, first of all, that the adoption of a recursive, time-domain approach to the incorporation of unequally-spaced data is essential in the medical monitoring context since we need to be able to detect the timepoints of model discontinuity. The dynamic linear model provides an ideal framework for this strategy. We recall, from (2.1) and (2.2) that the dynamic linear model involves two components: the observation equation and the system equation. Following Jones (1980) and Harvey and Pierse (1984), the essence of the state-space approach to missing observations is to bypass the observation-updating steps of the Kalman Filter recursion, and to complete the system-update steps by replacing the 'missing' estimates of $\underset{\sim}{\theta}$ and $\underset{\sim}{W}$ by their most recent estimates. In terms of the dynamic linear model, therefore, we disregard the observation equation (2.1), subject to the following provisos.

P4.1. The observation error at timepoint k, ε_k, is not dependent upon the length of the sampling interval.

P4.2. The length of the sampling interval is not dependent upon the magnitude, $\underset{\sim}{\theta}_k$, of $\underset{\sim}{\theta}$ at timepoint k.

Generally, it seems to us that P4.1 will be a reasonable assumption; if the interval between measurements is very large, however, there may be a chance of larger observation errors due, for example, to a lack of practise in the measurement procedures. We will assume that any lack-of-practise effects that might exist are negligible.

Assumption P4.2 implies that the level of $\underset{\sim}{\theta}_k$ does not determine the length of the sampling interval. This assumption may appear to be slightly questionable, in the medical setting, since the levels of $\underset{\sim}{\theta}$ represent the well-being of the patient. However, d_k is only affected by $\underset{\sim}{\theta}_{k-1}, \underset{\sim}{\theta}_{k-2}, \ldots$, etc. and <u>not</u> by $\underset{\sim}{\theta}_k$.

P4.2 also provides another insight into the accommodation of missing data. If the observation at timepoint k is unavailable, its associated error component is also missing. However, this error component is <u>unknown</u> even if the observation is present. Therefore, for those models where the error structure exhibits time-dependent behaviour (e.g. MA models) we will imagine that the error exists even if an observation has not been made, and we merely replace the variance of that error by the most recent variance estimate. If a measurement has been made, we can update this variance first and then use the updated estimate (see Section 4.3.5).

We also choose to adopt discrete-time, rather than continuous-time, state-space formulations, in direct contrast to the views of Jones (1985). The main motivation for our choice of approach concerns the modelling of between-timepoint behaviour. A continuous model implies that the behaviour between timepoints is (in some sense) smooth and akin to the timepoint-to-timepoint model. For instance, if the linear growth model is adequate to describe the behaviour of daily observations then a continuous form would imply that the within-day behaviour is also linear. However, in the medical setting, we will sometimes wish to use an alternative within-day sub-model, for instance the sinusoidal model (see Section 2.3.3). In general, the micro-kinetics governing the disease process may well be extremely complex and an approximation by the 'macro'-model is unlikely to represent the characteristics of these mechanisms.

We see, therefore, that our ability to incorporate sub-models, and to build up a global model from these sub-models would be severely restricted by the use of continuous models. However, there are genuine interpretational problems for some of the global models we shall consider when fractions of time intervals are permitted. Consider, for example, the case of a discrete AR(1) process with a negative autoregressive parameter. What model should we invoke if the time interval is suddenly halved?

In the light of these considerations, we shall adopt a discrete-time, recursive approach to the incorporation of unequally-spaced observations, and we shall make the further assumption that:

P4.3. We can conceive of a basic 'time unit' representing the smallest conceivable interval between observations, so that <u>no</u>

fractions of this basic interval are possible. Therefore, we have $d_k \in Z^+$.

NOTE: By the simple addition of a growth component, etc. to the standard DLM representations of ARMA models (see Section 2.3.5) we may avoid the complications associated with the standard approaches to ARIMA models and unequally-spaced timepoints (namely, that first differences, etc. are difficult to form when some of the observations are not present), but will still allow for non-stationarities in the steady-state time series.

4.2.2 THE DLM FOR UNEQUALLY-SPACED OBSERVATIONS

In this section we provide a general formulation of the dynamic linear model when the time series is obtained at unequally-spaced timepoints.

We rewrite the DLM described by (2.1) and (2.2) in the form:

$$\underset{\sim}{y}_k = \underset{\sim}{H}_{T_k} \underset{\sim}{\theta}_k + \underset{\sim}{\varepsilon}_{T_k} \tag{4.1}$$

$$\underset{\sim}{\theta}_k = \underset{\sim}{G}_k \, \underset{\sim}{\theta}_{k-1} + \underset{\sim}{\omega}_k \tag{4.2}$$

where

$\underset{\sim}{y}_k$ is the kth observation vector (made at time T_k)
$\underset{\sim}{\theta}_k$ is the vector of system parameters at time T_k
$\underset{\sim}{H}_{T_k}$ is a regression matrix, fully specified at time T_k
$\underset{\sim}{G}_k$ is a transition matrix, dependent on $d_k = T_k - T_{k-1}$
$\underset{\sim}{\varepsilon}_{T_k}, \underset{\sim}{\omega}_k$ are zero-mean, random vectors associated with time T_k

(4.3)

We make the following assumptions:

P4.4. $\underset{\sim}{\varepsilon}_{T_k}$ is independent of $\underset{\sim}{\varepsilon}_{T_j}$, $\forall j \neq k$

P4.5. $\underset{\sim}{\varepsilon}_{T_k}, \underset{\sim}{\omega}_{T_k}$ are independent of $\underset{\sim}{\theta}_{k-1}$, $\forall k$ given $\underset{\sim}{D}_{k-1}$ (i.e. the past k - 1 observations)

P4.6. $\underset{\sim}{\varepsilon}_{T_k}$ is independent of $\underset{\sim}{\omega}_k$, $\forall k$

P4.7. $\underset{\sim}{\varepsilon}_{T_k} \sim N(\underset{\sim}{0}, \lambda^{-1}\underset{\sim}{R}_{\varepsilon})$; $\underset{\sim}{\omega}_k \sim N(\underset{\sim}{0}, \lambda^{-1}\underset{\sim}{R}_{\omega}(k))$.

Notice that $\underset{\sim}{\varepsilon}_{T_k}$ is dependent only on the timepoint k and not on the time interval d_k, so its variance is equivalent to that given for the equally-spaced DLM (Equation (2.11)). Following on from the discussions of the previous section, the form of $\underset{\sim}{\omega}_k$ has yet to be determined from the system update for missing observations, and therefore its variance may not be equivalent to the equally-spaced case, in which $\mathrm{var}(\underset{\sim}{\omega}_t) = \lambda^{-1}\underset{\sim}{R}_{\omega}$.

Since (4.1) is completely specified, except for $\underset{\sim}{\theta}_k$, we deal only with the system equations (4.2), in order to find the form of $\underset{\sim}{\theta}_k$. Two distinct possibilities arise in the equally-spaced formulation: (a) $\underset{\sim}{G}$ is not time-dependent, as in (2.2), or (b) $\underset{\sim}{G}$ is time-dependent. We shall examine each of these situations in turn.

NOTE: When an observation is made at time T_k, we define $\underset{\sim}{\theta}_k \equiv \underset{\sim}{\theta}_{T_k}$.

4.2.2.1: Constant Transition Matrix. For equally-spaced (unit) intervals between successive observations, we restate (2.2):

$$\underset{\sim}{\theta}_t = \underset{\sim}{G}\underset{\sim}{\theta}_{t-1} + \underset{\sim}{\omega}_t \tag{4.4}$$

LEMMA 4.1: When the interval between successive observations is $d_k \in Z^+$ units, we may write:

$$\underset{\sim}{\theta}_k = \underset{\sim}{G}^{d_k}\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k-t}\,\underset{\sim}{\omega}_t \tag{4.5}$$

PROOF: Assume (4.5) to be true for k = j and assume that the observation actually occurs at time $T_j = T_k + 1$. Then, according to (4.4), we may write:

$$\underset{\sim}{\theta}_j = \underset{\sim}{G}\underset{\sim}{\theta}_{T_j-1} + \underset{\sim}{\omega}_{T_j}$$

$$= \underset{\sim}{G}\underset{\sim}{\theta}_k + \underset{\sim}{\omega}_{T_k+1} \qquad \text{(since } T_k = T_j - 1\text{)}$$

$$= \underset{\sim}{G}[\underset{\sim}{G}^{d_k}\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t] + \underset{\sim}{\omega}_{T_k+1} \qquad \text{(using (4.5))}$$

$$= \underset{\sim}{G}^{d_k+1}\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k+1-t}\underset{\sim}{\omega}_t + \underset{\sim}{\omega}_{T_k+1}$$

$$= \underset{\sim}{G}^{d_k+1}\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_k+1} \underset{\sim}{G}^{T_k+1-t}\,\underset{\sim}{\omega}_t \qquad \text{(since } \underset{\sim}{G}^o = \underset{\sim}{I}\text{)}$$

Since $T_j = T_k + 1$, and $d_j = d_k + 1$ we have:

$$\underset{\sim}{\theta}_j = \underset{\sim}{G}^{d_j}\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_j} \underset{\sim}{G}^{T_j-t}\underset{\sim}{\omega}_t.$$

Therefore if (4.5) is true for d_k it is also true for $d_k + 1$; but (4.5) holds for $d_k = 1$, since (4.5) then reduces to:

$$\underset{\sim}{\theta}_k = \underset{\sim}{G}\underset{\sim}{\theta}_{k-1} + \underset{\sim}{\omega}_{T_k} \quad \text{(c.f. (4.4))}$$

Hence (4.5) holds for any integer $d_k \geq 1$ by induction.

In terms of (4.2), we have

$$\underset{\sim}{G}_k = \underset{\sim}{G}^{d_k}. \tag{4.6}$$

VARIANCE CALCULATIONS

Let

$$\underset{\sim}{\omega}_k = \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k - t} \underset{\sim}{\omega}_t \ . \tag{4.7}$$

In order to calculate the variance of $\underset{\sim}{\omega}_k$ we consider two distinct situations:

(a) $\underset{\sim}{\omega}_t$ is independent of $\underset{\sim}{\omega}_s$, $\forall s \neq t$

(b) $\underset{\sim}{\omega}_t$ is not independent of $\underset{\sim}{\omega}_s$, for some $s \neq t$.

(a) Independent Errors. From (2.11) we have $\text{var}(\underset{\sim}{\omega}_t) = \lambda^{-1}\underset{\sim}{R}_\omega$, when the interval between observations is one unit.

If the interval between observations is d_k units then:

$$\text{var}(\underset{\sim}{\omega}_k) = \lambda^{-1}\underset{\sim}{R}_\omega(k) = \lambda^{-1} \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k - t} \underset{\sim}{R}_\omega (\underset{\sim}{G}^{T_k - t})^T . \tag{4.8}$$

Equation (4.8) follows directly from (4.7) assuming independence in the $\underset{\sim}{\omega}_t$ sequence. Notice that we have made use of the idea put forward in the previous section, in that we are proceeding as if these errors exist even when an observation does not, and have replaced their variance by the most recent estimate (in our case, the estimate of λ^{-1} made at timepoint $k - 1$).

(b) Error Dependence. For the case where the $\underset{\sim}{\omega}_t$'s are not all independent:

$$\text{var}(\underset{\sim}{\omega}_k) = \lambda^{-1}\underset{\sim}{R}_{\omega}(k)$$

$$= \lambda^{-1} \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k-t}\underset{\sim}{R}_{\omega}(\underset{\sim}{G}^{T_k-t})^T + \sum_{i=T_{k-1}+1}^{T_k} \sum_{\substack{j=T_{k-1}+1 \\ (i\neq j)}}^{T_k} \text{cov}(\underset{\sim}{G}^{T_k-i}\underset{\sim}{\omega}_i, \underset{\sim}{G}^{T_k-j}\underset{\sim}{\omega}_j) \tag{4.9}$$

This, too, follows directly from (4.7) when the $\underset{\sim}{\omega}_t$'s are not all independent. The exact form of the covariance term depends upon the extent of $\underset{\sim}{\omega}$ dependence and the precise structure of $\underset{\sim}{G}$.

4.2.2.2: Time-Dependent Transition Matrix. For equally-spaced (unit) intervals between successive observations, we have:

$$\underset{\sim}{\theta}_t = \underset{\sim}{G}_t\underset{\sim}{\theta}_{t-1} + \underset{\sim}{\omega}_t \tag{4.10}$$

where the transition matrix, $\underset{\sim}{G}_t$, is now permitted to be dependent on time.

LEMMA 4.2: When the interval between successive observations is d_k units, we may write:

$$\underset{\sim}{\theta}_k = \left(\prod_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}_t\right)\underset{\sim}{\theta}_{k-1} + \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} \left(\prod_{i=t+1}^{T_k} \underset{\sim}{G}_i\right)\underset{\sim}{\omega}_t + \underset{\sim}{\omega}_{T_k}. \tag{4.11}$$

PROOF: Assume (4.11) to be true and assume that the observation actually occurs at time $T_j = T_k + 1$. Then, according to (4.10), we may write

$$\underset{\sim}{\theta}_j = \underset{\sim}{G}_{T_j}\underset{\sim}{\theta}_{T_j-1} + \underset{\sim}{\omega}_{T_j}$$

$$= \underset{\sim}{G}_{T_k+1}\underset{\sim}{\theta}_k + \underset{\sim}{\omega}_{T_k+1} \qquad \text{(since } T_k = T_j - 1\text{)}$$

$$= \underset{\sim}{G}_{T_k+1}\left\{\left(\prod_{t=T_{k-1}+1}^{T_k}\underset{\sim}{G}_t\right)\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_k-1}\left(\prod_{i=t+1}^{T_k}\underset{\sim}{G}_i\right)\underset{\sim}{\omega}_t + \underset{\sim}{\omega}_{T_k}\right\} + \underset{\sim}{\omega}_{T_k+1}$$

(using (4.11))

$$= \left(\prod_{t=T_{k-1}+1}^{T_k+1}\underset{\sim}{G}_t\right)\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_k-1}\left(\prod_{i=t+1}^{T_k+1}\underset{\sim}{G}_i\right)\underset{\sim}{\omega}_t + \underset{\sim}{G}_{T_k+1}\underset{\sim}{\omega}_{T_k} + \underset{\sim}{\omega}_{T_k+1}$$

$$= \left(\prod_{t=T_{k-1}+1}^{T_j}\underset{\sim}{G}_t\right)\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_j-1}\left(\prod_{i=t+1}^{T_j}\underset{\sim}{G}_i\right)\underset{\sim}{\omega}_t + \underset{\sim}{\omega}_{T_j} .$$

Therefore, if (4.11) is true for d_k it is also true for $d_k + 1$; but (4.11) holds for $d_k = 1$ since (4.11) then reduces to

$$\underset{\sim}{\theta}_k = \underset{\sim}{G}_{T_k}\underset{\sim}{\theta}_{k-1} + \underset{\sim}{\omega}_{T_k} \qquad \text{(c.f. (4.10))}.$$

Hence (4.11) holds for any integer $d_k \geq 1$ by induction. In terms of (4.2), we have:

$$\underset{\sim}{G}_k = \prod_{t=T_{k-1}+1}^{T_k}\underset{\sim}{G}_t \tag{4.12}$$

VARIANCE CALCULATIONS: Let

$$\underset{\sim}{\omega}_k = \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1}\left(\prod_{i=t+1}^{T_k}\underset{\sim}{G}_i\right)\underset{\sim}{\omega}_t + \underset{\sim}{\omega}_{T_k} \tag{4.13}$$

The variances follow immediately from (4.13) for both the case of independent and dependent $\underset{\sim}{\omega}_t$'s.

(a) Independent Errors

$$\mathrm{var}(\underset{\sim}{\omega}_k) = \lambda^{-1}\underset{\sim}{R}_{\omega}(k)$$

$$= \lambda^{-1}\left(\underset{\sim}{R}_{\omega} + \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} \left(\prod_{i=t+1}^{T_k} \underset{\sim}{G}_i\right) \underset{\sim}{R}_{\omega} \left(\prod_{i=t+1}^{T_k} \underset{\sim}{G}_i\right)^T\right) \tag{4.14}$$

(b) Error Dependence

$$\mathrm{var}(\underset{\sim}{\omega}_k) = \lambda^{-1}\underset{\sim}{R}_{\omega}(k)$$

$$= \lambda^{-1}\left(\underset{\sim}{R}_{\omega} + \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} \left(\prod_{i=t+1}^{T_k} \underset{\sim}{G}_i\right) \underset{\sim}{R}_{\omega} \left(\prod_{i=t+1}^{T_k} \underset{\sim}{G}_i\right)^T\right)$$

$$+ \sum_{\substack{s=T_{k-1}+1 \\ (s\neq t,\, d_k>1)}}^{T_k-1} \; \sum_{t=T_{k-1}+1}^{T_k-1} \mathrm{cov}\left(\left(\prod_{m=s+1}^{T_k} \underset{\sim}{G}_m\right)\underset{\sim}{\omega}_s, \left(\prod_{n=t+1}^{T_k} \underset{\sim}{G}_n\right)\underset{\sim}{\omega}_t\right)$$

$$+ \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} \mathrm{cov}\left(\underset{\sim}{\omega}_{T_k}, \left(\prod_{n=t+1}^{T_k} \underset{\sim}{G}_n\right)\underset{\sim}{\omega}_t\right)$$

$$+ \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} \mathrm{cov}\left(\left(\prod_{n=t+1}^{T_k} \underset{\sim}{G}_n\right)\underset{\sim}{\omega}_t, \underset{\sim}{\omega}_{T_k}\right). \tag{4.15}$$

NOTES ON KALMAN FILTER RECURSION: The above is achieved by minor modifications to the Kalman Filter recursions given by (3.16). We replace:

(i) $\underset{\sim}{G}$ by $\underset{\sim}{G}_k$,

and (ii) $\underset{\sim}{R}_{\omega}^{(j)}$ by $\underset{\sim}{R}_{\omega}^{(j)}(k)$.

4.3 MODEL REFORMULATION

Using the general structure defined in Section 4.2 we now describe in detail a number of special cases; in particular, the univariate models outlined in Section 2.3 and extended in Section 3.3. We note that, for each of these models, the transition matrix, $\underset{\sim}{G}$, is time-independent for the equally-spaced, unit-interval case (henceforth referred to as the equally-spaced model). Therefore we use the result of Lemma 4.1 throughout in order to formulate the unequally-spaced model.

In the derivation of unequally-spaced growth models, the following identities will prove to be of use:

(i) $$\sum_{t=1}^{n} 1 = n \tag{4.16}$$

(ii) $$\sum_{t=1}^{n} t = \frac{n(n+1)}{2} \tag{4.17}$$

(iii) $$\sum_{t=1}^{n} t^2 = \frac{n(n+1)(2n+1)}{6} \tag{4.18}$$

(iv) $$\sum_{t=1}^{n} t^3 = \frac{n^2(n+1)^2}{4} \tag{4.19}$$

(v) $$\sum_{t=1}^{n} t^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30} \tag{4.20}$$

4.3.1 LINEAR GROWTH

Restatement of the Equally-Spaced Model

$$\underset{\sim}{\theta}_t = \underset{\sim}{G}\underset{\sim}{\theta}_{t-1} + \underset{\sim}{\omega}_t$$

i.e.

$$\begin{pmatrix}\mu_t \\ \beta_t\end{pmatrix} = \begin{pmatrix}1 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1} \\ \beta_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t + \delta\beta_t \\ \delta\beta_t\end{pmatrix} \qquad (4.21)$$

Derivation of the Unequally-Spaced Model

$$\underset{\sim}{\theta}_k = \underset{\sim}{G}^{d_k}\underset{\sim}{\theta}_{k-1} + \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t$$

i.e.

$$\begin{pmatrix}\mu_k \\ \beta_k\end{pmatrix} = \begin{pmatrix}1 & 1 \\ 0 & 1\end{pmatrix}^{d_k}\begin{pmatrix}\mu_{k-1} \\ \beta_{k-1}\end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k}\begin{pmatrix}1 & 1 \\ 0 & 1\end{pmatrix}^{T_k-t}\begin{pmatrix}\delta\mu_t + \delta\beta_t \\ \delta\beta_t\end{pmatrix} \qquad (4.22)$$

Now

$$\underset{\sim}{G}^i = \begin{pmatrix}1 & i \\ 0 & 1\end{pmatrix}$$

since

$$\underset{\sim}{G}^{i+1} = \underset{\sim}{G}\underset{\sim}{G}^i = \begin{pmatrix}1 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & i \\ 0 & 1\end{pmatrix} = \begin{pmatrix}1 & i+1 \\ 0 & 1\end{pmatrix}$$

So

$$\begin{pmatrix}\mu_k \\ \\ \beta_k\end{pmatrix} = \begin{pmatrix}1 & d_k \\ \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{k-1} \\ \\ \beta_{k-1}\end{pmatrix} + \begin{pmatrix}\displaystyle\sum_{t=T_{k-1}+1}^{T_k}\delta\mu_t + \sum_{t=T_{k-1}+1}^{T_k}(T_k - t + 1)\delta\beta_t \\ \displaystyle\sum_{t=T_{k-1}+1}^{T_k}\delta\beta_t\end{pmatrix} \qquad (4.23)$$

(Notice that β_k is interpreted as the <u>slope</u> at timepoint k; β_k can only be interpreted as the increment if the current interval is one unit.)

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu + R_\beta & R_\beta \\ R_\beta & R_\beta \end{pmatrix} \tag{4.24}$$

$$\underset{\sim}{R}_\omega(k) = \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k-t} \underset{\sim}{R}_\omega (\underset{\sim}{G}^{T_k-t})^T$$

$$= \sum_{t=T_{k-1}+1}^{T_k} \begin{pmatrix} 1 & T_k - t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_\mu + R_\beta & R_\beta \\ R_\beta & R_\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ T_k - t & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{t=T_{k-1}+1}^{T_k} R_\mu + \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1)^2 R_\beta & \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_\beta \\ \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_\beta & \sum_{t=T_{k-1}+1}^{T_k} R_\beta \end{pmatrix}$$

$$= \begin{pmatrix} d_k R_\mu + \frac{d_k(d_k + 1)(2d_k + 1)R_\beta}{6} & \frac{d_k(d_k + 1)R_\beta}{2} \\ \frac{d_k(d_k + 1)R_\beta}{2} & d_k R_\beta \end{pmatrix} , \tag{4.25}$$

(using (4.16), (4.17) and (4.18)).

4.3.2 QUADRATIC GROWTH

Restatement of the Equally-Spaced Model

$$\begin{pmatrix}\mu_t\\ \beta_t\\ \gamma_t\end{pmatrix} = \begin{pmatrix}1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1}\\ \beta_{t-1}\\ \gamma_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t + \delta\beta_t + \delta\gamma_t\\ \delta\beta_t + \delta\gamma_t\\ \delta\gamma_t\end{pmatrix} \tag{4.26}$$

Derivation of the Unequally-Spaced Model

$$\begin{pmatrix}\mu_k\\ \beta_k\\ \gamma_k\end{pmatrix} = \begin{pmatrix}1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{pmatrix}^{d_k}\begin{pmatrix}\mu_{k-1}\\ \beta_{k-1}\\ \gamma_{k-1}\end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k}\begin{pmatrix}1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{pmatrix}^{T_k-t}\begin{pmatrix}\delta\mu_t + \delta\beta_t + \delta\gamma_t\\ \delta\beta_t + \delta\gamma_t\\ \delta\gamma_t\end{pmatrix} \tag{4.27}$$

Now

$$\underset{\sim}{G}^i = \begin{pmatrix}1 & i & \frac{i(i+1)}{2}\\ 0 & 1 & i\\ 0 & 0 & 1\end{pmatrix}$$

since

$$\underset{\sim}{G}^{i+1} = \underset{\sim}{G}\underset{\sim}{G}^i = \begin{pmatrix}1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & i & \frac{i(i+1)}{2}\\ 0 & 1 & i\\ 0 & 0 & 1\end{pmatrix}$$

$$= \begin{pmatrix}1 & i+1 & \frac{(i+1)(i+2)}{2}\\ 0 & 1 & i+1\\ 0 & 0 & 1\end{pmatrix}$$

So

$$\begin{pmatrix}\mu_k\\ \beta_k\\ \gamma_k\end{pmatrix} = \begin{pmatrix}1 & d_k & \frac{d_k(d_k+1)}{2}\\ 0 & 1 & d_k\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{k-1}\\ \beta_{k-1}\\ \gamma_{k-1}\end{pmatrix}$$

$$+ \begin{pmatrix}\sum_{t=T_{k-1}+1}^{T_k} \delta\mu_t + \sum_{t=T_{k-1}+1}^{T_k} (T_k-t+1)\delta\beta_t + \sum_{t=T_{k-1}+1}^{T_k} \frac{(T_k-t+1)(T_k-t+2)\delta\gamma_t}{2}\\ \sum_{t=T_{k-1}+1}^{T_k} \delta\beta_t + \sum_{t=T_{k-1}+1}^{T_k} (T_k-t+1)\delta\gamma_t\\ \sum_{t=T_{k-1}+1}^{T_k} \delta\gamma_t\end{pmatrix}$$

(4.28)

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_\omega = \begin{pmatrix}R_\mu + R_\beta + R_\gamma & R_\beta + R_\gamma & R_\gamma\\ R_\beta + R_\gamma & R_\beta + R_\gamma & R_\gamma\\ R_\gamma & R_\gamma & R_\gamma\end{pmatrix} \tag{4.29}$$

$$\underset{\sim}{R}_\omega(k) = \sum_{t=T_{k-1}+1}^{T_k} \begin{pmatrix}1 & T_k-t & \frac{(T_k-t)(T_k-t+1)}{2}\\ 0 & 1 & T_k-t\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}R_\mu + R_\beta + R_\gamma & R_\beta + R_\gamma & R_\gamma\\ R_\beta + R_\gamma & R_\beta + R_\gamma & R_\gamma\\ R_\gamma & R_\gamma & R_\gamma\end{pmatrix}$$

$$\times \begin{pmatrix}1 & 0 & 0\\ T_k - t & 1 & 0\\ \frac{(T_k-t)(T_k-t+1)}{2} & T_k-t & 1\end{pmatrix}$$

$$= \begin{pmatrix}W_{11} & W_{12} & W_{13}\\ W_{12} & W_{22} & W_{23}\\ W_{13} & W_{23} & W_{33}\end{pmatrix}, \text{ say}$$

where

$$W_{11} = \sum_{t=T_{k-1}+1}^{T_k} R_\mu + \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1)^2 R_\beta$$

$$+ \sum_{t=T_{k-1}+1}^{T_k} \tfrac{1}{4}(T_k - t + 1)^2 (T_k - t + 2)^2 R_\gamma$$

$$W_{12} = \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_\beta + \sum_{t=T_{k-1}+1}^{T_k} \tfrac{1}{2}(T_k - t + 1)^2 (T_k - t + 2) R_\gamma$$

$$W_{13} = \sum_{t=T_{k-1}+1}^{T_k} \tfrac{1}{2}(T_k - t + 1)(T_k - t + 2) R_\gamma$$

$$W_{22} = \sum_{t=T_{k-1}+1}^{T_k} R_\beta + \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1)^2 R_\gamma$$

$$W_{23} = \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_\gamma$$

and

$$W_{33} = \sum_{t=T_{k-1}+1}^{T_k} R_\gamma .$$

Using (4.16) to (4.20) we can see that:

$$\text{(i)} \quad \sum_{t=T_{k-1}+1}^{T_k} \frac{(T_k - t + 1)(T_k - t + 2)}{2} = \tfrac{1}{2}\left[\frac{d_k(d_k + 1)(2d_k + 1)}{6} + \frac{d_k(d_k + 1)}{2}\right] = \frac{d_k(d_k + 1)(d_k + 2)}{6}$$

(ii) $$\sum_{t=T_{k-1}+1}^{T_k} \frac{(T_k - t + 1)^2(T_k - t + 2)}{2} = \tfrac{1}{2}\left[\frac{d_k^2(d_k + 1)^2}{4} + \frac{d_k(d_k + 1)(2d_k + 1)}{6}\right] = \frac{d_k(d_k + 1)(d_k + 2)(3d_k + 1)}{24}$$

(iii) $$\sum_{t=T_{k-1}+1}^{T_k} \frac{(T_k - t + 1)^2(T_k - t + 2)^2}{4}$$

$$= \tfrac{1}{4}\left[\frac{d_k(d_k + 1)(2d_k + 1)(3d_k^2 + 3d_k - 1)}{30} + \frac{d_k^2(d_k + 1)^2}{2} + \frac{d_k(d_k + 1)(2d_k + 1)}{6}\right]$$

$$= \frac{d_k(d_k + 1)(d_k + 2)(3d_k^2 + 6d_k + 1)}{60}$$

So

$$\left.\begin{aligned}
W_{11} &= d_k R_\mu + \frac{1}{6} d_k(d_k + 1)(2d_k + 1)R_\beta + \frac{1}{60} d_k(d_k + 1)(d_k + 2)(3d_k^2 + 6d_k + 1)R_\gamma \\
W_{12} &= \tfrac{1}{2}d_k(d_k + 1)R_\beta + \frac{1}{24} d_k(d_k + 1)(d_k + 2)(3d_k + 1)R_\gamma \\
W_{13} &= \frac{1}{6} d_k(d_k + 1)(d_k + 2)R_\gamma \\
W_{22} &= d_k R_\beta + \frac{1}{6} d_k(d_k + 1)(2d_k + 1)R_\gamma \\
W_{23} &= \tfrac{1}{2}d_k(d_k + 1)R_\gamma \\
&\text{and} \\
W_{33} &= d_k R_\gamma
\end{aligned}\right\} \quad (4.30)$$

4.3.3 SINUSOIDAL MODEL

Restatement of the Equally-Spaced Model

$$\begin{pmatrix} \mu_t \\ \alpha_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \alpha_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_t \\ \delta\alpha_t \end{pmatrix} \tag{4.31}$$

Derivation of the Unequally-Spaced Model

Clearly, $\underset{\sim}{G}^i = \underset{\sim}{I}^i = \underset{\sim}{I}$ and so

$$\begin{pmatrix} \mu_k \\ \alpha_k \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{k-1} \\ \alpha_{k-1} \end{pmatrix} + \begin{pmatrix} \sum_{t=T_{k-1}+1}^{T_k} \delta\mu_t \\ \sum_{t=T_{k-1}+1}^{T_k} \delta\alpha_t \end{pmatrix} \tag{4.32}$$

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu & 0 \\ 0 & R_\alpha \end{pmatrix} \tag{4.33}$$

Therefore,

$$\underset{\sim}{R}_\omega(k) = \begin{pmatrix} d_k R_\mu & 0 \\ 0 & d_k R_\alpha \end{pmatrix} \tag{4.34}$$

(using (4.16)).

4.3.4 SINUSOIDAL MODEL WITH LINEAR GROWTH

Restatement of the Equally-Spaced Model

$$\begin{pmatrix} \mu_t \\ \beta_t \\ \alpha_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \alpha_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_t + \delta\beta_t \\ \delta\beta_t \\ \delta\alpha_t \end{pmatrix} \tag{4.35}$$

Derivation of the Unequally-Spaced Model

$$\begin{pmatrix}\mu_k\\ \beta_k\\ \alpha_k\end{pmatrix} = \begin{pmatrix}1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}^{d_k}\begin{pmatrix}\mu_{k-1}\\ \beta_{k-1}\\ \alpha_{k-1}\end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k}\begin{pmatrix}1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}^{T_k - t}\begin{pmatrix}\delta\mu_t + \delta\beta_t\\ \delta\beta_t\\ \delta\alpha_t\end{pmatrix}$$

Now

$$\underset{\sim}{G}^i = \begin{pmatrix}1 & i & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

since

$$\underset{\sim}{G}^{i+1} = \underset{\sim}{G}\underset{\sim}{G}^i = \begin{pmatrix}1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & i & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} = \begin{pmatrix}1 & i+1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

So

$$\begin{pmatrix}\mu_k\\ \beta_k\\ \alpha_k\end{pmatrix} = \begin{pmatrix}1 & d_k & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{k-1}\\ \beta_{k-1}\\ \alpha_{k-1}\end{pmatrix} + \begin{pmatrix}\sum_{t=T_{k-1}+1}^{T_k}\delta\mu_t + \sum_{t=T_{k-1}+1}^{T_k}(T_k - t + 1)\delta\beta_t\\ \sum_{t=T_{k-1}+1}^{T_k}\delta\beta_t\\ \sum_{t=T_{k-1}+1}^{T_k}\delta\alpha_t\end{pmatrix} \tag{4.36}$$

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_\omega = \begin{pmatrix}R_\mu + R_\beta & R_\beta & 0\\ R_\beta & R_\beta & 0\\ 0 & 0 & R_\alpha\end{pmatrix} \tag{4.37}$$

$$\underset{\sim}{R}_\omega(k) = \sum_{t=T_{k-1}+1}^{T_k}\begin{pmatrix}1 & T_k - t & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}R_\mu + R_\beta & R_\beta & 0\\ R_\beta & R_\beta & 0\\ 0 & 0 & R_\alpha\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ T_k - t & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

$$\underset{\sim}{R}_\omega(k) = \begin{pmatrix} \sum_{t=T_{k-1}+1}^{T_k} R_\mu + \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1)^2 R_\beta & \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_\beta & 0 \\ \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_\beta & \sum_{t=T_{k-1}+1}^{T_k} R_\beta & 0 \\ 0 & 0 & \sum_{t=T_{k-1}+1}^{T_k} R_\alpha \end{pmatrix}$$

and, using the results of (4.16) to (4.18), we see that:

$$\underset{\sim}{R}_\omega(k) = \begin{pmatrix} d_k R_\mu + \frac{d_k(d_k + 1)(2d_k + 1)R_\beta}{6} & \frac{d_k(d_k + 1)R_\beta}{2} & 0 \\ \frac{d_k(d_k + 1)R_\beta}{2} & d_k R_\beta & 0 \\ 0 & 0 & d_k R_\alpha \end{pmatrix} \quad (4.38)$$

4.3.5. ARMA MODELS

4.3.5.1: *AR(1)*

Restatement of the Equally-Spaced Model

$$\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} = \begin{pmatrix} \phi & 1 - \phi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \nu_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_t + \delta\nu_t \\ \delta\nu_t \end{pmatrix} \quad (4.39)$$

Derivation of the Unequally-Spaced Model

$$\begin{pmatrix} \mu_k \\ \nu_k \end{pmatrix} = \begin{pmatrix} \phi & 1 - \phi \\ 0 & 1 \end{pmatrix}^{d_k} \begin{pmatrix} \mu_{k-1} \\ \nu_{k-1} \end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k} \begin{pmatrix} \phi & 1 - \phi \\ 0 & 1 \end{pmatrix}^{T_k - t} \begin{pmatrix} \delta\mu_t + \delta\nu_t \\ \delta\nu_t \end{pmatrix}$$

Now

$$\underset{\sim}{G}^i = \begin{pmatrix} \phi^i & 1 - \phi^i \\ 0 & 1 \end{pmatrix}$$

since

$$\underset{\sim}{G}^{i+1} = \underset{\sim\sim}{GG}^i = \begin{pmatrix} \phi & 1 - \phi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \phi^i & 1 - \phi^i \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \phi^{i+1} & 1 - \phi^{i+1} \\ 0 & 1 \end{pmatrix}$$

So

$$\begin{pmatrix} \mu_k \\ \nu_k \end{pmatrix} = \begin{pmatrix} \phi^{d_k} & 1 - \phi^{d_k} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{k-1} \\ \nu_{k-1} \end{pmatrix} + \begin{pmatrix} \sum_{t=T_{k-1}+1}^{T_k} \phi^{T_k - t} \delta\mu_t + \sum_{t=T_{k-1}+1}^{T_k} \delta\nu_t \\ \sum_{t=T_{k-1}+1}^{T_k} \delta\nu_t \end{pmatrix} \tag{4.40}$$

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu + R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix} \tag{4.41}$$

$$\underset{\sim}{R}_\omega(k) = \sum_{t=T_{k-1}+1}^{T_k} \begin{pmatrix} \phi^{T_k-t} & 1 - \phi^{T_k-t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_\mu + R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix} \begin{pmatrix} \phi^{T_k-t} & 0 \\ 1 - \phi^{T_k-t} & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{t=T_{k-1}+1}^{T_k} \phi^{2(T_k-t)} R_\mu + \sum_{t=T_{k-1}+1}^{T_k} R_\nu & \sum_{t=T_{k-1}+1}^{T_k} R_\nu \\ \sum_{t=T_{k-1}+1}^{T_k} R_\nu & \sum_{t=T_{k-1}+1}^{T_k} R_\nu \end{pmatrix}$$

But,

$$\sum_{t=T_{k-1}+1}^{T_k} \phi^{2(T_k-t)} = \sum_{s=0}^{d_k-1} [\phi^2]^s = \frac{1 - \phi^{2d_k}}{1 - \phi^2} \tag{4.42}$$

(as the sum of a geometric progression with $|\phi| < 1$). So, using (4.16) and (4.42), we have:

$$\underset{\sim}{R}_\omega(k) = \begin{pmatrix} \frac{1-\phi^{2d_k}}{1-\phi^2} R_\mu + d_k R_\nu & d_k R_\nu \\ d_k R_\nu & d_k R_\nu \end{pmatrix} \quad (4.43)$$

4.3.5.2: MA(1)

Restatement of the Equally-Spaced Model

$$\begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \nu_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_t - \eta\delta\mu_{t-1} + \delta\nu_t \\ \delta\nu_t \end{pmatrix} \quad (4.44)$$

Derivation of the Unequally-Spaced Model

$$\begin{pmatrix} \mu_k \\ \nu_k \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}^{d_k} \begin{pmatrix} \mu_{k-1} \\ \nu_{k-1} \end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k} \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}^{T_k - t} \begin{pmatrix} \delta\mu_t - \eta\delta\mu_{t-1} + \delta\nu_t \\ \delta\nu_t \end{pmatrix}$$

Now

$$\underset{\sim}{G}^i = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \ i \geq 1 \qquad \text{and} \qquad \underset{\sim}{G}^o = \underset{\sim}{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

since

$$\underset{\sim}{G}^{i+1} = \underset{\sim}{G}\underset{\sim}{G}^i = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

So

$$\begin{pmatrix} \mu_k \\ \nu_k \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{k-1} \\ \nu_{k-1} \end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k - 1} \begin{pmatrix} \delta\nu_t \\ \delta\nu_t \end{pmatrix} + \begin{pmatrix} \delta\mu_{T_k} - \eta\delta\mu_{T_k - 1} + \delta\nu_{T_k} \\ \delta\nu_{T_k} \end{pmatrix}$$

i.e.

$$\begin{pmatrix} \mu_k \\ \nu_k \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{k-1} \\ \nu_{k-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_{T_k} - \eta\delta\mu_{T_k - 1} + \sum_{t=T_{k-1}+1}^{T_k} \delta\nu_t \\ \sum_{t=T_{k-1}+1}^{T_k} \delta\nu_t \end{pmatrix} \quad (4.45)$$

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu(1+\eta^2)+R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix} \tag{4.46}$$

Since $\underset{\sim}{\omega}_t$ contains error terms common to $\underset{\sim}{\omega}_{t-1}$ we use (4.9) to calculate $\underset{\sim}{R}_\omega(k)$, obtaining

$$\lambda^{-1}\underset{\sim}{R}_\omega(k) = \lambda^{-1}\{\sum_{t=T_{k-1}+1}^{T_k-1}\begin{pmatrix}0&1\\0&1\end{pmatrix}\begin{pmatrix} R_\mu(1+\eta^2)+R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix}\begin{pmatrix}0&0\\1&1\end{pmatrix} + \begin{pmatrix} R_\mu(1+\eta^2)+R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix}\}$$

$$+ \sum_{\substack{i=T_{k-1}+1 \\ (i\neq j,\, d_k>1)}}^{T_k-1}\;\sum_{j=T_{k-1}+1}^{T_k-1} \text{cov}(\begin{pmatrix}\delta\nu_i\\ \delta\nu_i\end{pmatrix},\begin{pmatrix}\delta\nu_j\\ \delta\nu_j\end{pmatrix})$$

$$+ \sum_{\substack{t=T_{k-1}+1\\(d_k>1)}}^{T_k-1}\text{cov}(\underset{\sim}{\omega}_{T_k},\begin{pmatrix}\delta\nu_t\\ \delta\nu_t\end{pmatrix}) + \sum_{\substack{t=T_{k-1}+1\\(d_k>1)}}^{T_k-1}\text{cov}(\begin{pmatrix}\delta\nu_t\\ \delta\nu_t\end{pmatrix},\underset{\sim}{\omega}_{T_k})$$

Due to the structure of $\underset{\sim}{G}$, the covariance terms are all zero, since $\text{cov}(\delta\nu_i,\delta\nu_j) = 0 \;\forall i \neq j$, and using (4.16) we have:

$$\underset{\sim}{R}_\omega(k) = \begin{pmatrix} R_\mu(1+\eta^2)+d_kR_\nu & d_kR_\nu \\ d_kR_\nu & d_kR_\nu \end{pmatrix} \tag{4.47}$$

4.3.5.3: AR(2).

Restatement of the Equally-Spaced Model

$$\begin{pmatrix}\mu_t\\ \mu_{t-1}\\ \nu_t\end{pmatrix} = \begin{pmatrix}\phi_1 & \phi_2 & 1-\phi_1-\phi_2\\ 1&0&0\\ 0&0&1\end{pmatrix}\begin{pmatrix}\mu_{t-1}\\ \mu_{t-2}\\ \nu_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t+\phi_2\delta\nu_{t-1}+\delta\nu_t\\ 0\\ \delta\nu_t\end{pmatrix} \tag{4.48}$$

Derivation of the Unequally-Spaced Model

$$\begin{pmatrix}\mu_k\\ \mu_{k-1}\\ \nu_k\end{pmatrix} = \begin{pmatrix}\phi_1 & \phi_2 & 1-\phi_1-\phi_2\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}^{d_k}\begin{pmatrix}\mu_{k-1}\\ \mu_{k-2}\\ \nu_{k-1}\end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k}\begin{pmatrix}\phi_1 & \phi_2 & 1-\phi_1-\phi_2\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}^{T_k-t}$$

$$\times\begin{pmatrix}\delta\mu_t+\phi_2\delta\nu_{t-1}+\delta\nu_t\\ 0\\ \delta\nu_t\end{pmatrix}$$

In order to calculate $\underset{\sim}{G}^i$, we note that $\underset{\sim}{G}$ can be written in the form:

$$\underset{\sim}{G} = \underset{\sim}{\Omega}\underset{\sim}{\Lambda}\underset{\sim}{\Omega}^{-1} \qquad (4.49)$$

where $\underset{\sim}{\Lambda}$ has eigenvalues of $\underset{\sim}{G}$ on its diagonal and zeroes elsewhere, and where $\underset{\sim}{\Omega}$ is the matrix of corresponding eigenvectors.

The eigenvalues of $\underset{\sim}{G}$, denoted by λ_1, λ_2, λ_3, satisfy:

$$(\lambda - \phi_1)\lambda(\lambda - 1) + \phi_2(1 - \lambda) = 0$$

i.e.

$$(\lambda - 1)(\lambda^2 - \phi_1\lambda - \phi_2) = 0 \qquad (4.50)$$

(see, for example, Cox and Miller, 1965), i.e.

$$\left.\begin{aligned}\lambda_1 &= 1\\ \lambda_2 &= \frac{\phi_1 + \sqrt{(\phi_1^2 + 4\phi_2)}}{2}\\ \lambda_3 &= \frac{\phi_1 - \sqrt{(\phi_1^2 + 4\phi_2)}}{2}\end{aligned}\right\} \qquad (4.51)$$

and we find

$$\underset{\sim}{\Omega} = \begin{pmatrix} 1 & \lambda_2 & \lambda_3 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{4.52}$$

so that

$$\underset{\sim}{\Omega}^{-1} = \begin{pmatrix} 1 & 0 & \lambda_3 - \lambda_2 \\ -1 & \lambda_3 & 1 - \lambda_3 \\ 1 & -\lambda_2 & -(1 - \lambda_2) \end{pmatrix} / (\lambda_3 - \lambda_2) \tag{4.53}$$

From (4.49) we see that

$$\underset{\sim}{G}^i = \underset{\sim}{\Omega}\underset{\sim}{\Lambda}^i\underset{\sim}{\Omega}^{-1} \tag{4.54}$$

i.e.

$$\underset{\sim}{G}^i = \frac{1}{\lambda_3 - \lambda_2}\begin{pmatrix} \lambda_3^{i+1} - \lambda_2^{i+1} & \lambda_3\lambda_2^{i+1} - \lambda_2\lambda_3^{i+1} & \lambda_3 - \lambda_2 + (1 - \lambda_3)\lambda_2^{i+1} - (1 - \lambda_2)\lambda_3^{i+1} \\ \lambda_3^i - \lambda_2^i & \lambda_3\lambda_2^i - \lambda_2\lambda_3^i & \lambda_3 - \lambda_2 + (1 - \lambda_3)\lambda_2^i - (1 - \lambda_2)\lambda_3^i \\ 0 & 0 & \lambda_3 - \lambda_2 \end{pmatrix} \tag{4.55}$$

where λ_2 and λ_3 are defined by (4.51). So

$$\begin{pmatrix}\mu_k \\ \mu_{k-1} \\ \nu_k\end{pmatrix} = \frac{1}{\lambda_3 - \lambda_2}\begin{pmatrix} \lambda_3^{d_k+1} - \lambda_2^{d_k+1} & \lambda_3\lambda_2^{d_k+1} - \lambda_2\lambda_3^{d_k+1} & \lambda_3 - \lambda_2 + (1-\lambda_3)\lambda_2^{d_k+1} - (1-\lambda_2)\lambda_3^{d_k+1} \\ \lambda_3^{d_k} - \lambda_2^{d_k} & \lambda_3\lambda_2^{d_k} - \lambda_2\lambda_3^{d_k} & \lambda_3 - \lambda_2 + (1-\lambda_3)\lambda_2^{d_k} - (1-\lambda_2)\lambda_3^{d_k} \\ 0 & 0 & \lambda_3 - \lambda_2 \end{pmatrix}$$

$$\times \begin{pmatrix}\mu_{k-1} \\ \mu_{k-2} \\ \nu_{k-1}\end{pmatrix}$$

$$+ \frac{1}{\lambda_3 - \lambda_2}\begin{pmatrix} \sum_{t=T_{k-1}+1}^{T_k} (\lambda_3^{T_k-t+1} - \lambda_2^{T_k-t+1})\delta\mu_t + \phi_2 \sum_{t=T_{k-1}+1}^{T_k} (\lambda_3^{T_k-t+1} - \lambda_2^{T_k-t+1})\delta\nu_{t-1} + \sum_{t=T_{k-1}+1}^{T_k} (\lambda_3 - \lambda_2 + \lambda_2\lambda_3^{T_k-t+1} - \lambda_3\lambda_2^{T_k-t+1})\delta\nu_t \\ \sum_{t=T_{k-1}+1}^{T_k} (\lambda_3^{T_k-t} - \lambda_2^{T_k-t})\delta\mu_t + \phi_2 \sum_{t=T_{k-1}+1}^{T_k} (\lambda_3^{T_k-t} - \lambda_2^{T_k-t})\delta\nu_{t-1} + \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} (\lambda_3 - \lambda_2 + \lambda_2\lambda_3^{T_k-t} - \lambda_3\lambda_2^{T_k-t})\delta\nu_t \\ (\lambda_3 - \lambda_2) \sum_{t=T_{k-1}+1}^{T_k} \delta\nu_t \end{pmatrix}$$

(4.56)

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu + R_\nu(1+\phi_2^2) & 0 & R_\nu \\ 0 & 0 & 0 \\ R_\nu & 0 & R_\nu \end{pmatrix} \tag{4.57}$$

$$\lambda^{-1}\underset{\sim}{R}_\omega(k) = \lambda^{-1}\sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} \underset{\sim}{G}^{T_k-t}\underset{\sim}{R}_\omega(\underset{\sim}{G}^{T_k-t})^T + \lambda^{-1}\underset{\sim}{R}_\omega$$

$$+ \sum_{i=T_{k-1}+1}^{T_k-1}\sum_{\substack{j=T_{k-1}+1 \\ (i\neq j,\, d_k>1)}}^{T_k-1} \mathrm{cov}(\underset{\sim}{G}^{T_k-i}\underset{\sim}{\omega}_i, \underset{\sim}{G}^{T_k-j}\underset{\sim}{\omega}_j)$$

$$+ \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} \mathrm{cov}(\underset{\sim}{\omega}_{T_k}, \underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t)$$

$$+ \sum_{\substack{t=T_{k-1}+1 \\ (d_k>1)}}^{T_k-1} \mathrm{cov}(\underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t, \underset{\sim}{\omega}_{T_k})$$

Now

$$\mathrm{cov}(\underset{\sim}{\omega}_{T_k}, \underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t) = \mathrm{cov}\left(\begin{pmatrix} \delta\mu_{T_k} + \phi_2\delta\nu_{T_k-1} + \delta\nu_{T_k} \\ 0 \\ \delta\nu_{T_k} \end{pmatrix}, \frac{1}{\lambda_3-\lambda_2}\right.$$

$$\left. \times \begin{pmatrix} (\lambda_3^{T_k-t+1} - \lambda_2^{T_k-t+1})(\delta\mu_t + \phi_2\delta\nu_{t-1}) + (\lambda_3 - \lambda_2 + \lambda_2\lambda_3^{T_k-t+1} - \lambda_3\lambda_2^{T_k-t+1})\delta\nu_t \\ (\lambda_3^{T_k-t} - \lambda_2^{T_k-t})(\delta\mu_t + \phi_2\delta\nu_{t-1}) + (\lambda_3 - \lambda_2 + \lambda_2\lambda_3^{T_k-t} - \lambda_3\lambda_2^{T_k-t})\delta\nu_t \\ (\lambda_3 - \lambda_2)\delta\nu_t \end{pmatrix}\right)$$

This covariance is only non-zero for $t = T_k - 1$, in which case

$$\frac{1}{\lambda^{-1}R_\nu}\text{ covariance } = \frac{1}{\lambda_3 - \lambda_2}\begin{pmatrix} \phi_2(\lambda_3 - \lambda_2 + \lambda_2\lambda_3^2 - \lambda_3\lambda_2^2) & \phi_2(\lambda_3 - \lambda_2) & \phi_2(\lambda_3 - \lambda_2) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \phi_2(1 + \lambda_2\lambda_3) & \phi_2 & \phi_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \phi_2(1 - \phi_2) & \phi_2 & \phi_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

(since $\lambda_2\lambda_3 = -\phi_2$ from (4.50).) By symmetry, $\text{cov}(\underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t, \underset{\sim}{\omega}_{T_k})$ is also non-zero only for $t = T_k - 1$, and is equal to

$$\lambda^{-1}R_\nu\begin{pmatrix} \phi_2(1 - \phi_2) & 0 & 0 \\ \phi_2 & 0 & 0 \\ \phi_2 & 0 & 0 \end{pmatrix}$$

Therefore

$$\sum_{t=T_{k-1}+1}^{T_k-1} \text{cov}(\underset{\sim}{\omega}_{T_k}, \underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t) + \sum_{t=T_{k-1}+1}^{T_k-1} \text{cov}(\underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t, \underset{\sim}{\omega}_{T_k})$$

$$= \lambda^{-1}R_\nu\begin{pmatrix} 2\phi_2(1 - \phi_2) & \phi_2 & \phi_2 \\ \phi_2 & 0 & 0 \\ \phi_2 & 0 & 0 \end{pmatrix}_{d_k > 1} \qquad (4.58)$$

Let

$$\left.\begin{aligned} &\underset{\sim}{G}^{T_k - t} = \begin{pmatrix} g(t) & \phi_2 g(t+1) & 1 - g(t) - \phi_2 g(t+1) \\ g(t+1) & \phi_2 g(t+2) & 1 - g(t+1) - \phi_2 g(t+2) \\ 0 & 0 & 1 \end{pmatrix} \\ &\text{where} \\ &g(t) = \frac{\lambda_3^{T_k - t+1} - \lambda_2^{T_k - t+1}}{\lambda_3 - \lambda_2} \end{aligned}\right\} \quad (4.59)$$

Then

$$\mathrm{cov}(\underset{\sim}{G}^{T_k - i}\underset{\sim}{\omega}_i, \underset{\sim}{G}^{T_k - j}\underset{\sim}{\omega}_j) = \mathrm{cov}\left(\begin{pmatrix} g(i)\delta\mu_i + \phi_2 g(i)\delta\nu_{i-1} + (1 - \phi_2 g(i+1))\delta\nu_i \\ g(i+1)\delta\mu_i + \phi_2 g(i+1)\delta\nu_{i-1} + (1 - \phi_2 g(i+2))\delta\nu_i \\ \delta\nu_i \end{pmatrix}, \begin{pmatrix} g(j)\delta\mu_j + \phi_2 g(j)\delta\nu_{j-1} + (1 - \phi_2 g(j+1))\delta\nu_j \\ g(j+1)\delta\mu_j + \phi_2 g(j+1)\delta\nu_{j-1} + (1 - \phi_2 g(j+2))\delta\nu_j \\ \delta\nu_j \end{pmatrix}\right)$$

This covariance is zero unless $j = i - 1$ or $j = i + 1$ $(i \neq j)$, and if $j = i - 1$ is equal to:

$$\lambda^{-1} R_\nu \begin{pmatrix} \phi_2 g(i)(1 - \phi_2 g(i)) & \phi_2 g(i)(1 - \phi_2 g(i+1)) & \phi_2 g(i) \\ \phi_2 g(i+1)(1 - \phi_2 g(i)) & \phi_2 g(i+1)(1 - \phi_2 g(i+1)) & \phi_2 g(i+1) \\ 0 & 0 & 0 \end{pmatrix}$$

By symmetry, for $j = i + 1$ the covariance term is:

$$\lambda^{-1} R_\nu \begin{pmatrix} \phi_2 g(i+1)(1 - \phi_2 g(i+1)) & \phi_2 g(i+2)(1 - \phi_2 g(i+1)) & 0 \\ \phi_2 g(i+1)(1 - \phi_2 g(i+2)) & \phi_2 g(i+2)(1 - \phi_2 g(i+2)) & 0 \\ \phi_2 g(i+1) & \phi_2 g(i+2) & 0 \end{pmatrix}$$

Therefore

$$\sum_{i=T_{k-1}+1}^{T_k-1} \sum_{j=T_{k-1}+1}^{T_k-1} \operatorname{cov}(\underset{\sim}{G}^{T_k-i}\underset{\sim}{\omega}_i, \underset{\sim}{G}^{T_k-j}\underset{\sim}{\omega}_j) = \lambda^{-1}R_\nu \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \end{pmatrix}_{d_k > 2} \tag{4.60}$$

say, where

$$U_{11} = \phi_2(\sum_{i=T_{k-1}+2}^{T_k-1} g(i) - \phi_2 \sum_{i=T_{k-1}+2}^{T_k-1} [g(i)]^2)$$
$$+ \phi_2(\sum_{i=T_{k-1}+1}^{T_k-2} g(i+1) - \phi_2 \sum_{i=T_{k-1}+1}^{T_k-2} [g(i+1)]^2)$$

$$U_{12} = \phi_2(\sum_{i=T_{k-1}+2}^{T_k-1} g(i) - \phi_2 \sum_{i=T_{k-1}+2}^{T_k-1} g(i).g(i+1))$$
$$+ \phi_2(\sum_{i=T_{k-1}+1}^{T_k-2} g(i+2) - \phi_2 \sum_{i=T_{k-1}+1}^{T_k-2} g(i+1).g(i+2))$$

$$U_{13} = \phi_2 \sum_{i=T_{k-1}+2}^{T_k-1} g(i)$$

$$U_{21} = \phi_2(\sum_{i=T_{k-1}+2}^{T_k-1} g(i+1) - \phi_2 \sum_{i=T_{k-1}+2}^{T_k-1} g(i).g(i+1))$$

$$+ \phi_2(\sum_{i=T_{k-1}+1}^{T_k-2} g(i+1) - \phi_2 \sum_{i=T_{k-1}+1}^{T_k-2} g(i+1).g(i+2))$$

$$U_{22} = \phi_2 \sum_{i=T_{k-1}+2}^{T_k-1} g(i+1) - \phi_2 \sum_{i=T_{k-1}+2}^{T_k-1} [g(i+1)]^2)$$

$$+ \phi_2(\sum_{i=T_{k-1}+1}^{T_k-2} g(i+2) - \phi_2 \sum_{i=T_{k-1}+1}^{T_k-2} [g(i+2)]^2)$$

$$U_{23} = \phi_2 \sum_{i=T_{k-1}+2}^{T_k-1} g(i+1)$$

$$U_{31} = \phi_2 \sum_{i=T_{k-1}+1}^{T_k-2} g(i + 1)$$

$$U_{32} = \phi_2 \sum_{i=T_{k-1}+1}^{T_k-2} g(i + 2)$$

and

$$U_{33} = 0$$

Also,

$$\sum_{t=T_{k-1}+1}^{T_k-1} \underset{\sim}{G}^{T_k-t} \underset{\sim\omega}{R} (\underset{\sim}{G}^{T_k-t})^T = \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{12} & V_{22} & V_{23} \\ V_{13} & V_{23} & V_{33} \end{pmatrix}_{d_k > 1} \tag{4.61}$$

say, where

$$V_{11} = R_\mu \sum_{t=T_{k-1}+1}^{T_k-1} [g(t)]^2$$

$$+ R_\nu\{\phi_2^2 \sum_{t=T_{k-1}+1}^{T_k-1} [g(t)]^2 - 2\phi_2 \sum_{t=T_{k-1}+1}^{T_k-1} g(t + 1)$$

$$+ \phi_2^2 \sum_{t=T_{k-1}+1}^{T_k-1} [g(t + 1)]^2 + d_k - 1\}$$

$$V_{12} = R_\mu \sum_{t=T_{k-1}+1}^{T_k-1} g(t).g(t + 1)$$

$$+ R_\nu\{\phi_2^2 \sum_{t=T_{k-1}+1}^{T_k-1} g(t).g(t + 1) - \phi_2 \sum_{t=T_{k-1}+1}^{T_k-1} g(t + 1)$$

$$- \phi_2 \sum_{t=T_{k-1}+1}^{T_k-1} g(t + 2) + \phi_2^2 \sum_{t=T_{k-1}+1}^{T_k-1} g(t + 1).g(t + 2) + d_k - 1\}$$

$$V_{13} = R_\nu\{d_k - 1 - \phi_2 \sum_{t=T_{k-1}+1}^{T_k-1} g(t+1)\}$$

$$V_{22} = R_\mu \sum_{t=T_{k-1}+1}^{T_k-1} [g(t+1)]^2$$

$$+ R_\nu\{\phi_2^2 \sum_{t=T_{k-1}+1}^{T_k-1} [g(t+1)]^2 - 2\phi_2 \sum_{t=T_{k-1}+1}^{T_k-1} g(t+2)$$

$$+ \phi_2^2 \sum_{t=T_{k-1}+1}^{T_k} [g(t+2)]^2 + d_k - 1\}$$

$$V_{23} = R_\nu\{d_k - 1 - \phi_2 \sum_{t=T_{k-1}+1}^{T_k-1} g(t+2)\}$$

and

$$V_{33} = R_\nu\{d_k - 1\}.$$

$\underset{\sim}{R}_\omega(k)$ is then formed by summing the matrices given in (4.58), (4.60) and (4.61) and adding this sum to $\underset{\sim}{R}_\omega$ (given by (4.57)) when $d_k > 1$.

NOTE: If $\phi_2 = 0$ (AR(1) model), all covariance terms vanish and (dropping the second row/column) we have:

$$\underset{\sim}{R}_\omega(k) = \begin{pmatrix} R_\mu + R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix} + \begin{pmatrix} R_\mu \sum_{t=T_{k-1}+1}^{T_k-1} \phi_1^{2(T_k-t)} + R_\nu(d_k - 1) & R_\nu(d_k - 1) \\ R_\nu(d_k - 1) & R_\nu(d_k - 1) \end{pmatrix}$$

In order to define the quantities in (4.60) and (4.61) explicitly we note that:

(a) $$\sum_{t=T_{k-1}+2}^{T_k-1} g(t) = \sum_{t=T_{k-1}+1}^{T_k-2} g(t+1)$$

$$= \frac{1}{\lambda_3 - \lambda_2}\left(\frac{1 - \lambda_3^{d_k}}{1 - \lambda_3} - \frac{1 - \lambda_2^{d_k}}{1 - \lambda_2}\right) - 1$$

(b) $$\sum_{t=T_{k-1}+2}^{T_k-1} g(t+1) = \sum_{t=T_{k-1}+1}^{T_k-2} g(t+2)$$

$$= \frac{1}{\lambda_3 - \lambda_2}\left(\frac{1 - \lambda_3^{d_k-1}}{1 - \lambda_3} - \frac{1 - \lambda_2^{d_k-1}}{1 - \lambda_2}\right)$$

(c) $$\sum_{t=T_{k-1}+2}^{T_k-1} [g(t)]^2 = \sum_{t=T_{k-1}+1}^{T_k-2} [g(t+1)]^2$$

$$= \frac{1}{(\lambda_3 - \lambda_2)^2}\left(\frac{1 - \lambda_3^{2d_k}}{1 - \lambda_3^2} - \frac{2(1 - (\lambda_2\lambda_3)^{d_k})}{1 - \lambda_2\lambda_3} + \frac{1 - \lambda_2^{2d_k}}{1 - \lambda_2^2}\right) - 1$$

(d) $$\sum_{t=T_{k-1}+2}^{T_k-1} [g(t+1)]^2 = \sum_{t=T_{k-1}+1}^{T_k-2} [g(t+2)]^2$$

$$= \frac{1}{(\lambda_3 - \lambda_2)^2}\left(\frac{1 - \lambda_3^{2(d_k-1)}}{1 - \lambda_3^2} - \frac{2(1 - (\lambda_2\lambda_3)^{d_k-1})}{1 - \lambda_2\lambda_3} + \frac{1 - \lambda_2^{2(d_k-1)}}{1 - \lambda_2^2}\right)$$

(e) $$\sum_{t=T_{k-1}+2}^{T_k-1} g(t).g(t+1) = \sum_{t=T_{k-1}+1}^{T_k-2} g(t+1).g(t+2)$$

$$= \frac{1}{(\lambda_3 - \lambda_2)^2}\left\{\frac{\lambda_3(1 - \lambda_3^{2(d_k-1)})}{1 - \lambda_3^2} - \frac{\phi_1(1 - (\lambda_2\lambda_3)^{d_k-1})}{1 - \lambda_2\lambda_3} + \frac{\lambda_2(1 - \lambda_2^{2(d_k-1)})}{1 - \lambda_2^2}\right\}$$

4.3.5.4: MA(2).

Restatement of the Equally-Spaced Model

$$\begin{pmatrix}\mu_t \\ \nu_t\end{pmatrix} = \begin{pmatrix}0 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1} \\ \nu_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t - \eta_1\delta\mu_{t-1} - \eta_2\delta\mu_{t-2} + \delta\nu_t \\ \delta\nu_t\end{pmatrix} \tag{4.62}$$

Derivation of the Unequally-Spaced Model

Recall from Section 4.3.5.2 that $\underset{\sim}{G}^i = \underset{\sim}{G}\ \forall i \geq 1$ with $\underset{\sim}{G}^o = \underset{\sim}{I}$.

So

$$\begin{pmatrix}\mu_k \\ \nu_k\end{pmatrix} = \begin{pmatrix}0 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{k-1} \\ \nu_{k-1}\end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k-1}\begin{pmatrix}0 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\delta\mu_t - \eta_1\delta\mu_{t-1} - \eta_2\delta\mu_{t-2} + \delta\nu_t \\ \delta\nu_t\end{pmatrix}$$

$$+ \begin{pmatrix}\delta\mu_{T_k} - \eta_1\delta\mu_{T_k-1} - \eta_2\delta\mu_{T_k-2} + \delta\nu_{T_k} \\ \delta\nu_{T_k}\end{pmatrix}$$

$$= \begin{pmatrix}0 & 1 \\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{k-1} \\ \nu_{k-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_{T_k} - \eta_1\delta\mu_{T_k-1} - \eta_2\delta\mu_{T_k-2} + \sum_{t=T_{k-1}+1}^{T_k}\delta\nu_t \\ \sum_{t=T_{k-1}+1}^{T_k}\delta\nu_t\end{pmatrix} \tag{4.63}$$

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_\omega = \begin{pmatrix} R_\mu(1 + \eta_1^2 + \eta_2^2) + R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix} \tag{4.64}$$

$$\lambda^{-1}\underset{\sim}{R}_\omega(k) = \lambda^{-1}\{ \sum_{t=T_{k-1}+1}^{T_k-1} \begin{pmatrix} R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix} + \begin{pmatrix} R_\mu(1 + \eta_1^2 + \eta_2^2) + R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix} \}$$

$$+ \sum_{i=T_{k-1}+1}^{T_k-1} \sum_{j=T_{k-1}+1}^{T_k-1} \mathrm{cov}\left(\begin{pmatrix} \delta\nu_i \\ \delta\nu_i \end{pmatrix}, \begin{pmatrix} \delta\nu_j \\ \delta\nu_j \end{pmatrix} \right)$$

$$+ \sum_{t=T_{k-1}+1}^{T_k-1} \mathrm{cov}\left(\underset{\sim}{\omega}_{T_k}, \begin{pmatrix} \delta\nu_t \\ \delta\nu_t \end{pmatrix}\right) + \sum_{t=T_{k-1}+1}^{T_k-1} \mathrm{cov}\left(\begin{pmatrix} \delta\nu_t \\ \delta\nu_t \end{pmatrix}, \underset{\sim}{\omega}_{T_k} \right)$$

and it is clear that all covariance terms vanish since $\mathrm{cov}(\delta\nu_i, \delta\nu_j) = 0 \;\forall i \neq j$. So

$$\underset{\sim}{R}_\omega(k) = \begin{pmatrix} R_\mu(1 + \eta_1^2 + \eta_2^2) + \sum_{t=T_{k-1}+1}^{T_k} R_\nu & \sum_{t=T_{k-1}+1}^{T_k} R_\nu \\ \sum_{t=T_{k-1}+1}^{T_k} R_\nu & \sum_{t=T_{k-1}+1}^{T_k} R_\nu \end{pmatrix}$$

$$= \begin{pmatrix} R_\mu(1 + \eta_1^2 + \eta_2^2) + d_k R_\nu & d_k R_\nu \\ d_k R_\nu & d_k R_\nu \end{pmatrix} \tag{4.65}$$

(using (4.16)).

4.3.5.5: ARMA(1,1).

Restatement of the Equally-Spaced Model

$$\begin{pmatrix}\mu_t\\ \nu_t\end{pmatrix} = \begin{pmatrix}\phi & 1-\phi\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{t-1}\\ \nu_{t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_t - \eta\delta\mu_{t-1} + \delta\nu_t\\ \delta\nu_t\end{pmatrix} \qquad (4.66)$$

Derivation of the Unequally-Spaced Model

$$\begin{pmatrix}\mu_k\\ \nu_k\end{pmatrix} = \begin{pmatrix}\phi & 1-\phi\\ 0 & 1\end{pmatrix}^{d_k}\begin{pmatrix}\mu_{k-1}\\ \nu_{k-1}\end{pmatrix} + \sum_{t=T_{k-1}+1}^{T_k}\begin{pmatrix}\phi & 1-\phi\\ 0 & 1\end{pmatrix}^{T_k-t}\begin{pmatrix}\delta\mu_t - \eta\delta\mu_{t-1} + \delta\nu_t\\ \delta\nu_t\end{pmatrix}$$

We have shown in Section 4.3.5.1 that:

$$\begin{pmatrix}\phi & 1-\phi\\ 0 & 1\end{pmatrix}^{i} = \begin{pmatrix}\phi^i & 1-\phi^i\\ 0 & 1\end{pmatrix}$$

So

$$\begin{pmatrix}\mu_k\\ \nu_k\end{pmatrix} = \begin{pmatrix}\phi^{d_k} & 1-\phi^{d_k}\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{k-1}\\ \nu_{k-1}\end{pmatrix} + \begin{pmatrix}\sum_{t=T_{k-1}+1}^{T_k-1}\phi^{T_k-t}\delta\mu_t - \eta\sum_{t=T_{k-1}+1}^{T_k-1}\phi^{T_k-t}\delta\mu_{t-1} + \sum_{t=T_{k-1}+1}^{T_k}\delta\nu_t + \delta\mu_{T_k} - \eta\delta\mu_{T_k-1}\\ \sum_{t=T_{k-1}+1}^{T_k}\delta\nu_t\end{pmatrix} \qquad (4.67)$$

Variance Calculation

Equally-spaced model:

$$\underset{\sim}{R}_{\omega} = \begin{pmatrix}R_\mu(1+\eta^2) + R_\nu & R_\nu\\ R_\nu & R_\nu\end{pmatrix} \qquad (4.68)$$

$$\lambda^{-1}\underset{\sim}{R}_{\omega}(k) = \lambda^{-1}\sum_{t=T_{k-1}+1}^{T_k}\begin{pmatrix}\phi^{T_k-t} & 1-\phi^{T_k-t}\\ 0 & 1\end{pmatrix}\begin{pmatrix}R_\mu(1+\eta^2)+R_\nu & R_\nu\\ R_\nu & R_\nu\end{pmatrix} \times \begin{pmatrix}\phi^{T_k-t} & 0\\ 1-\phi^{T_k-t} & 1\end{pmatrix}$$

$$+ \sum_{\substack{i=T_{k-1}+1\\(i\neq j,\, d_k>1)}}^{T_k-1}\ \sum_{j=T_{k-1}+1}^{T_k-1} \operatorname{cov}(\underset{\sim}{G}^{T_k-i}\underset{\sim}{\omega}_i, \underset{\sim}{G}^{T_k-j}\underset{\sim}{\omega}_j)$$

$$+ \sum_{\substack{t=T_{k-1}+1\\(d_k>1)}}^{T_k-1} \operatorname{cov}(\underset{\sim}{\omega}_{T_k}, \underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t) + \sum_{\substack{t=T_{k-1}+1\\(d_k>1)}}^{T_k-1} \operatorname{cov}(\underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t, \underset{\sim}{\omega}_{T_k}).$$

Now

$$\underset{\sim}{G}^{T_k-t}\underset{\sim}{\omega}_t = \begin{pmatrix}\phi^{T_k-t}\delta\mu_t - \eta\phi^{T_k-t}\delta\mu_{t-1} + \delta\nu_t\\ \delta\nu_t\end{pmatrix}$$

Therefore the only non-zero terms in the final two summations occur at time $t = T_k - 1$, with

$$\operatorname{cov}(\underset{\sim}{\omega}_{T_k}, \underset{\sim}{G}\underset{\sim}{\omega}_{T_k-1}) = \operatorname{cov}(\underset{\sim}{G}\underset{\sim}{\omega}_{T_k-1}, \underset{\sim}{\omega}_{T_k}) = \lambda^{-1}\begin{pmatrix}-\eta\phi R_\mu & 0\\ 0 & 0\end{pmatrix} \tag{4.69}$$

In the double summation, non-zero terms arise when either $j = i - 1$ or $j = i + 1$.

If $j = i - 1$, we have:

$$\sum_{i=T_{k-1}+2}^{T_k-1} \operatorname{cov}\left(\begin{pmatrix}\phi^{T_k-i}\delta\mu_i - \eta\phi^{T_k-i}\delta\mu_{i-1} + \delta\nu_i\\ \delta\nu_i\end{pmatrix}, \begin{pmatrix}\phi^{T_k-i+1}\delta\mu_{i-1} - \eta\phi^{T_k-i+1}\delta\mu_{i-2} + \delta\nu_{i-1}\\ \delta\nu_{i-1}\end{pmatrix}\right)$$

$$= \lambda^{-1}R_\mu \sum_{i=T_{k-1}+2}^{T_k-1}\begin{pmatrix}-\eta\phi^{T_k-i}.\phi^{T_k-i+1} & 0\\ 0 & 0\end{pmatrix}$$

$$= \begin{pmatrix} -\lambda^{-1}R_\mu\eta\phi \sum_{i=T_{k-1}+1}^{T_k-1} \phi^{2(T_k-i)} & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \lambda^{-1}\begin{pmatrix} \frac{-\eta\phi^3(1 - \phi^{2(d_k-2)})R_\mu}{1 - \phi^2} & 0 \\ 0 & 0 \end{pmatrix} \tag{4.70}$$

(as sum of geometric progression). Also,

$$\sum_{t=T_{k-1}+1}^{T_k} \begin{pmatrix} \phi^{T_k-t} & 1 - \phi^{T_k-t} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} R_\mu(1 + \eta^2) + R_\nu & R_\nu \\ R_\nu & R_\nu \end{pmatrix}\begin{pmatrix} \phi^{T_k-t} & 0 \\ 1 - \phi^{T_k-t} & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{R_\mu(1 + \eta^2)(1 - \phi^{2d_k})}{1 - \phi^2} + d_kR_\nu & d_kR_\nu \\ d_kR_\nu & d_kR_\nu \end{pmatrix} \tag{4.71}$$

So

$$R_\omega(k) = \begin{pmatrix} [\frac{(1 + \eta^2)(1 - \phi^{2d_k}) - 2\eta\phi^3(1 - \phi^{2(d_k-2)}) - 2\eta\phi(1 - \phi^2)}{1 - \phi^2}]R_\mu + d_kR_\nu & d_kR_\nu \\ d_kR_\nu & d_kR_\nu \end{pmatrix}$$

(using (4.69), (4.70) and (4.71))

$$= \begin{pmatrix} [\frac{(1 + \eta^2)(1 - \phi^{2d_k}) - 2\eta\phi(1 - \phi^{2(d_k-1)})}{1 - \phi^2}]R_\mu + d_kR_\nu & d_kR_\nu \\ d_kR_\nu & d_kR_\nu \end{pmatrix}$$

(4.72)

4.4 PERFORMANCE AND SENSITIVITY

We now examine the performance of multistate dynamic linear models with unequally-spaced data, using the measures outlined in Section 3.4. In order to do so, we again restrict our attention to the linear growth, sinusoidal and AR(1) models, extended to the unequally-spaced case in Sections 4.3.1, 4.3.3 and 4.3.5.1, respectively. In performance terms, we wish to compare our results with the equally-spaced models, and so the data sets adopted are identical to those used in Section 3.4 except that observations have been removed at a number of timepoints.

In order to examine the sensitivity of these models to the 'degree of unequal-spacing', we adapt each of the original series to produce four extra series.

SERIES 1: Original data set (see Appendix 3), with observations removed at times: 22, 24, 26, 28, 43, 45, 46, 47, 52, 53, i.e. 10% of the series removed.

SERIES 2: As Series 1, with additional observations removed at times: 55, 56, 57, 58, 59, 60, 62, 63, 68, 69, 70, 81, 83, 84, 91, i.e. 25% of the series removed.

SERIES 3: As Series 2, with additional observations removed at times: 9, 10, 11, 15, 18, 20, 65, 66, 67, 73, 74, 77, 78, 79, 85, 86, 87, 89, 92, 94, 95, 96, 97, 98, 99, i.e. 50% of the series removed.

SERIES 4: Original data set, with observations removed at times: 1, 2, 3, 4, 22, 24, 26, 28, 43, 45, i.e. 10% of the series removed.

Series 4 has been created in order to examine whether or not an initial 'blank' period has any serious effects on performance; in many medical monitoring contexts, however, this situation is unlikely to arise. A more typical sampling pattern might be as follows: an initial intense monitoring period, followed by a gradual decrease in sampling rate as the patient is seen to improve; this decrease in rate might well be interrupted from time to time by clinically interesting events which would prompt a return to more intense observation.

NOTES:

(i) To calculate all the quantities below, the prior values given in Sections 3.4.1.1, 3.4.2.1 and 3.4.3.1 (for the linear growth, sinusoidal and AR(1) models respectively) have been used.

(ii) Since SSFE would increase purely on the number of observations, we use only the MAD for comparisons of forecasting ability.

(iii) In terms of event detection, we retain the use of one-step-back probabilities, $O_t^{(i)}$, and the number of false positives, NFP (for which $O_i^{(i)} > 0.2$, $i \neq 1$), in order to evaluate performance, and we use the final estimate of $\underset{\sim}{\theta}$, $\underset{\sim}{m}_{100}$, in order to assess estimation capabilities.

4.4.1 LINEAR GROWTH

TABLE 4.1

	$O_{26}^{(3)}$	$O_{36}^{(4)}$	$O_{51}^{(2)}$	$O_{81}^{(4)}$	NFP	$\underset{\sim}{m}_{100}$	MAD
Original Time Series	0.799	1.000	1.000	1.000	2	$\begin{pmatrix}-116.9\\-7.8\end{pmatrix}$	7.9
Series 1	0.339*	1.000	0.999	1.000	3	$\begin{pmatrix}-116.9\\-7.8\end{pmatrix}$	8.8
Series 2	0.339*	1.000	0.999	0.999†	2	$\begin{pmatrix}-116.9\\-7.8\end{pmatrix}$	10.2
Series 3	0.688*	1.000	1.000	0.856†	1	$\begin{pmatrix}-119.4\\-5.7\end{pmatrix}$	15.5
Series 4	0.375*	1.000	1.000	1.000	4	$\begin{pmatrix}-117.0\\-7.8\end{pmatrix}$	8.5
		Theoretical Values:				$\begin{pmatrix}-117.5\\-5.0\end{pmatrix}$	

Recall: i = 1 - steady state
i = 2 - level change
i = 3 - slope change
i = 4 - transient

*Observation not available at t = 26; $O_{27}^{(3)}$ used.

†Observation not available at t = 81; $O_{82}^{(4)}$ used.

See Figues 4.1 to 4.4 for Kalman Filter results along with one-step-ahead forecasts (asterisks) for Series 1 to 4 respectively.

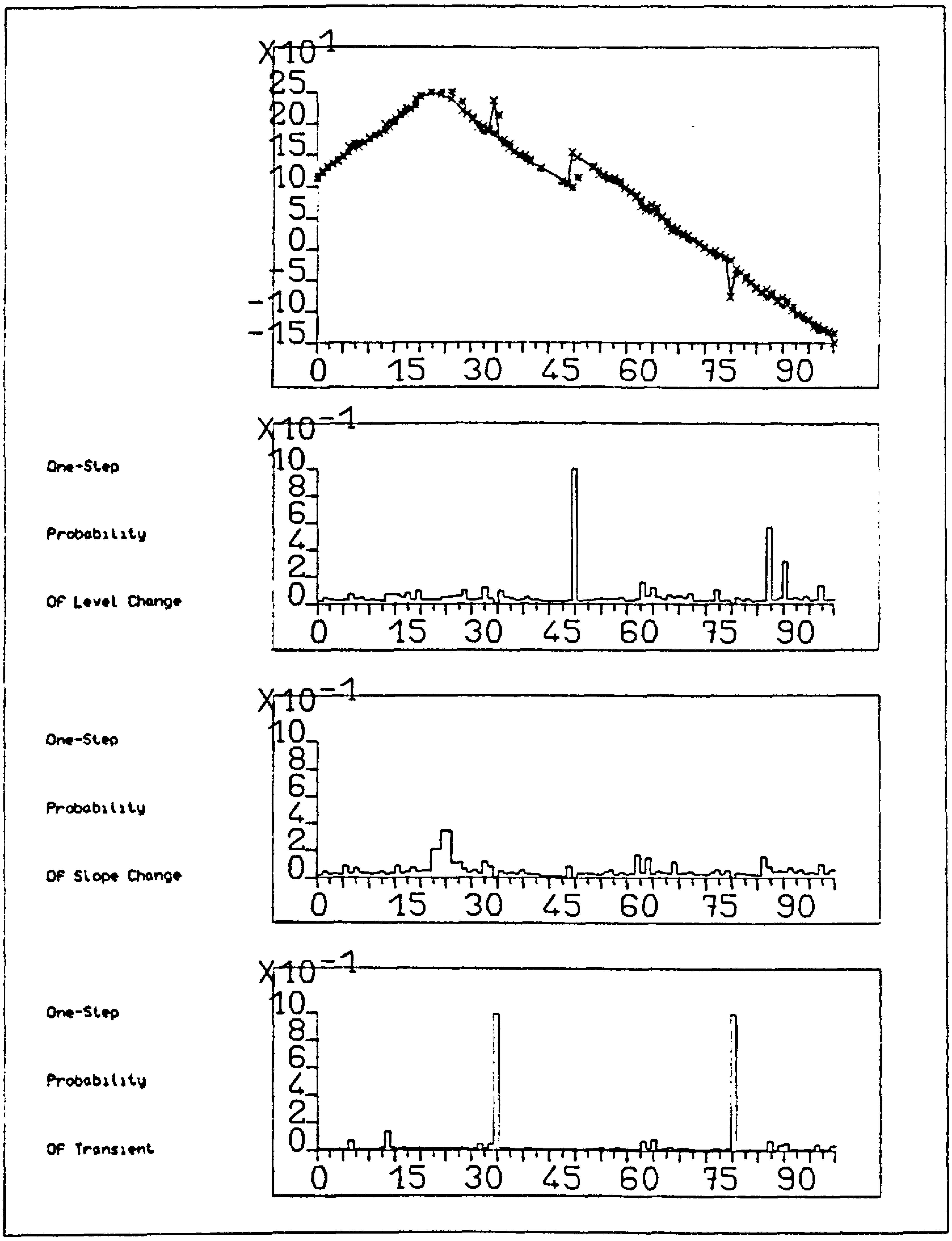
NOTE: In order to calculate one-step-ahead forecasts we now have:

$$f_k = \underset{\sim}{H}_{T_k}^{T} \underset{\sim}{G}_k \underset{\sim}{m}_{k-1} \tag{4.73}$$

where

$$\underset{\sim}{G}_k = \underset{\sim}{G}^{d_k}.$$

(c.f. (2.14))



One-Step
Probability
Of Level Change
One-Step
Probability
Of Slope Change
One-Step
Probability
Of Transient

FIGURE 4.1

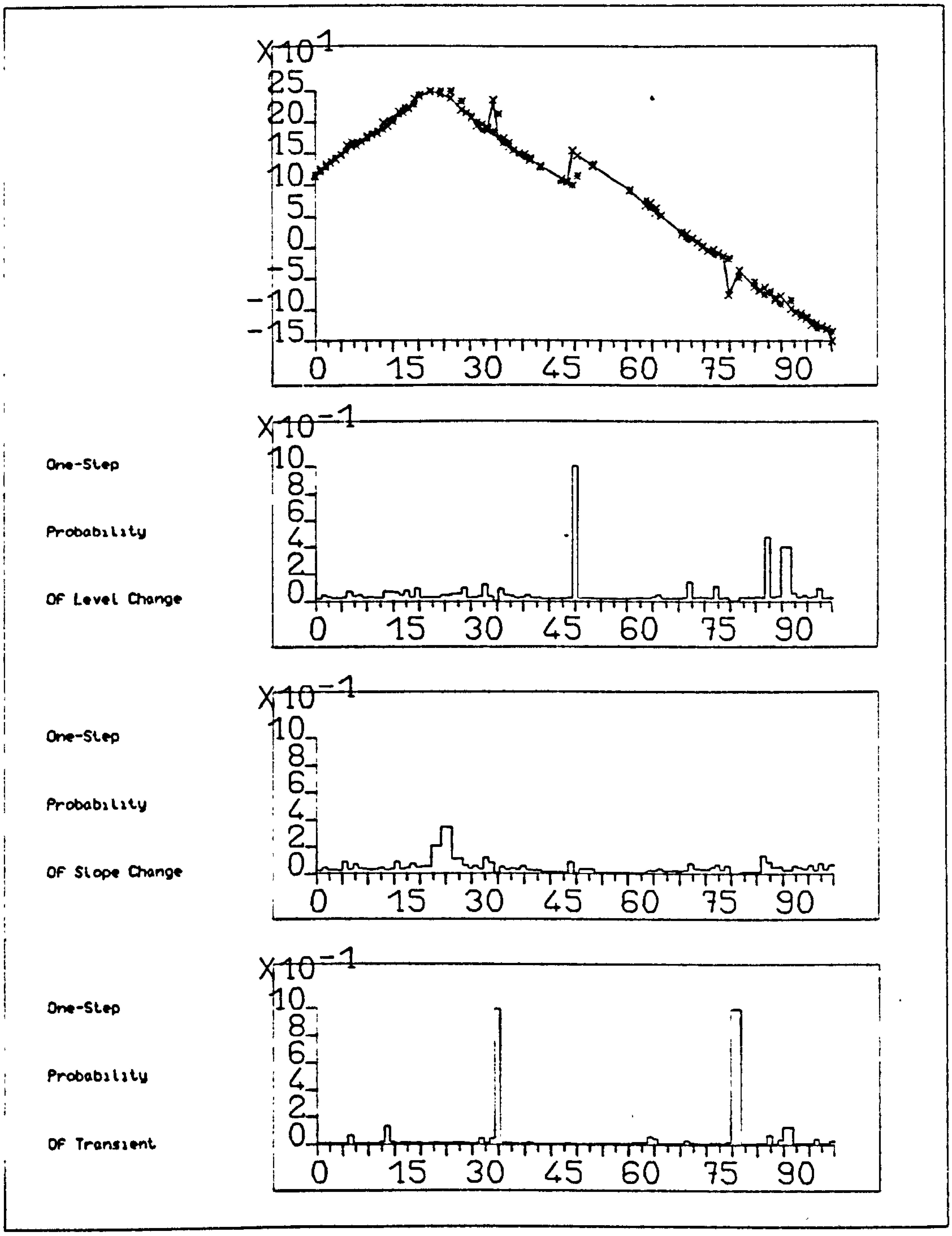

X10^1
25
20
15
10
5
0
-5
-10
-15
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Slope Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 4.2

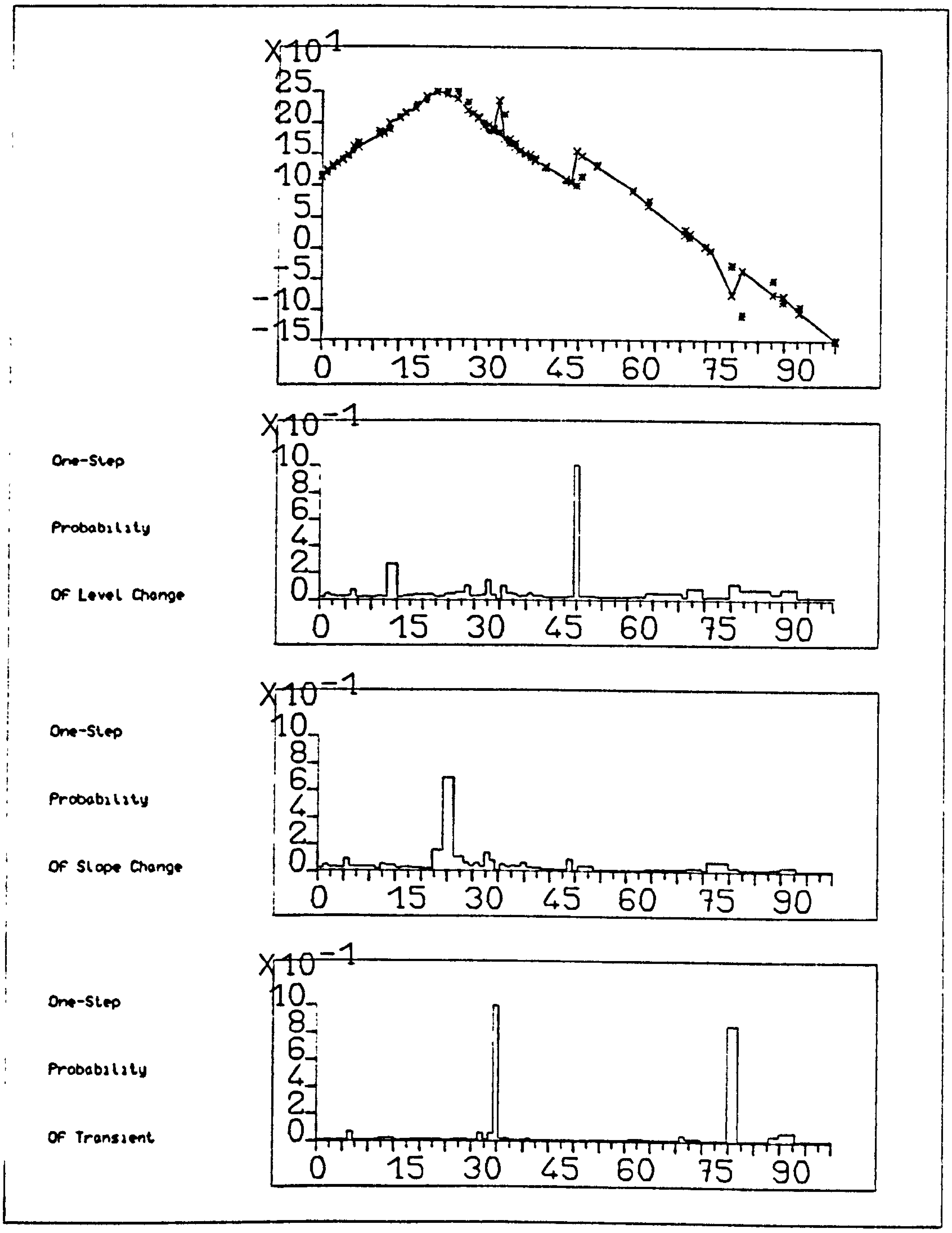

One-Step
Probability
Of Level Change
One-Step
Probability
Of Slope Change
One-Step
Probability
Of Transient
X10^1
X10^-1
0 15 30 45 60 75 90

FIGURE 4.3

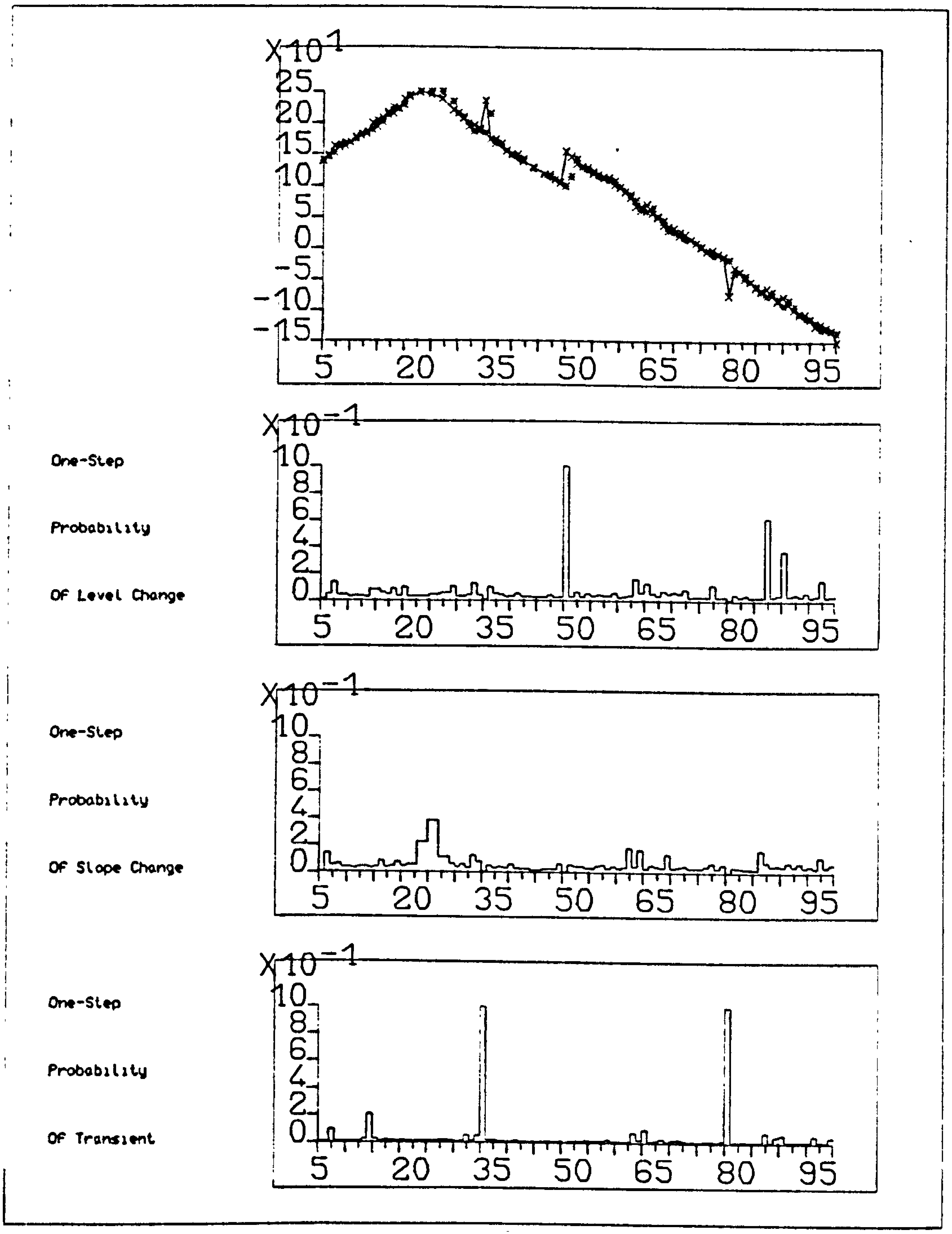

X10^1
25
20
15
10
5
0
-5
-10
-15
5
20
35
50
65
80
95
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
One-Step
Probability
Of Slope Change
One-Step
Probability
Of Transient

FIGURE 4.4

4.4.2 SINUSOIDAL MODEL

TABLE 4.2

	$0_{26}^{(3)}$	$0_{36}^{(4)}$	$0_{51}^{(2)}$	$0_{81}^{(4)}$	NFP	$\underset{\sim}{m}_{100}$	ϕ	MAD
Original Series	0.213	0.997	0.992	1.000	0	$\begin{pmatrix}150.0\\15.4\end{pmatrix}$	90.0	9.1
Series 1	0.238*	0.998	0.987	1.000	1	$\begin{pmatrix}150.0\\15.4\end{pmatrix}$	90.0	9.8
Series 2	0.238*	0.998	0.987	1.000†	1	$\begin{pmatrix}150.0\\15.3\end{pmatrix}$	90.0	11.0
Series 3	0.158*	0.997	0.984	0.960†	1	$\begin{pmatrix}148.5\\13.8\end{pmatrix}$	91.1	15.8
Series 4	0.250*	0.998	0.990	1.000	1	$\begin{pmatrix}150.0\\15.4\end{pmatrix}$	90.0	9.4
		Theoretical Values:				$\begin{pmatrix}150.0\\15.0\end{pmatrix}$	90.0	

*Observation not available at t = 26; $0_{27}^{(3)}$ used.

†Observation not available at t = 81; $0_{82}^{(4)}$ used.

See Figures 4.5 to 4.8 for Kalman Filter results along with one-step-ahead forecasts for Series 1 to 4 respectively; Figures 4.9 to 4.12 show on-line estimation of the ϕ-grid for these series.

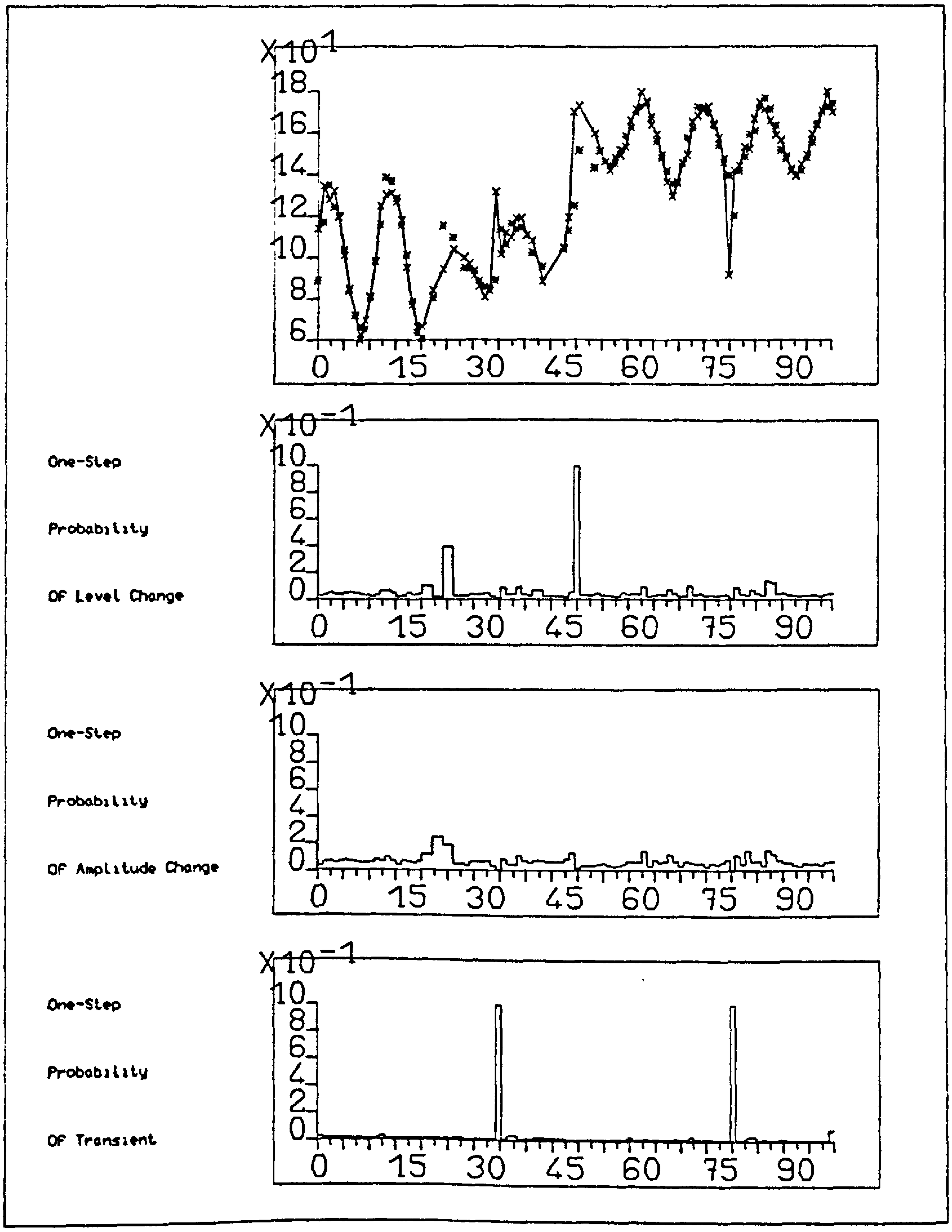

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 4.5

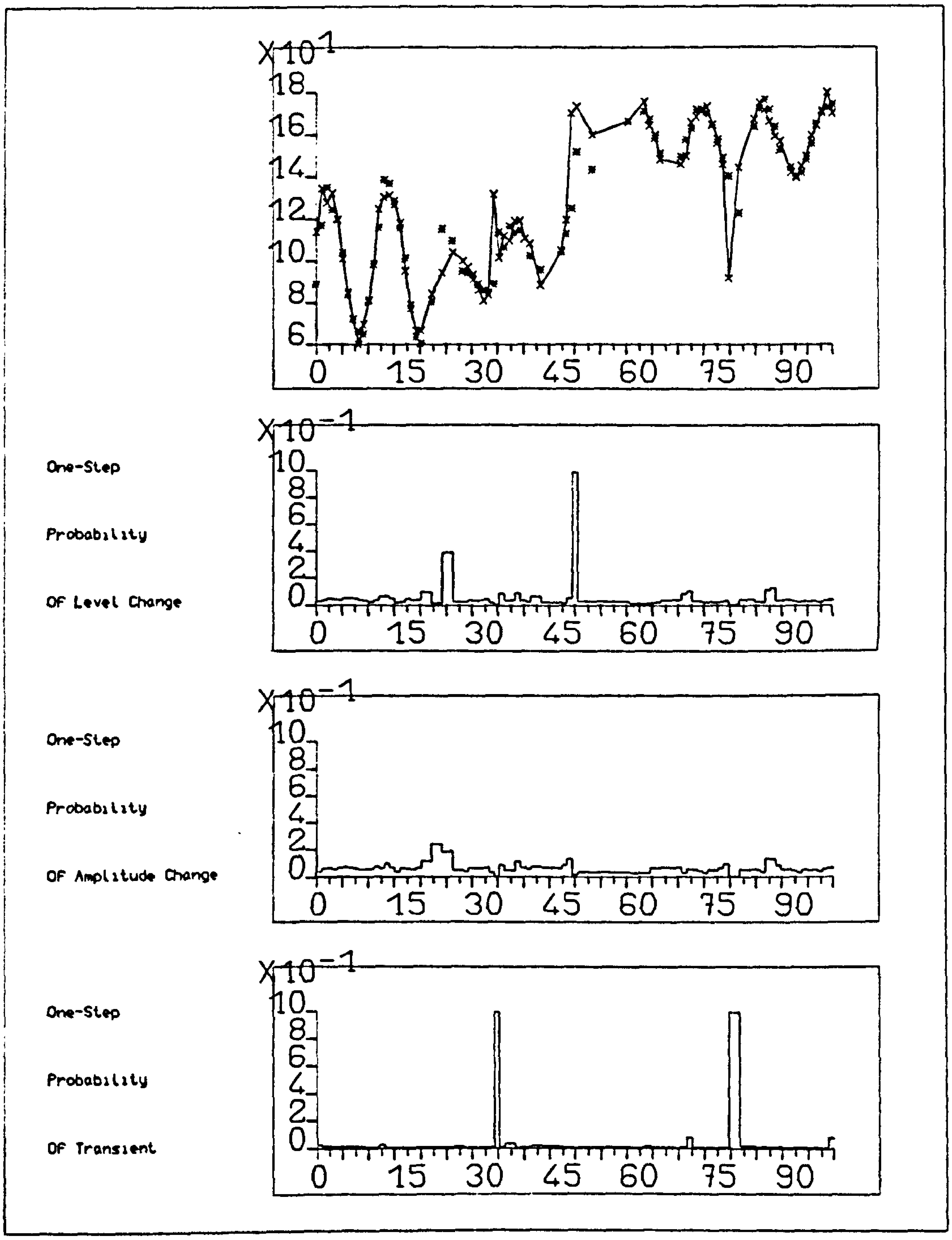

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 4.6

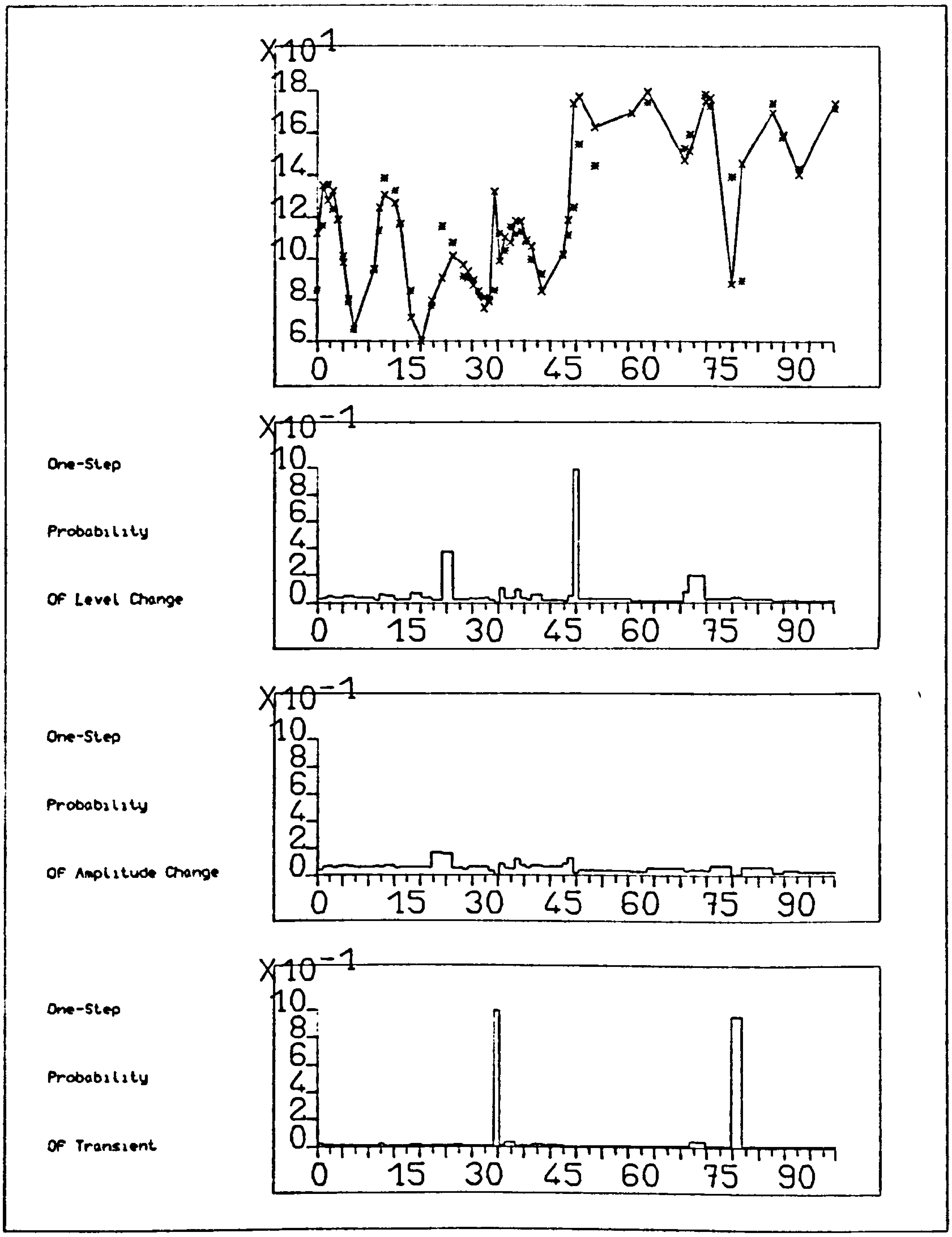

X10^1
18
16
14
12
10
8
6
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Amplitude Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 4.7

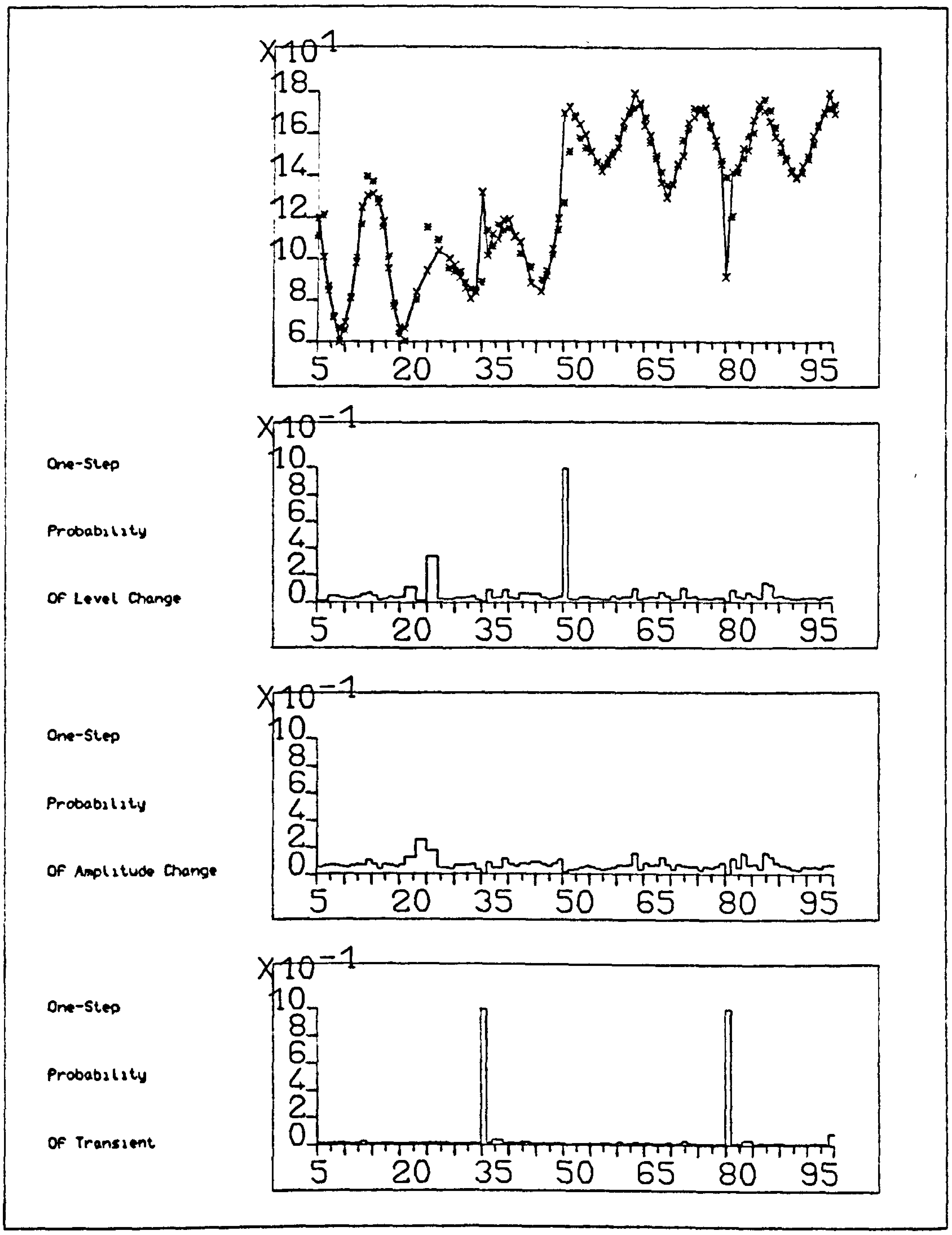

X10^1
18
16
14
12
10
8
6
5
20
35
50
65
80
95
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
One-Step
Probability
Of Amplitude Change
One-Step
Probability
Of Transient

FIGURE 4.8

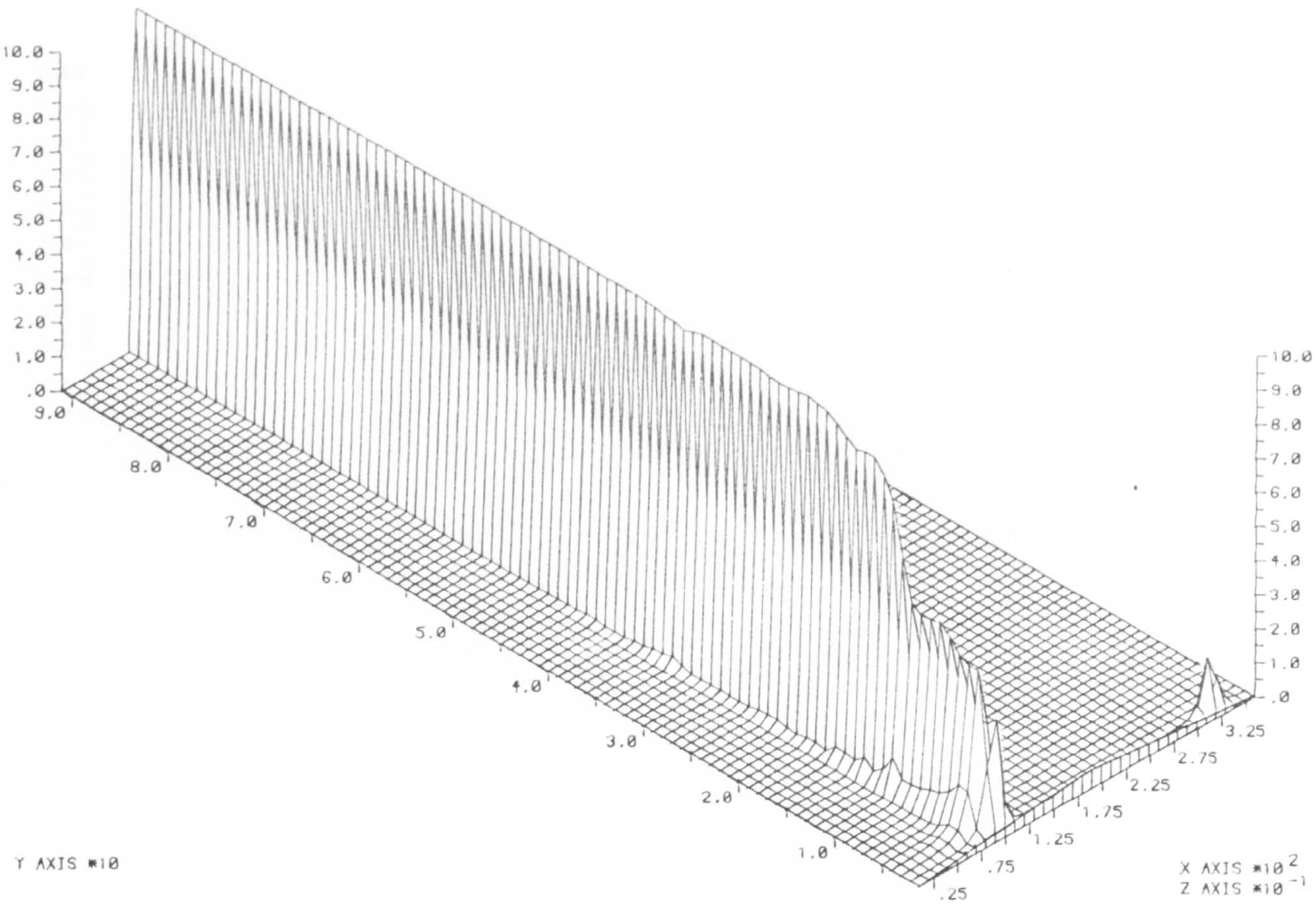

10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.25
.75
1.25
1.75
2.25
2.75
3.25
Y AXIS *10
X AXIS *10 2
Z AXIS *10 -1

FIGURE 4.9

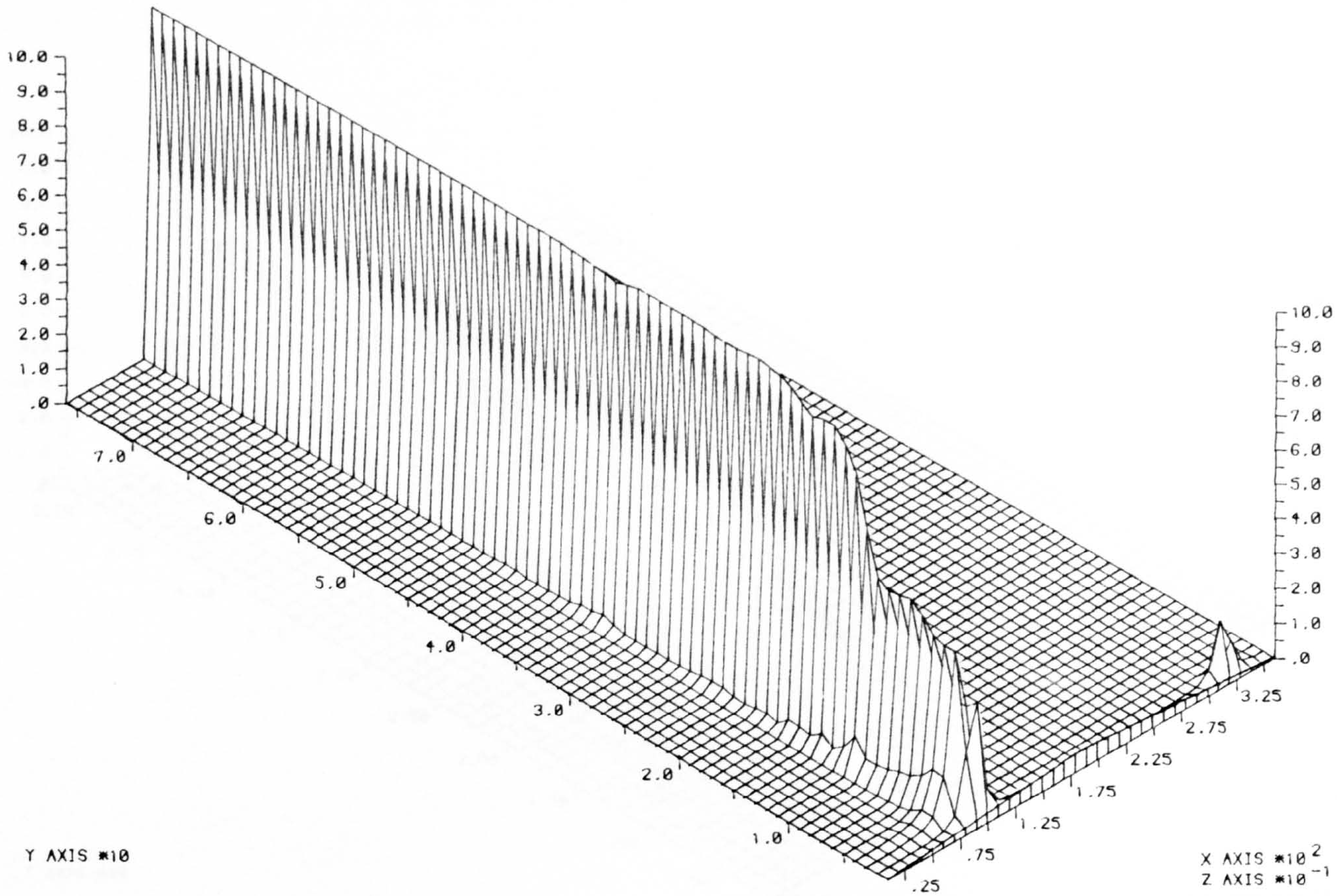

10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.25
.75
1.25
1.75
2.25
2.75
3.25
Y AXIS *10
X AXIS *10 2
Z AXIS *10 -1

FIGURE 4.10

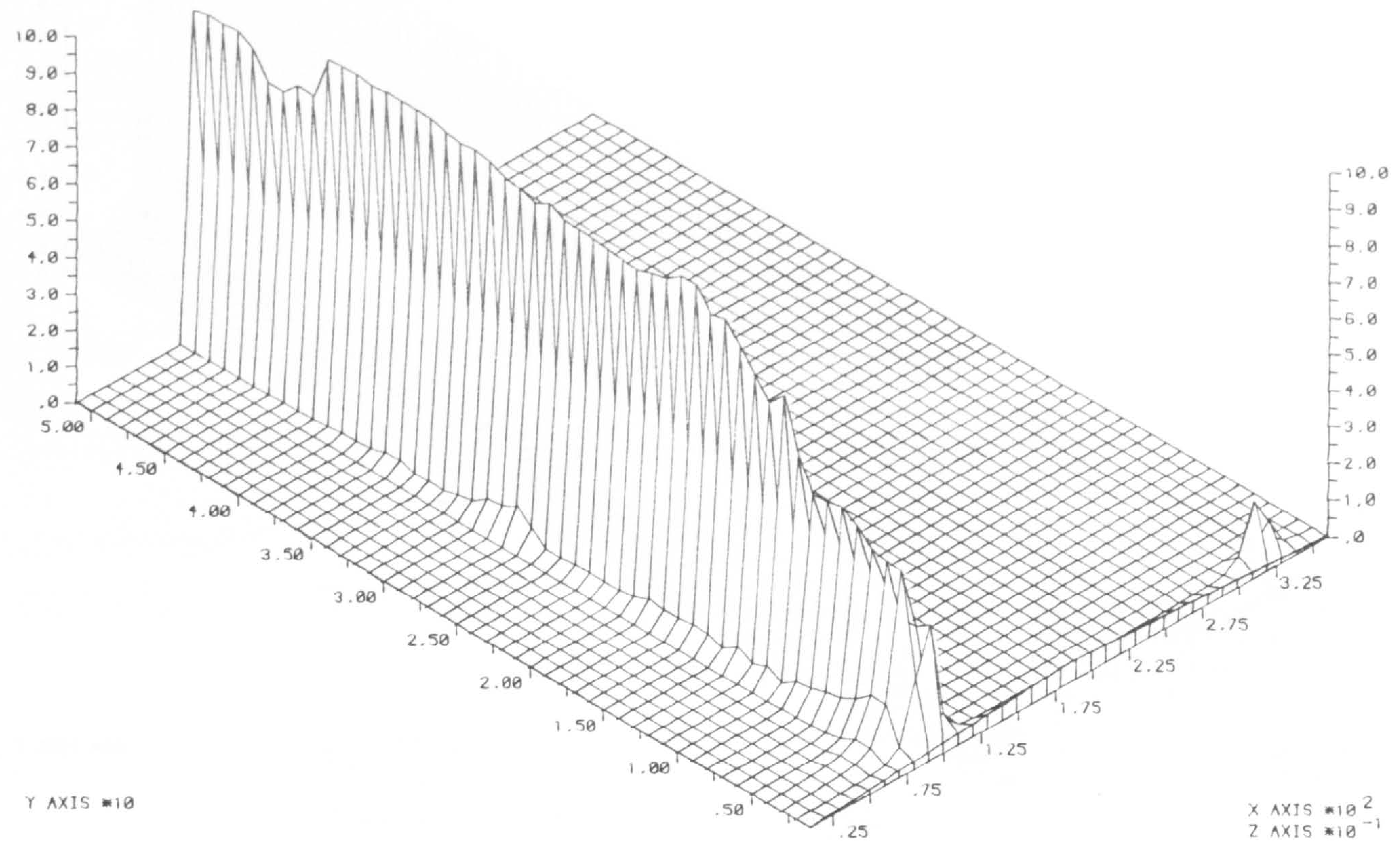

10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.25
.75
1.25
1.75
2.25
2.75
3.25
Y AXIS *10
X AXIS *10 2
Z AXIS *10 -1

FIGURE 4.11

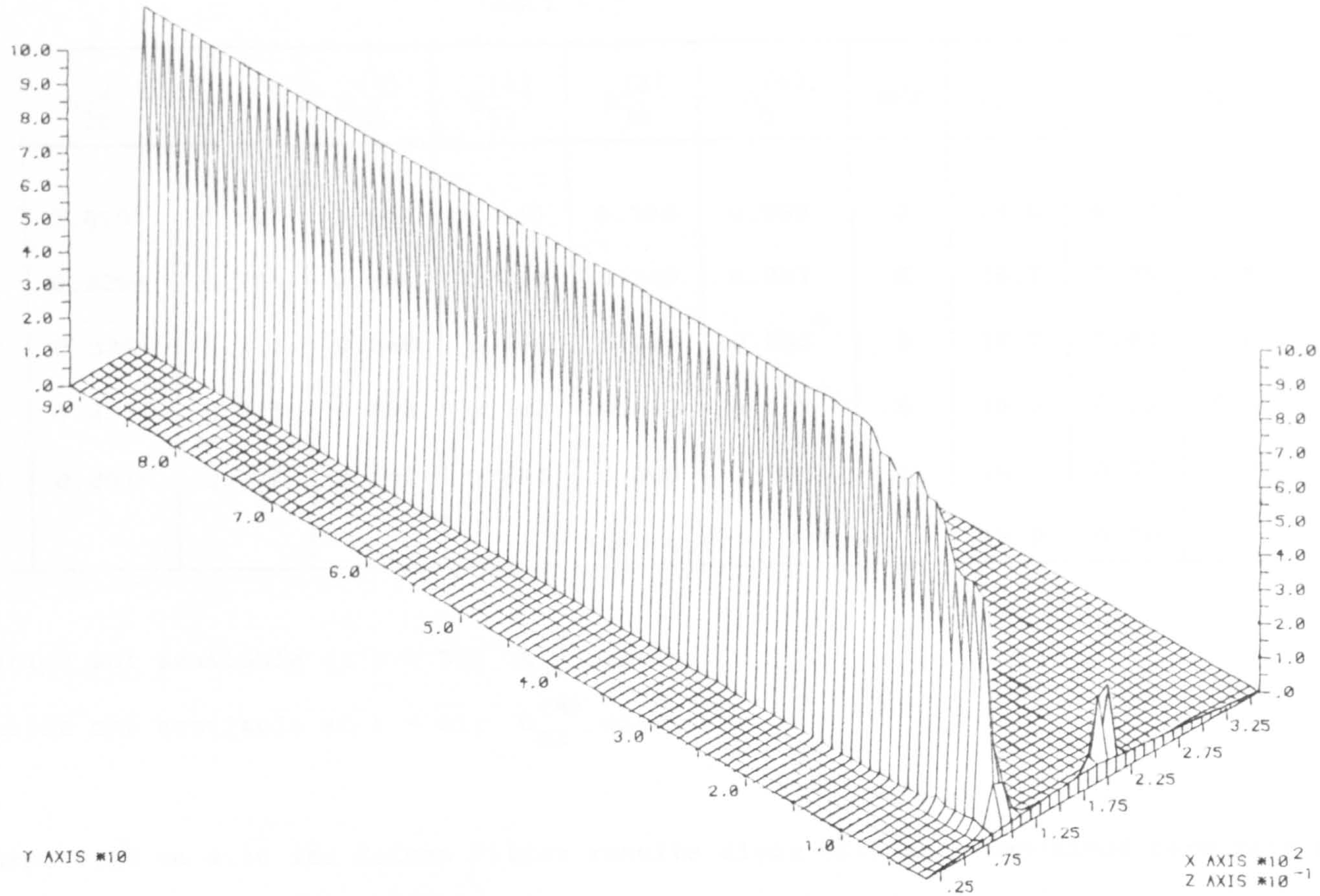

10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.25
.75
1.25
1.75
2.25
2.75
3.25
Y AXIS *10
X AXIS *10 2
Z AXIS *10 -1

FIGURE 4.12

4.4.3 AR(1)

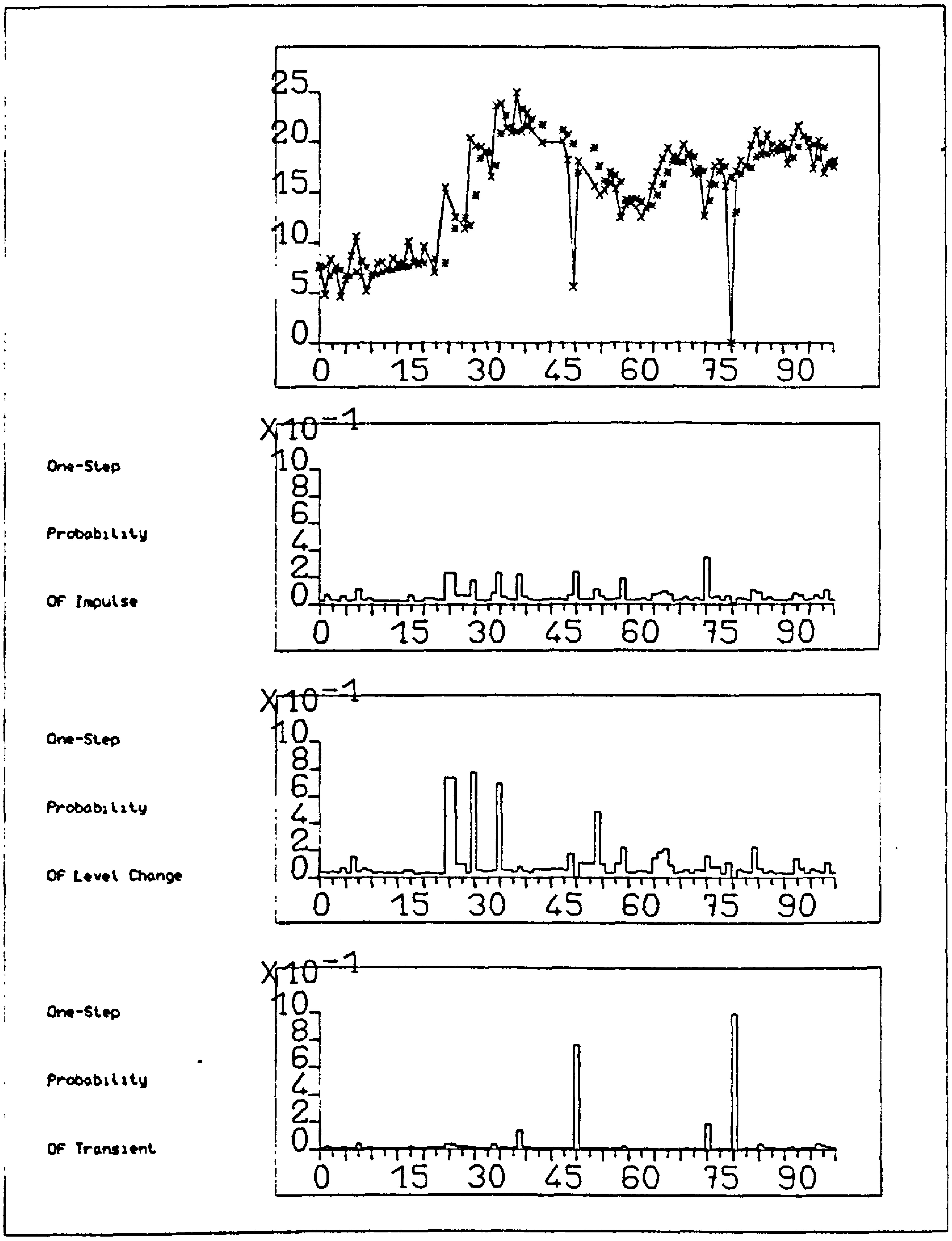
TABLE 4.3

	$O_{26}^{(2)}$	$O_{31}^{(3)}$	$O_{36}^{(3)}$	$O_{51}^{(4)}$	$O_{76}^{(2)}$	$O_{81}^{(4)}$	NFP	v_{100}	ϕ	MAD
Original Series	0.679	0.842	0.742	0.840	0.298	0.999	3	18.6	0.72	1.8
Series 1	0.229*	0.771	0.686	0.764	0.340	0.997	5	18.7	0.68	1.8
Series 2	0.229*	0.771	0.686	0.764	0.344	0.893†	5	18.7	0.46	2.0
Series 3	0.235*	0.773	0.698	0.763	0.313	0.419†	6	19.0	0.32	2.6
Series 4	0.261*	0.776	0.643	0.894	0.309	1.000	5	18.6	0.75	1.8
				Theoretical Values:				18.9	0.70	

*Observation not available at t = 26; $O_{27}^{(2)}$ used.

†Observation not available at t = 81; $O_{82}^{(4)}$ used.

See Figures 4.13 to 4.16 for Kalman Filter results along with one-step-ahead forecasts for Series 1 to 4 respectively; Figures 4.17 to 4.20 show on-line estimation of the ϕ-grid for these series.



25
20
15
10
5
0
0 15 30 45 60 75 90
One-Step
Probability
Of Impulse
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 15 30 45 60 75 90

FIGURE 4.13

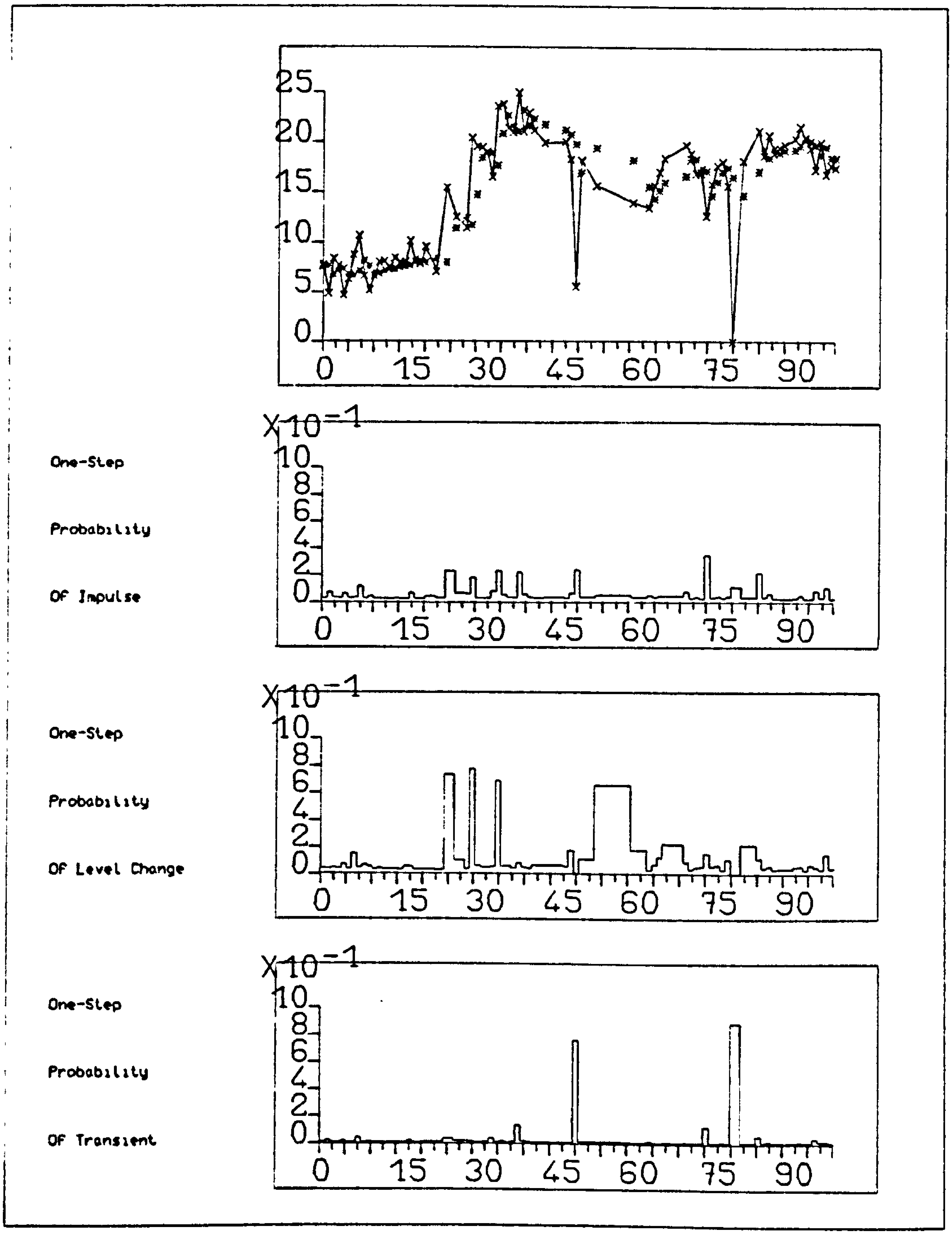
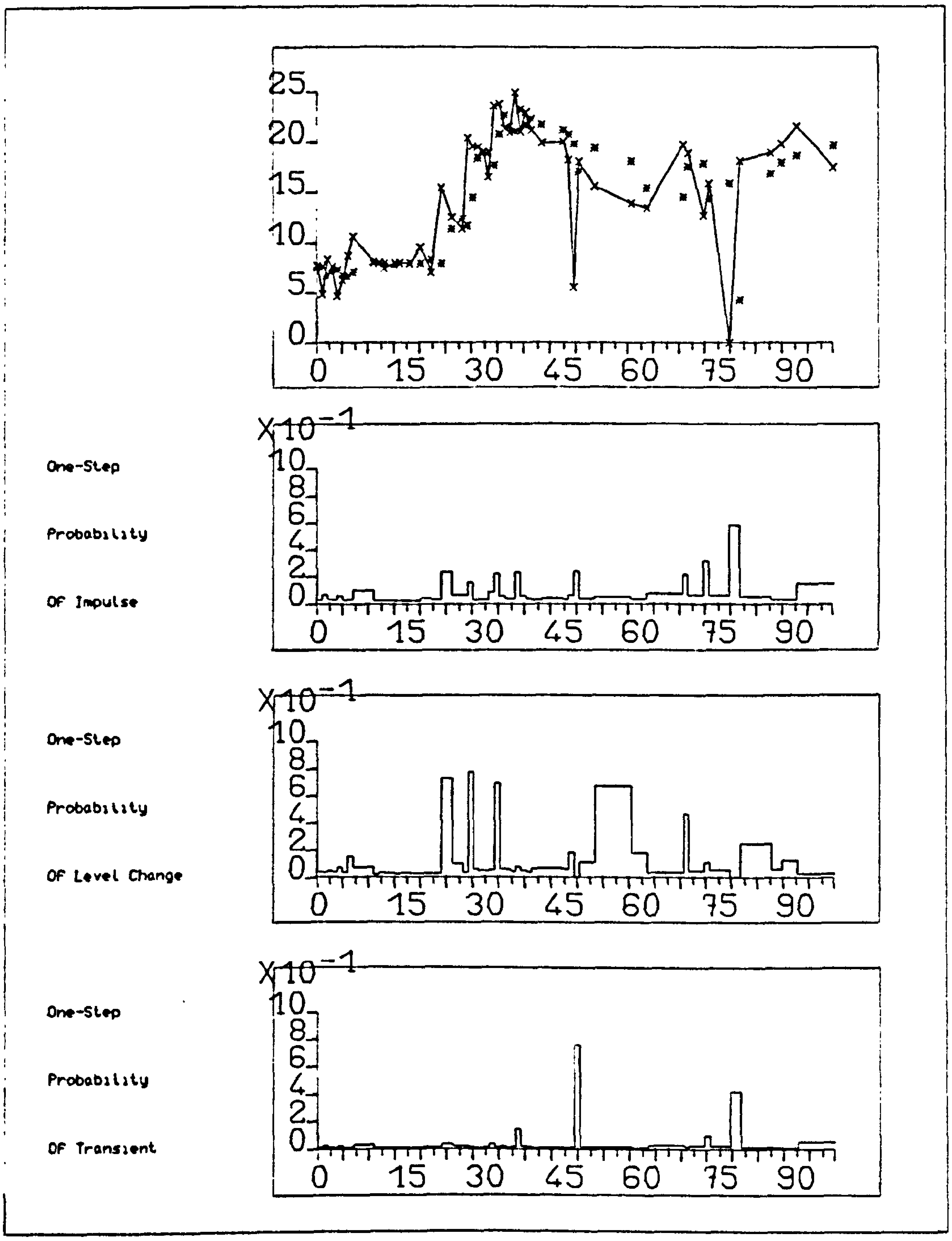

One-Step
Probability
Of Impulse
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient

FIGURE 4.14



One-Step
Probability
Of Impulse
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient

FIGURE 4.15

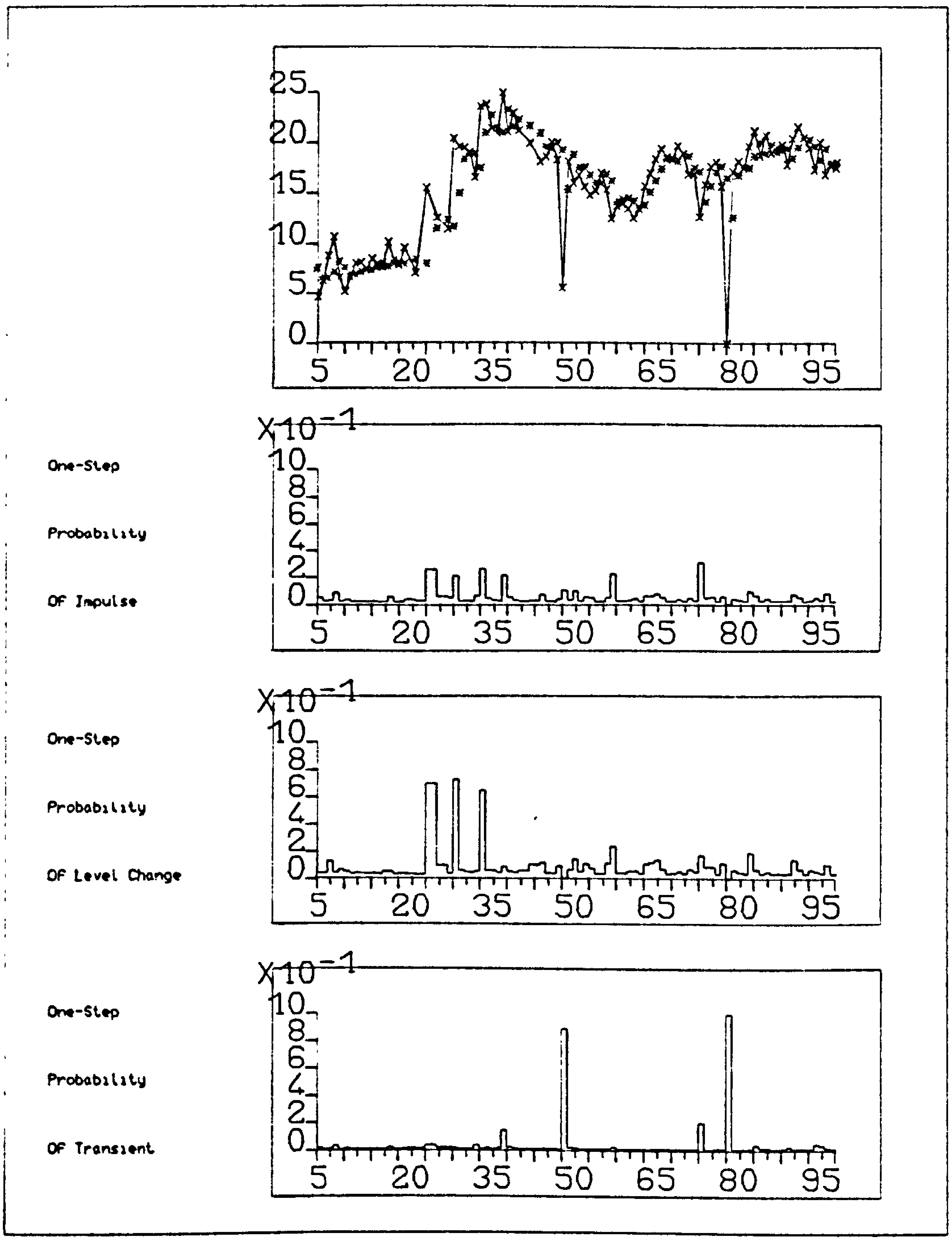

One-Step
Probability
Of Impulse
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient
X10^-1

FIGURE 4.16

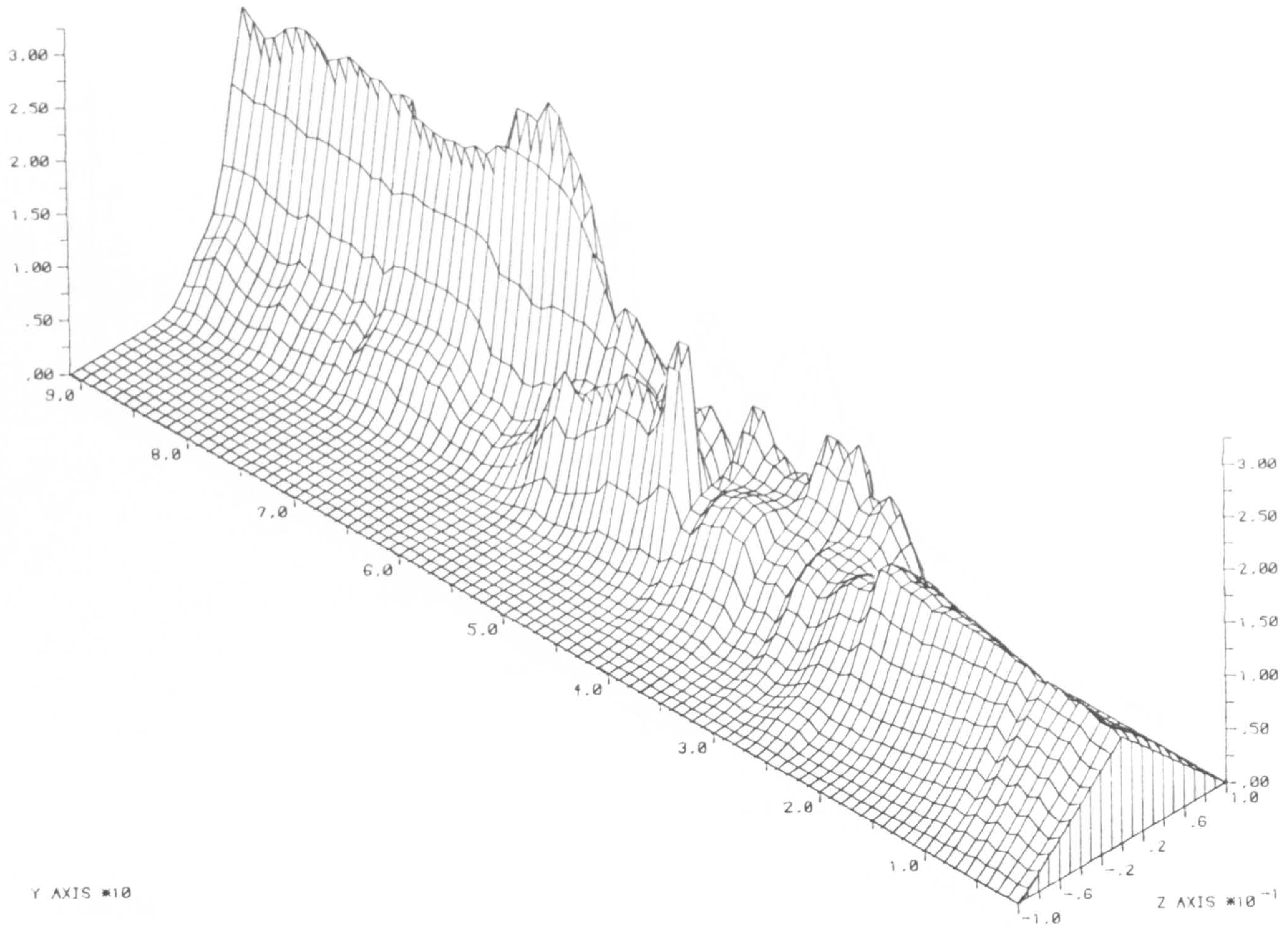

3.00
2.50
2.00
1.50
1.00
.50
.00
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
1.0
.6
.2
-.2
-.6
-1.0
Z AXIS *10 -1

FIGURE 4.17

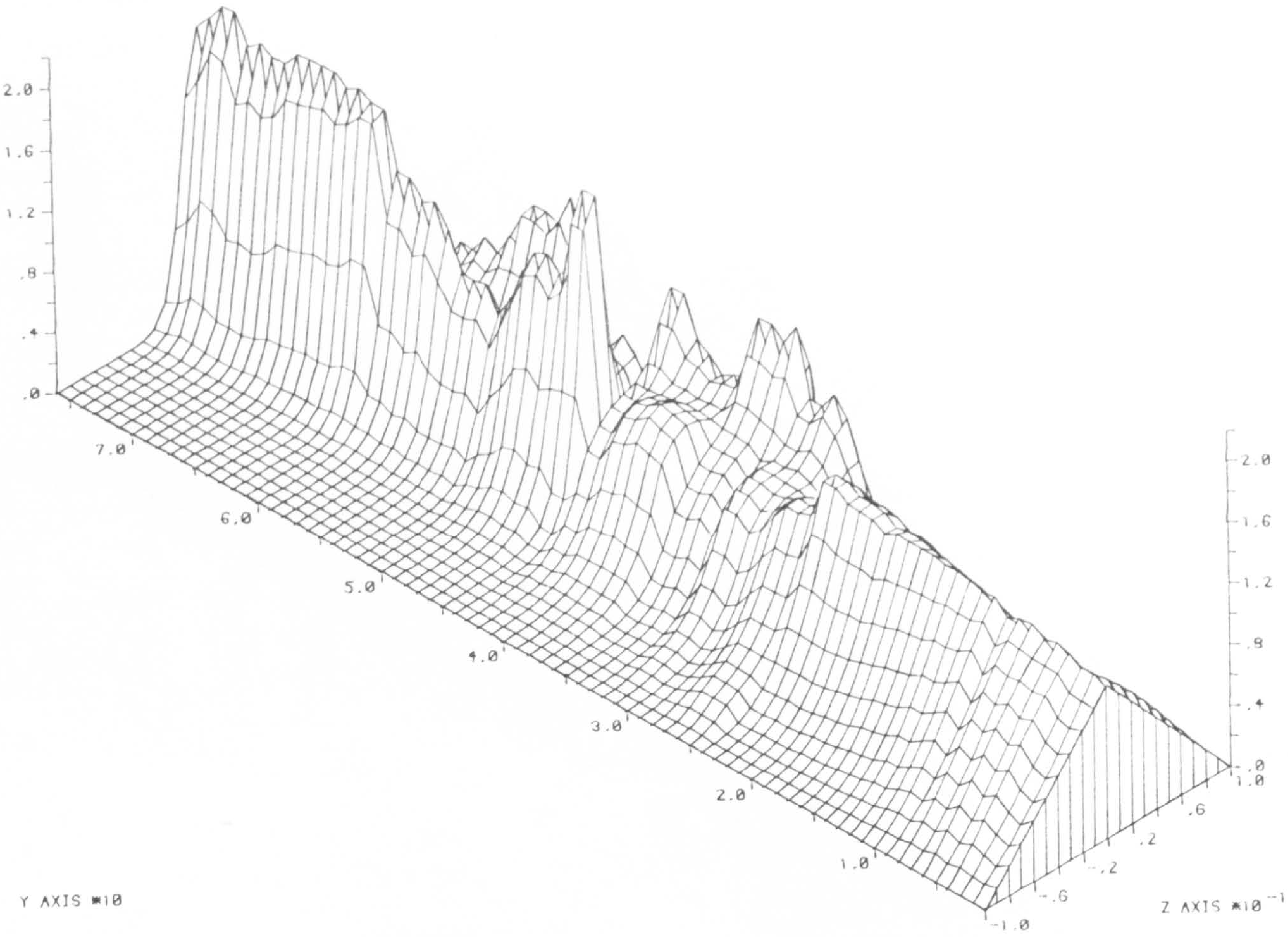

2.0
1.6
1.2
.8
.4
.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
2.0
1.6
1.2
.8
.4
.0
1.0
.6
.2
-.2
-.6
-1.0
Z AXIS *10^-1

FIGURE 4.18

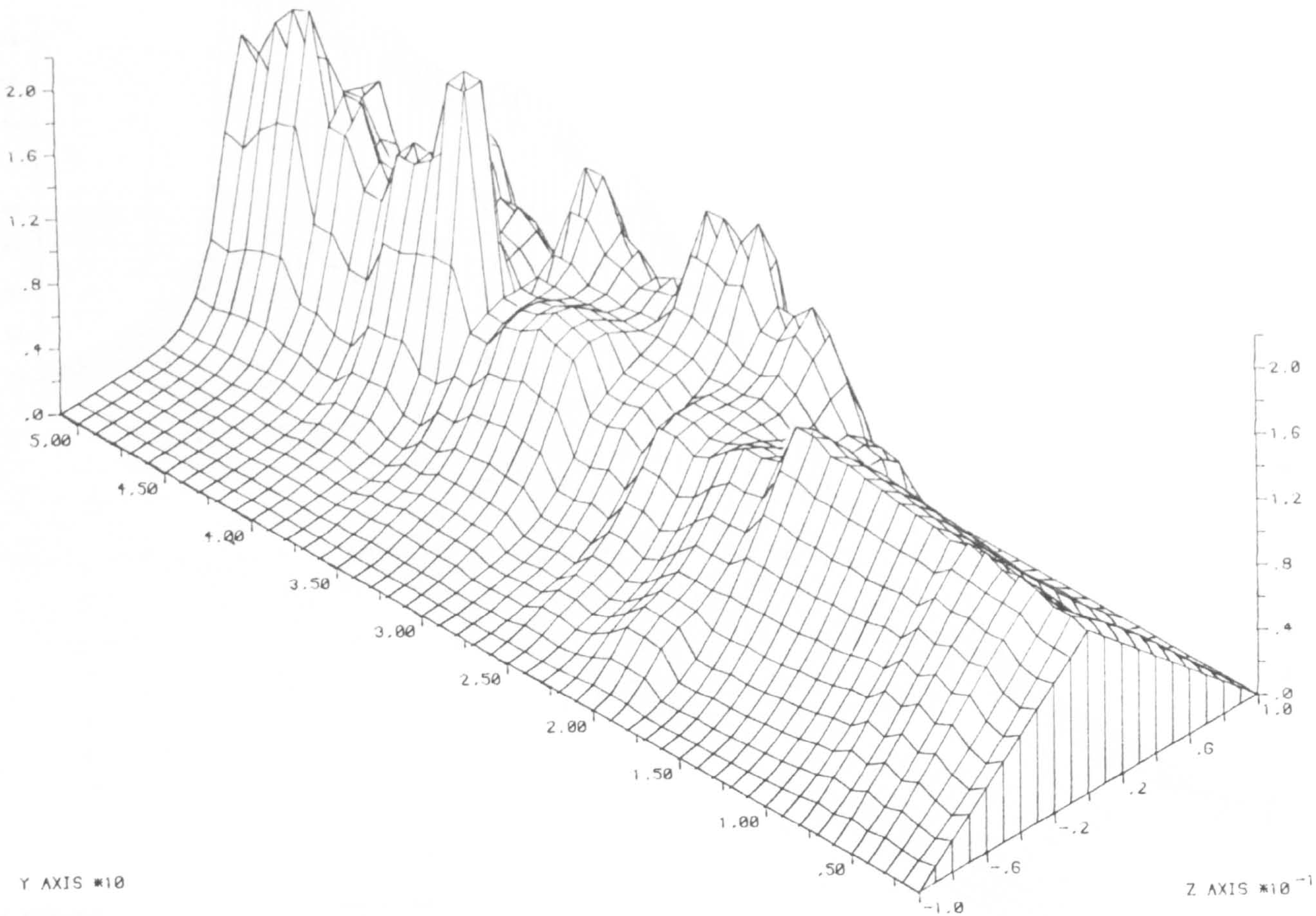

2.0
1.6
1.2
.8
.4
.0
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
Y AXIS *10
2.0
1.6
1.2
.8
.4
.0
1.0
.6
.2
-.2
-.6
-1.0
Z AXIS *10^-1

FIGURE 4.19

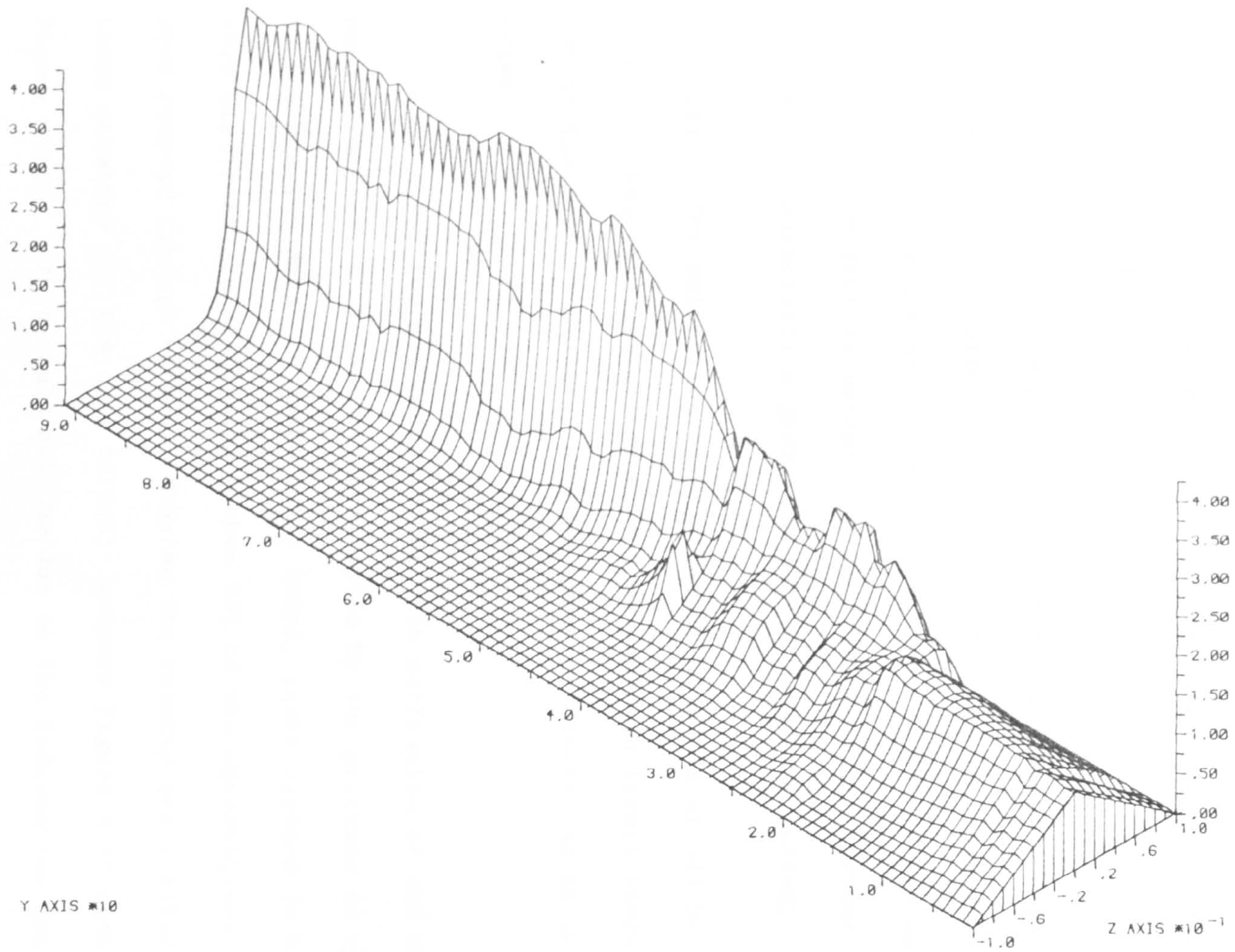

4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.00
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
-1.0
-.6
-.2
.2
.6
1.0
Z AXIS *10⁻¹

FIGURE 4.20

4.4.4 CONCLUSIONS

The results presented in the previous section suggest the following conclusions:

(i) Early unequal-spacing in the time series has no obvious detrimental effects on estimation, forecasting or event detection when compared with the more typical sampling patterns (Series 1 and Series 4).

(ii) An increase in the 'proportion' of unequally-spaced data has an adverse effect on forecasting ability, represented by the mean absolute deviation, although with $\leq$ 25% of the data missing this effect is minimal. With > 25% missing, however, the MAD is appreciably higher than for the complete series.

(iii) The estimation of $\underset{\sim}{\theta}$ is hardly affected at all by unequally-spaced data, although when we approach levels where 50% of the data is missing, minimal adverse effects can be detected.

(iv) For the sinusoidal model, the estimation of the nuisance parameter, ϕ, is largely unaffected by the presence of up to 50% missing data. For the AR(1) model, there appears to be more difficulty in estimating ϕ when 25% of the observations have been removed (though with 10% missing the results are similar to those obtained for the full series; compare Figure 4.17 with Figure (3.73). This could well be due to the information lost as a result of a misinterpretation of changepoint type (see next paragraph) which leads to poor estimates of first-order autocorrelations (see Appendix A4.2).

(v) The results show that event detection is largely unaffected when the event itself is of a sustained type. If, however, the event has only a 'short-term' influence (e.g. impulses in ARMA models) there is a greater chance of misinterpreting the changepoint-type when observations are missing. For instance, an impulse may look very much like a transient if the immediately subsequent observations showing a gradual return to the steady state are not available. In this situation the information about ϕ, which would have been available during the return-to-stability period, is now unavailable (see, for instance in Figure 3.73, the sudden shift in both location and height of the ϕ-grid around $t = 26$, i.e. immediately following the induced impulse). Moreover it is possible that we might miss the real signal (or, at least, the magnitude of the relevant probability may be lower) and, instead, signal a different changepoint-type resulting in a false positive. The number of false positives may also be increased by apparent discontinuities which are merely a feature of long gaps between recordings.

oOo

APPENDIX FOUR

Further Notes on Unequal-Spacing

A4.1 GENERAL FORMS OF $\tilde{G}$ FOR UNEQUALLY-SPACED DATA

A4.1.1 POLYNOMIAL GROWTH

The linear and quadratic growth models of Sections 4.3.1 and 4.3.2 fall within the general framework of polynomial growth models; linear growth being the first-order model, and quadratic growth the second-order model.

For unequally-spaced data:

$$\tilde{G}_k = \tilde{G}^{d_k} = \begin{pmatrix} 1 & d_k \\ 0 & 1 \end{pmatrix} \quad \text{for the linear growth model}$$

and

$$\tilde{G}_k = \begin{pmatrix} 1 & d_k & \frac{d_k(d_k+1)}{2} \\ 0 & 1 & d_k \\ 0 & 0 & 1 \end{pmatrix}$$

for the quadratic growth model.

Lemma A4.1. For a general polynomial growth model of order n, we have

$$\tilde{G}_k = \tilde{G}^{d_k} = \begin{pmatrix} 1 & d_k & \frac{d_k(d_k+1)}{2} & \cdots & \frac{d_k(d_k+1)(d_k+2)\cdots(d_k+n-1)}{n!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{d_k(d_k+1)}{2} \\ & \mathbb{0} & & \ddots & d_k \\ & & & & 1 \end{pmatrix} \tag{4.74}$$

where d_k is the current time interval in units, $d_k \in Z^+$.

Proof: Suppose (4.74) holds for interval d_k. Then, for an interval of $d_k + 1$ we have:

$$\underset{\sim}{G}'_k = \underset{\sim}{G}\underset{\sim}{G}_k = \begin{pmatrix} 1 & & \underset{\sim}{1} \\ & \ddots & \\ \underset{\sim}{0} & & 1 \end{pmatrix} \begin{pmatrix} 1 & d_k & \frac{d_k(d_k+1)}{2} & \cdots & \frac{d_k(d_k+1)\ldots(d_k+n-1)}{n!} \\ & \ddots & \ddots & & \vdots \\ & & & \ddots & d_k \\ \underset{\sim}{0} & & & & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & d_k+1 & \frac{d_k(d_k+1)}{2}+d_k+1 & \cdots & \frac{d_k(d_k+1)\ldots(d_k+n-1)}{n!}+\ldots+d_k+1 \\ & \ddots & d_k+1 & & \vdots \\ & & \ddots & \ddots & d_k+1 \\ \underset{\sim}{0} & & & & 1 \end{pmatrix}$$

The (p + 1)th element of the first row of this matrix is:

$$1 + d_k + \frac{d_k(d_k+1)}{2!} + \frac{d_k(d_k+1)(d_k+2)}{3!} + \ldots + \frac{d_k(d_k+1)(d_k+2)\ldots(d_k+p-1)}{p!} \qquad (n \geq p \geq 1)$$

$$= (1+d_k)\{1 + \frac{d_k}{2} + \frac{d_k(d_k+2)}{3!} + \ldots + \frac{d_k(d_k+2)\ldots(d_k+p-1)}{p!}\}$$

$$= (d_k+1)\frac{(2+d_k)}{2}\{1 + \frac{d_k}{3} + \ldots + \frac{d_k(d_k+3)\ldots(d_k+p-1)}{(p!/2!)}\}$$

$$= (d_k+1)\frac{(d_k+2)}{2}\,\frac{(d_k+3)}{3}\ldots\frac{(d_k+p-1)}{p-1}(1 + \frac{d_k}{(p!/(p-1)!)})$$

$$= \frac{(d_k+1)(d_k+2)(d_k+3)\ldots(d_k+p)}{p!}$$

i.e. $\underset{\sim}{G}'_k$ is of the form of $\underset{\sim}{G}_k$ with d_k replaced by $d_k + 1$. So if $\underset{\sim}{G}_k$

is correct for d_k, it is also correct for $d_k + 1$; but for $d_k = 1$ we have:

$$\underset{\sim}{G}_k = \begin{pmatrix} 1 & & \\ & \ddots & \underset{\sim}{1} \\ \underset{\sim}{0} & & \ddots \; 1 \end{pmatrix} = \underset{\sim}{G}$$

and, by induction, $\underset{\sim}{G}_k$ holds for any integer $d_k \geq 1$.

A4.1.2 ARMA MODELS

Extending the structure of the ARMA models described in Section 2.3.5 we see that the $\underset{\sim}{G}$ matrix for an ARMA(p,q) model is of the form:

$$\underset{\sim}{G} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots\cdots & \phi_p & 1 - \sum_{i=1}^{p} \phi_i \\ 1 & \ddots & & \underset{\sim}{0} & \vdots \\ \vdots & & \ddots\; 1 & \ddots\; 0 & 0 \\ 0 & \underset{\sim}{0} & & 0 \quad \ddots & 1 \end{pmatrix} \tag{4.75}$$

for the equally-spaced case.

Lemma A4.2.

$$\underset{\sim}{G}_k = \underset{\sim}{G}^{d_k} = \underset{\sim}{\Omega}\underset{\sim}{\Lambda}^{d_k}\underset{\sim}{\Omega}^{-1} = \underset{\sim}{\Omega}\begin{pmatrix} 1 & & \\ & \lambda_2^{d_k} & & \underset{\sim}{0} \\ & & \ddots & \\ \underset{\sim}{0} & & & \lambda_{p+1}^{d_k} \end{pmatrix}\underset{\sim}{\Omega}^{-1} \tag{4.76}$$

where the λ's satisfy:

$$\{(-\lambda)^p + \sum_{j=1}^{p} (-1)^{j+1}(-\lambda)^{p-j}.\phi_j\} = 0 \tag{4.77}$$

and where

$$\underset{\sim}{\Omega} = \begin{pmatrix} 1 & \lambda_2^{p-1} & \cdots & \lambda_{p+1}^{p-1} \\ 1 & \lambda_2^{p-2} & \cdots & \lambda_{p+1}^{p-2} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_2 & \cdots & \lambda_{p+1} \\ 1 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \end{pmatrix} \tag{4.78}$$

Proof: $\underset{\sim}{G}^{d_k}$ can be written in the form $\underset{\sim}{\Omega}\underset{\sim}{\Lambda}^{d_k}\underset{\sim}{\Omega}^{-1}$ where $\underset{\sim}{\Lambda}$ is a matrix with the eigenvalues of $\underset{\sim}{G}$ on its diagonal and zeroes elsewhere, and $\underset{\sim}{\Omega}$ is a matrix of corresponding eigenvectors.

The eigenvalues, λ, of $\underset{\sim}{G}$ satisfy:

$$|\underset{\sim}{G} - \lambda\underset{\sim}{I}| = 0$$

i.e.

$$\begin{vmatrix} \phi_1 - \lambda & \phi_2 & \cdots & \phi_p & 1 - \sum_{i=1}^{p} \phi_i \\ 1 & -\lambda & 0 & \cdots & 0 \\ & 1 & -\lambda & 0 \cdots & 0 \\ & & \ddots & \ddots & \underset{\sim}{0} \\ \underset{\sim}{0} & & 1 & -\lambda & 0 \\ & & & 0 & 1 - \lambda \end{vmatrix} = 0$$

$$\Rightarrow (1 - \lambda) \begin{vmatrix} \phi_1 - \lambda & \phi_2 & \cdots & \phi_p \\ 1 & -\lambda & & \underset{\sim}{0} \\ & \ddots & \ddots & \\ \underset{\sim}{0} & & 1 & -\lambda \end{vmatrix} = 0$$

$$\Rightarrow (1 - \lambda)\left\{(\phi_1 - \lambda) \begin{vmatrix} -\lambda & & \underset{\sim}{0} \\ 1 & \ddots & \\ \underset{\sim}{0} & 1 & -\lambda \end{vmatrix} - \begin{vmatrix} \phi_2 & \cdots & \phi_p \\ 1 & -\lambda & \underset{\sim}{0} \\ \underset{\sim}{0} & 1 & -\lambda \end{vmatrix}\right\} = 0 \tag{4.79}$$

Let

$$D_{p-1} = \begin{vmatrix} -\lambda & & & \underset{\sim}{0} \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ \underset{\sim}{0} & & 1 & -\lambda \end{vmatrix}_{(p-1) \times (p-1)} \tag{4.80}$$

and

$$M_1 = \begin{vmatrix} \phi_1 & \cdots & \cdots & \phi_p \\ 1 & -\lambda & & \underset{\sim}{0} \\ & \ddots & \ddots & \\ \underset{\sim}{0} & & 1 & -\lambda \end{vmatrix} \tag{4.81}$$

Then

$$D_{p-1} = -\lambda \begin{vmatrix} -\lambda & & & \underset{\sim}{0} \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ \underset{\sim}{0} & & 1 & -\lambda \end{vmatrix}_{(p-2) \times (p-2)}$$

$$= -\lambda D_{p-2}$$

Clearly, $D_1 = -\lambda$, so that

$$D_{p-1} = (-\lambda)^{p-1} \tag{4.82}$$

Also

$$M_2 = \phi_2 \begin{vmatrix} -\lambda & & & \underset{\sim}{0} \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ \underset{\sim}{0} & & 1 & -\lambda \end{vmatrix}_{(p-2) \times (p-2)}$$

$$- \begin{vmatrix} \phi_3 & \cdots & \cdots & \phi_p \\ 1 & -\lambda & & \underset{\sim}{0} \\ & \ddots & \ddots & \\ \underset{\sim}{0} & & 1 & -\lambda \end{vmatrix}$$

$$= \phi_2 D_{p-2} - M_3$$

$$= \phi_2 D_{p-2} - \phi_3 D_{p-3} + \ldots$$

$$= \phi_2(-\lambda)^{p-2} - \phi_3(-\lambda)^{p-3} + \phi_4(-\lambda)^{p-4} + \ldots$$

$$= \sum_{j=2}^{p} \phi_j(-\lambda)^{p-j}(-1)^j \tag{4.83}$$

Substituting (4.82) and (4.83) into (4.79) we have:

$$|\underset{\sim}{G} - \lambda \underset{\sim}{I}| = (1 - \lambda)\{(\phi_1 - \lambda)(-\lambda)^{p-1} - \sum_{j=2}^{p} \phi_j(-\lambda)^{p-j}(-1)^j\}$$

$$= (1 - \lambda)\{(-\lambda)^p + \sum_{j=1}^{p} \phi_j(-\lambda)^{p-j}(-1)^{j+1}\}.$$

In order to see that $\underset{\sim}{\Omega}$, given by (4.78), contains eigenvectors of $\underset{\sim}{G}$ we note that:

$$\underset{\sim}{G}\underset{\sim}{\Omega} = \begin{pmatrix} \phi_1 & \cdots & \phi_p & 1 - \sum_{i=1}^{p} \phi_i \\ 1 & \ddots & \underset{\sim}{0} & \\ & \ddots & & \\ & 1 & 0 & \\ \underset{\sim}{0} & & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \lambda_2^{p-1} & \cdots & \lambda_{p+1}^{p-1} \\ \vdots & \vdots & & \vdots \\ \vdots & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \phi_1\lambda_2^{p-1} + \phi_2\lambda_2^{p-2} + \ldots + \phi_p & \cdots & \phi_1\lambda_{p+1}^{p-1} + \ldots + \phi_p \\ 1 & \lambda_2^{p-1} & \cdots & \lambda_{p+1}^{p-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_2 & \cdots & \lambda_{p+1} \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

But, from (4.77),

$$\sum_{j=1}^{p} \phi_j \lambda_i^{p-j} = \lambda_i^p \quad \forall i$$

So,

$$\underset{\sim}{G}\underset{\sim}{\Omega} = \begin{pmatrix} 1 & \lambda_2^p & \cdots & \lambda_{p+1}^p \\ 1 & \lambda_2^{p-1} & \cdots & \lambda_{p+1}^{p-1} \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \end{pmatrix} = \underset{\sim}{\Omega}\underset{\sim}{\Lambda}.$$

A4.2 PROBLEMS ASSOCIATED WITH IDENTIFICATION OF ARMA MODELS

When $R_\varepsilon = 0$ in the ARMA models described earlier we obtain an ARMA structure for the observations themselves. In this case we may wish to use an observed time series to identify which ARMA model is most suitable for the variable under study. For the conventional, equally-spaced case, sample autocorrelations have proved to be popular tools for preliminary identification of ARMA order. We examine the problems associated with this approach when the time series are unequally-spaced, using the AR(1) model for illustration.

Before looking at sample autocorrelations, we note that the theoretical autocorrelations have an anticipated form, e.g. AR(1):

$$\begin{aligned} y_k &= \mu_k + \varepsilon_k \\ \mu_k - \nu_k &= \phi^{d_k}(\mu_{k-1} - \nu_{k-1}) + \delta\mu_k \\ \nu_k &= \nu_{k-1} + \delta\nu_k \end{aligned}$$

where

$$\delta\mu_k = \sum_{t=T_{k-1}+1}^{T_k} \phi^{T_k - t}\delta\mu_t \quad \text{and} \quad \delta\nu_k = \sum_{t=T_{k-1}+1}^{T_k} \delta\nu_t .$$

For a classical, steady-state, time series we have $\varepsilon_k = \delta\nu_k = 0, \forall k$.

So

$$y_k - \nu = \phi^{d_k}(y_{k-1} - \nu) + \delta\mu_k . \tag{4.84}$$

Let $X_k = y_k - \nu$. Then $X_k = \phi^{d_k} X_{k-1} + \delta\mu_k$.

(i) $$E(X_k) = E(\phi^{d_k}X_{k-1} + \delta\mu_k)$$
$$= \phi^{d_k}E(X_{k-1}) \qquad \text{(since } E(\delta\mu_k) = 0, \forall k)$$
$$= \phi^{d_k+d_{k-1}}E(X_{k-2})$$

As $i \to -\infty$,

$$E(X_k) \to \prod_{i=-\infty}^{k} \phi^{d_i} = 0 \tag{4.85}$$

since $|\phi| < 1$.

(ii) $$\text{Var}(X_k) = E(X_k^2) = E\{(\phi^{d_k}X_{k-1} + \delta\mu_k)^2\}$$
$$= \phi^{2d_k}E(X_{k-1}^2) + 2\phi^{d_k}E(X_{k-1}\cdot\delta\mu_k) + E(\delta\mu_k^2)$$

$$= \phi^{2d_k}E(X_{k-1}^2) + \lambda^{-1}R_\mu \frac{(1 - \phi^{2d_k})}{1 - \phi^2}$$

(since $E(X_s, \delta\mu_t) = 0 \; \forall s < t$ and

$$E(\delta\mu_k^2) = \text{Var}(\delta\mu_k) = \lambda^{-1}R_\mu \frac{(1 - \phi^{2d_k})}{1 - \phi^2})$$

i.e. $$\text{Var}(X_k) = \phi^{2d_k}[\phi^{2d_{k-1}}E(X_{k-2}^2) + \lambda^{-1}R_\mu \frac{(1 - \phi^{2d_{k-1}})}{1 - \phi^2}]$$
$$+ \lambda^{-1}R_\mu \frac{(1 - \phi^{2d_k})}{1 - \phi^2}$$
$$= \phi^{2(d_k+d_{k-1})}E(X_{k-2}^2) + \frac{\lambda^{-1}R_\mu}{1 - \phi^2}(1 - \phi^{2(d_k+d_{k-1})})$$

As $i \to -\infty$,

$$\text{Var}(X_k) \to \phi^{2\sum_{i=-\infty}^{k} d_i} + \frac{\lambda^{-1}R_\mu}{1 - \phi^2}(1 - \phi^{2\sum_{i=-\infty}^{k} d_i})$$
$$= \frac{\lambda^{-1}R_\mu}{1 - \phi^2} \tag{4.86}$$

since $\sum_{i=-\infty}^{k} d_i = \infty$ (as $d_i \geq 1, \forall i$) $\Rightarrow \phi^{2\sum_{i=-\infty}^{k} d_i} = 0$ ($|\phi| < 1$).

(iii) $\gamma_k(h) = E(X_k . X_{k-h})$

$$= E\{(\phi^{T_k - T_{k-h}} X_{k-h} + \sum_{t=T_{k-h}+1}^{T_k} \phi^{T_k - t} \delta\mu_t) . X_{k-h}\}$$

$$= \phi^{T_k - T_{k-h}} E(X_{k-h}^2)$$

(since $E(X_{k-h} . \delta\mu_t) = 0$ for $t > T_{k-h}$)

$$= \frac{\phi^{T_k - T_{k-h}} \lambda^{-1} R_\mu}{1 - \phi^2} \quad \text{(using (4.86))} \tag{4.87}$$

So

$$\rho_k(h) = \frac{\gamma_k(h)}{Var(X_k)} = \phi^{T_k - T_{k-h}}. \tag{4.88}$$

Notice that the 'correlation' depends on k as well as h; for the equally-spaced case: $T_k - T_{k-h} = h, \forall k$, i.e. $\rho_k(h) = \phi^h = \rho(h)$.

Sample autocorrelations have, conventionally, been used to help in the identification of ARMA order. These statistics are based on correlations between successive observations, etc. assuming that these observations are one unit apart. If this is not necessarily true, the meaning of, say, the standard first-order autocorrelation statistic, i.e.

$$\hat{\rho}_1 = \frac{1}{n} \frac{\sum_{i=2}^{n} (y_i - \bar{y})(y_{i+1} - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{4.89}$$

is somewhat meaningless since i is not now an index of time in units. A replacement for (4.89) is not immediately obvious, and it has been found that using merely those observations that <u>are</u> one unit apart in the calculation of $\hat{\rho}_1$ does not produce good estimates of ρ_1.

This problem is, as yet, unresolved. We note, though, that the three-dimensional ϕ-grid plots we obtain from recursive model-fitting may help to provide some clues as to the goodness-of-fit of a specific ARMA model. For instance, Figure 4.21 shows how the estimation of ϕ is 'confused' when the model is inappropriate. In this case we tried to fit an AR(1) model to data simulated from an MA(1)-type mechanism. The outcome suggests natural uncertainty as to the true location of ϕ, and the resulting pattern is clearly very different from that which obtains when the model is appropriate (see, for instance, Figure 4.17).

Similarly, Figure 4.22 shows how the estimation of ϕ, the phase, progresses for the sinusoidal model applied to the same data set. The 'switching' location of ϕ through 180° suggests that the amplitude is wobbling around zero (since $\cos(\phi + 180) = -\cos(\phi)$), i.e. the sine wave is inappropriate.

Jones (1980) has pointed out, however, that likelihood-based techniques for model identification (e.g. Akaike's Information Criterion; Akaike (1974)) can still be used within a state-space framework, by calculating likelihood contributions recursively. In the context of the work described here, we note that these contributions are of the form given by (3.44), even when the time series is unequally-spaced, and so these techniques are to be preferred to autocorrelation-based criteria.

oOo

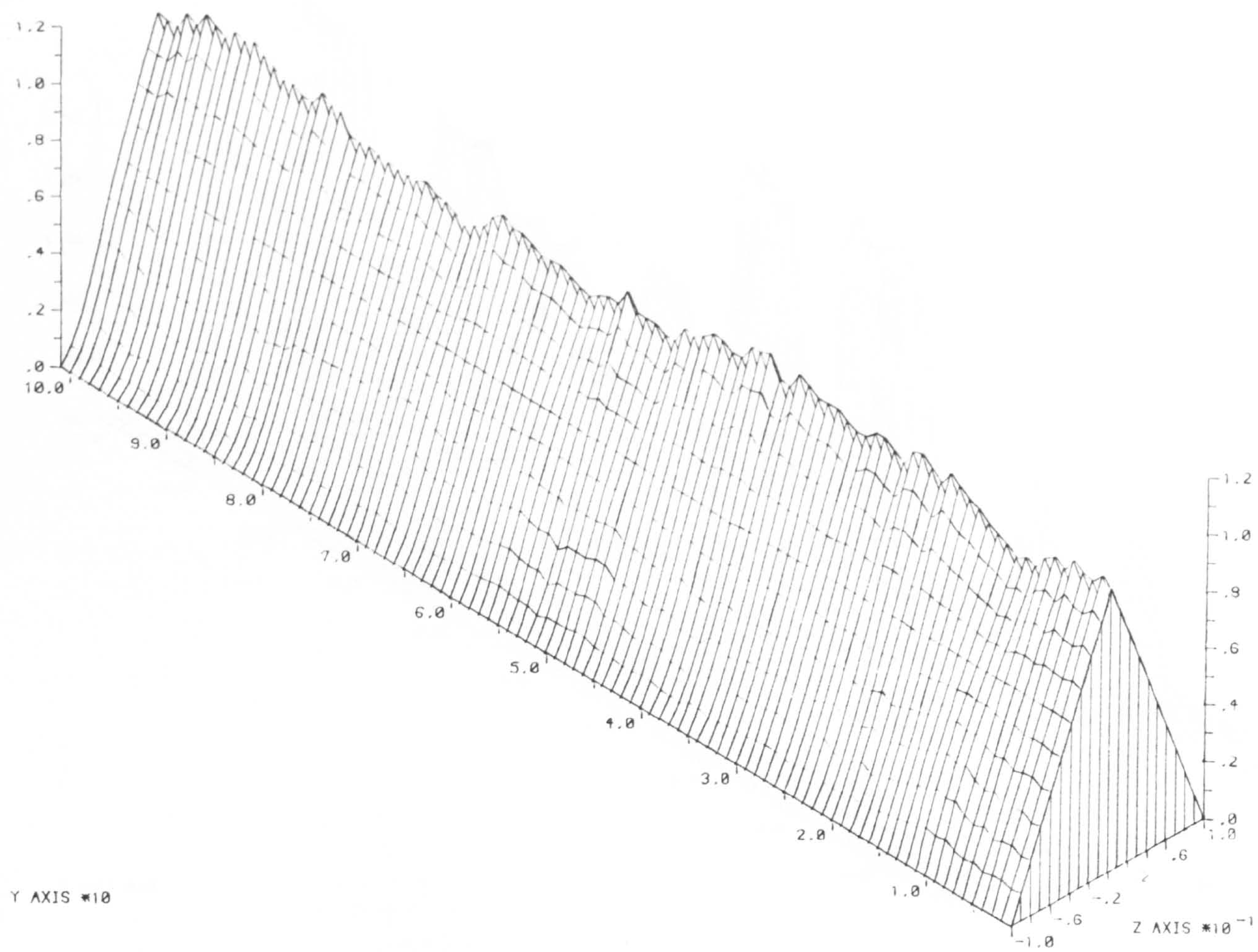


FIGURE 4.21

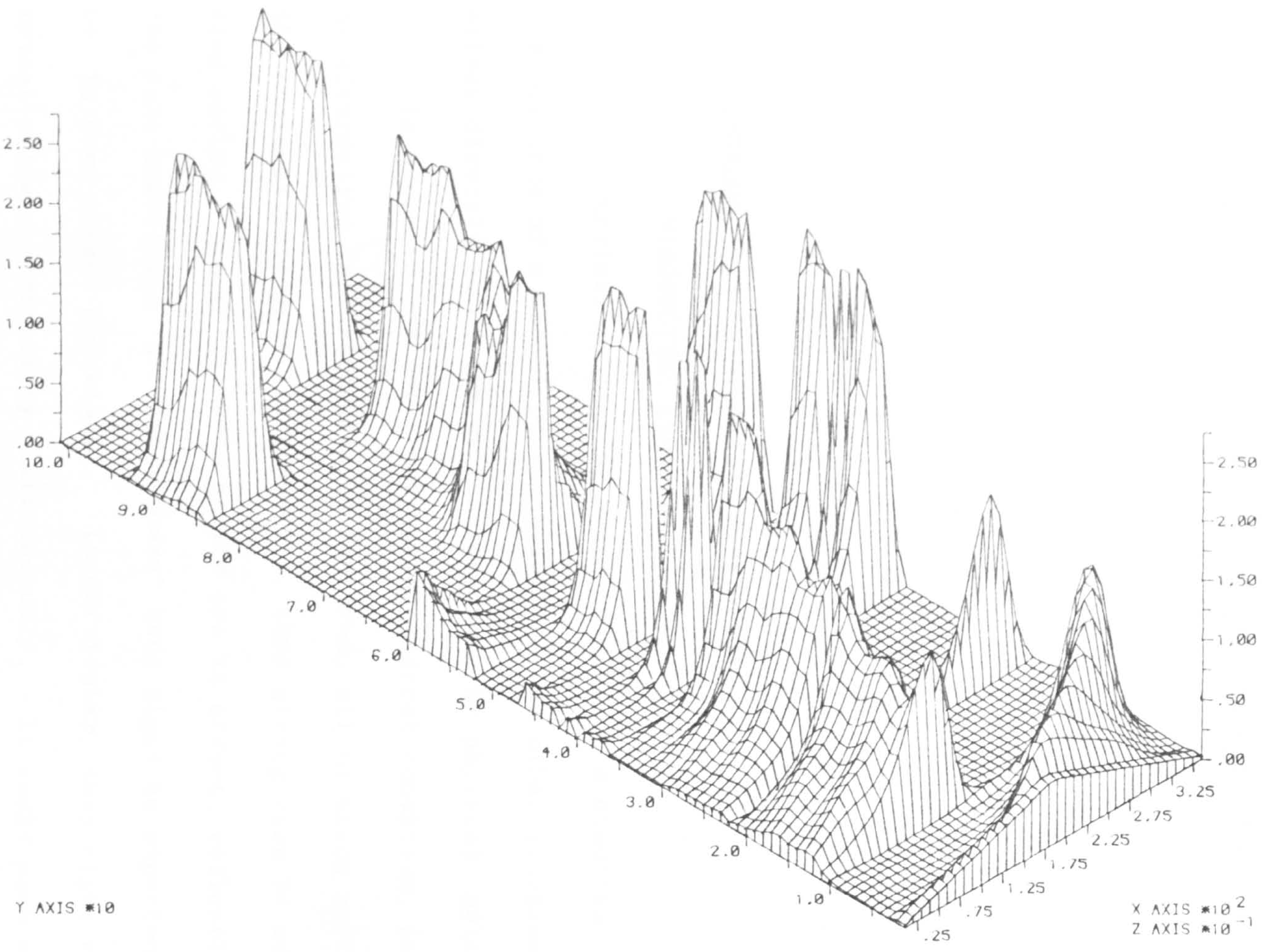

2.50
2.00
1.50
1.00
.50
.00
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.25
.75
1.25
1.75
2.25
2.75
3.25
Y AXIS *10
X AXIS *10 2
Z AXIS *10 -1

FIGURE 4.22

CHAPTER FIVE

BIVARIATE MODELS FOR UNEQUALLY-SPACED DATA

5.1 INTRODUCTION TO BIVARIATE TIME SERIES

5.1.1 INTRODUCTORY REMARKS

Medical monitoring often involves the simultaneous surveillance of a number of physiological and other functions, either directly or by way of a biochemical or physical indicator. In order to monitor a particular medical condition, several alternative indicators may be measured, all of which reflect the state of this specific condition, thus giving rise to multiple time series. Since all these series are in effect, reflecting the same underlying 'disease process' they might be expected to be, in some sense, correlated. In particular, they might each reveal process instability simultaneously. It would seem sensible, therefore, to try to incorporate the possible interconnections between the series into one overall monitoring model (rather than to apply the univariate models to each of the individual series), thereby hoping to refine the discrimination between biological and statistical variability.

In this chapter we shall concentrate on the formulation of recursive bivariate time-series models. Even for the bivariate time series problem (as opposed to the, more general, multiple time series case), the process of model formulation is much more complex than for the univariate case. Interpretation of model parameters, for instance, may not be quite so straightforward (a slope vector does not have as much intuitive appeal as a single slope parameter) and this may be a hindrance in the formulation of changepoint definitions. For example, we might be interested in a change in a slope vector, which is much more difficult to visualize than a change in a single slope parameter. Additionally, we will require much more knowledge about the system under study than in the univariate case, so that the interactions between the two series may be properly modelled.

In line with the 'compartmental' approach to modelling, we shall attempt to combine knowledge of the univariate time series characteristics together with plausible relationships between the two series, including the possibility of causality and feedback (where one, or each, of the series is directly dependent upon some characteristic of the other, perhaps involving a time-lag; Newbold 1979). We must sound a cautionary note, however. The univariate sub-models adopted may bear very little resemblance to the governing mechanism when there is substantial feedback between the series (see Section 5.3.1).

We are interested, too, in constructing a framework that can incorporate missing or, more generally, unequally-spaced data, if the models are to be of practical use in the on-line medical monitoring context. Our main objective is the identification of changepoint phenomena and so, in the following sections, we shall extend the Dynamic Linear Model structure, outlined in the preceding chapters, so that this structure may be used for generally unequally-spaced bivariate time series. The next section gives a brief summary of previous literature related to this topic.

5.1.2 BACKGROUND TO BIVARIATE TIME SERIES MODELLING

Statistical analysis of multiple time series dates back to the work of Whittle (1953,1963) and Quenouille (1957), in which theory was developed to enable the fitting of vector autoregressions to sets of equally-spaced time-related data. This topic has subsequently attracted considerably more attention since the publication of the book by Box and Jenkins (1970), in which the authors described the use of the ARIMA class of models for univariate time series analysis. Many attempts have been made to adapt these techniques to the case of multiple time series, both from the point of view of <u>selecting</u> the appropriate multivariate ARMA model and also <u>fitting</u> the chosen model. Model identification has been examined by, for instance, Haugh and Box (1977), Parzen (1977) and Quinn (1980), while model fitting and estimation have been discussed by, among others, Osborn (1977), Hillmer and Tiao (1979), Nicholls and Hall (1979) and by Anderson (1980). Also of note is the work of Tiao and Box (1981), in which the whole scope of multiple time series analysis, from model identification through to diagnostic checking, is presented for the general class of multivariate ARMA models, while

Newbold (1979) proposes a model-building strategy for bivariate time series involving causality and feedback.

We are more concerned, however, with recursive, time-domain procedures for fitting pre-selected time series models, as has been noted in previous sections. The dynamic linear model framework, as presented by Harrison and Stevens (1976), incorporates the case of multiple observations at a single timepoint, although little attention has been given to the specification of between-series relationships. The technique, however, is recursive and allows for the individual observation series to be modelled very differently: for example, one series might be modelled as a polynomial growth, while the other could be autoregressive in nature. This flexibility is not offered by previous techniques for multiple time series analysis. In addition, the capacity provided by the dynamic linear model to incorporate changepoint phenomena is vital, although an attempt has also been made to allow for 'interventions' using Box-Jenkins techniques by Abraham (1980), where he extends the ideas proposed initially by Box and Tiao (1975) for univariate series.

We would also like to be able to take into account unequally-spaced, or missing, data. There has been very little attention indeed, in the literature, given to the topic of unequally-spaced data in multiple time series. One or two articles, however, have emerged in recent years, most notably those of Robinson (1984) and Jones (1984). Robinson (1984) is mainly concerned with inferences, using a non-recursive approach, about fitted model parameters (i.e. parameter estimation and hypothesis testing) when the time series are irregularly-spaced. Jones (1984), on the other hand, studies the fitting of continuous time autoregressions to irregularly-spaced time series from a recursive viewpoint, extending the ideas presented initially by Jones (1981) whereby model fitting, parameter estimation and transfer function estimation may be carried out for unequally-spaced time series data, using a state-space formulation involving the Kalman Filter. Both of these papers generally assume that, although the data may be irregularly observed, either <u>all</u> components of the observation vector are available at any particular timepoint or <u>none</u> of them are available (though the paper by Mehta and Swamy (1974) does <u>not</u> make this assumption, when the authors examine a Bayesian analysis of a bivariate normal distribution with missing observations). Clearly this is a restriction we do not wish to impose, since component observations from multiple medical time series need not be measured simultaneously.

Finally, on the question of unequally-spaced data, it is worth reiterating a remark made by Robinson (1984), who points out that certain patterns of unequal spacing will result in the unidentifiability of certain models. For example, if an appropriate model for a set of equally-spaced data is the first order moving average model, this model will be unidentifiable if every other observation is missing.

In the next section we outline a general framework within which the dynamic linear model for unequally-spaced univariate time series (as described in Chapter 4) can be extended to the case of generally unequally-spaced bivariate time series. These models may be used for parameter estimation, prediction or change-point detection, though it is the latter which is the main focus of interest in this thesis.

5.2 GENERAL FRAMEWORK

5.2.1 UNDERLYING ASSUMPTIONS

Recall the dynamic linear model given by (2.1) and (2.2):

$$\underset{\sim}{y}_t = \underset{\sim}{H}_t \underset{\sim}{\theta}_t + \underset{\sim}{\varepsilon}_t \tag{5.1}$$

$$\underset{\sim}{\theta}_t = \underset{\sim}{G} \underset{\sim}{\theta}_{t-1} + \underset{\sim}{\omega}_t \tag{5.2}$$

where $\underset{\sim}{y}_t$ is now the observation vector $\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$.

In order to utilize knowledge of the univariate time series characteristics, we shall find it useful to partition many of the vectors and matrices involved into sub-matrices of suitable dimension. We shall write:

$$\left.\begin{array}{l} \underset{\sim}{y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}, \quad \underset{\sim}{\theta}_t = \begin{pmatrix} \underset{\sim}{\theta}_{1t} \\ \underset{\sim}{\theta}_{2t} \end{pmatrix}, \quad \underset{\sim}{H}_t = \begin{pmatrix} \underset{\sim}{h}_{11t} & \underset{\sim}{h}_{12t} \\ \underset{\sim}{h}_{21t} & \underset{\sim}{h}_{22t} \end{pmatrix}, \\ \qquad\qquad \underset{\sim}{G} = \begin{pmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{G}_{12} \\ \underset{\sim}{G}_{21} & \underset{\sim}{G}_{22} \end{pmatrix} \\ \text{and} \\ \underset{\sim}{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \quad \underset{\sim}{\omega}_t = \begin{pmatrix} \underset{\sim}{\omega}_{1t} \\ \underset{\sim}{\omega}_{2t} \end{pmatrix}, \quad \underset{\sim}{R}_\varepsilon = \begin{pmatrix} R_{11} & R_{12} \\ R_{12} & R_{22} \end{pmatrix} \end{array}\right\} \tag{5.3}$$

so that many of the characteristics of the y_1 series might be described by $\underset{\sim}{G}_{11}$, etc. Notice that we need not restrict our attention to the case where y_1 and y_2 arise from the same class of models, since $\underset{\sim}{G}_{11}$ may have an entirely different structure to $\underset{\sim}{G}_{22}$: e.g. $\underset{\sim}{G}_{11}$ could represent the linear growth model (see Section 2.3.1), while $\underset{\sim}{G}_{22}$ could represent the AR(1) model (see Section 2.3.5.1). In what follows we shall assume that the system transition matrix, $\underset{\sim}{G}$, is <u>time-independent</u> for the equally-spaced model, so that the results of Section 4.2.2.1 are applicable. Moreover, in accordance with the philosophy outlined in Section 2.1.2, we shall assume the regression matrix, $\underset{\sim}{H}_t$, to be of the form

$$\underset{\sim}{H}_t = \begin{pmatrix} \underset{\sim}{h}_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{h}_2 \end{pmatrix} \tag{5.4}$$

where $\underset{\sim}{h}_1$ and $\underset{\sim}{h}_2$ are possibly time-dependent. In this way, we restrict y_{1t}, for example, to depend solely upon $\underset{\sim}{\theta}_{1t}$ and not upon $\underset{\sim}{\theta}_{2t}$; any <u>steady-state</u> interrelationships between the two series will be introduced via the system equation, (5.2).

The assumptions P4.4 to P4.6, given in Section 4.2.2, will be retained:

P4.4: $\underset{\sim}{\varepsilon}_t$ is independent of $\underset{\sim}{\varepsilon}_s$, $\forall s \neq t$

P4.5: $\underset{\sim}{\varepsilon}_t, \underset{\sim}{\omega}_t$ are independent of $\underset{\sim}{\theta}_{t-1}$, $\forall t$ (given $\underset{\sim}{D}_{t-1} = (\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_{t-1})$)

P4.6: $\underset{\sim}{\varepsilon}_t$ is independent of $\underset{\sim}{\omega}_t$, $\forall t$.

We shall not, however, assume independence between error <u>components</u>: e.g. ε_{1t} is not necessarily independent of ε_{2t}, etc.

The assumption given by P4.7 will be discussed further in Section 5.2.2.

With regard to unequally-spaced observations, we shall also retain the assumptions outlined in Section 4.2.1:

P4.1: The observation error at timepoint k, $\underset{\sim}{\varepsilon}_k$, is not dependent upon the sampling interval, d_k.

P4.2: d_k is not dependent upon $\underset{\sim}{\theta}_k$

P4.3: The time 'unit' is the smallest conceivable interval, $d_k \in Z^+$, between observations.

Some care must be taken, however, with the definition of d_k. Existing literature on bivariate time series with unequally-spaced data generally assumes that y_{1t} and y_{2t} are either observed together or not at all. We wish to avoid this restriction since, very often in the medical context, only one of the observations may be available at any time t. We, therefore, define d_k as the time in units between the current observation (or observation pair), made at time T_k, and the previous observation (or observation pair), made at time T_{k-1}. In this manner, 'k' denotes the kth observation vector regardless of whether one or both components are actually observed, i.e.

$$\left.\begin{aligned} \underset{\sim}{y}_k &= \begin{pmatrix} y_{1T_k} \\ y_{2T_k} \end{pmatrix} && \text{if both observations are available at time } T_k \\ &= y_{1T_k} && \text{if only } y_1 \text{ is available at time } T_k \\ &= y_{2T_k} && \text{if only } y_2 \text{ is available at time } T_k \end{aligned}\right\} \quad (5.5)$$

e.g. if $\underset{\sim}{y}_{k-1} = y_{1T_{k-1}}$ (i.e. only y_1 observed at time T_{k-1}) and $\underset{\sim}{y}_k = y_{2T_k}$ (i.e. only y_2 observed at time T_k), the current interval will still be $d_k = T_k - T_{k-1}$. If $\underset{\sim}{D}_{k-1} = (\underset{\sim}{y}_1, \underset{\sim}{y}_2, \ldots, \underset{\sim}{y}_{k-1})$ represents the past k - 1 observation vectors (or part-vectors) we may write

$$\underset{\sim}{D}_k = (\underset{\sim}{D}_{k-1}, \underset{\sim}{y}_k) \quad (5.6)$$

with $\underset{\sim}{y}_k$ defined by (5.5).

In terms of handling missing observations in the timepoint-to-timepoint recursion, we retain the ideas proposed in Chapter 4 only if the full observation vector is observed at time T_k: i.e. we bypass the observation equation, (5.1), and update purely the system equation, (5.2), using previous variance estimates where necessary (see Section 4.2.1 for details). The procedure when only one observation component is available at time T_k is outlined in Section 5.2.2.

Finally, in this section, we examine the implications of the introduction of bivariate time series for the multistate structure. Consider, for example, the case where each of the univariate series, y_1 and y_2, can be represented by the linear growth model (see Section 3.3.1). Then, for each series, the simple multistate structure (disregarding changepoint combinations) involves the following four states: steady state, level change, slope change and transient (as described in Section 3.3.1). Therefore, for the bivariate model, we have 16 possible states at any time t: e.g. $M_t^{(1)}$ denotes y_{1t} steady state, y_{2t} steady state; $M_t^{(2)}$ denotes y_{1t} steady state, y_{2t} level change; ...; $M_t^{(16)}$ denotes y_{1t} transient, y_{2t} transient.

In the most general case, if there are J_1 states associated with the univariate model for the y_1 series, and J_2 states associated with the univariate model for the y_2 series, there will be $J_1 \times J_2$ states associated with the corresponding bivariate model. However, problems may arise with the size of the overall model framework when this approach is adopted. Firstly, we note that, at each recursion, $J_1 \times J_2$ prior distributions for $\underset{\sim}{\theta}$ need to be updated to form $J_1^2 \times J_2^2$ posterior distributions, which must then be 'collapsed' back to $J_1 \times J_2$ posterior distributions (see Section 3.2 for details). For the bivariate linear growth model, for example, 256 model possibilities must be incorporated at each recursion. This problem is aggravated further if the model contains nuisance parameters, for example $\underset{\sim}{\phi}$ for the bivariate AR(1) model (in which the y_1 series is AR(1) with autoregressive parameter ϕ_1, and the y_2 series is AR(1) with autoregressive parameter ϕ_2, with $J_1 = J_2 = 4$, see Section 3.3.5.1). If we use a grid with 11 nodes for ϕ_1 and for ϕ_2, there are then 30,976 model possibilities at each recursion! The difficulties associated with this are threefold.

(i) If probabilities are attached to each of the possible models (see Section 3.2 for details), most of these probabilities will be very small indeed. More to the point, if we wish to select the most plausible model at a particular timepoint we must choose one from 30,976 possibilities (in the AR(1) case). It seems likely, therefore, that discrimination between competing models may be poor, implying not only poor changepoint discrimination but also poor $\underset{\sim}{\phi}$ estimation; this conjecture will be investigated in Section 5.4.

(ii) Compounding the problem in (i), we have the possible inaccuracies imposed by the collapsing procedures. For the univariate case, it has been demonstrated that collapsing procedures based on the Kullback-Leibler divergence criterion are reasonably accurate and effective both in the context of forecasting and estimation (see Harrison and Stevens 1975) and in the context of changepoint detection (see Smith and West 1983, and Section 3.4 of this thesis). It is not at all clear, however, how well these procedures will perform given that we need to collapse from 30,976 models down to 16 models at each recursion (as for the AR(1) example).

(iii) The power and accuracy of the computer that is used for implementing these algorithms may be crucial; it is not clear whether or not the computing time necessary will be too long for the models to be of practical use (bearing in mind that we were hoping to use the techniques for on-line detection of time series discontinuities in, perhaps, critical care situations, where d_k might be in the order of a few minutes).

In the next sections, however, we disregard these problems of size and proceed with the development of theoretical results, on the assumption that problems of implementation will eventually be overcome by developments in computing resources. It will be seen, in Section 5.4, that much of our concern about model size can be largely dismissed, when we investigate the performance of specific models on a number of data sets.

5.2.2 RECURSIVE ESTIMATION

In accordance with assumption P4.7 and equations (3.7) and (3.8), concerning error distributions, we shall assume that:

$$\underset{\sim}{\varepsilon}_k \sim N(0,\lambda^{-1}\underset{\sim}{R}_{\varepsilon}) \qquad (5.7)$$

$$\underset{\sim}{\omega}_k \sim N(0,\lambda^{-1}\underset{\sim}{R}_{\omega}(k)) \qquad (5.8)$$

$$\underset{\sim}{\theta}_o \sim N(\underset{\sim}{m}_o, \lambda^{-1}\underset{\sim}{C}_o) \tag{5.9}$$

and that

$$p(\underset{\sim}{\theta}_{k-1}|\underset{\sim}{D}_{k-1}, M_{k-1}^{(i)}) \sim N(\underset{\sim}{m}_{k-1}^{(i)}, \lambda^{-1}\underset{\sim}{C}_{k-1}^{(i)}) \tag{5.10}$$

where

$$\underset{\sim}{m}_{k-1}^{(i)} = \begin{bmatrix} \underset{\sim}{m}_{1k-1}^{(i)} \\ \underset{\sim}{m}_{2k-1}^{(i)} \end{bmatrix} \tag{5.11}$$

By adopting this formulation we restrict ourselves to one of two situations:

(i) λ is known, so that the variances can be completely specified;

(ii) λ is unknown and must be estimated recursively, in which case the ratio of variances, $var(\varepsilon_{1k})/var(\varepsilon_{2k})$, must be known.

In order to appreciate the second alternative we note, from (5.7), using the definition of $\underset{\sim}{R}_\varepsilon$ given by (5.3), that

$$\left.\begin{array}{l} var(\varepsilon_{1k}) = \lambda^{-1}R_{11},\ var(\varepsilon_{2k}) = \lambda^{-1}R_{22} \\ \text{and} \quad cov(\varepsilon_{1k}, \varepsilon_{2k}) = \lambda^{-1}R_{12} \end{array}\right\} \tag{5.12}$$

Since R_{11} and R_{22} are fixed and prespecified, we must also be able to specify their ratio $R_{11}/R_{22} = var(\varepsilon_{1k})/var(\varepsilon_{2k})$.

For the case where the variances are completely unknown, i.e. when their ratio cannot be specified, we must adopt an alternative formulation to that described in (5.7) to (5.12). The natural extension to this approach, for the case of unknown variances, is discussed in Section 5.2.2.3. Firstly, though, we shall consider each of the situations described by (i) and (ii), in turn, in the context of generally unequally-spaced bivariate data.

5.2.2.1 Known Variances

Case 1: y_{1k} and y_{2k} Available at Time T_k

From (5.1) and (5.2), we have:

$$p(\underset{\sim}{y}_k|\underset{\sim}{\theta}_k, M_k^{(j)}) \sim N(\underset{\sim}{H}_{T_k}\underset{\sim}{\theta}_k, \lambda^{-1}\underset{\sim}{R}_\varepsilon^{(j)}) \tag{5.13}$$

and

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{\theta}_{k-1}, M_k^{(j)}) \sim N(\underset{\sim}{G}_k\underset{\sim}{\theta}_{k-1}, \lambda^{-1}\underset{\sim}{R}_\omega^{(j)}(k)) \tag{5.14}$$

where

$$\left.\begin{array}{l} \underset{\sim}{G}_k = \underset{\sim}{G},\ \underset{\sim}{R}_\omega(k) = \underset{\sim}{R}_\omega, \text{ if } d_k = 1 \\ \underset{\sim}{G}_k = \underset{\sim}{G}^{d_k},\ \underset{\sim}{R}_\omega(k) = \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k-t}\underset{\sim}{R}_\omega(\underset{\sim}{G}^{T_k-t})^T, \text{ if } d_k > 1 \end{array}\right\} \tag{5.15}$$

(using (4.6) and (4.8), when the current interval is d_k units). From (5.10) and (5.14), we have:

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_{k-1}, M_{k-1}^{(i)}, M_k^{(j)}) \sim N(\underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)}, \lambda^{-1}[\underset{\sim}{G}_k\underset{\sim}{C}_{k-1}^{(i)}\underset{\sim}{G}_k^T + \underset{\sim}{R}_\omega^{(j)}(k)]) \tag{5.16}$$

Let

$$\underset{\sim}{P}_k^{(ij)} = \underset{\sim}{G}_k\underset{\sim}{C}_{k-1}^{(i)}\underset{\sim}{G}_k^T + \underset{\sim}{R}_\omega^{(j)}(k) \tag{5.17}$$

and

$$\underset{\sim}{D}_k = (\underset{\sim}{D}_{k-1}, \underset{\sim}{y}_k) \text{ with } \underset{\sim}{y}_k = \begin{pmatrix} y_{1k} \\ y_{2k} \end{pmatrix} \tag{5.18}$$

In order to proceed with the recursion we must calculate the posterior distribution for $\underset{\sim}{\theta}_k$:

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)}) \propto p(\underset{\sim}{y}_k|\underset{\sim}{\theta}_k,\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)})p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)})$$

$$\propto \exp[-\frac{\lambda}{2}\{(\underset{\sim}{y}_k - \underset{\sim}{H}_{T_k}\underset{\sim}{\theta}_k)^T(\underset{\sim}{R}_\varepsilon^{(j)})^{-1}(\underset{\sim}{y}_k - \underset{\sim}{H}_{T_k}\underset{\sim}{\theta}_k)$$

$$+ (\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)})^T(\underset{\sim}{P}_k^{(ij)})^{-1}$$

$$\times (\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)})\}]$$

This is the recursion outlined by Harrison and Stevens (1976), for vector $\underset{\sim}{y}_k$, for which the Kalman Filter equations yield the following result:

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)}) \sim N(\underset{\sim}{m}_k^{(ij)},\lambda^{-1}\underset{\sim}{C}_k^{(ij)}) \tag{5.19}$$

where

$$\underset{\sim}{m}_k^{(ij)} = \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)} + \underset{\sim}{S}_k^{(ij)}(\underset{\sim}{y}_k - \underset{\sim}{f}_k^{(i)}) \tag{5.20}$$

$$\underset{\sim}{C}_k^{(ij)} = \underset{\sim}{P}_k^{(ij)} - \underset{\sim}{S}_k^{(ij)}\underset{\sim}{F}_k^{(ij)}(\underset{\sim}{S}_k^{(ij)})^T \tag{5.21}$$

and where $\underset{\sim}{f}_k^{(i)}$, $\underset{\sim}{F}_k^{(ij)}$ and $\underset{\sim}{S}_k^{(ij)}$ are defined by (3.16) with $\underset{\sim}{G}$ replaced by $\underset{\sim}{G}_k$ and $\underset{\sim}{R}_\omega^{(j)}$ replaced by $\underset{\sim}{R}_\omega^{(j)}(k)$. Also

$$p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)}) \sim N(\underset{\sim}{f}_k^{(i)},\lambda^{-1}\underset{\sim}{F}_k^{(ij)}) \tag{5.22}$$

gives the predictive density used for the calculation of multi-state probabilities as well as for providing forecasts.

Notice that the calculation of $\underset{\sim}{m}_k^{(ij)}$, in (5.20), is dependent upon $\underset{\sim}{y}_k = \begin{pmatrix} y_{1k} \\ y_{2k} \end{pmatrix}$.

Case 2: Only y_{1k} Available at Time T_k

Theorem 5.2.1

Consider the dynamic linear model described by

$$\underset{\sim}{y}_k = \underset{\sim}{H}_{T_k}\underset{\sim}{\theta}_k + \underset{\sim}{\varepsilon}_k \tag{5.23}$$

$$\underset{\sim}{\theta}_k = \underset{\sim}{G}_k\underset{\sim}{\theta}_{k-1} + \underset{\sim}{\omega}_k \tag{5.24}$$

where

$$\underset{\sim}{y}_k = \begin{pmatrix} y_{1k} \\ y_{2k} \end{pmatrix}, \ \underset{\sim}{\theta}_k = \begin{pmatrix} \underset{\sim}{\theta}_{1k} \\ \underset{\sim}{\theta}_{2k} \end{pmatrix}, \ \underset{\sim}{H}_{T_k} = \begin{pmatrix} \underset{\sim}{h}_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{h}_2 \end{pmatrix}, \ \underset{\sim}{G}_k = \begin{pmatrix} \underset{\sim}{G}_{11k} & \underset{\sim}{G}_{12k} \\ \underset{\sim}{G}_{21k} & \underset{\sim}{G}_{22k} \end{pmatrix}, \tag{5.25}$$

$$\left.\begin{aligned} p(\underset{\sim}{\varepsilon}_k|M_k^{(j)}) &= p(\begin{pmatrix} \varepsilon_{1k} \\ \varepsilon_{2k} \end{pmatrix}|M_k^{(j)}) \sim N(\underset{\sim}{0},\lambda^{-1}\underset{\sim}{R}_{\varepsilon}^{(j)}), \\ p(\underset{\sim}{\omega}_k|M_k^{(j)}) &\sim N(\underset{\sim}{0},\lambda^{-1}\underset{\sim}{R}_{\omega}^{(j)}(k)) \end{aligned}\right\} \tag{5.26}$$

and

$$\underset{\sim}{R}_{\varepsilon}^{(j)} = \begin{pmatrix} E_{11}^{(j)} & E_{12}^{(j)} \\ E_{12}^{(j)} & E_{22}^{(j)} \end{pmatrix}, \ \underset{\sim}{R}_{\omega}^{(j)}(k) = \begin{pmatrix} \underset{\sim}{W}_{11}^{(j)} & \underset{\sim}{W}_{12}^{(j)} \\ (\underset{\sim}{W}_{12}^{(j)})^T & \underset{\sim}{W}_{22}^{(j)} \end{pmatrix}. \tag{5.27}$$

Assume that λ is fixed and known, and that

$$p(\underset{\sim}{\theta}_{k-1}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)}) \sim N(\underset{\sim}{m}_{k-1}^{(i)},\lambda^{-1}\underset{\sim}{C}_{k-1}^{(i)}) \tag{5.28}$$

where $\underset{\sim}{D}_{k-1} = (\underset{\sim}{D}_{k-2},\underset{\sim}{y}_{k-1})$, $M_{k-1}^{(i)}$ denotes that the system is in state i at timepoint k - 1, and

$$\underset{\sim}{m}_{k-1}^{(i)} = \begin{pmatrix} \underset{\sim}{m}_{1k-1}^{(i)} \\ \underset{\sim}{m}_{2k-1}^{(i)} \end{pmatrix} \tag{5.29}$$

Then, <u>if only y_{1k}, and not y_{2k}, is available at time T_k</u> (i.e. $\underset{\sim}{D}_k = (\underset{\sim}{D}_{k-1},y_{1k})$), the posterior distribution for $\underset{\sim}{\theta}_k$ is given by:

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)}) \sim N(\underset{\sim}{m}_k^{(ij)},\lambda^{-1}\underset{\sim}{C}_k^{(ij)}) \tag{5.30}$$

where

$$\underset{\sim}{m}_k^{(ij)} = \underset{\sim}{G}_k \underset{\sim}{m}_{k-1}^{(i)} + \underset{\sim}{S}_k^{(ij)} (y_{1k} - f_{1k}^{(i)}) \tag{5.31}$$

and

$$\underset{\sim}{C}_k^{(ij)} = \underset{\sim}{P}_k^{(ij)} - \underset{\sim}{S}_k^{(ij)} F_{11k}^{(ij)} (\underset{\sim}{S}_k^{(ij)})^T \tag{5.32}$$

with

$$\left.\begin{aligned}
f_{1k}^{(i)} &= [1 \quad 0] \underset{\sim}{H}_{T_k} \underset{\sim}{G}_k \underset{\sim}{m}_{k-1}^{(i)} \\
\underset{\sim}{P}_k^{(ij)} &= \begin{bmatrix} \underset{\sim}{P}_{11k}^{(ij)} & \underset{\sim}{P}_{12k}^{(ij)} \\ (\underset{\sim}{P}_{12k}^{(ij)})^T & \underset{\sim}{P}_{22k}^{(ij)} \end{bmatrix} = \underset{\sim}{G}_k \underset{\sim}{C}_{k-1}^{(i)} \underset{\sim}{G}_k^T + \underset{\sim}{R}_\omega^{(j)}(k) \\
F_{11k}^{(ij)} &= [1 \quad 0] \underset{\sim}{H}_{T_k}^T \underset{\sim}{P}_k^{(ij)} \underset{\sim}{H}_{T_k} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + E_{11}^{(j)} \\
&= \underset{\sim}{h}_1^T \underset{\sim}{P}_{11k}^{(ij)} \underset{\sim}{h}_1 + E_{11}^{(j)} \\
\text{and} \quad & \\
\underset{\sim}{S}_k^{(ij)} &= \underset{\sim}{P}_k^{(ij)} \underset{\sim}{H}_{T_k}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} (F_{11k}^{(ij)})^{-1}
\end{aligned}\right\} \tag{5.33}$$

[c.f. (2.9), (2.10) and (2.14)].

Proof: Using (5.23) and (5.25), we have:

$$y_{1k} = \underset{\sim}{h}_1 \underset{\sim}{\theta}_{1k} + \varepsilon_{1k}, \tag{5.34}$$

i.e.

$$p(y_{1k} | \underset{\sim}{\theta}_k, M_k^{(j)}) \sim N(\underset{\sim}{h}_1 \underset{\sim}{\theta}_{1k}, \lambda^{-1} E_{11}^{(j)}) \tag{5.35}$$

From (5.24) and (5.28):

$$p(\underset{\sim}{\theta}_k | \underset{\sim}{D}_{k-1}, M_{k-1}^{(i)}, M_k^{(j)}) \sim N(\underset{\sim}{G}_k \underset{\sim}{m}_{k-1}^{(i)}, \lambda^{-1} \underset{\sim}{P}_k^{(ij)}) \tag{5.36}$$

where $\underset{\sim}{P}_k^{(ij)}$ is defined by (5.33).

Using Bayes theorem:

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_k, M_{k-1}^{(i)}, M_k^{(j)}) \propto p(y_{1k}|\underset{\sim}{\theta}_k, \underset{\sim}{D}_{k-1}, M_{k-1}^{(i)}, M_k^{(j)})\, p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_{k-1}, M_{k-1}^{(i)}, M_k^{(j)})$$

$$\propto \exp\{-\frac{\lambda}{2}[(y_{1k} - \underset{\sim}{h}_1\underset{\sim}{\theta}_{1k})^T (E_{11}^{(j)})^{-1} (y_{1k} - \underset{\sim}{h}_1\underset{\sim}{\theta}_{1k})$$

$$+ (\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)})^T (\underset{\sim}{P}_k^{(ij)})^{-1} (\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)})]\}$$

$$= \exp\{-\frac{\lambda}{2} A_k^{(ij)}\}, \text{ say.}$$

Let

$$\underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)} = \begin{bmatrix} \underset{\sim}{g}_{1k}^{(i)} \\ \underset{\sim}{g}_{2k}^{(i)} \end{bmatrix}, \quad \underset{\sim}{P}_k^{(ij)} = \begin{bmatrix} \underset{\sim}{P}_{11k}^{(ij)} & \underset{\sim}{P}_{12k}^{(ij)} \\ (\underset{\sim}{P}_{12k}^{(ij)})^T & \underset{\sim}{P}_{22k}^{(ij)} \end{bmatrix}$$

and

$$\underset{\sim}{Q}_k^{(ij)} = (\underset{\sim}{P}_k^{(ij)})^{-1} = \begin{bmatrix} \underset{\sim}{Q}_{11k}^{(ij)} & \underset{\sim}{Q}_{12k}^{(ij)} \\ (\underset{\sim}{Q}_{12k}^{(ij)})^T & \underset{\sim}{Q}_{22k}^{(ij)} \end{bmatrix}$$

For notational convenience, we may drop the superscripts (ij) and the suffix k when appropriate, and write, for example, $\underset{\sim}{Q}_{11}$ in place of $\underset{\sim}{Q}_{11k}^{(ij)}$, etc. Then

$$A_k^{(ij)} = \frac{(\underset{\sim}{\theta}_1^T\underset{\sim}{h}_1^T - y_1)(\underset{\sim}{h}_1\underset{\sim}{\theta}_1 - y_1)}{E_{11}} + [\underset{\sim}{\theta}_1^T - \underset{\sim}{g}_1^T \quad \underset{\sim}{\theta}_2^T - \underset{\sim}{g}_2^T] \begin{bmatrix} \underset{\sim}{Q}_{11} & \underset{\sim}{Q}_{12} \\ \underset{\sim}{Q}_{12}^T & \underset{\sim}{Q}_{22} \end{bmatrix} \begin{bmatrix} \underset{\sim}{\theta}_1 - \underset{\sim}{g}_1 \\ \underset{\sim}{\theta}_2 - \underset{\sim}{g}_2 \end{bmatrix}$$

$$\left.\begin{aligned} &= \underset{\sim}{\theta}_1^T(\underset{\sim}{Q}_{11} + \frac{\underset{\sim}{h}_1^T\underset{\sim}{h}_1}{E_{11}})\underset{\sim}{\theta}_1 + \underset{\sim}{\theta}_2^T\underset{\sim}{Q}_{22}\underset{\sim}{\theta}_2 + \underset{\sim}{\theta}_2^T\underset{\sim}{Q}_{12}^T\underset{\sim}{\theta}_1 + \underset{\sim}{\theta}_1^T\underset{\sim}{Q}_{12}\underset{\sim}{\theta}_2 \\ &- \underset{\sim}{\theta}_1^T(\underset{\sim}{Q}_{11}\underset{\sim}{g}_1 + \underset{\sim}{Q}_{12}\underset{\sim}{g}_2 + \frac{\underset{\sim}{h}_1^T y_1}{E_{11}}) - (\underset{\sim}{g}_1^T\underset{\sim}{Q}_{11} + \underset{\sim}{g}_2^T\underset{\sim}{Q}_{12}^T + \frac{\underset{\sim}{h}_1 y_1}{E_{11}})\underset{\sim}{\theta}_1 \\ &- \underset{\sim}{\theta}_2^T(\underset{\sim}{Q}_{12}^T\underset{\sim}{g}_1 + \underset{\sim}{Q}_{22}\underset{\sim}{g}_2) - (\underset{\sim}{g}_1^T\underset{\sim}{Q}_{12} + \underset{\sim}{g}_2^T\underset{\sim}{Q}_{22})\underset{\sim}{\theta}_2 \end{aligned}\right\} \quad (5.37)$$

$$+ \{\text{terms not involving } \underset{\sim}{\theta}\}.$$

Assume that

$$p(\underset{\sim}{\theta}_k | \underset{\sim}{D}_k, M_{k-1}^{(i)}, M_k^{(j)}) \sim N(\underset{\sim}{m}_k^{(ij)}, \lambda^{-1}\underset{\sim}{C}_k^{(ij)})$$

$$\propto \exp\{-\frac{\lambda}{2}(\underset{\sim}{\theta}_k - \underset{\sim}{m}_k^{(ij)})^T(\underset{\sim}{C}_k^{(ij)})^{-1}(\underset{\sim}{\theta}_k - \underset{\sim}{m}_k^{(ij)})\}$$

$$\propto \exp\{-\frac{\lambda}{2} B_k^{(ij)}\}, \text{ say.}$$

Write

$$(\underset{\sim}{C}_k^{(ij)})^{-1} = \begin{pmatrix} \underset{\sim}{D}_{11k}^{(ij)} & \underset{\sim}{D}_{12k}^{(ij)} \\ (\underset{\sim}{D}_{12k}^{(ij)})^T & \underset{\sim}{D}_{22k}^{(ij)} \end{pmatrix}$$

Then

$$B_k^{(ij)} = [\underset{\sim}{\theta}_1^T - \underset{\sim}{m}_1^T \quad \underset{\sim}{\theta}_2^T - \underset{\sim}{m}_2^T]\begin{pmatrix} \underset{\sim}{D}_{11} & \underset{\sim}{D}_{12} \\ \underset{\sim}{D}_{12}^T & \underset{\sim}{D}_{22} \end{pmatrix}\begin{pmatrix} \underset{\sim}{\theta}_1 - \underset{\sim}{m}_1 \\ \underset{\sim}{\theta}_2 - \underset{\sim}{m}_2 \end{pmatrix}$$

$$\left.\begin{aligned} &= \underset{\sim}{\theta}_1^T\underset{\sim}{D}_{11}\underset{\sim}{\theta}_1 + \underset{\sim}{\theta}_2^T\underset{\sim}{D}_{22}\underset{\sim}{\theta}_2 + \underset{\sim}{\theta}_2^T\underset{\sim}{D}_{12}^T\underset{\sim}{\theta}_1 + \underset{\sim}{\theta}_1^T\underset{\sim}{D}_{12}\underset{\sim}{\theta}_2 \\ &- \underset{\sim}{\theta}_1^T(\underset{\sim}{D}_{11}\underset{\sim}{m}_1 + \underset{\sim}{D}_{12}\underset{\sim}{m}_2) - (\underset{\sim}{m}_1^T\underset{\sim}{D}_{11} + \underset{\sim}{m}_2^T\underset{\sim}{D}_{12}^T)\underset{\sim}{\theta}_1 \\ &- \underset{\sim}{\theta}_2^T(\underset{\sim}{D}_{12}^T\underset{\sim}{m}_1 + \underset{\sim}{D}_{22}\underset{\sim}{m}_2) - (\underset{\sim}{m}_1^T\underset{\sim}{D}_{12} + \underset{\sim}{m}_2^T\underset{\sim}{D}_{22})\underset{\sim}{\theta}_2 \\ &+ \{\text{terms not involving } \underset{\sim}{\theta}\}. \end{aligned}\right\} \quad (5.38)$$

Equating (5.37) and (5.38) we get:

$$\left.\begin{aligned} \underset{\sim}{D}_{11} &= \underset{\sim}{Q}_{11} + \underset{\sim}{h}_1^T\underset{\sim}{h}_1/E_{11} \\ \underset{\sim}{D}_{22} &= \underset{\sim}{Q}_{22} \\ \underset{\sim}{D}_{12} &= \underset{\sim}{Q}_{12} \end{aligned}\right\} \quad (5.39)$$

$$\left.\begin{aligned} &\underset{\sim}{D}_{11}\underset{\sim}{m}_1 + \underset{\sim}{D}_{12}\underset{\sim}{m}_2 = \underset{\sim}{Q}_{11}\underset{\sim}{g}_1 + \underset{\sim}{Q}_{12}\underset{\sim}{g}_2 + \underset{\sim}{h}_1^T y_1/E_{11} \\ \text{and} \quad &\underset{\sim}{D}_{12}^T\underset{\sim}{m}_1 + \underset{\sim}{D}_{22}\underset{\sim}{m}_2 = \underset{\sim}{Q}_{12}^T\underset{\sim}{g}_1 + \underset{\sim}{Q}_{22}\underset{\sim}{g}_2 \end{aligned}\right\} \quad (5.40)$$

i.e.

$$(\underset{\sim}{C}_k^{(ij)})^{-1} = \begin{pmatrix} \underset{\sim}{Q}_{11} & \underset{\sim}{Q}_{12} \\ \underset{\sim}{Q}_{12}^T & \underset{\sim}{Q}_{22} \end{pmatrix} + \begin{pmatrix} \underset{\sim}{h}_1^T \underset{\sim}{h}_1 / E_{11} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}$$

$$= (\underset{\sim}{P}_k^{(ij)})^{-1} + \underset{\sim}{H}_{T_k}^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} (E_{11}^{(j)})^{-1} [1 \quad 0] \underset{\sim}{H}_{T_k} \tag{5.41}$$

(c.f. $(\underset{\sim}{C}_k^{(ij)})^{-1} = (\underset{\sim}{P}_k^{(ij)})^{-1} + \underset{\sim}{H}_{T_k}^T (\underset{\sim}{R}_\varepsilon^{(j)})^{-1} \underset{\sim}{H}_{T_k}$

when

$$\underset{\sim}{y}_k = \begin{pmatrix} y_{1k} \\ y_{2k} \end{pmatrix}.)$$

Therefore

$$\underset{\sim}{C}_k^{(ij)} = \{(\underset{\sim}{P}_k^{(ij)})^{-1} + \underset{\sim}{H}_{T_k}^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} (E_{11}^{(j)})^{-1} [1 \quad 0] \underset{\sim}{H}_{T_k}\}^{-1}$$

$$= \underset{\sim}{P}_k^{(ij)} - \underset{\sim}{P}_k^{(ij)} \underset{\sim}{H}_{T_k}^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} \{[1 \quad 0] \underset{\sim}{H}_{T_k}^T \underset{\sim}{P}_k^{(ij)} \underset{\sim}{H}_{T_k} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + E_{11}^{(j)}\}^{-1}$$

$$\times [1 \quad 0] \underset{\sim}{H}_{T_k} \underset{\sim}{P}_k^{(ij)}$$

(see, for example, Lindley and Smith 1972)

$$= \underset{\sim}{P}_k^{(ij)} - \underset{\sim}{S}_k^{(ij)} F_{11k}^{(ij)} (\underset{\sim}{S}_k^{(ij)})^T \tag{5.42}$$

using the definitions given in (5.33). From (5.40),

$$(\underset{\sim}{C}_k^{(ij)})^{-1} \underset{\sim}{m}_k^{(ij)} = (\underset{\sim}{P}_k^{(ij)})^{-1} \begin{pmatrix} \underset{\sim}{g}_{1k}^{(i)} \\ \underset{\sim}{g}_{2k}^{(i)} \end{pmatrix} + \begin{pmatrix} \underset{\sim}{h}_1^T y_{1k} / E_{11}^{(j)} \\ \underset{\sim}{0} \end{pmatrix}$$

which implies that, using (5.41),

$$\left((\underset{\sim}{P}_k^{(ij)})^{-1} + \begin{pmatrix} \underset{\sim}{h}_1^T \underset{\sim}{h}_1 / E_{11}^{(j)} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} \right) \underset{\sim}{m}_k^{(ij)} = (\underset{\sim}{P}_k^{(ij)})^{-1} \begin{pmatrix} \underset{\sim}{g}_{1k}^{(i)} \\ \underset{\sim}{g}_{2k}^{(i)} \end{pmatrix} + \begin{pmatrix} \underset{\sim}{h}_1^T y_{1k} / E_{11}^{(j)} \\ \underset{\sim}{0} \end{pmatrix} \tag{5.43}$$

Substituting the definition of $\underset{\sim}{m}_k^{(ij)}$, given by the theorem, into the left-hand-side of (5.43), we obtain:

$$\underset{\sim}{P}_k^{-1}\begin{pmatrix}\underset{\sim}{g}_1\\ \underset{\sim}{g}_2\end{pmatrix} + \begin{pmatrix}\underset{\sim}{h}_1^T\\ \underset{\sim}{0}\end{pmatrix}F_{11}^{-1}(y_1 - \underset{\sim}{h}_1\underset{\sim}{g}_1) + \begin{pmatrix}\underset{\sim}{h}_1^T\underset{\sim}{h}_1\underset{\sim}{g}_1/E_{11}\\ \underset{\sim}{0}\end{pmatrix}$$

$$+ \begin{pmatrix}\underset{\sim}{h}_1^T\underset{\sim}{h}_1/E_{11} & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{0}\end{pmatrix}\underset{\sim}{P}_k\begin{pmatrix}\underset{\sim}{h}_1^T\\ \underset{\sim}{0}\end{pmatrix}F_{11}^{-1}(y_1 - \underset{\sim}{h}_1\underset{\sim}{g}_1)$$

$$= \underset{\sim}{P}_k^{-1}\begin{pmatrix}\underset{\sim}{g}_1\\ \underset{\sim}{g}_2\end{pmatrix} + \{\underset{\sim}{I} + \begin{pmatrix}\underset{\sim}{h}_1^T\underset{\sim}{h}_1/E_{11} & \underset{\sim}{0}\\ \underset{\sim}{0} & \underset{\sim}{0}\end{pmatrix}\underset{\sim}{P}_k\}\begin{pmatrix}\underset{\sim}{h}_1^T\\ \underset{\sim}{0}\end{pmatrix}F_{11}^{-1}(y_1 - \underset{\sim}{h}_1\underset{\sim}{g}_1) + \begin{pmatrix}\underset{\sim}{h}_1^T\underset{\sim}{h}_1\underset{\sim}{g}_1/E_{11}\\ \underset{\sim}{0}\end{pmatrix}$$

$$= \underset{\sim}{P}_k^{-1}\begin{pmatrix}\underset{\sim}{g}_1\\ \underset{\sim}{g}_2\end{pmatrix} + \begin{pmatrix}F_{11}/E_{11} & \underset{\sim}{h}_1^T\underset{\sim}{P}_{12}\underset{\sim}{h}_1/E_{11}\\ \underset{\sim}{0} & \underset{\sim}{I}\end{pmatrix}\begin{pmatrix}\underset{\sim}{h}_1^T\\ \underset{\sim}{0}\end{pmatrix}F_{11}^{-1}(y_1 - \underset{\sim}{h}_1\underset{\sim}{g}_1) + \begin{pmatrix}\underset{\sim}{h}_1^T\underset{\sim}{h}_1\underset{\sim}{g}_1/E_{11}\\ \underset{\sim}{0}\end{pmatrix}$$

$$= \underset{\sim}{P}_k^{-1}\begin{pmatrix}\underset{\sim}{g}_1\\ \underset{\sim}{g}_2\end{pmatrix} + \begin{pmatrix}(\underset{\sim}{h}_1^T/E_{11})(y_1 - \underset{\sim}{h}_1\underset{\sim}{g}_1) + (\underset{\sim}{h}_1^T/E_{11})\underset{\sim}{h}_1\underset{\sim}{g}_1\\ \underset{\sim}{0}\end{pmatrix}$$

$$= \underset{\sim}{P}_k^{-1}\begin{pmatrix}\underset{\sim}{g}_1\\ \underset{\sim}{g}_2\end{pmatrix} + \begin{pmatrix}\underset{\sim}{h}_1^T y_1/E_{11}\\ \underset{\sim}{0}\end{pmatrix}$$

and thus the identity in (5.43) is proven.

Notice that the calculation of $\underset{\sim}{m}_k^{(ij)}$, in (5.31), is now dependent only upon y_{1k} and not y_{2k}.

Predictive Density

In this case, the required predictive density is $p(y_{1k}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)})$. From (5.36), we have:

$$p(\underset{\sim}{\theta}_{1k}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)}) \sim N(\underset{\sim}{g}_{1k}^{(i)},\lambda^{-1}\underset{\sim}{P}_{11k}^{(ij)}) \tag{5.44}$$

and so (5.35) and (5.44) give:

$$p(y_{1k}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)}) \sim N(\underset{\sim}{h}_1\underset{\sim}{g}_{1k}^{(i)},\lambda^{-1}(\underset{\sim}{h}_1^T\underset{\sim}{P}_{11k}^{(ij)}\underset{\sim}{h}_1 + E_{11}^{(j)}))$$

$$\sim N(f_{1k}^{(i)},\lambda^{-1}F_{11k}^{(ij)}) \quad (5.45)$$

NOTE: For the case when only y_{2k} is available at time T_k, the results are clearly of a similar form, due to symmetry, i.e.

$$\underset{\sim}{m}_k^{(ij)} = \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)} + \underset{\sim}{S}_k^{(ij)}(y_{2k} - f_{2k}^{(i)}) \quad (5.46)$$

and

$$\underset{\sim}{C}_k^{(ij)} = \underset{\sim}{P}_k^{(ij)} - \underset{\sim}{S}_k^{(ij)}F_{22k}^{(ij)}(\underset{\sim}{S}_k^{(ij)})^T \quad (5.47)$$

where

$$\left.\begin{aligned} f_{2k}^{(i)} &= [0 \quad 1]\underset{\sim}{H}_{T_k}\underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)} \\ F_{22k}^{(ij)} &= [0 \quad 1]\underset{\sim}{H}_{T_k}^T\underset{\sim}{P}_k^{(ij)}\underset{\sim}{H}_{T_k}\begin{pmatrix}0\\1\end{pmatrix} + E_{22}^{(j)} \\ \text{and} \quad \underset{\sim}{S}_k^{(ij)} &= \underset{\sim}{P}_k^{(ij)}\underset{\sim}{H}_{T_k}^T\begin{pmatrix}0\\1\end{pmatrix}(F_{22k}^{(ij)})^{-1} \end{aligned}\right\} \quad (5.48)$$

5.2.2.2 Known Variance Ratio

Case 1: y_{1k} and y_{2k} Available at Time T_k

For the case where λ is unknown, we have:

$$p(\underset{\sim}{y}_k|\underset{\sim}{\theta}_k,M_k^{(j)},\lambda) \sim N(\underset{\sim}{H}_{T_k}\underset{\sim}{\theta}_k,\lambda^{-1}\underset{\sim}{R}_\varepsilon^{(j)}) \quad (5.49)$$

and

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)},\lambda) \sim N(\underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)},\lambda^{-1}\underset{\sim}{P}_k^{(ij)}) \quad (5.50)$$

Assume that

$$p(\lambda|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)}) \sim G(\alpha_{k-1},\beta_{k-1}^{(i)}) \quad (5.51)$$

where $U \sim G(a,b)$ denotes the gamma distribution with parameters a and b. Using Bayes theorem,

$$p(\underset{\sim}{\theta}_k,\lambda|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)}) \propto p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1},\underset{\sim}{\theta}_k,M_{k-1}^{(i)},M_k^{(j)},\lambda)p(\underset{\sim}{\theta}_k,\lambda|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)})$$

$$\propto \lambda \exp[-\frac{\lambda}{2}(\underset{\sim}{H}_{T_k}\underset{\sim}{\theta}_k - \underset{\sim}{y}_k)^T(\underset{\sim}{R}_\varepsilon^{(j)})^{-1}(\underset{\sim}{H}_{T_k}\underset{\sim}{\theta}_k - \underset{\sim}{y}_k)].\lambda^{\alpha_{k-1}-1}\exp[-\lambda\beta_{k-1}^{(i)}]$$

$$\times \lambda^{d_\theta/2}\exp[-\frac{\lambda}{2}(\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)})^T(\underset{\sim}{P}_k^{(ij)})^{-1}(\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)})]$$

$$\propto \lambda^{d_\theta/2}\exp[-\frac{\lambda}{2}(\underset{\sim}{\theta}_k - \underset{\sim}{m}_k^{(ij)})^T(\underset{\sim}{C}_k^{(ij)})^{-1}(\underset{\sim}{\theta}_k - \underset{\sim}{m}_k^{(ij)})].\lambda^{\alpha_k-1}\exp[-\lambda\beta_k^{(ij)}]$$

where $(d_\theta \times 1)$ is the dimension of $\underset{\sim}{\theta}_k$, and $\underset{\sim}{m}_k^{(ij)}$ and $\underset{\sim}{C}_k^{(ij)}$ are defined by (5.20) and (5.21), with

$$\alpha_k = \alpha_{k-1} + 1 \qquad (5.52)$$

$$\beta_k^{(ij)} = \beta_k^{(i)} + \tfrac{1}{2}(\underset{\sim}{y}_k - \underset{\sim}{f}_k^{(i)})^T(\underset{\sim}{F}_k^{(ij)})^{-1}(\underset{\sim}{y}_k - \underset{\sim}{f}_k^{(i)}) \qquad (5.53)$$

and $\underset{\sim}{f}_k^{(i)}$ and $\underset{\sim}{F}_k^{(ij)}$ as defined by (3.16), so that

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)},\lambda) \sim N(\underset{\sim}{m}_k^{(ij)},\lambda^{-1}\underset{\sim}{C}_k^{(ij)}) \qquad (5.54)$$

and

$$p(\lambda|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)}) \sim G(\alpha_k,\beta_k^{(ij)}). \qquad (5.55)$$

This is the standard result (see, for example, DeGroot 1970 for details) involving, as before, the Kalman Filter update for $\underset{\sim}{m}_k^{(ij)}$ and $\underset{\sim}{C}_k^{(ij)}$; the predictive density, $p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)})$, can be shown to be proportional to

$$(\underset{\sim}{F}_k^{(ij)})^{-\frac{1}{2}}(\beta_{k-1}^{(i)})^{\alpha_{k-1}}(\beta_k^{(ij)})^{-\alpha_k} \qquad (5.56)$$

(see Equation (3.18)).

Case 2: Only y_{1k} Available at Time T_k

Theorem 5.2.2

Consider the dynamic linear model given by (5.23) to (5.27), and assume that:

$$p(\underset{\sim}{\theta}_{k-1}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},\lambda) \sim N(\underset{\sim}{m}_{k-1}^{(i)},\lambda^{-1}\underset{\sim}{C}_{k-1}^{(i)}) \qquad (5.57)$$

and

$$p(\lambda|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)}) \sim G(\alpha_{k-1},\beta_{k-1}^{(i)}) \qquad (5.58)$$

Then, if only y_{1k} is available at time T_k (i.e. $\underset{\sim}{D}_k = (\underset{\sim}{D}_{k-1},y_{1k})$), the posterior distribution for $\underset{\sim}{\theta}_k$ and λ is given by:

$$p(\underset{\sim}{\theta}_k|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)},\lambda) \sim N(\underset{\sim}{m}_k^{(ij)},\lambda^{-1}\underset{\sim}{C}_k^{(ij)}) \qquad (5.59)$$

and

$$p(\lambda|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)}) \sim G(\alpha_k,\beta_k^{(ij)}) \qquad (5.60)$$

where $\underset{\sim}{m}_k^{(ij)}$ and $\underset{\sim}{C}_k^{(ij)}$ are given, from Theorem 5.2.1, by (5.31) and (5.32), respectively, and where

$$\alpha_k = \alpha_{k-1} + \tfrac{1}{2} \qquad (5.61)$$

and

$$\beta_k^{(ij)} = \beta_{k-1}^{(i)} + \tfrac{1}{2}(y_{1k} - f_{1k}^{(i)})^2/F_{11k}^{(ij)} \qquad (5.62)$$

[c.f. (5.52) and (5.53)].

NOTE: An important special case, that of the static normal linear model, is examined in Appendix A5.1.

Proof

Using Bayes theorem and (5.34), (5.36) and (5.58),

$$p(\underset{\sim}{\theta}_k,\lambda|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)}) \propto p(y_{1k}|\underset{\sim}{\theta}_k,\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)},\lambda)p(\underset{\sim}{\theta}_k,\lambda|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)})$$

$$\propto \lambda^{d_\theta/2}\lambda^{\alpha_{k-1}+\frac{1}{2}-1}\exp\{-\frac{\lambda}{2}[\frac{(y_{1k}-\underset{\sim}{h}_1\underset{\sim}{\theta}_{1k})^2}{E_{11}^{(j)}} + (\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)})^T(\underset{\sim}{P}_k^{(ij)})^{-1} \times (\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)}) + 2\beta_{k-1}^{(i)}]\} \tag{5.63}$$

From (5.59) and (5.60),

$$p(\underset{\sim}{\theta}_k,\lambda|\underset{\sim}{D}_k,M_{k-1}^{(i)},M_k^{(j)}) \propto \lambda^{d_\theta/2}\exp[-\frac{\lambda}{2}(\underset{\sim}{\theta}_k - \underset{\sim}{m}_k^{(ij)})^T(\underset{\sim}{C}_k^{(ij)})^{-1}(\underset{\sim}{\theta}_k - \underset{\sim}{m}_k^{(ij)})] \times \lambda^{\alpha_k-1}\exp[-\lambda\beta_k^{(ij)}] \tag{5.64}$$

Equating (5.63) and (5.64), we obtain:

$$\alpha_k = \alpha_{k-1} + \tfrac{1}{2}$$

and

$$(\underset{\sim}{\theta}_k - \underset{\sim}{m}_k^{(ij)})^T(\underset{\sim}{C}_k^{(ij)})^{-1}(\underset{\sim}{\theta}_k - \underset{\sim}{m}_k^{(ij)}) + 2\beta_k^{(ij)}$$

$$= \frac{(y_{1k}-\underset{\sim}{h}_1\underset{\sim}{\theta}_{1k})^2}{E_{11}^{(j)}} + (\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)})^T(\underset{\sim}{P}_k^{(ij)})^{-1}(\underset{\sim}{\theta}_k - \underset{\sim}{G}_k\underset{\sim}{m}_{k-1}^{(i)}) + 2\beta_{k-1}^{(i)}$$

Since $\beta_{k-1}^{(i)}$ and $\beta_k^{(ij)}$ do not contain terms involving $\underset{\sim}{\theta}$, the solutions for $\underset{\sim}{m}_k^{(ij)}$ and $\underset{\sim}{C}_k^{(ij)}$ are identical to those given in Theorem 5.2.1. Upon equating the terms not involving $\underset{\sim}{\theta}$ (using the notation from Theorem 5.2.1, and dropping superscripts, etc. where convenient) we find:

$$\underset{\sim}{m}_1^T(\underset{\sim}{D}_{11}\underset{\sim}{m}_1 + \underset{\sim}{D}_{12}\underset{\sim}{m}_2) + \underset{\sim}{m}_2^T(\underset{\sim}{D}_{12}^T\underset{\sim}{m}_1 + \underset{\sim}{D}_{22}\underset{\sim}{m}_2) + 2\beta_k = \underset{\sim}{g}_1^T(\underset{\sim}{Q}_{11}\underset{\sim}{g}_1 + \underset{\sim}{Q}_{12}\underset{\sim}{g}_2) + \underset{\sim}{g}_2^T(\underset{\sim}{Q}_{12}^T\underset{\sim}{g}_1 + \underset{\sim}{Q}_{22}\underset{\sim}{g}_2) + \frac{y_1^2}{E_{11}} + 2\beta_{k-1}$$

$$\Rightarrow \underset{\sim}{m}_1^T(\underset{\sim}{Q}_{11}\underset{\sim}{g}_1 + \underset{\sim}{Q}_{12}\underset{\sim}{g}_2 + \frac{\underset{\sim}{h}_1^T y_1}{E_{11}}) + \underset{\sim}{m}_2^T(\underset{\sim}{Q}_{12}^T\underset{\sim}{g}_1 + \underset{\sim}{Q}_{22}\underset{\sim}{g}_2) + 2\beta_k$$

$$= \underset{\sim}{g}_1^T(\underset{\sim}{Q}_{11}\underset{\sim}{g}_1 + \underset{\sim}{Q}_{12}\underset{\sim}{g}_2) + \underset{\sim}{g}_2^T(\underset{\sim}{Q}_{12}^T\underset{\sim}{g}_1 + \underset{\sim}{Q}_{22}\underset{\sim}{g}_2) + \frac{y_1^2}{E_{11}} + 2\beta_{k-1}$$

(using (5.40)), i.e.

$$\beta_k = \beta_{k-1} + \tfrac{1}{2}\{(\underset{\sim}{g}_1^T - \underset{\sim}{m}_1^T)(\underset{\sim}{Q}_{11}\underset{\sim}{g}_1 + \underset{\sim}{Q}_{12}\underset{\sim}{g}_2) + (\underset{\sim}{g}_2^T - \underset{\sim}{m}_2^T)(\underset{\sim}{Q}_{12}^T\underset{\sim}{g}_1 + \underset{\sim}{Q}_{22}\underset{\sim}{g}_2) + \frac{y_1^2}{E_{11}} - \frac{\underset{\sim}{m}_1^T\underset{\sim}{h}_1^T y_1}{E_{11}}\}$$

$$= \beta_{k-1} + \tfrac{1}{2}\{[\underset{\sim}{g}_1^T - \underset{\sim}{m}_1^T \quad \underset{\sim}{g}_2^T - \underset{\sim}{m}_2^T]\underset{\sim}{P}_k^{-1}\begin{pmatrix}\underset{\sim}{g}_1\\ \underset{\sim}{g}_2\end{pmatrix} + \frac{y_1^2}{E_{11}} - \frac{\underset{\sim}{m}_1^T\underset{\sim}{h}_1^T y_1}{E_{11}}\}$$

$$= \beta_{k-1} + \tfrac{1}{2}\{-\frac{(y_1 - \underset{\sim}{h}_1\underset{\sim}{g}_1)}{F_{11}}\underset{\sim}{h}_1\underset{\sim}{g}_1 + \frac{y_1^2}{E_{11}} - \frac{(y_1 - \underset{\sim}{h}_1\underset{\sim}{g}_1)}{F_{11}}\underset{\sim}{h}_1\underset{\sim}{P}_{11}\frac{\underset{\sim}{h}_1^T y_1}{E_{11}} - \underset{\sim}{g}_1^T\frac{\underset{\sim}{h}_1^T y_1}{E_{11}}\}$$

(using (5.31) and (5.33))

$$= \beta_{k-1} + \frac{1}{2F_{11}}\{(\underset{\sim}{h}_1\underset{\sim}{g}_1)^2 - y_1\underset{\sim}{h}_1\underset{\sim}{g}_1 + \frac{F_{11}}{E_{11}}y_1^2 - \frac{y_1\underset{\sim}{h}_1\underset{\sim}{P}_{11}\underset{\sim}{h}_1^T y_1}{E_{11}} + \frac{\underset{\sim}{h}_1\underset{\sim}{g}_1\underset{\sim}{h}_1\underset{\sim}{P}_{11}\underset{\sim}{h}_1^T y_1}{E_{11}} - \frac{F_{11}\underset{\sim}{g}_1^T\underset{\sim}{h}_1^T y_1}{E_{11}}\}$$

$$= \beta_{k-1} + \frac{1}{2F_{11}}\{(\underset{\sim}{h}_1\underset{\sim}{g}_1)^2 - y_1\underset{\sim}{h}_1\underset{\sim}{g}_1 + y_1^2 - \underset{\sim}{g}_1^T\underset{\sim}{h}_1^T y_1)\}$$

(since

$$\frac{F_{11}}{E_{11}} = 1 + \frac{\underset{\sim}{h}_1 P_{11}\underset{\sim}{h}_1^T}{E_{11}})$$

$$= \beta_{k-1} + \frac{1}{2F_{11}}(y_1 - \underset{\sim}{h}_1\underset{\sim}{g}_1)^2$$

i.e.

$$\beta_k^{(ij)} = \beta_{k-1}^{(i)} + \tfrac{1}{2}\frac{(y_{1k} - f_{1k}^{(i)})^2}{F_{11k}^{(ij)}}$$

Predictive Density

Clearly,

$$p(y_{1k}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)},\lambda) \sim N(f_{1k}^{(i)},\lambda^{-1}F_{11k}^{(ij)}) \tag{5.65}$$

and, from (5.58),

$$p(\lambda|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)}) \sim G(\alpha_{k-1},\beta_{k-1}^{(i)}).$$

Now

$$p(y_{1k}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)}) = \int p(y_{1k}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)},\lambda)p(\lambda|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)})d\lambda$$

$$\propto (F_{11k}^{(ij)})^{-\frac{1}{2}}(\beta_{k-1}^{(i)})^{\alpha_{k-1}}\int \lambda^{\frac{1}{2}+\alpha_{k-1}-1}\exp\{-\lambda[\frac{\frac{1}{2}(y_1 - f_{1k}^{(i)})^2}{F_{11k}^{(ij)}} + \beta_{k-1}^{(i)}]\}d\lambda$$

$$= (F_{11k}^{(ij)})^{-\frac{1}{2}}(\beta_{k-1}^{(i)})^{\alpha_{k-1}}\int \lambda^{\alpha_k-1}\exp[-\lambda\beta_k]d\lambda$$

$$\propto (F_{11k}^{(ij)})^{-\frac{1}{2}}(\beta_{k-1}^{(i)})^{\alpha_{k-1}}(\beta_k^{(ij)})^{-\alpha_k} \tag{5.66}$$

[c.f. (5.56)].

NOTE: For the case where only y_{2k} is available at time T_k, the result is clearly of a similar form to that given by Theorem 5.2.2, i.e.

$$\beta_k^{(ij)} = \beta_{k-1}^{(i)} + \tfrac{1}{2}\,\frac{(y_{2k} - f_{2k}^{(i)})^2}{F_{22k}^{(ij)}} \tag{5.67}$$

5.2.2.3 Unknown Variances

When considering the problem of sampling from a bivariate normal distribution, for which the mean vector, $\underset{\sim}{\theta}$, and covariance matrix, $\underset{\sim}{C}$, are unknown, the standard Bayesian conjugate analysis (see, for example, DeGroot 1970) proposes a joint Normal-Wishart prior distribution for $\underset{\sim}{\theta}$ and $\underset{\sim}{C}$. This approach poses a number of difficulties when put in the context of multistate dynamic linear models for unequally-spaced data. Recall, from (5.1) and (5.4), that:

$$y_{1t} = \underset{\sim}{h}_1\underset{\sim}{\theta}_{1t} + \varepsilon_{1t} \tag{5.68}$$

and

$$y_{2t} = \underset{\sim}{h}_2\underset{\sim}{\theta}_{2t} + \varepsilon_{2t} \tag{5.69}$$

If we suppose that:

$$\left.\begin{array}{l} \mathrm{var}(\varepsilon_{1t}) = \lambda_1^{-1}E_{11},\ \mathrm{var}(\varepsilon_{2t}) = \lambda_2^{-1}E_{22} \\ \text{and} \qquad \mathrm{cov}(\varepsilon_{1t},\varepsilon_{2t}) = \lambda_{12}^{-1}E_{12} \end{array}\right\} \tag{5.70}$$

then

$$\underset{\sim}{R}_\varepsilon = \mathrm{var}\begin{pmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\end{pmatrix} = \begin{pmatrix}\lambda_1^{-1}E_{11} & \lambda_{12}^{-1}E_{12}\\ \lambda_{12}^{-1}E_{12} & \lambda_2^{-1}E_{22}\end{pmatrix} \tag{5.71}$$

Let us attempt to write $\underset{\sim}{R}_{\varepsilon}$ in the form

$$\underset{\sim}{R}_{\varepsilon} = \underset{\sim}{\Lambda}\underset{\sim}{E} \tag{5.72}$$

where

$$\underset{\sim}{\Lambda} = \begin{pmatrix} \lambda_1^{-1} & \lambda_{12}^{-1} \\ \lambda_{12}^{-1} & \lambda_2^{-1} \end{pmatrix} \tag{5.73}$$

and

$$\underset{\sim}{E} = \begin{pmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{pmatrix} \tag{5.74}$$

Then

$$\underset{\sim}{\Lambda}\underset{\sim}{E} = \begin{pmatrix} \lambda_1^{-1}e_{11} + \lambda_{12}^{-1}e_{21} & \lambda_1^{-1}e_{12} + \lambda_{12}^{-1}e_{22} \\ \lambda_{12}^{-1}e_{11} + \lambda_2^{-1}e_{21} & \lambda_{12}^{-1}e_{12} + \lambda_2^{-1}e_{22} \end{pmatrix}$$

If $\underset{\sim}{\Lambda}\underset{\sim}{E}$ is to be a covariance matrix for $\underset{\sim}{\varepsilon}_t$, it must be symmetical, i.e.

$$\lambda_{12}^{-1}e_{11} + \lambda_2^{-1}e_{21} = \lambda_1^{-1}e_{12} + \lambda_{12}^{-1}e_{22} \tag{5.75}$$

Now, $e_{21} = e_{12} = 0 \Rightarrow e_{11} = e_{22} = e$, say, so that

$$\underset{\sim}{E} = \begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix} = e\underset{\sim}{I}. \tag{5.76}$$

If $e_{21} \neq 0$ or $e_{12} \neq 0$, the elements of $\underset{\sim}{E}$ must be constrained to satisfy (5.75). For instance, if we set $\lambda_{12}^{-1} = 0$ (i.e. $\mathrm{cov}(\varepsilon_{1t}, \varepsilon_{2t}) = 0$), (5.75) yields:

$$\lambda_2^{-1}e_{21} = \lambda_1^{-1}e_{12} \Rightarrow \frac{e_{12}}{e_{21}} = \frac{\lambda_1}{\lambda_2} .$$

But, since we need to pre-specify the elements of $\underset{\sim}{E}$, this amounts to having prior knowledge of the ratio $\text{var}(\varepsilon_{1t})/\text{var}(\varepsilon_{2t})$, in which case the results of Section 5.2.2.2 are applicable. We see, therefore, that $\underset{\sim}{R}_\varepsilon$ cannot be expressed in the form $\underset{\sim}{\Lambda}\underset{\sim}{E}$ (nor $\underset{\sim}{E}\underset{\sim}{\Lambda}$) unless $\underset{\sim}{E} = e\underset{\sim}{I}$, where e is a scalar. In this case,

$$\underset{\sim}{R}_\varepsilon = \underset{\sim}{\Lambda}e = \begin{pmatrix} \lambda_1^{-1}e & \lambda_{12}^{-1}e \\ \lambda_{12}^{-1}e & \lambda_2^{-1}e \end{pmatrix} \tag{5.77}$$

i.e.

$$\text{var}(\varepsilon_{1t}) = \lambda_1^{-1}e, \quad \text{var}(\varepsilon_{2t}) = \lambda_2^{-1}e$$

and

$$\text{cov}(\varepsilon_{1t},\varepsilon_{2t}) = \lambda_{12}^{-1}e \tag{5.78}$$

Suppose, however, that we wish to incorporate a multi-state structure into our overall error structure, in order to accommodate model discontinuities. We must use the scalar, e, for this purpose, and so we must specify $e^{(j)}$, $j = 1, \ldots, J$. In this case, though, if $e^{(j)}$ is large this will cause an aberration in ε_{1t} and ε_{2t}, so that simultaneous transients for y_1 and y_2 will be induced. It is impossible to manipulate $e^{(j)}$ so that a change is induced in only one of the series. In other words, the multi-state structure (in its present form, at least) cannot be used when we assume an unknown covariance matrix, $\underset{\sim}{\Lambda}$.

Aside from the difficulties encountered by the imposition of a multi-state structure, there is a further problem associated with the use of a joint Normal-Wishart prior distribution when the data is unequally-spaced; namely, that this joint distribution is not conjugate unless both observations are available at a particular timepoint (see Appendix A5.2). This is mainly a consequence of the fact that if $\underset{\sim}{\Lambda}^{-1}$ has a Wishart distribution, then $(\lambda_1^{-1})^{-1} = \lambda_1$ does not have a Wishart distribution (see, for example, DeGroot 1970 for details). Mehta and Swamy (1974), however, show how prior-to-posterior analysis can proceed with alternative choices of prior distribution.

The problem of unknown variances within the unequally-spaced bivariate multi-state dynamic linear model context is, as yet, unresolved. In Section 5.4, though, we examine how sensitive the models are, in practice, to a misspecification of the variance ratio, $var(\varepsilon_{1t})/var(\varepsilon_{2t})$, having assumed that this ratio must be specified a priori.

5.3 BIVARIATE DYNAMIC LINEAR MODELS

5.3.1 DERIVATION OF UNEQUALLY-SPACED BIVARIATE MODELS

5.3.1.1 General Comments

Pairs of time series arising from periods of medical monitoring are likely to be correlated if each of these series reflects the progress of the same medical condition. That is to say, in the steady state, when the patient's condition is following a stable course, there will be some form of correlation between the two series. The situation, however, may be much more complex than this. It may be that the level, say, of the first series is directly dependent upon the level, say, of the second series, so that if the first level were to rise, the second level would rise too, either simultaneously or, perhaps, after some time lag. Using the phraseology of Newbold (1979), this results in unidirectional causality. Bidirectional causality (feedback) would arise if the level of the second series is also dependent on the level of the first so that, for example, a rise in the second series causes a rise in the first which, in turn, causes a further rise in the second, and so on.

There are a number of ways in which we can incorporate such behaviour into the dynamic linear model framework. Firstly, we note that there is a distinction between causality (either uni- or bidirectional) in the steady state (henceforth referred to as steady-state causality) and causality with respect to changepoints (henceforth referred to as changepoint causality). The former is of the type described in the previous paragraph, whereas the latter implies that a discontinuity in one series would induce a discontinuity into the other. In this section, we shall examine each of these types of causality, in turn, also taking into account the possibility of unequally-spaced data.

By way of introduction, we note that the concept of steady-state causality can be introduced through the off-diagonal elements of the bivariate system transition matrix. Recall the DLM described by (5.1) to (5.4):

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{h}_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{h}_2 \end{pmatrix} \begin{pmatrix} \underset{\sim}{\theta}_{1t} \\ \underset{\sim}{\theta}_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{5.79}$$

$$\begin{pmatrix} \underset{\sim}{\theta}_{1t} \\ \underset{\sim}{\theta}_{2t} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{G}_{12} \\ \underset{\sim}{G}_{21} & \underset{\sim}{G}_{22} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\theta}_{1t-1} \\ \underset{\sim}{\theta}_{2t-1} \end{pmatrix} + \begin{pmatrix} \underset{\sim}{\omega}_{1t} \\ \underset{\sim}{\omega}_{2t} \end{pmatrix} \tag{5.80}$$

where $\underset{\sim}{G}_{11}$ is the system matrix for the univariate series, y_{1t}, and $\underset{\sim}{G}_{22}$ is the system matrix for the univariate series, y_{2t}. Then, if $\underset{\sim}{G}_{12} = \underset{\sim}{0}$ and $\underset{\sim}{G}_{21} = \underset{\sim}{0}$, this would indicate that $\underset{\sim}{\theta}_{1t}$ and $\underset{\sim}{\theta}_{2t}$ are independent processes and that no steady-state causality exists between them. If, on the other hand, $\underset{\sim}{G}_{12} = \underset{\sim}{0}$, but $\underset{\sim}{G}_{21} \neq \underset{\sim}{0}$ this would indicate that $\underset{\sim}{\theta}_{1t-1}$ has a direct influence on $\underset{\sim}{\theta}_{2t}$ (i.e. unidirectional causality exists) and, if both $\underset{\sim}{G}_{12}$ and $\underset{\sim}{G}_{21}$ are non-zero matrices, there will be feedback (bidirectional causality) between the two processes.

We have assumed, here, that we are able to model each of the univariate series independently of one another, via $\underset{\sim}{G}_{11}$ and $\underset{\sim}{G}_{22}$, and <u>then</u> stipulate the interaction between the series using $\underset{\sim}{G}_{12}$ and $\underset{\sim}{G}_{21}$. Some care must be taken at this stage of the modelling procedure, however, since non-negligible feedback may make the modelling of the univariate series very tricky. Figure 5.1 shows a typical example of this problem. At first glance, it may seem reasonable to model each of these series by the sinusoidal model, outlined in Sections 2.3.3, 3.3.3 and 4.3.3, and then to investigate the interrelationships between the two series. However, these (simulated) series were, in fact, generated from a bivariate <u>linear growth</u> model with feedback, i.e.

$$\underset{\sim}{G}_{11} = \underset{\sim}{G}_{22} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

with

$$\underset{\sim}{G}_{12} = \begin{pmatrix} 0.2 & 0.2 \\ 0 & 0 \end{pmatrix}$$

and

$$\underset{\sim}{G}_{21} = \begin{pmatrix} -0.1 & -0.1 \\ 0 & 0 \end{pmatrix}.$$

It could, of course, be argued that even if we had mis-modelled the individual series by assuming that they were sinusoidal, the effect this would have in the monitoring context might be minimal.

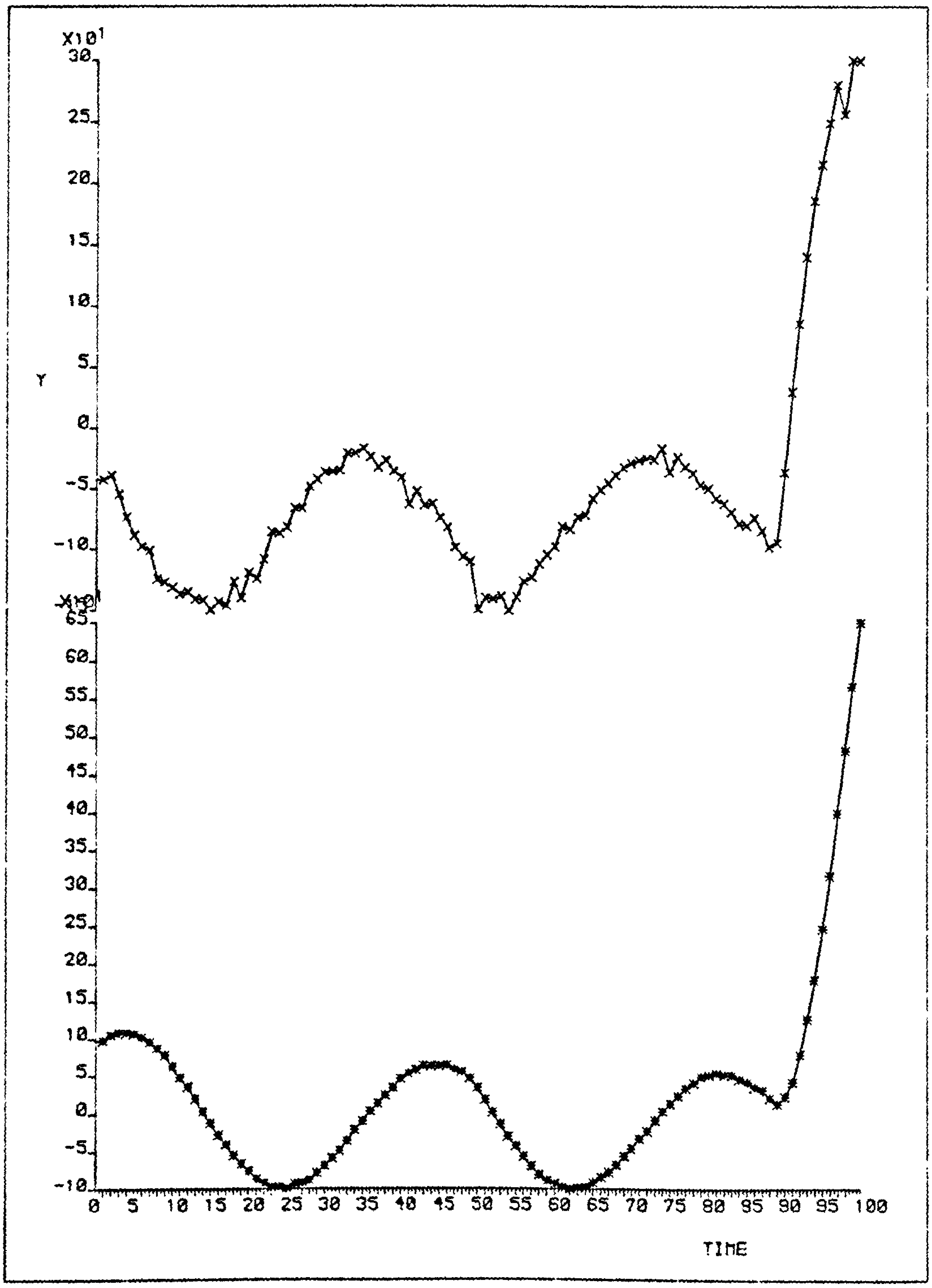


FIGURE 5.1

It would be very misleading, though, if we were to suggest that the mechanism which generated these series was sinusoidal in character.

The sudden change in the pattern of behaviour that can be seen towards the end of the series shown in Figure 5.1, is due to an induced change in the slope of the uppermost series at timepoint t = 90. Notice that this discontinuity has caused the lower series to be discontinuous after a time lag of one (i.e. at t = 91). This is a feature of steady-state causality, in that changes in one series tend to lag changes in the other by one timepoint. This restriction can, in fact, be overcome by suitable manipulation of the matrix $\underset{\sim}{G}$ and the vector $\underset{\sim}{\omega}_t$. Consider, for example, the simple bivariate steady DLM given by:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{5.81}$$

$$\begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} = \begin{pmatrix} \mu_{1t-1} \\ \mu_{2t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_{1t} \\ \delta\mu_{2t} \end{pmatrix} \tag{5.82}$$

for which

$$\underset{\sim}{G} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{5.83}$$

Suppose that we discover that the level of the y_1 series may also be influenced by the level of the y_2 series with a time lag of three units, i.e. μ_{1t} is dependent upon μ_{2t-3}. Then we may write

$$\mu_{1t} = \mu_{1t-1} + c\mu_{2t-3} + \delta\mu_{1t} \tag{5.84}$$

$$\mu_{2t} = \mu_{2t-1} + \delta\mu_{2t} \tag{5.85}$$

where the scalar c indicates the strength (and direction) of the dependence of μ_{1t} on μ_{2t-3}. From (5.85), we can see that:

$$\mu_{2t-3} = \mu_{2t-1} - (\delta\mu_{2t-2} + \delta\mu_{2t-1}) \tag{5.86}$$

Substituting this result into (5.84) gives:

$$\mu_{1t} = \mu_{1t-1} + c\mu_{2t-1} - c(\delta\mu_{2t-2} + \delta\mu_{2t-1}) \tag{5.87}$$

and hence

$$\begin{pmatrix}\mu_{1t}\\ \mu_{2t}\end{pmatrix} = \begin{pmatrix}1 & c\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{1t-1}\\ \mu_{2t-1}\end{pmatrix} + \begin{pmatrix}\delta\mu_{1t-1} - c(\delta\mu_{2t-2} + \delta\mu_{2t-1})\\ \delta\mu_{2t}\end{pmatrix} \tag{5.88}$$

so that causality has been introduced via the system matrix, $\underset{\sim}{G}$, and the time lag involved has been introduced by manipulation of $\underset{\sim}{\omega}_t$.

An alternative way of introducing changepoint causality involves a relaxation of the assumption given by Equation (3.6) in Section 3.1.2, i.e.

$$p(M_t^{(j)}|h) = p_o^{(j)}, \quad j = 1, \ldots, J, \forall t,$$

which defines the state occupied at time t to be independent of all process history. In Section 5.3.1.3 we investigate the case where the state occupied at time t is assumed dependent upon the state occupied at time t - 1:

$$p(M_t^{(j)}|M_{t-1}^{(i)}) = {}^{(i)}p_o^{(j)} \tag{5.89}$$

and we show how further lagging may be incorporated into this Markovian state-transition structure.

As a final comment, we note that concurrent changepoint causality (when the series exhibit discontinuities simultaneously) may also be incorporated into the dynamic linear model framework by allowing the off-diagonal elements of $\underset{\sim}{R}_{\varepsilon}^{(j)}$ and $\underset{\sim}{R}_{\omega}^{(j)}$ to be non-zero and, possibly, state-dependent. For example, for the DLM given by (5.81) and (5.82) we have:

$$\text{var}\left(\begin{bmatrix} \delta\mu_{1t} \\ \delta\mu_{2t} \end{bmatrix} \middle| M_t^{(j)}\right) = \lambda^{-1}\underset{\sim}{R}_{\omega}^{(j)} = \lambda^{-1}\begin{bmatrix} R_{\mu 1}^{(j)} & R_{\mu\mu}^{(j)} \\ R_{\mu\mu}^{(j)} & R_{\mu 2}^{(j)} \end{bmatrix}, \text{ say.} \tag{5.90}$$

If we believe that simultaneous level changes were possible, we would set $R_{\mu\mu}^{(j)} > 0$ for some, if not all, j.

5.3.1.2 *Steady-State Causality*

We now examine the way in which

$$\underset{\sim}{G}_k = \begin{bmatrix} \underset{\sim}{G}_{11}(d_k) & \underset{\sim}{G}_{12}(d_k) \\ \underset{\sim}{G}_{21}(d_k) & \underset{\sim}{G}_{22}(d_k) \end{bmatrix}$$

is formed, when the interval between successive observations is d_k units ($d_k \in Z^+$), given that we have knowledge of the univariate transition matrices, $\underset{\sim}{G}_{11}(d_k)$ and $\underset{\sim}{G}_{22}(d_k)$. We shall consider, in turn, the cases where there is (i) no causality, (ii) unidirectional causality, (iii) feedback in the equally-spaced system transition matrix,

$$\underset{\sim}{G} = \begin{bmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{G}_{12} \\ \underset{\sim}{G}_{21} & \underset{\sim}{G}_{22} \end{bmatrix}.$$

Block-Diagonal Transition Matrix

Assume that there is no causality between $\underset{\sim}{\theta}_{1t}$ and $\underset{\sim}{\theta}_{2t}$, $\forall t$. Then we may write:

$$\underset{\sim}{G} = \begin{bmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{G}_{22} \end{bmatrix} \tag{5.91}$$

According to Equation (4.6) given in Section 4.2, when the interval between successive observation vectors (or part-vectors) is d_k units, we have:

$$\underset{\sim}{G}_k = \begin{bmatrix} \underset{\sim}{G}_{11}(d_k) & \underset{\sim}{G}_{12}(d_k) \\ \underset{\sim}{G}_{21}(d_k) & \underset{\sim}{G}_{22}(d_k) \end{bmatrix} = \underset{\sim}{G}^{d_k} \tag{5.92}$$

and, using (5.91), it is clear that:

$$\underset{\sim}{G}_k = \begin{bmatrix} \underset{\sim}{G}_{11}^{d_k} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{G}_{22}^{d_k} \end{bmatrix} \tag{5.93}$$

so that

$$\underset{\sim}{G}_{11}(d_k) = \underset{\sim}{G}_{11}^{d_k},\ \underset{\sim}{G}_{12}(d_k) = \underset{\sim}{0},\ \underset{\sim}{G}_{21}(d_k) = \underset{\sim}{0} \text{ and } \underset{\sim}{G}_{22}(d_k) = \underset{\sim}{G}_{22}^{d_k} \tag{5.94}$$

The usefulness of this particularly simple identity lies in the fact that, since we have derived $\underset{\sim}{G}_{11}^{d_k}$ and $\underset{\sim}{G}_{22}^{d_k}$ for a number of univariate models (see Section 4.3), we need not derive any additional results for the specification of $\underset{\sim}{G}_k$ for the bivariate model. Note that in this instance, and in the following instances, $\underset{\sim}{R}_{\omega}(k)$ (the system perturbation covariance matrix) may be calculated using either (4.8) or (4.9) once $\underset{\sim}{G}_k$ has been derived.

Unidirectional Causality

Assume that $\underset{\sim}{\theta}_{2t}$ is dependent upon $\underset{\sim}{\theta}_{1t-1}$ (or upon some earlier value of $\underset{\sim}{\theta}_1$, using the arguments put forward in Section 5.3.1.1) and that $\underset{\sim}{\theta}_{1t}$ is not dependent on $\underset{\sim}{\theta}_{2t}$, $\forall t$, so that

$$\underset{\sim}{G} = \begin{pmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{0} \\ \underset{\sim}{G}_{21} & \underset{\sim}{G}_{22} \end{pmatrix} \tag{5.95}$$

where $\underset{\sim}{G}_{21} \neq \underset{\sim}{0}$.

Lemma 5.3.1

For any $d_k \in \mathbb{Z}^+$,

$$\underset{\sim}{G}_k = \underset{\sim}{G}^{d_k} = \begin{pmatrix} \underset{\sim}{G}_{11}^{d_k} & \underset{\sim}{0} \\ \underset{\sim}{G}_{21}(d_k) & \underset{\sim}{G}_{22}^{d_k} \end{pmatrix} \tag{5.96}$$

where $\underset{\sim}{G}_{21}(d_k)$ satisfies the recurrence relationship:

$$\underset{\sim}{G}_{21}(d_k) = \underset{\sim}{G}_{21} \cdot \underset{\sim}{G}_{11}^{d_k - 1} + \underset{\sim}{G}_{22} \cdot \underset{\sim}{G}_{21}(d_k - 1) \tag{5.97}$$

Proof

Suppose (5.96) and (5.97) hold for $d_k = d_j - 1$. Then

$$\underset{\sim}{G}_j = \underset{\sim}{G}^{d_j} = \underset{\sim}{G}.\underset{\sim}{G}^{d_j - 1} = \begin{pmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{0} \\ \underset{\sim}{G}_{21} & \underset{\sim}{G}_{22} \end{pmatrix} \begin{pmatrix} \underset{\sim}{G}_{11}^{d_j - 1} & \underset{\sim}{0} \\ \underset{\sim}{G}_{21}(d_j - 1) & \underset{\sim}{G}_{22}^{d_j - 1} \end{pmatrix}$$

$$= \begin{pmatrix} \underset{\sim}{G}_{11}^{d_j} & \underset{\sim}{0} \\ \underset{\sim}{G}_{21}(d_j) & \underset{\sim}{G}_{22}^{d_j} \end{pmatrix}$$

where

$$\underset{\sim}{G}_{21}(d_j) = \underset{\sim}{G}_{21}.\underset{\sim}{G}_{11}^{d_j-1} + \underset{\sim}{G}_{22}.\underset{\sim}{G}_{21}(d_j - 1).$$

Therefore, if (5.96) and (5.97) are true for $d_k = d_j - 1$ they are also true for $d_k = d_j$ and, since they clearly hold for $d_k = 1$, in which case $\underset{\sim}{G}_k = \underset{\sim}{G}$, they must therefore be true for all $d_k \in Z^+$, by induction.

Notice that the block-diagonal part of $\underset{\sim}{G}_k$ may, once again, be identified directly in any particular modelling application from the results of Section 4.3. Notice, too, that the difference equation given by (5.97) involves <u>matrices</u> and its general solution, therefore, is dependent upon the elements of these matrices. As long as $\underset{\sim}{G}$ is completely specified, however, the solution of (5.97) is relatively trivial. For the case where G_{11}, G_{22} and G_{21} are scalars, (5.97) yields:

$$G_{21}(d_k) = G_{21}.G_{22}^{d_k-1} \sum_{t=0}^{d_k-1} \left(\frac{G_{11}}{G_{22}}\right)^t \tag{5.98}$$

Moreover, if $|G_{11}/G_{22}| < 1$ this relationship simplifies, still further, to:

$$G_{21}(d_k) = \frac{G_{21}(G_{22}^{d_k} - G_{11}^{d_k})}{G_{22} - G_{11}}, \quad (G_{11} \neq G_{22}) \tag{5.99}$$

N.B. For the case when $\underset{\sim}{G}_{21} = \underset{\sim}{0}$ but $\underset{\sim}{G}_{12} \neq \underset{\sim}{0}$, it is clear that similar results may be derived, due to symmetry.

Example: For the bivariate steady model with unidirectional causality, given by (5.88), we have:

$$\underset{\sim}{G} = \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \Rightarrow \underset{\sim}{G}^{d_k} = \begin{pmatrix} 1 & cd_k \\ 0 & 1 \end{pmatrix}$$

using (5.96) and (5.98).

A couple of specific models involving unidirectional causality will be examined in more detail in Section 5.3.2.

Feedback

Assume that $\underset{\sim}{\theta}_{2t}$ is dependent upon $\underset{\sim}{\theta}_{1t-1}$ and that $\underset{\sim}{\theta}_{1t}$ is dependent upon $\underset{\sim}{\theta}_{2t-1}$, $\forall t$, so that

$$\underset{\sim}{G} = \begin{pmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{G}_{12} \\ \underset{\sim}{G}_{21} & \underset{\sim}{G}_{22} \end{pmatrix} \tag{5.100}$$

where $\underset{\sim}{G}_{12} \neq \underset{\sim}{0}$ and $\underset{\sim}{G}_{21} \neq \underset{\sim}{0}$.

In this case there is no explicit general form for $\underset{\sim}{G}_k$ (except for the conventional eigenvalue/eigenvector representation) and it is certainly true that $\underset{\sim}{G}_{11}(d_k) \neq \underset{\sim}{G}_{11}^{d_k}$, etc., in general, since it is easy to show that

$$\underset{\sim}{G}_k = \begin{pmatrix} \underset{\sim}{G}_{11}(d_k) & \underset{\sim}{G}_{12}(d_k) \\ \underset{\sim}{G}_{21}(d_k) & \underset{\sim}{G}_{22}(d_k) \end{pmatrix}$$

where

$$\left.\begin{aligned} \underset{\sim}{G}_{11}(d_k) &= \underset{\sim}{G}_{11}\cdot\underset{\sim}{G}_{11}(d_k - 1) + \underset{\sim}{G}_{12}\cdot\underset{\sim}{G}_{21}(d_k - 1) \\ \underset{\sim}{G}_{12}(d_k) &= \underset{\sim}{G}_{11}\cdot\underset{\sim}{G}_{12}(d_k - 1) + \underset{\sim}{G}_{12}\cdot\underset{\sim}{G}_{22}(d_k - 1) \\ \underset{\sim}{G}_{21}(d_k) &= \underset{\sim}{G}_{21}\cdot\underset{\sim}{G}_{11}(d_k - 1) + \underset{\sim}{G}_{22}\cdot\underset{\sim}{G}_{21}(d_k - 1) \\ \underset{\sim}{G}_{22}(d_k) &= \underset{\sim}{G}_{21}\cdot\underset{\sim}{G}_{12}(d_k - 1) + \underset{\sim}{G}_{22}\cdot\underset{\sim}{G}_{22}(d_k - 1) \end{aligned}\right\} \tag{5.101}$$

We note, once more, however, that this set of recurrence relationships can be solved as long as $\underset{\sim}{G}$ is completely specified.

In this case, however, the results from Section 4.3 are of little value.

Example: Consider the bivariate steady model with feedback, for which

$$\underset{\sim}{G} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = \begin{pmatrix} 1 & c_1 \\ c_2 & 1 \end{pmatrix}$$

Then (5.101) implies that

$$\underset{\sim}{G}_k = \begin{pmatrix} G_{11}(d_k) & G_{12}(d_k) \\ G_{21}(d_k) & G_{22}(d_k) \end{pmatrix},$$

where

$$G_{11}(d_k) = G_{11}(d_k - 1) + c_1G_{21}(d_k - 1) \tag{5.102}$$

$$G_{12}(d_k) = G_{12}(d_k - 1) + c_1G_{22}(d_k - 1) \tag{5.103}$$

$$G_{21}(d_k) = G_{21}(d_k - 1) + c_2G_{11}(d_k - 1) \tag{5.104}$$

$$G_{22}(d_k) = G_{22}(d_k - 1) + c_2G_{12}(d_k - 1) \tag{5.105}$$

From (5.102) and (5.104), we obtain:

$$G_{11}(d_k + 1) - 2G_{11}(d_k) + (1 - c_1c_2)G_{11}(d_k - 1) = 0$$

which implies that

$$G_{11}(d_k) = \tfrac{1}{2}\{(1 + \sqrt{c_1c_2})^{d_k} + (1 - \sqrt{c_1c_2})^{d_k}\} \tag{5.106}$$

and that, using (5.102),

$$G_{21}(d_k) = \frac{\sqrt{c_2}}{2\sqrt{c_1}}\{(1 + \sqrt{c_1c_2})^{d_k} - (1 - \sqrt{c_1c_2})^{d_k}\} \qquad (5.107)$$

Similarly,

$$G_{22}(d_k) = \tfrac{1}{2}\{(1 + \sqrt{c_1c_2})^{d_k} + (1 - \sqrt{c_1c_2})^{d_k}\} \qquad (5.108)$$

and

$$G_{12}(d_k) = \frac{\sqrt{c_1}}{2\sqrt{c_2}}\{(1 + \sqrt{c_1c_2})^{d_k} - (1 - \sqrt{c_1c_2})^{d_k}\} \qquad (5.109)$$

NOTE: For the bivariate steady model <u>without</u> causality, $\underset{\sim}{G}_k = \underset{\sim}{I}$.

5.3.1.3 Changepoint Causality: Markovian State-Transition

We now examine the way in which Markovian state dependence can be used to incorporate changepoint causality, taking unequally-spaced data into account. Recall the definition given by (5.89) that ${}^{(i)}p_o^{(j)} = p(M_t^{(j)}|M_{t-1}^{(i)})$; i.e. the state, j, at time t is dependent upon the state, i, at time t - 1, $\forall t$. Thus, for instance, if state i represents a change in the y_{1t} series and state j represents a change in the y_{2t} series, we have unidirectional changepoint causality.

We could make this definition even more flexible by allowing the state-dependence to change with time, i.e.

$$p(M_t^{(j)}|M_{t-1}^{(i)}) = {}^{(i)}p_{t,t-1}^{(j)} \qquad (5.110)$$

For most purposes, however, (5.89) is adequate, bearing in mind that we could always redefine ${}^{(i)}p_o^{(j)}$, at any particular timepoint, if we had knowledge of an intervention which might change our beliefs about the ensuing state. The Markovian formulation is, in fact, very useful for dealing with interventions. Suppose, for example, that an instantly active drug is given to a patient at time t - 1 and that, just after infusion, we observe an unusually low value (say) in the series being monitored. Then we would be much more inclined to believe that another low value will follow, at time t, rather than a return to the previous steady state, since the drug is believed to have an immediate and prolonged effect; i.e.

p(Sustained Effect,t|Low Value,t - 1) > p(Steady State,t|Low Value,t - 1).

This type of information cannot be incorporated when we use the non-dependent $p_o^{(j)}$ formulation unless

p(Sustained Effect,t) > p(Steady State,t) ∀t, which is highly undesirable.

Consider the case when the interval between successive observations is d_k units, where $d_k > 1$, and let

$$ {}^{(i)}p_k^{(j)} = p(M_k^{(j)} | M_{k-1}^{(i)}) \tag{5.111} $$

denote the probability that state j obtains at timepoint k (at time T_k) given that state i obtains at timepoint k - 1 (at time T_{k-1}). In order to calculate ${}^{(i)}p_k^{(j)}$, we shall introduce some further notation. Since observations are available at times T_{k-1} and T_k, we may write

$$ M_k^{(j)} = M_{T_k}^{(j)} \qquad \text{and} \qquad M_{k-1}^{(i)} = M_{T_{k-1}}^{(i)} \tag{5.112} $$

Let

$$ j(o) = j \qquad \text{and} \qquad j(d_k) = i \tag{5.113} $$

and suppose that there are J possible states. Then:

$$^{(i)}p_k^{(j)} = {}^{(j(d_k))}p_k^{(j(o))} = p(M_{T_k}^{(j(o))} | M_{T_{k-1}}^{(j(d_k))})$$

$$= \sum_{j(1)=1}^{J} p(M_{T_k}^{(j(o))} | M_{T_k-1}^{(j(1))}, M_{T_{k-1}}^{(j(d_k))}) p(M_{T_k-1}^{(j(1))} | M_{T_{k-1}}^{(j(d_k))})$$

$$= \sum_{j(1)=1}^{J} {}^{(j(1))}p_o^{(j(o))} \cdot \sum_{j(2)=1}^{J} p(M_{T_k-1}^{(j(1))} | M_{T_k-2}^{(j(2))}, M_{T_{k-1}}^{(j(d_k))})$$

$$\times\, p(M_{T_k-2}^{(j(2))} | M_{T_{k-1}}^{(j(d_k))})$$

(using (5.89))

$$= \sum_{j(1)=1}^{J} \cdots \sum_{j(d_k-1)=1}^{J} \prod_{i=1}^{d_k} {}^{j(i)}p_o^{j(i-1)} \qquad (5.114)$$

i.e. when the current interval is d_k (> 1) units, we use (5.114) in place of $^{(i)}p_o^{(j)}$; if $d_k = 1$, we use $^{(i)}p_o^{(j)}$ as before.

A much neater way of arriving at the above relationship involves the use of a state transition (J x J) matrix, $\{p_o\}$, whose (i,j)th element is $^{(i)}p_o^{(j)}$, i.e.

$$\{p_o\} = \begin{bmatrix} ^{(1)}p_o^{(1)} & ^{(1)}p_o^{(2)} & \cdots & ^{(1)}p_o^{(J)} \\ ^{(J)}p_o^{(1)} & \cdots\cdots & & ^{(J)}p_o^{(J)} \end{bmatrix} \qquad (5.115)$$

Then, for $d_k \in Z^+$, $^{(i)}p_k^{(j)}$ is the (i,j)th element of $\{p_o\}^{d_k}$. (5.116)

It then becomes clear that if we wish to incorporate a time lag, ℓ, into our Markovian state dependence, we merely set the (i,j)th element of $\{p_o\}^{\ell} = {}^{(i)}p_{\ell}^{(j)} = p(M_t^{(j)} | M_{t-\ell}^{(i)})$.

. There will clearly be a significant interaction between event detection and the sampling interval, since a longer interval may make the chance of some changepoint-types less likely. We note that by using (3.6) (i.e. non-dependent state-transition) we have no way of modelling this interaction, whereas (5.116) provides a sensible method for adjusting our beliefs about the state occupied according to the size of the sampling interval. Note, too, that since $\{p_o\}$ is a stochastic transition matrix, it may well equilibriate once the interval is large enough; i.e., once the interval is greater than a certain length, our beliefs remain unchanged.

By way of completeness, we shall derive the multistate probabilities needed for event detection (re. Section 3.2), for the case when ${}^{(i)}p_k^{(j)}$ is preferred to $p_o^{(j)}$. In order to discriminate fully between alternative changepoint-types, we still require quantities of the form $p(M_k^{(j)}|\underset{\sim}{D}_k)$, etc., which are not dependent upon previous states. Let

$$p_{T_k}^{(j)} = p(M_k^{(j)}|\underset{\sim}{D}_k) \tag{5.117}$$

$$o_{T_k}^{(i)} = p(M_{k-1}^{(i)}|\underset{\sim}{D}_k) \tag{5.118}$$

$$T_{T_k}^{(h)} = p(M_{k-2}^{(h)}|\underset{\sim}{D}_k) \tag{5.119}$$

and

$$Z_k^{(ij)} = p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)},M_k^{(j)}) \tag{5.120}$$

where $Z_k^{(ij)}$ is the predictive density (see Section 5.2), and where $\underset{\sim}{y}_k$ and $\underset{\sim}{D}_k$ are defined by (5.5) and (5.6), respectively. Then, using Bayes theorem and (5.111),

$$\begin{aligned}
p_{T_k}^{(j)} &= p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1},M_k^{(j)})p(M_k^{(j)}|\underset{\sim}{D}_{k-1})/p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1}) \\
&= \sum_{i=1}^{J} Z_k^{(ij)} p(M_{k-1}^{(i)}|\underset{\sim}{D}_{k-1},M_k^{(j)}).p(M_k^{(j)}|\underset{\sim}{D}_{k-1})/p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1}) \\
&= \sum_{i=1}^{J} Z_k^{(ij)} \frac{p(M_k^{(j)}|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)}).p(M_{k-1}^{(i)}|\underset{\sim}{D}_{k-1})}{p(M_k^{(j)}|\underset{\sim}{D}_{k-1})}.\frac{p(M_k^{(j)}|\underset{\sim}{D}_{k-1})}{p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1})} \\
&= \sum_{i=1}^{J} Z_k^{(ij)}\,{}^{(i)}p_k^{(j)}\,p_{T_{k-1}}^{(i)}/p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1})
\end{aligned} \tag{5.121}$$

[c.f. (3.28)], where

$$p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1}) = \sum_{j=1}^{J}\sum_{i=1}^{J} Z_k^{(ij)(i)} p_k^{(j)} p_{T_{k-1}}^{(i)} \tag{5.122}$$

[c.f. (3.21)]. Also

$$\begin{aligned} O_{T_k}^{(i)} &= p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1},M_{k-1}^{(i)})p(M_{k-1}^{(i)}|\underset{\sim}{D}_{k-1})/p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1}) \\ &= \sum_{j=1}^{J} Z_k^{(ij)(i)} p_k^{(j)} p_{T_{k-1}}^{(i)} / p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1}) \end{aligned} \tag{5.123}$$

[c.f. (3.31)], where $p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1})$ is given by (5.122). Similarly,

$$\begin{aligned} T_{T_K}^{(h)} &= p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1},M_{k-2}^{(h)})p(M_{k-2}^{(h)}|\underset{\sim}{D}_{k-1})/p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1}) \\ &= \sum_{i=1}^{J} p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1},M_{k-2}^{(h)},M_{k-1}^{(i)})p(M_{k-1}^{(i)}|\underset{\sim}{D}_{k-1},M_{k-2}^{(h)}).O_{T_{k-1}}^{(h)}/p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1}) \\ &= \sum_{i=1}^{J}\sum_{j=1}^{J} Z_k^{(ij)(i)} p_k^{(j)} . \frac{Z_{k-1}^{(hi)\ (h)} p_{k-1}^{(i)}}{\sum_{i=1}^{J} Z_{k-1}^{(hi)(h)} p_{k-1}^{(i)}} . O_{T_{k-1}}^{(h)} / p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1}) \end{aligned} \tag{5.124}$$

[c.f. (3.32)], where $p(\underset{\sim}{y}_k|\underset{\sim}{D}_{k-1})$ is given by (5.122) and where $Z_{k-1}^{(hi)}$ and $O_{T_{k-1}}^{(h)}$ have been calculated at the previous recursion.

NOTES: (i) If the model contains nuisance parameters, ϕ, these can be incorporated into the above calculations by referring to (3.47) to (3.49) in Section 3.2.

(ii) The idea of Markovian state-transition is not restricted to the bivariate case, and may prove to be useful for univariate models as well. For instance, a patient on kidney dialysis would expect the concentration levels of certain blood chemicals to drop immediately following treatment. In the absence of further treatment, however, the concentrations of these chemicals would soon return to their original levels, since dialysis is not a curative treatment. In this case we would like to set p(Level Change,t | Level Change,t - 1) quite high.

5.3.2 SPECIFIC BIVARIATE MODELS

In order to show how bivariate models for unequally-spaced data are formed, we present, here, two particular examples: the bivariate linear growth model and the bivariate AR(1)/linear growth model. In each case, we have allowed for unidirectional steady-state causality, but not feedback. We have also allowed for concurrent changepoint causality, by letting the perturbations on the level (say) of each series have a non-zero correlation. Once we have derived the correct form for $\underset{\sim}{G}_k$ and $\underset{\sim}{R}_{\omega}(k)$, the unequally-spaced bivariate model can be completely specified, and the results from Section 5.2.2 may be used for updating the system, on the receipt of successive observation vectors (or part-vectors).

For the most part, notation will conform with that used in Section 2.3, except that additional subscripts '1' and '2' will be adopted, when appropriate, in order to distinguish between those parameters associated with the first series and those associated with the second. In addition, we shall write:

$$\text{cov}(\varepsilon_{1t},\varepsilon_{2t}) = \lambda^{-1}R_{\varepsilon\varepsilon} \quad (5.125)$$

$$\text{cov}(\delta\mu_{1t},\delta\mu_{2t}) = \lambda^{-1}R_{\mu\mu} \quad (5.126)$$

$$\text{cov}(\delta\beta_{1t},\delta\beta_{2t}) = \lambda^{-1}R_{\beta\beta} \quad (5.127)$$

where $\delta\mu_1$ is the perturbation associated with the level of the first series, etc.

5.3.2.1 *Bivariate Linear Growth Model*

Consider the bivariate model for which each of the univariate series is thought to be reasonably well modelled by a linear growth model (as described in Sections 2.3.1, 3.3.1 and 4.3.1). Using the notation from Section 5.3.1 we have, for the equally-spaced case,

$$\underset{\sim}{G}_{11} = \underset{\sim}{G}_{22} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{5.128}$$

Assume also that the level of the second series, μ_2, is believed to be dependent upon the level of the first series, μ_1, but that no feedback exists, so that

$$\left.\begin{aligned} \mu_{2t} &= \mu_{2t-1} + \beta_{2t} + \delta\mu_{2t} + c\mu_{1t} \\ \mu_{1t} &= \mu_{1t-1} + \beta_{1t} + \delta\mu_{1t} \end{aligned}\right\} \tag{5.129}$$

where c is a scalar (assumed known). Then,

$$\underset{\sim}{G} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ c & c & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.130}$$

i.e.

$$\underset{\sim}{G}_{12} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \underset{\sim}{G}_{21} = \begin{pmatrix} c & c \\ 0 & 0 \end{pmatrix} \tag{5.131}$$

The full equally-spaced model is:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_{1t} \\ \beta_{1t} \\ \mu_{2t} \\ \beta_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{5.132}$$

$$\begin{pmatrix} \mu_{1t} \\ \beta_{1t} \\ \mu_{2t} \\ \beta_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ c & c & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{1t-1} \\ \beta_{1t-1} \\ \mu_{2t-1} \\ \beta_{2t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_{1t} + \delta\beta_{1t} \\ \delta\beta_{1t} \\ \delta\mu_{2t} + \delta\beta_{2t} + c(\delta\mu_{1t} + \delta\beta_{1t}) \\ \delta\beta_{2t} \end{pmatrix} \tag{5.133}$$

where

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N(\underset{\sim}{0}, \lambda^{-1} \begin{pmatrix} R_{\varepsilon 1} & R_{\varepsilon\varepsilon} \\ R_{\varepsilon\varepsilon} & R_{\varepsilon 2} \end{pmatrix}).$$

In addition to the above, we shall assume that:

$$\left.\begin{array}{l} \text{var}(\delta\mu_{1t}) = \lambda^{-1}R_{\mu 1},\ \text{var}(\delta\mu_{2t}) = \lambda^{-1}R_{\mu 2},\ \text{var}(\delta\beta_{1t}) = \lambda^{-1}R_{\beta 1}, \\ \text{var}(\delta\beta_{2t}) = \lambda^{-1}R_{\beta 2},\ \text{cov}(\delta\mu_{1t},\delta\mu_{2t}) = \lambda^{-1}R_{\mu\mu}, \\ \text{cov}(\delta\beta_{1t},\delta\beta_{2t}) = \lambda^{-1}R_{\beta\beta}, \\ \text{and that all other covariances involved are zero.} \end{array}\right\} \tag{5.134}$$

Then, for the equally-spaced model,

$$\underset{\sim}{R}_{\omega} = \begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ R_{12} & R_{22} & R_{23} & R_{24} \\ R_{13} & R_{23} & R_{33} & R_{34} \\ R_{14} & R_{24} & R_{34} & R_{44} \end{pmatrix}, \text{ say} \tag{5.135}$$

where

$$R_{11} = R_{\mu 1} + R_{\beta 1} \tag{5.136}$$

$$R_{12} = R_{\beta 1} \tag{5.137}$$

$$R_{13} = R_{\mu\mu} + R_{\beta\beta} + c(R_{\mu 1} + R_{\beta 1}) \tag{5.138}$$

$$R_{14} = R_{\beta\beta} \tag{5.139}$$

$$R_{22} = R_{\beta 1} \tag{5.140}$$

$$R_{23} = R_{\beta\beta} + cR_{\beta 1} \tag{5.141}$$

$$R_{24} = R_{\beta\beta} \tag{5.142}$$

$$R_{33} = R_{\mu 2} + R_{\beta 2} + c^2(R_{\mu 1} + R_{\beta 1}) + 2c(R_{\mu\mu} + R_{\beta\beta}) \tag{5.143}$$

$$R_{34} = R_{\beta 2} + cR_{\beta\beta} \tag{5.144}$$

and

$$R_{44} = R_{\beta 2} \tag{5.145}$$

In order to form the unequally-spaced equivalent of (5.130) and (5.135) we must first derive $\underset{\sim}{G}_k$, and then $\underset{\sim}{R}_\omega(k)$, when the interval is d_k units. Using the results from the previous section, we see that $\underset{\sim}{G}_k$ must be of the form:

$$\underset{\sim}{G}_k = \begin{pmatrix} \underset{\sim}{G}_{11}^{d_k} & \underset{\sim}{0} \\ \underset{\sim}{G}_{21}(d_k) & \underset{\sim}{G}_{22}^{d_k} \end{pmatrix}$$

and, from Section 4.3.1,

$$\underset{\sim}{G}_{11}^{d_k} = \underset{\sim}{G}_{22}^{d_k} = \begin{pmatrix} 1 & d_k \\ 0 & 1 \end{pmatrix} \tag{5.146}$$

Moreover,

$$\underset{\sim}{G}_{21}(d_k) = \underset{\sim}{G}_{21}\underset{\sim}{G}_{11}^{d_k - 1} + \underset{\sim}{G}_{22}\underset{\sim}{G}_{21}(d_k - 1) \tag{5.147}$$

Let

$$\underset{\sim}{G}_{21}(d_k) = \begin{pmatrix} g_{211}(d_k) & g_{212}(d_k) \\ 0 & 0 \end{pmatrix} \tag{5.148}$$

Then, from (5.146) and (5.147),

$$\begin{pmatrix} g_{211}(d_k) & g_{212}(d_k) \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} c & c \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & d_k - 1 \\ 0 & 1 \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} g_{211}(d_k - 1) & g_{212}(d_k - 1) \\ 0 & 0 \end{pmatrix}$$

i.e.

$$g_{211}(d_k) = c + g_{211}(d_k - 1) \Rightarrow g_{211}(d_k) = cd_k \tag{5.149}$$

and

$$\begin{aligned} g_{212}(d_k) &= cd_k + g_{212}(d_k - 1) \\ &= c(d_k + (d_k - 1) + \ldots + 1) \\ &= \tfrac{1}{2}cd_k(d_k + 1) \end{aligned} \tag{5.150}$$

i.e.

$$\underset{\sim}{G}_k = \begin{pmatrix} 1 & d_k & 0 & 0 \\ 0 & 1 & 0 & 0 \\ cd_k & \tfrac{1}{2}cd_k(d_k + 1) & 1 & d_k \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.151}$$

From (4.8), given in Section 4.2, we have

$$\underset{\sim}{R}_\omega(k) = \sum_{t=T_{k-1}+1}^{T_k} \underset{\sim}{G}^{T_k - t} \underset{\sim}{R}_\omega (\underset{\sim}{G}^{T_k - t})^T$$

$$= \sum_{t=T_{k-1}+1}^{T_k} \begin{pmatrix} 1 & T_k-t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ c(T_k-t) & \tfrac{1}{2}c(T_k-t)(T_k-t+1) & 1 & T_k-t \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ R_{12} & R_{22} & R_{23} & R_{24} \\ R_{13} & R_{23} & R_{33} & R_{34} \\ R_{14} & R_{24} & R_{34} & R_{44} \end{pmatrix}$$

$$\times \begin{pmatrix} 1 & 0 & c(T_k-t) & 0 \\ T_k-t & 1 & \tfrac{1}{2}c(T_k-t)(T_k-t+1) & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & T_k-t & 1 \end{pmatrix} = \begin{pmatrix} R_{11k} & R_{12k} & \cdots & R_{14k} \\ \vdots & & & \vdots \\ R_{14k} & \cdots & \cdots & R_{44k} \end{pmatrix}, \tag{5.152}$$

say, where

$$R_{11k} = \sum_{t=T_{k-1}+1}^{T_k} (R_{\mu 1} + (T_k - t + 1)^2 R_{\beta 1})$$

$$= d_k R_{\mu 1} + \frac{1}{6} d_k (d_k + 1)(2d_k + 1) R_{\beta 1} \tag{5.153}$$

$$R_{12k} = \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_{\beta 1} = \tfrac{1}{2} d_k (d_k + 1) R_{\beta 1} \tag{5.154}$$

$$R_{13k} = \sum_{t=T_{k-1}+1}^{T_k} \{R_{\mu\mu} + (T_k - t + 1)^2 R_{\beta\beta} + c((T_k - t + 1) R_{\mu 1} + \tfrac{1}{2}(T_k - t + 1)^2 (T_k - t + 2) R_{\beta 1})\}$$

$$= d_k R_{\mu\mu} + \frac{1}{6} d_k (d_k + 1)(2d_k + 1) R_{\beta\beta} + c\{\tfrac{1}{2} d_k (d_k + 1) R_{\mu 1} + \frac{1}{24} d_k (d_k + 1)(d_k + 2)(3d_k + 1) R_{\beta 1}\} \tag{5.155}$$

$$R_{14k} = \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_{\beta\beta} = \tfrac{1}{2} d_k (d_k + 1) R_{\beta\beta} \tag{5.156}$$

$$R_{22k} = \sum_{t=T_{k-1}+1}^{T_k} R_{\beta 1} = d_k R_{\beta 1} \tag{5.157}$$

$$R_{23k} = \sum_{t=T_{k-1}+1}^{T_k} \{(T_k - t + 1) R_{\beta\beta} + \tfrac{1}{2} c (T_k - t + 1)(T_k - t + 2) R_{\beta 1}\}$$

$$= \tfrac{1}{2} d_k (d_k + 1) R_{\beta\beta} + \frac{1}{6} c d_k (d_k + 1)(d_k + 2) R_{\beta 1} \tag{5.158}$$

$$R_{24} = \sum_{t=T_{k-1}+1}^{T_k} R_{\beta\beta} = d_k R_{\beta\beta} \tag{5.159}$$

$$R_{33k} = \sum_{t=T_{k-1}+1}^{T_k} \{R_{\mu 2} + (T_k - t + 1)^2 R_{\beta 2} + c^2[(T_k - t + 1)^2 R_{\mu 1} + \tfrac{1}{4}(T_k - t + 1)^2 (T_k - t + 2)^2 R_{\beta 1}] + 2c[(T_k - t + 1) R_{\mu\mu} + \tfrac{1}{2}(T_k - t + 1)^2 \times (T_k - t + 2) R_{\beta\beta}]\}$$

$$= d_k R_{\mu 2} + \frac{1}{6} d_k(d_k + 1)(2d_k + 1)R_{\beta 2}$$
$$+ c^2\{\frac{1}{6} d_k(d_k + 1)(2d_k + 1)R_{\mu 1} + \frac{1}{60} d_k(d_k + 1)(d_k + 2) \times (3d_k^2 + 6d_k + 1)R_{\beta 1}\}$$
$$+ 2c\{\tfrac{1}{2}d_k(d_k + 1)R_{\mu\mu} + \frac{1}{24} d_k(d_k + 1)(d_k + 2) \times (3d_k + 1)R_{\beta\beta}\} \quad (5.160)$$

$$R_{34k} = \sum_{t=T_{k-1}+1}^{T_k} \{(T_k - t + 1)R_{\beta 2} + \tfrac{1}{2}c(T_k - t + 1)(T_k - t + 2)R_{\beta\beta}\}$$
$$= \tfrac{1}{2}d_k(d_k + 1)R_{\beta 2} + \frac{1}{6} cd_k(d_k + 1)(d_k + 2)R_{\beta\beta} \quad (5.161)$$

and

$$R_{44k} = \sum_{t=T_{k-1}+1}^{T_k} R_{\beta 2} = d_k R_{\beta 2} \quad (5.162)$$

(making use of the results stated by (4.16) to (4.20)).

5.3.2.2 *AR(1)/Linear Growth Model*

Consider the bivariate model for which the first series may be modelled satisfactorily by a first-order autoregressive model (as described in Sections 2.3.5.1, 3.3.5.1 and 4.3.5.1), whereas a linear growth model is thought to be appropriate for the second series. We make two notes here. Firstly, the model contains a nuisance parameter, ϕ (by way of the autoregressive parameter, which is treated as such). This means that we may run into 'size' problems when it comes to applying the model to actual data (see Section 5.2.1). Secondly, this bivariate model differs from most of the bivariate models discussed in the literature, in that we have assumed that the individual series are generated by different classes of model (one being a member of the ARMA class of models, and the other being a form of polynomial growth).

For the equally-spaced case, we have

$$\underset{\sim}{G}_{11} = \begin{pmatrix} \phi & 1 - \phi \\ 0 & 1 \end{pmatrix} \text{ and } \underset{\sim}{G}_{22} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{5.163}$$

Assume, once more, that the level of the second series, μ_2, is dependent upon the true level of the first series, ν, but that no feedback exists, so that

$$\left.\begin{aligned} \mu_{2t} &= \mu_{2t-1} + \beta_t + \delta\mu_{2t} + c\nu_t \\ \nu_t &= \nu_{t-1} + \delta\nu_t \end{aligned}\right\} \tag{5.164}$$

where c is a scalar (assumed known). Then,

$$\underset{\sim}{G} = \begin{pmatrix} \phi & 1 - \phi & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & c & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.165}$$

i.e.

$$\underset{\sim}{G}_{12} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \underset{\sim}{G}_{21} = \begin{pmatrix} 0 & c \\ 0 & 0 \end{pmatrix} \tag{5.166}$$

The full equally-spaced model is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_{1t} \\ \nu_t \\ \mu_{2t} \\ \beta_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{5.167}$$

$$\begin{pmatrix} \mu_{1t} \\ \nu_t \\ \mu_{2t} \\ \beta_t \end{pmatrix} = \begin{pmatrix} \phi & 1 - \phi & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & c & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{1t-1} \\ \nu_{t-1} \\ \mu_{2t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_{1t} + \delta\nu_t \\ \delta\nu_t \\ \delta\mu_{2t} + \delta\beta_t + c\,\delta\nu_t \\ \delta\beta_t \end{pmatrix} \tag{5.168}$$

where

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N(\underset{\sim}{0}, \lambda^{-1} \begin{pmatrix} R_{\varepsilon 1} & R_{\varepsilon\varepsilon} \\ R_{\varepsilon\varepsilon} & R_{\varepsilon 2} \end{pmatrix}).$$

In addition to the above, we shall assume that

$$\left.\begin{array}{l} var(\delta\mu_{1t}) = \lambda^{-1}R_{\mu 1},\ var(\delta_{\mu 2t}) = \lambda^{-1}R_{\mu 2},\ var(\delta\nu_t) = \lambda^{-1}R_{\nu}, \\ var(\delta\beta_t) = \lambda^{-1}R_{\beta},\ cov(\delta\mu_{1t},\delta\mu_{2t}) = \lambda^{-1}R_{\mu\mu}, \\ \text{and that all other covariances involved are zero} \end{array}\right\} \quad (5.169)$$

Then

$$\underset{\sim}{R}_{\omega} = \begin{pmatrix} R_{\mu 1} + R_{\nu} & R_{\nu} & R_{\mu\mu} + cR_{\nu} & 0 \\ R_{\nu} & R_{\nu} & cR_{\nu} & 0 \\ R_{\mu\mu} + cR_{\nu} & cR_{\nu} & R_{\mu 2} + R_{\beta} + c^2R_{\nu} & R_{\beta} \\ 0 & 0 & R_{\beta} & R_{\beta} \end{pmatrix} \quad (5.170)$$

If the interval between successive observations is d_k units, using Lemma 5.3.1 we see that

$$\underset{\sim}{G}_k = \underset{\sim}{G}^{d_k} = \begin{pmatrix} \underset{\sim}{G}_{11}^{d_k} & \underset{\sim}{0} \\ \underset{\sim}{G}_{21}(d_k) & \underset{\sim}{G}_{22}^{d_k} \end{pmatrix}$$

where

$$\underset{\sim}{G}_{21}(d_k) = \underset{\sim}{G}_{21}\underset{\sim}{G}_{11}^{d_k-1} + \underset{\sim}{G}_{22}\underset{\sim}{G}_{21}(d_k - 1)$$

Now, from Section 4.3.1,

$$\underset{\sim}{G}_{22}^{d_k} = \begin{pmatrix} 1 & d_k \\ 0 & 1 \end{pmatrix} \quad (5.171)$$

and, from Section 4.3.5.1,

$$\underset{\sim}{G}_{11}^{d_k} = \begin{pmatrix} \phi^{d_k} & 1 - \phi^{d_k} \\ 0 & 1 \end{pmatrix} \quad (5.172)$$

Let

$$\underset{\sim}{G}_{21}(d_k) = \begin{pmatrix} 0 & g_{212}(d_k) \\ 0 & 0 \end{pmatrix} \quad (5.173)$$

Then

$$\begin{pmatrix} 0 & g_{212}(dk) \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & c \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \phi^{d_k-1} & 1 - \phi^{d_k-1} \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & g_{212}(d_k - 1) \\ 0 & 0 \end{pmatrix}$$

i.e.

$$g_{212}(d_k) = c + g_{212}(d_k - 1) \Rightarrow g_{212}(d_k) = cd_k. \tag{5.174}$$

Therefore

$$\underset{\sim}{G}_k = \begin{pmatrix} \phi^{d_k} & 1 - \phi^{d_k} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & cd_k & 1 & d_k \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.175}$$

Using (4.8),

$$\underset{\sim}{R}_\omega(k) = \sum_{t=T_{k-1}+1}^{T_k} \begin{pmatrix} \phi^{T_k-t} & 1 - \phi^{T_k-t} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & c(T_k - t) & 1 & T_k - t \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\times \begin{pmatrix} R_{\mu 1} + R_\nu & R_\nu & R_{\mu\mu} + cR_\nu & 0 \\ R_\nu & R_\nu & cR_\nu & 0 \\ R_{\mu\mu} + cR_\nu & cR_\nu & R_{\mu 2} + R_\beta + c^2 R_\nu & R_\beta \\ 0 & 0 & R_\beta & R_\beta \end{pmatrix} \begin{pmatrix} \phi^{T_k-t} & 0 & 0 & 0 \\ 1 - \phi^{T_k-t} & 1 & c(T_k - t) & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & T_k - t & 1 \end{pmatrix}$$

$$= \begin{pmatrix} R_{11k} & R_{12k} & \cdots & R_{14k} \\ \vdots & & & \vdots \\ R_{14k} & \cdots & \cdots & R_{44k} \end{pmatrix}, \text{ say} \tag{5.176}$$

where

$$R_{11k} = \sum_{t=T_{k-1}+1}^{T_k} \{\phi^{2(T_k-t)} R_{\mu 1} + R_\nu\} = \frac{1 - \phi^{2d_k}}{1 - \phi^2} R_{\mu 1} + d_k R_\nu \tag{5.177}$$

$$R_{12k} = \sum_{t=T_{k-1}+1}^{T_k} R_\nu = d_k R_\nu \tag{5.178}$$

$$R_{13k} = \sum_{t=T_{k-1}+1}^{T_k} \{\phi^{T_k - t} R_{\mu\mu} + c(T_k - t + 1)R_\nu\}$$

$$= \frac{1 - \phi^{d_k}}{1 - \phi} R_{\mu\mu} + \tfrac{1}{2} c d_k (d_k + 1) R_\nu \tag{5.179}$$

$$R_{14k} = 0 \tag{5.180}$$

$$R_{22k} = \sum_{t=T_{k-1}+1}^{T_k} R_\nu = d_k R_\nu \tag{5.181}$$

$$R_{23k} = \sum_{t=T_{k-1}+1}^{T_k} c(T_k - t + 1)R_\nu = \tfrac{1}{2} c d_k (d_k + 1) R_\nu \tag{5.182}$$

$$R_{24k} = 0 \tag{5.183}$$

$$R_{33k} = \sum_{t=T_{k-1}+1}^{T_k} \{c^2(T_k - t + 1)^2 R_\nu + R_{\mu 2} + (T_k - t + 1)^2 R_\beta\}$$

$$= \frac{1}{6} c^2 d_k (d_k + 1)(2d_k + 1) R_\nu + d_k R_{\mu 2}$$
$$+ \frac{1}{6} d_k (d_k + 1)(2d_k + 1) R_\beta \tag{5.184}$$

$$R_{34k} = \sum_{t=T_{k-1}+1}^{T_k} (T_k - t + 1) R_\beta = \tfrac{1}{2} d_k (d_k + 1) R_\beta \tag{5.185}$$

and

$$R_{44k} = \sum_{t=T_{k-1}+1}^{T_k} R_\beta = d_k R_\beta \tag{5.186}$$

5.4 PERFORMANCE AND SENSITIVITY

In this section we examine the performance of the two models described in Section 5.3.2, i.e. the bivariate linear growth model and the bivariate AR(1)/linear growth model. In order to do so, we retain the performance measures outlined in Section 3.4, i.e.

(i) For assessment of event detection, we shall assume that $0_t^{(i)} > 0.2$, $i = 2, \ldots, J$ is a positive signal (where $0_t^{(i)}$ is the one-step-back probability of a discontinuity). If $0_t^{(i)} > 0.2$ when no change has actually been induced in the series, this will count as a false positive; we shall also use the number of false positives (NFP) as a performance measure.

(ii) In terms of estimation, we shall compare the parameter estimates with the true parameter values.

(iii) We shall use the two measures, SSFE and MAD (see Section 3.4) for assessment of forecasting ability by applying these measures to each of the univariate series in turn, so that we shall adopt the notation $SSFE_1$, etc.

As far as sensitivity is concerned, we shall only attempt to vary those parameters in the model that are peculiar to a bivariate formulation, such as $cov(\varepsilon_{1t},\varepsilon_{2t})$, etc. In particular, we will be interested to see how sensitive the models are to the correct specification of the ratio $var(\varepsilon_{1t})/var(\varepsilon_{2t})$, which will be preset and fixed throughout the analysis.

Multistate Structure:

Recall from Section 5.2.1, that each of the bivariate models involves 16 possible states; these are given in Table 5.1 where, for example, 'Level/Slope' denotes a level change in the 'first' series <u>and</u> a slope change in the 'second' series at the same time-point. Note that in the Figures which illustrate these analyses, only a few 'interesting' one-step-back state-probabilities are presented (rather than all fifteen types of changepoint-probability) and that the <u>lower</u> of the two time series represents the 'first' series, with the uppermost plot representing the 'second'.

TABLE 5.1

STATE NUMBER i	STATE DEFINITION BIVARIATE LINEAR GROWTH	STATE DEFINITION AR(1)/LINEAR GROWTH
1	Steady/Steady	Steady/Steady
2	Steady/Level	Steady/Level
3	Steady/Slope	Steady/Slope
4	Steady/Transient	Steady/Transient
5	Level/Steady	Impulse/Steady
6	Level/Level	Impulse/Level
7	Level/Slope	Impulse/Slope
8	Level/Transient	Impulse/Transient
9	Slope/Steady	Level/Steady
10	Slope/Level	Level/Level
11	Slope/Slope	Level/Slope
12	Slope/Transient	Level/Transient
13	Transient/Steady	Transient/Steady
14	Transient/Level	Transient/Level
15	Transient/Slope	Transient/Slope
16	Transient/Transient	Transient/Transient

5.4.1 BIVARIATE LINEAR GROWTH MODEL

In order to assess the performance of the bivariate linear growth model, we generated another series according to the univariate linear growth model (see Section 2.3.1). Details of this series, which will become the first series in the bivariate model, can be found in Appendix A5.3. For the second series, we use the data set (given in Appendix A3.1) which was also generated from a univariate linear growth model, and which was examined in Sections 3.4 and 4.4. Using this bivariate time series, we created an unequally-spaced bivariate time series by removing observations at certain timepoints. For the second series, observations were deleted at the timepoints given in Section 4.4 (corresponding to Series 2, in which 25% of the series has been removed). For the first series, observations were deleted at the following times:

t = 22, 26, 28, 33, 43, 45, 47, 48, 53, 54, 55, 56, 58, 59, 60, 61, 63, 64, 65, 68, 69, 70, 89, 90, 91,

i.e. 25% of this series has also been removed.

Notice that, for the bivariate time series, there will be some timepoints when both y_1 and y_2 are unavailable (e.g. $t = 22$), some when only y_1 is present (e.g. $t = 24$) and some when only y_2 is present (e.g. $t = 33$). Reference to Appendices A3.1 and A5.3, and to Table 5.1, shows that the following changepoints have been induced in the bivariate series.

TABLE 5.2

TIME t	CHANGEPOINT-TYPE	SIGNAL OF INTEREST
24	Level/Steady	$0_{25}^{(5)}$
25	Steady/Slope	$0_{26}^{(3)}$
35	Steady/Transient	$0_{36}^{(4)}$
50	Transient/Level	$0_{51}^{(14)}$
75	Slope/Steady	$0_{76}^{(9)}$
80	Transient/Transient	$0_{81}^{(16)}$

5.4.1.1 *Initial Setting*

The following prior values were employed (see Section 3.4.1.1 for those parameters associated with the second series):

$$\underset{\sim}{m}_o = \begin{pmatrix} 500 \\ 4 \\ 100 \\ 5 \end{pmatrix}; \quad \underset{\sim}{C}_o = \begin{pmatrix} 25 & 0 & 0.02 & 0 \\ 0 & 0.2 & 0 & 0.01 \\ 0.02 & 0 & 10 & 0 \\ 0 & 0.01 & 0 & 0.5 \end{pmatrix}$$

$n_o = 3$, $r_o = 30$ (i.e. $E(\lambda^{-1}) = r_o/(n_o - 1) = 15$; see Theorem 2.1.2)

$p_o^{(j)} = 0.85$, $j = 1$

$\quad\;\; = 0.01$, $j = 2, \ldots, 16$

$R_{\varepsilon 1}^{(j)} = 2$, $R_{\mu 1}^{(j)} = 0$, $R_{\beta 1}^{(j)} = 0$, $j = 1, \ldots, 4$

$R_{\varepsilon 1}^{(j)} = 2$, $R_{\mu 1}^{(j)} = 40$, $R_{\beta 1}^{(j)} = 0$, $j = 5, \ldots, 8$

$R_{\varepsilon 1}^{(j)} = 2$, $R_{\mu 1}^{(j)} = 0$, $R_{\beta 1}^{(j)} = 20$, $j = 9, \ldots, 12$

$R_{\varepsilon 1}^{(j)} = 60$, $R_{\mu 1}^{(j)} = 0$, $R_{\beta 1}^{(j)} = 0$, $j = 13, \ldots, 16$

$R_{\varepsilon 2}^{(j)} = 1$, $R_{\mu 2}^{(j)} = 0$, $R_{\beta 2}^{(j)} = 0$, $j = 1, 5, 9, 13$

$R_{\varepsilon 2}^{(j)} = 1$, $R_{\mu 2}^{(j)} = 20$, $R_{\beta 2}^{(j)} = 0$, $j = 2, 6, 10, 14$

$R_{\varepsilon 2}^{(j)} = 1$, $R_{\mu 2}^{(j)} = 0$, $R_{\beta 2}^{(j)} = 10$ $j = 3, 7, 11, 15$

$R_{\varepsilon 2}^{(j)} = 30$, $R_{\mu 2}^{(j)} = 0$, $R_{\beta 2}^{(j)} = 0$, $j = 4, 8, 12, 16$

so that $R_{\varepsilon 1}^{(1)}/R_{\varepsilon 2}^{(1)} = \text{var}(\varepsilon_{1t})/\text{var}(\varepsilon_{2t}) = 2$ (since $\lambda^{-1} = 30$ for the first series and $\lambda^{-1} = 15$ for the second series - see Appendices A5.3 and A3.1). Also $R_{\varepsilon\varepsilon}^{(j)} = R_{\mu\mu}^{(j)} = R_{\beta\beta}^{(j)} = 0.01 \; \forall j$ (the two univariate series were, in fact, generated independently, but many of the discontinuities were induced at coincident timepoints).

Using these values, the analysis was carried out on both the full data set and the unequally-spaced data set (as described above), and the results are presented in Table 5.3.

5.4.1.2 *Sensitivity Analysis*

The following changes to the initial setting were examined, with the remaining parameters unaltered from their original values in each case:

Prior 1: $R_{\varepsilon\varepsilon}^{(j)} = R_{\mu\mu}^{(j)} = R_{\beta\beta}^{(j)} = 0.1, \; \forall j$

Prior 2: $R_{\varepsilon\varepsilon}^{(j)} = R_{\mu\mu}^{(j)} = R_{\beta\beta}^{(j)} = 0, \; \forall j$

Prior 3: $\text{cov}(\mu_{1t},\mu_{2t}) = 0.2$ (previously 0.02);
$\text{cov}(\beta_{1t},\beta_{2t}) = 0.1$ (previously 0.01), at $t = 0$

Prior 4: $\text{cov}(\mu_{1t},\mu_{2t}) = \text{cov}(\beta_{1t},\beta_{2t}) = 0$, at $t = 0$

Prior 5: $R_{\varepsilon 1}^{(1)} = R_{\varepsilon 2}^{(1)}$, etc. (i.e. pre-set $\text{var}(\varepsilon_{1t}) = \text{var}(\varepsilon_{2t})$)

Prior 6: $R_{\varepsilon 1}^{(1)} = \frac{1}{2}R_{\varepsilon 2}^{(1)}$, etc. (i.e. pre-set $\text{var}(\varepsilon_{2t})/\text{var}(\varepsilon_{1t}) = 2$; the reverse of the true situation)

Prior 7: All parameters unchanged, except that we adopt a 16 x 16 Markovian state-transition matrix, $\{p_o\}$, in place of $p_o^{(j)}$, $j = 1, \ldots, 16$ (see Section 5.3.1.3) for which

$^{(i)}p_o^{(j)} = p_o^{(j)}, \; i = 1, \ldots, 4, 6, \ldots, 16, \; \forall j$

$^{(5)}p_o^{(1)} = 0.76; \quad {}^{(5)}p_o^{(3)} = 0.1; \quad {}^{(5)}p_o^{(j)} = 0.01,$

$j = 2, 4, 5, \ldots, 16$

TABLE 5.3

	$0_{25}^{(5)}$	$0_{26}^{(3)}$	$0_{36}^{(4)}$	$0_{51}^{(14)}$	$0_{76}^{(9)}$	$0_{81}^{(16)}$	NFP	$\hat{\mu}_1$	$\hat{\beta}_1$	$\hat{\mu}_2$	$\hat{\beta}_2$	$SSFE_1$	MAD_1	$SSFE_2$	MAD_2
Full Data Set	0.18	0.06*	0.41	0.99	0.21	0.99	6	3538.1	104.8	-115.4	-7.1	864688	59.8	13730	7.9
Unequally-Spaced Data Set	0.22	0.01†	0.37	1.00	0.11	0.71	4	3538.0	104.7	-115.4	-7.2	-	72.5	-	9.4

*$0_{26}^{(7)}$ = 0.12 (Level/Slope)

†$0_{27}^{(3)}$ used, since no observation is present at t = 26.

NOTES: (i) $\mu_1 = 3527.8$, $\beta_1 = 104.8$, $\mu_2 = -117.5$ and $\beta_2 = -5.0$ are the true values for $\underset{\sim}{\theta}$ at t = 100.

(ii) $\hat{\mu}_1$ is the final estimate of μ_1 at time t = 100, etc.

(iii) $SSFE_1$ and $SSFE_2$ have not been calculated for the unequally-spaced series, since this measure of forecasting ability is dependent upon the number of observations (see Section 4.4).

(iv) Compare forecasting and estimation values, for the second series, with those obtained in Sections 3.4.1.1 and 4.4.1.

These results are presented graphically in Figures 5.2 and 5.3.

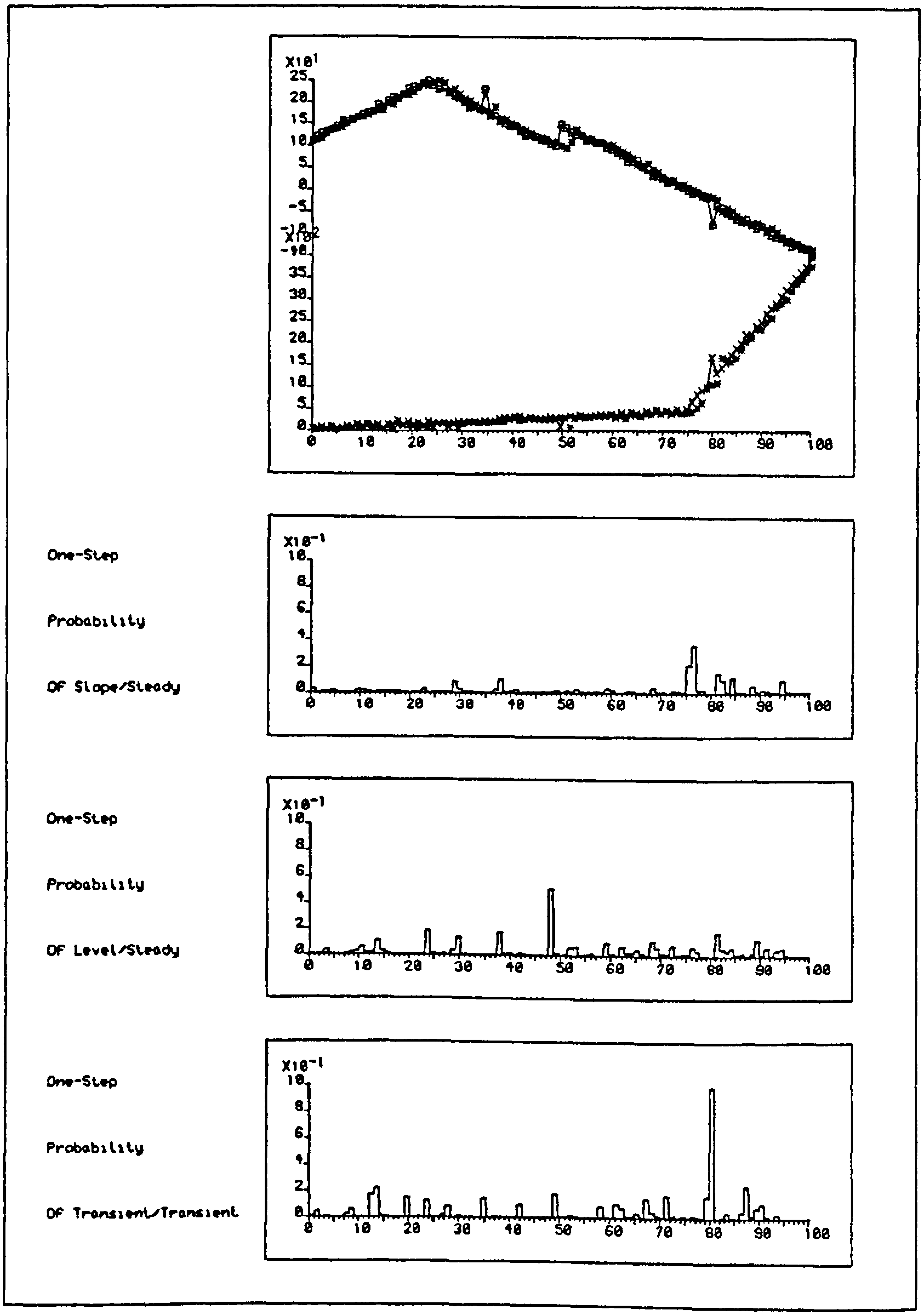
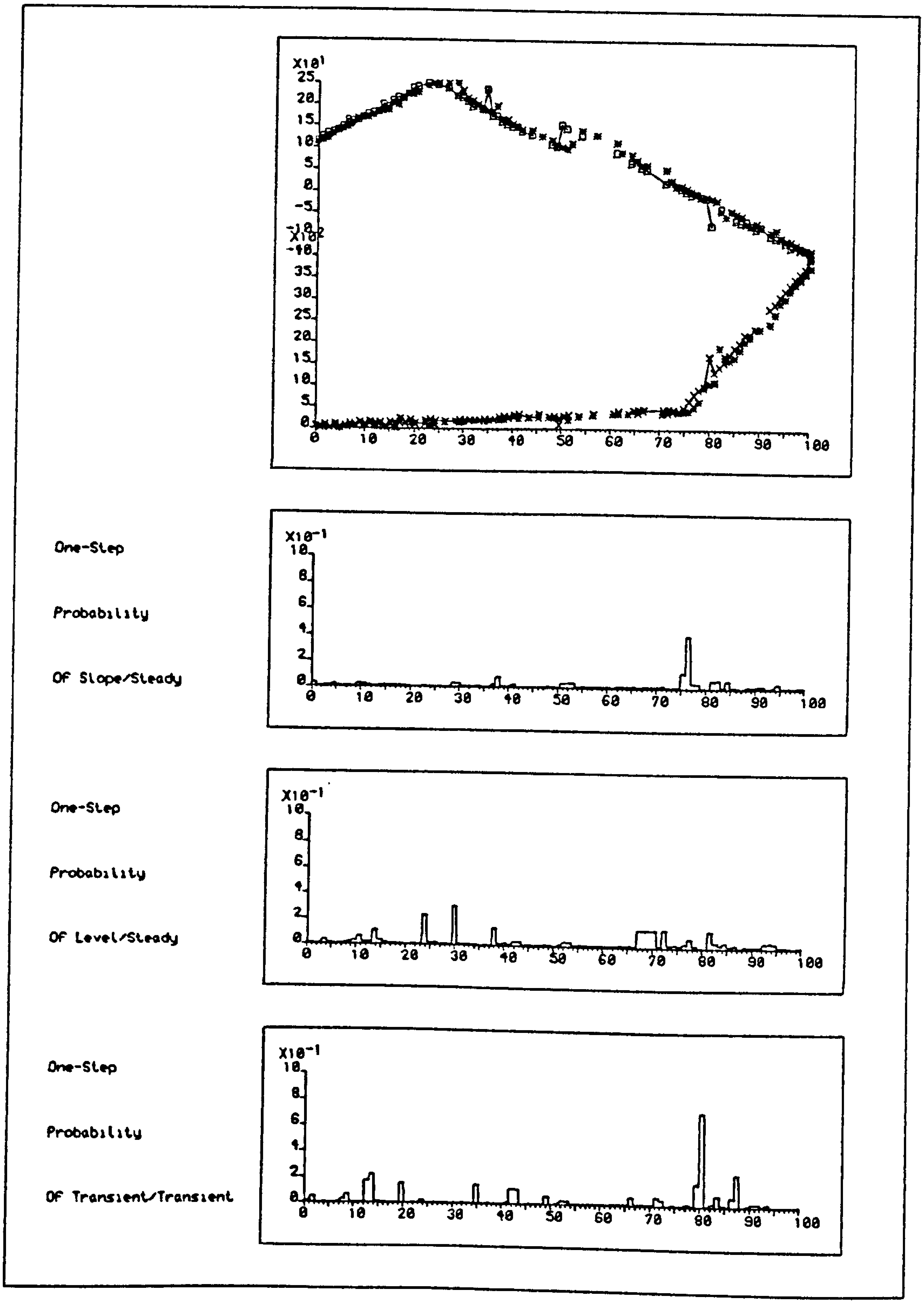

One-Step
Probability
Of Slope/Steady
One-Step
Probability
Of Level/Steady
One-Step
Probability
Of Transient/Transient

FIGURE 5.2



One-Step
Probability
Of Slope/Steady
One-Step
Probability
Of Level/Steady
One-Step
Probability
Of Transient/Transient

FIGURE 5.3

where $p_o^{(j)}$ has the original setting given in the previous section.

In other words, we have merely increased p(Steady/Slope, t|Level/Steady, t - 1) so that we might be able to identify the changes, occurring at t = 24 and t = 25, more easily.

Using each of these priors, in turn, the results obtained are shown in Table 5.4.

5.4.2 AR(1)/LINEAR GROWTH MODEL

In order to assess the performance of the AR(1)/linear growth model, we joined together the series described by Appendices A3.3 and A3.1 (previously examined separately in Sections 3.4 and 4.4) to form a bivariate time series, of which the first series is autoregressive, while the second is linear growth. We created an unequally-spaced bivariate time series by removing observations at certain timepoints. For the first series, observations were deleted at the times listed in the previous section; for the second series, observations were deleted at the times listed in Section 4.4 (corresponding to Series 2). Once again, therefore, there will be some timepoints where both observations are unavailable and some when only one observation is available.

Reference to Appendices A3.1 and A3.3, and to Table 5.1, shows that the following changepoints have been induced in the bivariate series.

TABLE 5.4

	$0_{25}^{(5)}$	$0_{26}^{(3)}$	$0_{36}^{(4)}$	$0_{51}^{(14)}$	$0_{76}^{(9)}$	$0_{81}^{(16)}$	NFP	$\hat{\mu}_1$	$\hat{\beta}_1$	$\hat{\mu}_2$	$\hat{\beta}_2$	$SSFE_1$	MAD_1	$SSFE_2$	MAD_2
Initial Setting	0.18	0.06	0.41	0.99	0.21	0.99	6	3538.1	104.8	-115.4	-7.1	864688	59.8	13730	7.9
Prior 1	0.21	0.06	0.42	0.99	0.22	0.99	8	3538	104.9	-115.5	-7.1	856682	59.5	13735	7.9
Prior 2	0.18	0.06[1]	0.41	0.99	0.21	0.99	6	3538.1	104.8	-115.4	-7.1	864700	59.8	13730	7.9
Prior 3	0.18	0.06[1]	0.41	0.99	0.21	0.99	6	3538.1	104.8	-115.4	-7.1	864604	59.8	13729	7.9
Prior 4	0.18	0.06[1]	0.41	0.99	0.21[4]	0.99	6	3538.1	104.8	-115.4	-7.1	864692	59.8	13732	7.9
Prior 5	0.00[2]	0.04	0.12[3]	0.78	0.05[4]	0.69	15	3525.3	104.0	-115.7	-7.3	849749	59.6	14324	8.2
Prior 6	0.02	0.01	0.09[3]	0.33	0.00[4]	0.48	16	3530.7	106.5	-115.5	-7.2	1014379	65.5	16245	8.4
Prior 7	0.24	0.26	0.38	0.99	0.21	1.00	6	3543.8	105.3	-117.2	-7.0	870530	59.7	13875	8.0

1. $0_{26}^{(7)} = 0.12$ (Level/Slope)
2. $0_{25}^{(13)} = 0.31$ (Transient/Steady)
3. $0_{36}^{(8)} > 0.5$ (Level/Transient)
4. $0_{76}^{(13)} > 0.3$ (Transient/Steady)

Notice that changepoint-types may be confused when the error variance-ratio is incorrectly specified. The dual changes occurring at $t = 24$ and $t = 25$ also lead to some confusion when a non-Markovian state-transition structure is used. The identification of these changepoints is clarified to some extent by the adoption of a Markovian state-transition matrix. See Figures 5.4 to 5.10 for a graphical presentation of (some) results for Prior 1 to Prior 7, respectively.

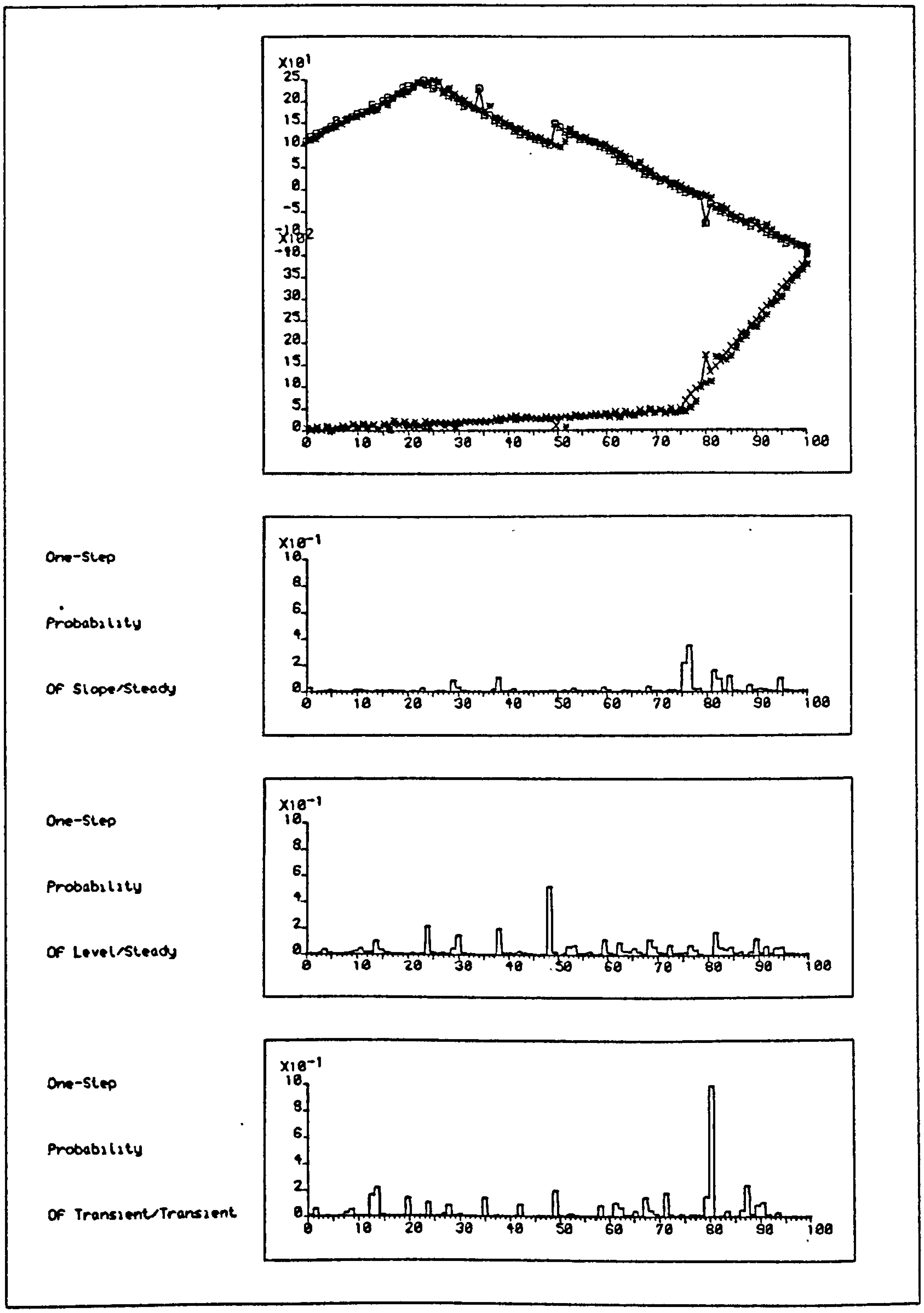

X10^1
X10^2
X10^-1
One-Step
Probability
OF Slope/Steady
X10^-1
One-Step
Probability
OF Level/Steady
X10^-1
One-Step
Probability
OF Transient/Transient

FIGURE 5.4

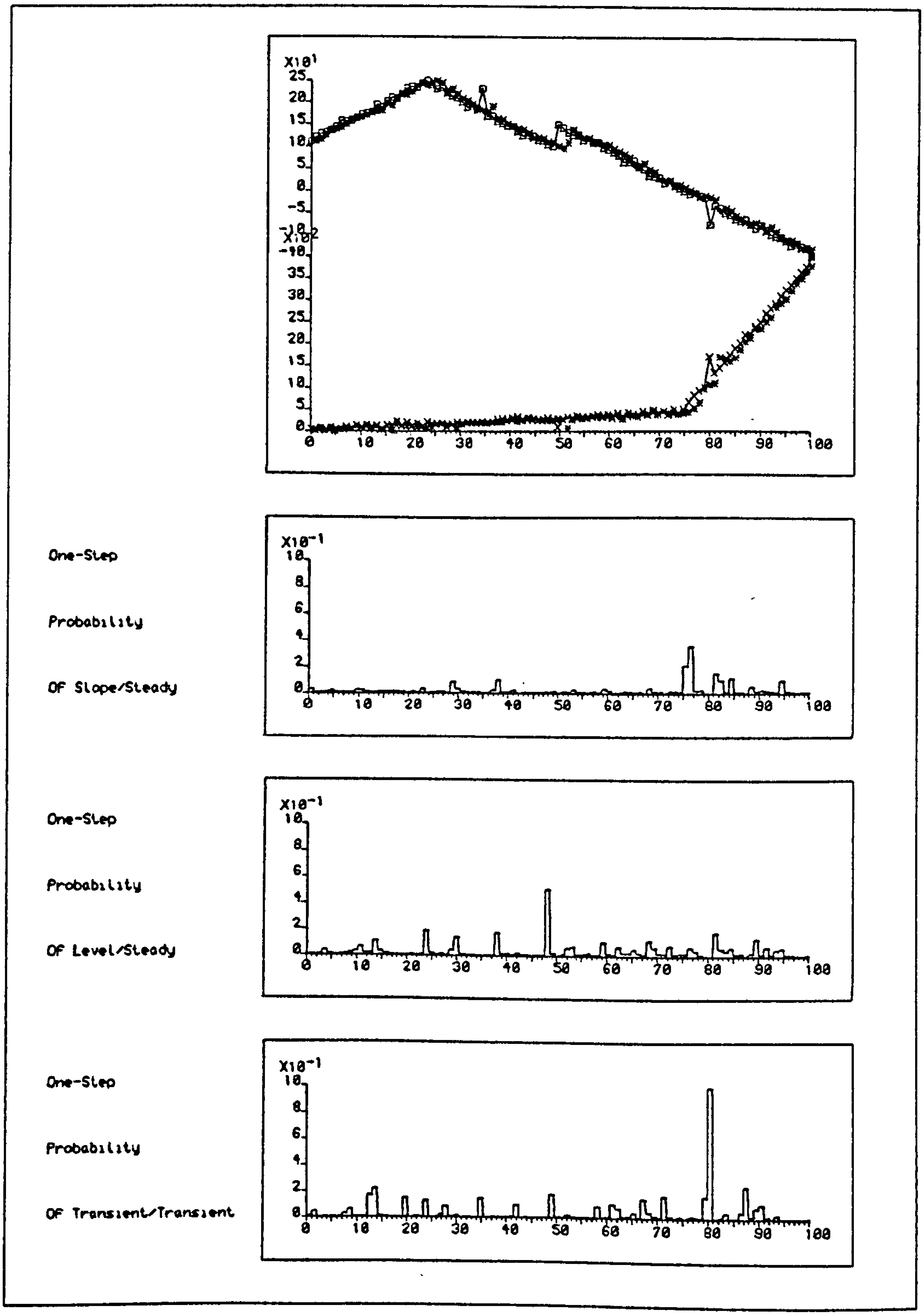
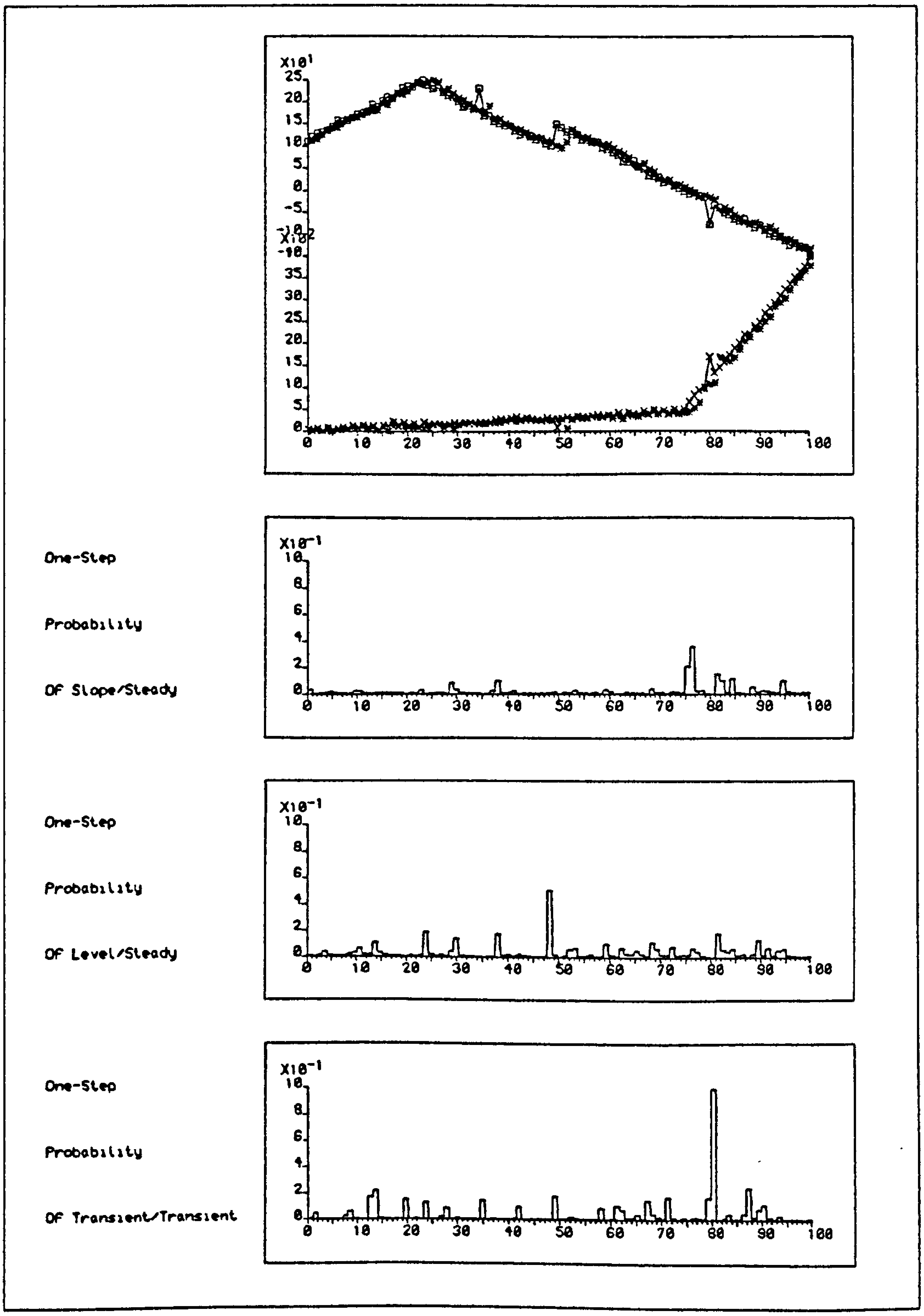

One-Step
Probability
Of Slope/Steady
One-Step
Probability
Of Level/Steady
One-Step
Probability
Of Transient/Transient

FIGURE 5.5



One-Step
Probability
Of Slope/Steady
One-Step
Probability
Of Level/Steady
One-Step
Probability
Of Transient/Transient

FIGURE 5.6

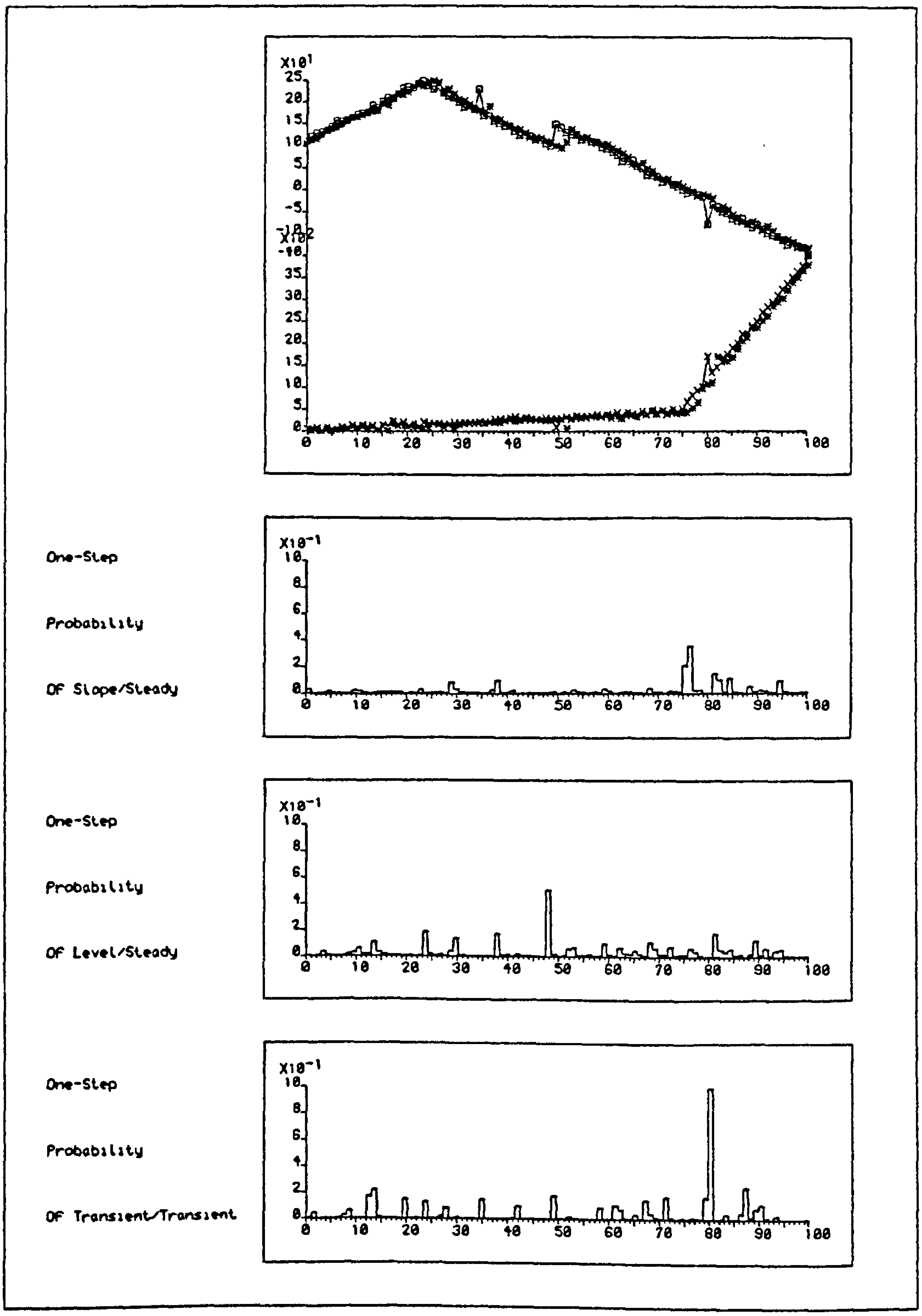
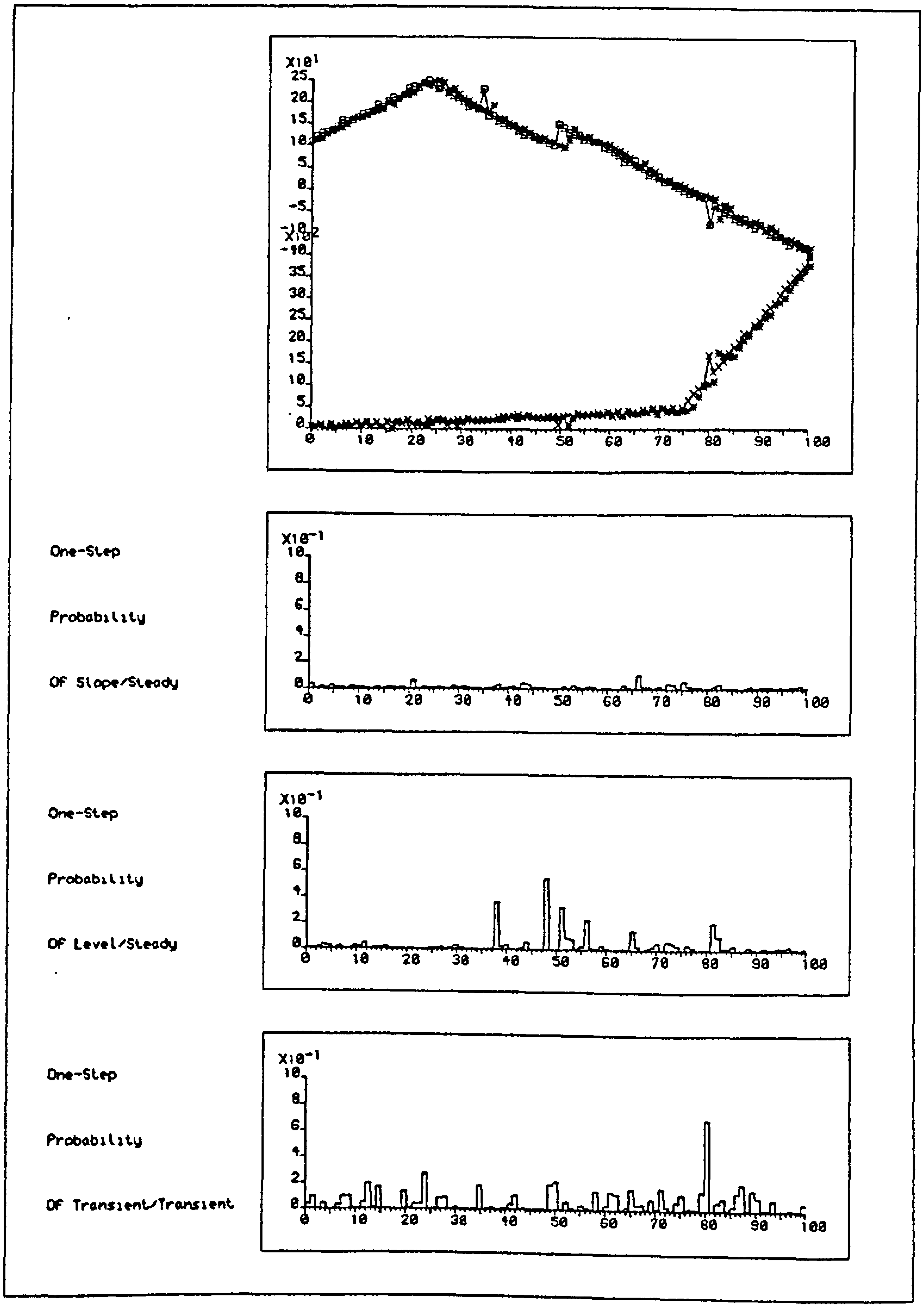

One-Step
Probability
Of Slope/Steady
One-Step
Probability
Of Level/Steady
One-Step
Probability
Of Transient/Transient

FIGURE 5.7



X10¹
X10²
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Slope/Steady
X10⁻¹
One-Step
Probability
Of Level/Steady
X10⁻¹
One-Step
Probability
Of Transient/Transient
X10⁻¹

FIGURE 5.8

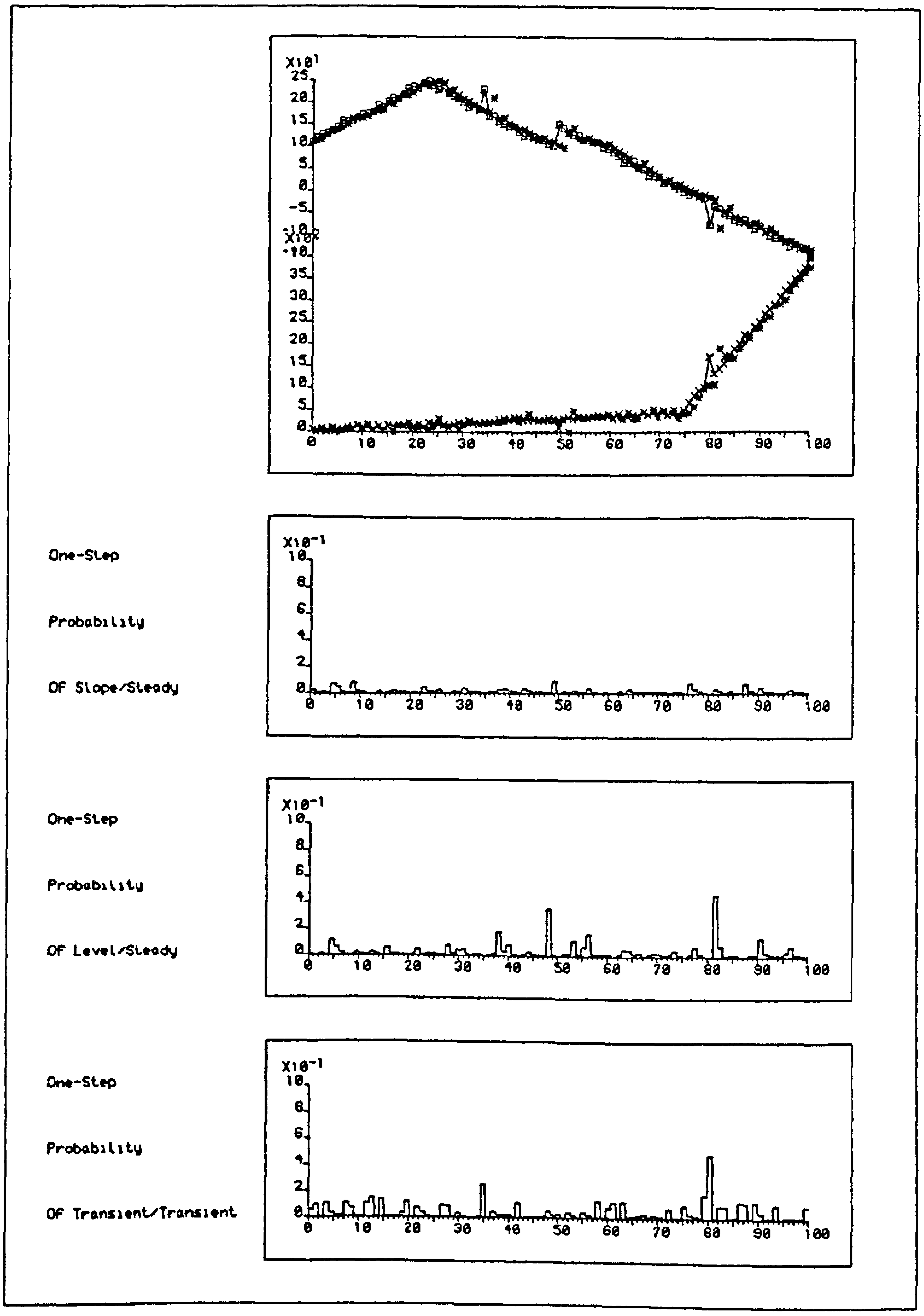

X10[1]
25
20
15
10
5
0
-5
X10[2]
35
30
25
20
15
10
5
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Slope/Steady
X10[-1]
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Level/Steady
X10[-1]
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Transient/Transient
X10[-1]
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100

FIGURE 5.9

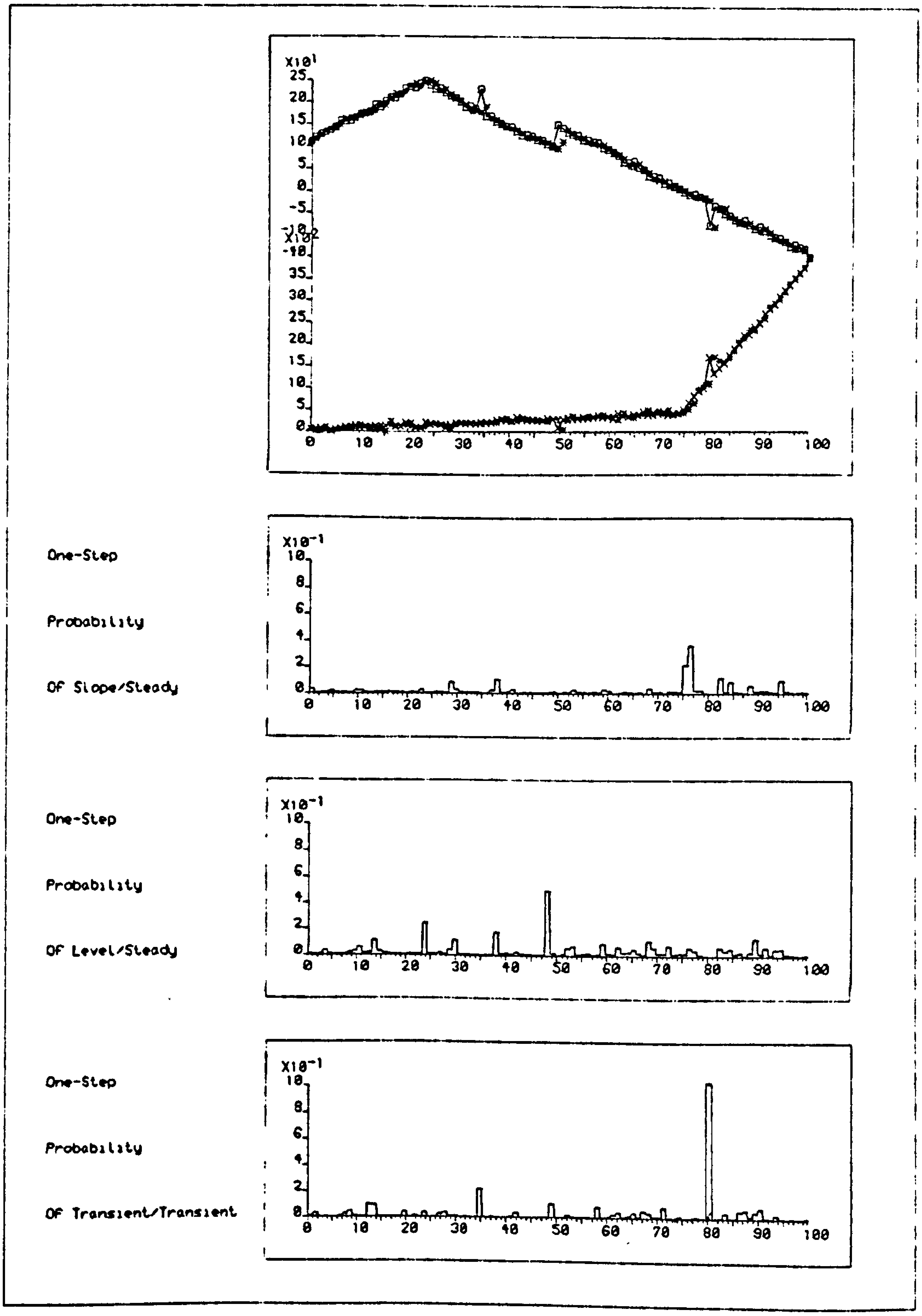

X10^1
25
20
15
10
5
0
-5
-10
X10^2
-10
35
30
25
20
15
10
5
0
0
10
20
30
40
50
60
70
80
90
100
One-Step
Probability
Of Slope/Steady
X10^-1
10
8
6
4
2
0
One-Step
Probability
Of Level/Steady
X10^-1
One-Step
Probability
Of Transient/Transient
X10^-1

FIGURE 5.10

TABLE 5.5

TIME, t	CHANGEPOINT-TYPE	SIGNAL OF INTEREST
25	Impulse/Slope	$0_{26}^{(7)}$
30	Level/Steady	$0_{31}^{(9)}$
35	Level/Transient	$0_{36}^{(12)}$
50	Transient/Level	$0_{51}^{(14)}$
75	Impulse/Steady	$0_{76}^{(5)}$
80	Transient/Transient	$0_{81}^{(16)}$

5.4.2.1 *Initial Setting*

The following prior values were employed (see Sections 3.4.1.1 and 3.4.3.1):

$$\underset{\sim}{m}_o = \begin{pmatrix} 10 \\ 10 \\ 100 \\ 5 \end{pmatrix}; \qquad \underset{\sim}{C}_o = \begin{pmatrix} 1 & 0 & 0.02 & 0 \\ 0 & 1 & 0 & 0.01 \\ 0.02 & 0 & 10 & 0 \\ 0 & 0.01 & 0 & 0.5 \end{pmatrix}$$

$n_o = 3,\ r_o = 30 \quad (\text{i.e. } E(\lambda^{-1}) = 15)$

$p_o^{(j)} = 0.85,\ j = 1$

$\quad = 0.01,\ j = 2, \ldots, 16$

$R_{\varepsilon 1}^{(j)} = 1/15;\ R_{\mu 1}^{(j)} = 0;\ R_{\nu}^{(j)} = 0,\ j = 1, \ldots, 4$

$R_{\varepsilon 1}^{(j)} = 1/15;\ R_{\mu 1}^{(j)} = 4/3;\ R_{\nu}^{(j)} = 0,\ j = 5, \ldots, 8$

$R_{\varepsilon 1}^{(j)} = 1/15;\ R_{\mu 1}^{(j)} = 0;\ R_{\nu}^{(j)} = 2/3,\ j = 9, \ldots, 12$

$R_{\varepsilon 1}^{(j)} = 2;\ R_{\mu 1}^{(j)} = 0;\ R_{\nu}^{(j)} = 0,\ j = 13, \ldots, 16$

with $R_{\varepsilon 2}^{(j)}$, $R_{\mu 2}^{(j)}$ and $R_{\beta}^{(j)}$ as given in the previous section, so that $R_{\varepsilon 2}^{(1)}/R_{\varepsilon 1}^{(1)} = \text{var}(\varepsilon_{2t})/\text{var}(\varepsilon_{1t}) = 15$ (since $\lambda^{-1} = 1$ for the first series, and $\lambda^{-1} = 15$ for the second series - see Appendices A3.3 and A3.1). Also

$$R_{\varepsilon\varepsilon}^{(j)} = R_{\mu\mu}^{(j)} = R_{\nu\beta}^{(j)} = 0.01, \forall j$$

and NN = number of nodes in the grid for ϕ, the autoregressive parameter = 11.

Using these starting values, the analysis was carried out on both the full and the unequally-spaced data sets, and the results are given in Table 5.6.

5.4.2.2 *Sensitivity Analysis*

The following changes to the initial setting were examined, with the remaining parameters unaltered from their original values in each case:

Prior 1: $R_{\varepsilon\varepsilon}^{(j)} = R_{\mu\mu}^{(j)} = R_{\nu\beta}^{(j)} = 0.1, \forall j$

Prior 2: $R_{\varepsilon\varepsilon}^{(j)} = R_{\mu\mu}^{(j)} = R_{\nu\beta}^{(j)} = 0, \forall j$

Prior 3: $\mathrm{cov}(\mu_{1t},\mu_{2t}) = 0.2$; $\mathrm{cov}(\nu_t,\beta_t) = 0.1$, at $t = 0$

Prior 4: $\mathrm{cov}(\mu_{1t},\mu_{2t}) = \mathrm{cov}(\nu_t,\beta_t) = 0$, at $t = 0$

Prior 5: $R_{\varepsilon 1}^{(1)} = R_{\varepsilon 2}^{(1)}$, etc. (i.e. pre-set $\mathrm{var}(\varepsilon_{1t}) = \mathrm{var}(\varepsilon_{2t})$)

Prior 6: $R_{\varepsilon 1}^{(1)} = 15R_{\varepsilon 2}^{(1)}$, etc. (i.e. pre-set $\mathrm{var}(\varepsilon_{1t})/\mathrm{var}(\varepsilon_{2t}) = 15$; the reverse of the true situation).

Using each of these priors, in turn, the results obtained are presented in Table 5.7.

TABLE 5.6

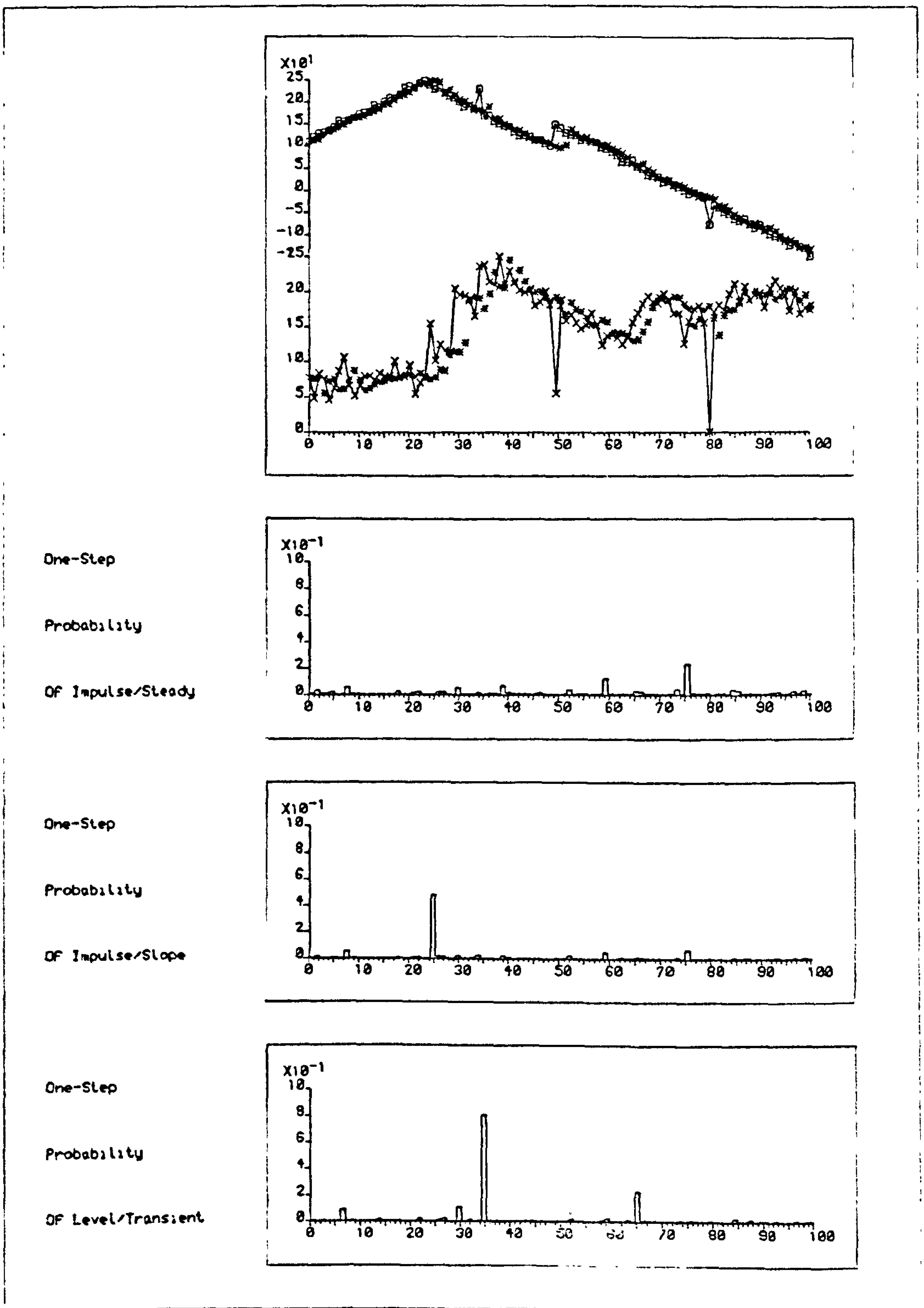
	$0_{26}^{(7)}$	$0_{31}^{(9)}$	$0_{36}^{(12)}$	$0_{51}^{(14)}$	$0_{76}^{(5)}$	$0_{81}^{(16)}$	NFP	$\hat{\nu}$	$\hat{\phi}$	$\hat{\mu}_2$	$\hat{\beta}$	$SSFE_1$	MAD_1	$SSFE_2$	MAD_2
Full Data Set	0.47	0.44	0.80	0.97	0.23	1.00	1	18.0	0.56	-115.1	-6.6	766	1.8	13495	7.7
Unequally-Spaced Data Set	0.30*	0.04	0.71	0.98	0.22	0.80	0	18.1	0.54	-114.8	-6.6	-	2.0	-	9.5

*$0_{27}^{(7)}$ used, since no observation is present at t = 26.

NOTES: (i) ν = 18.9, μ_2 = -117.5 and β = -5.0 are the true values for $\underset{\sim}{\theta}$ at t = 100; also ϕ = 0.7.

(ii) Compare forecasting and estimation values with those obtained in Sections 3.4.1.3, 4.4.3 (remembering that NN = 11 in this model), 3.4.1.1 and 4.4.1.

The full and the unequally-spaced series, along with (some) multistate probabilities and corresponding one-step-ahead forecasts, can be seen in Figures 5.11 and 5.12, respectively, while Figures 5.13 and 5.14 show on-line estimation of the ϕ-grid, for each of these series.



X10^1
One-Step
Probability
Of Impulse/Steady
One-Step
Probability
Of Impulse/Slope
One-Step
Probability
Of Level/Transient

FIGURE 5.11

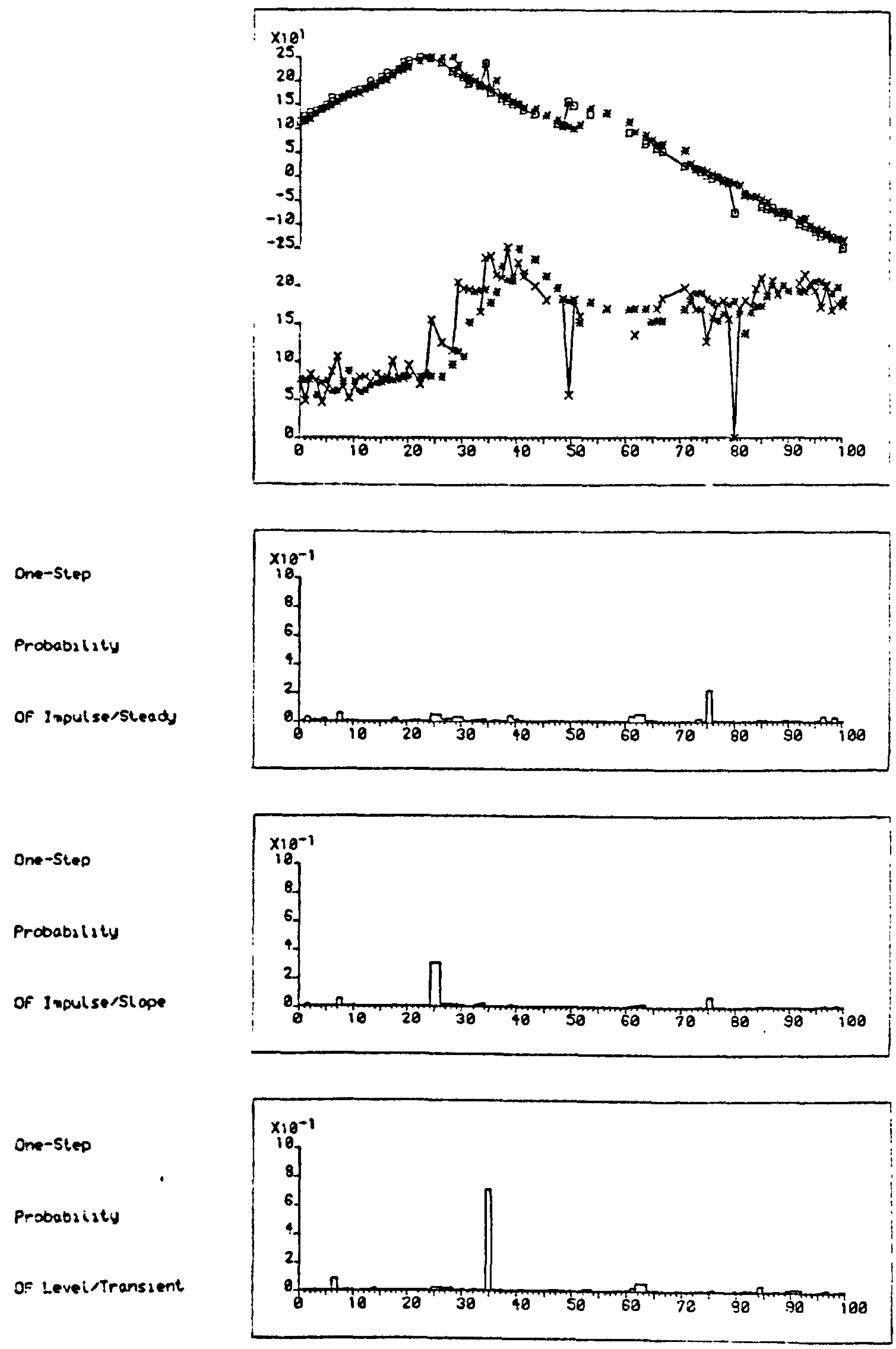

X10¹
25
20
15
10
5
0
-5
-10
-25
20
15
10
5
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Impulse/Steady
X10⁻¹
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Impulse/Slope
X10⁻¹
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Level/Transient
X10⁻¹
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100

FIGURE 5.12

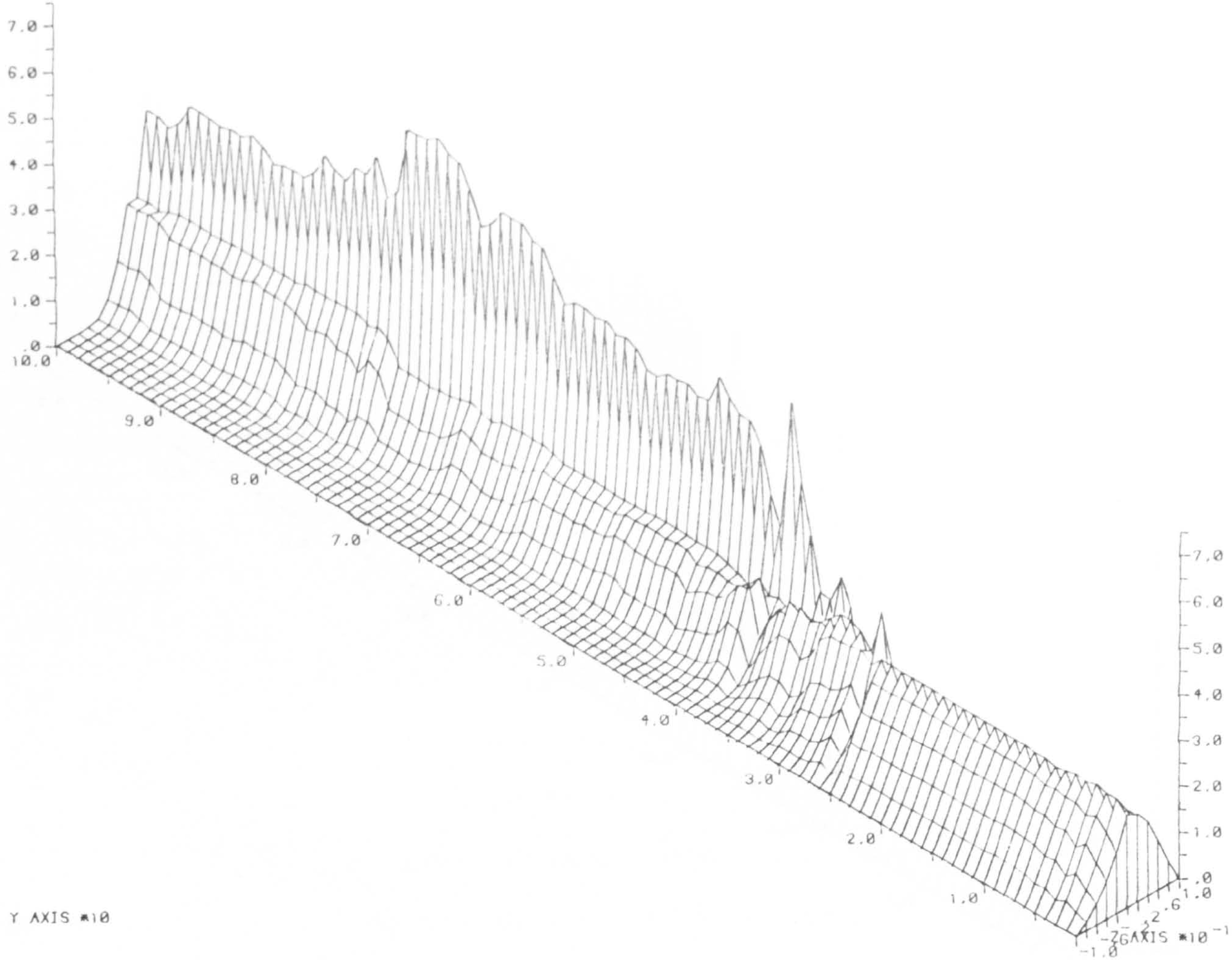

Y AXIS *10
Z AXIS *10⁻¹
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
-1.0
-.2
.2
.6
1.0

FIGURE 5.13

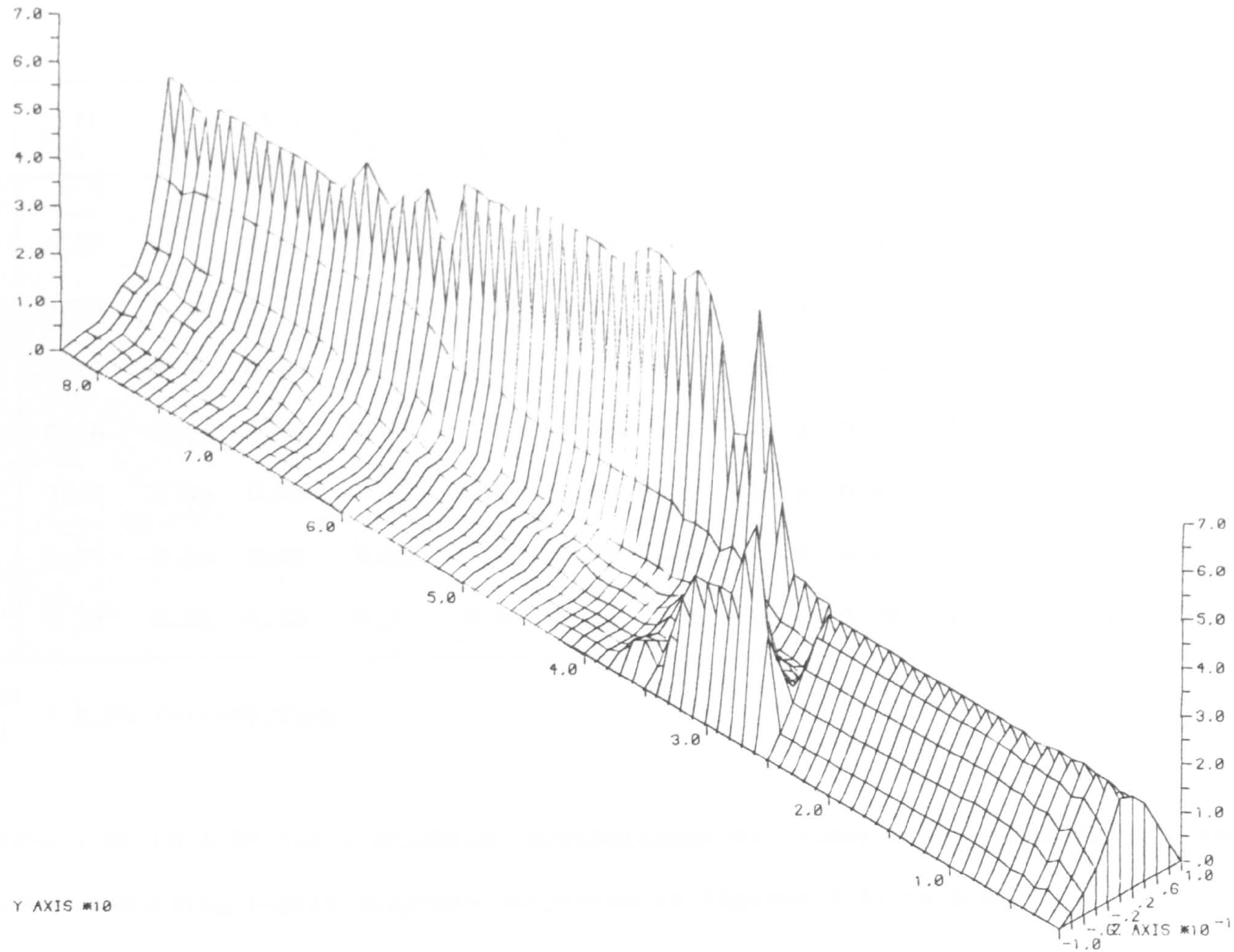

7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
Z AXIS *10⁻¹
-1.0
-.6
-.2
.2
.6
1.0

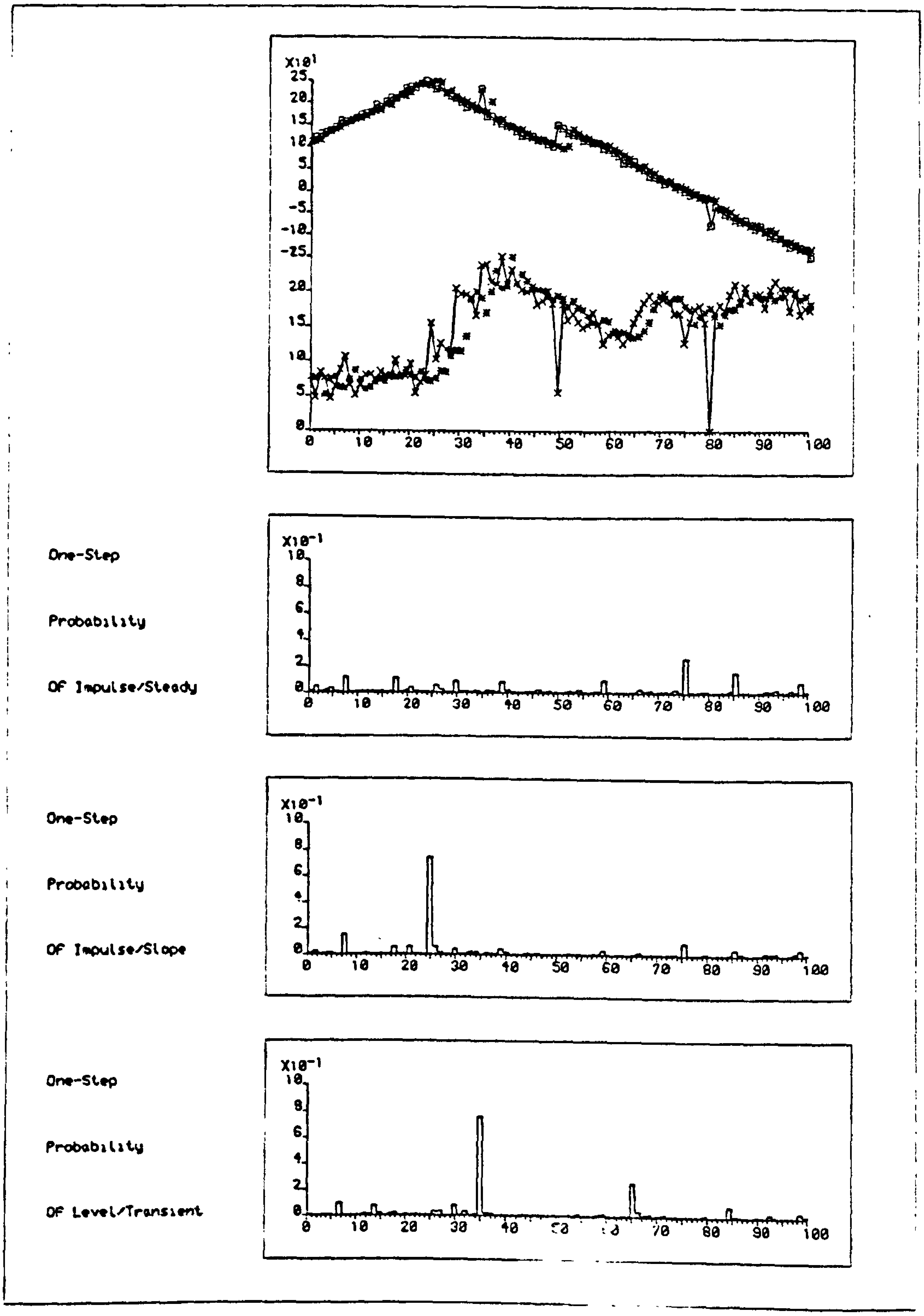
FIGURE 5.14

TABLE 5.7

	$0_{26}^{(7)}$	$0_{31}^{(9)}$	$0_{36}^{(12)}$	$0_{51}^{(14)}$	$0_{76}^{(5)}$	$0_{81}^{(16)}$	NFP	$\hat{\nu}$	$\hat{\phi}$	$\hat{\mu}_1$	$\hat{\beta}_1$	$SSFE_1$	MAD_1	$SSFE_2$	MAD_2
Initial Setting	0.47	0.44	0.80	0.97	0.23	1.00	1	18.0	0.56	-115.1	-6.6	766	1.8	13495	7.7
Prior 1	0.73	0.42	0.75	0.96	0.26	0.97	5	18.3	0.14	-115.8	-7.3	758	1.8	14089	7.9
Prior 2	0.47	0.44	0.80	0.97	0.23	1.00	1	18.0	0.56	-115.1	-6.6	766	1.8	13496	7.7
Prior 3	0.48	0.44	0.80	0.97	0.23	1.00	1	18.0	0.57	-115.1	-6.6	765	1.8	13498	7.7
Prior 4	0.47	0.44	0.80	0.97	0.23	1.00	1	18.0	0.56	-115.1	-6.6	766	1.8	13495	7.7
Prior 5	0.34	0.24	0.51	0.63	0.01	0.79	2	18.8	0.29	-115.0	-6.5	777	1.9	13468	7.6
Prior 6	0.15[1]	0.01	0.23	0.21	0.01	0.29	5	18.8	0.06	-114.4	-6.1	917	2.1	13466	7.6

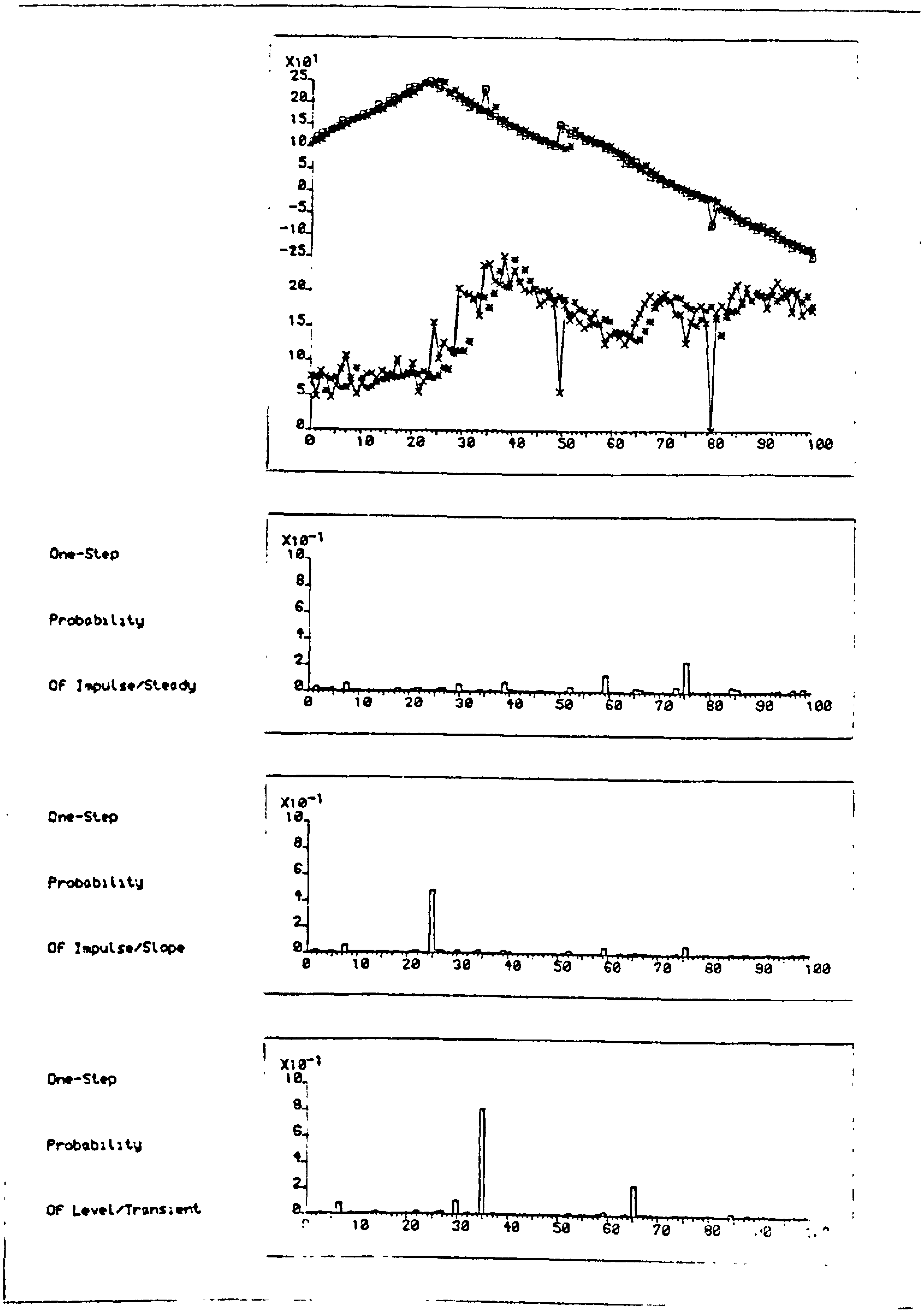
1. $0_{27}^{(3)} = 0.53$ (Steady/Slope)

See Figures 5.15 to 5.20 for a graphical presentation of (some) results for Prior 1 to Prior 6, respectively, with the corresponding ϕ-grid diagrams displayed in Figures 5.21 to 5.26.



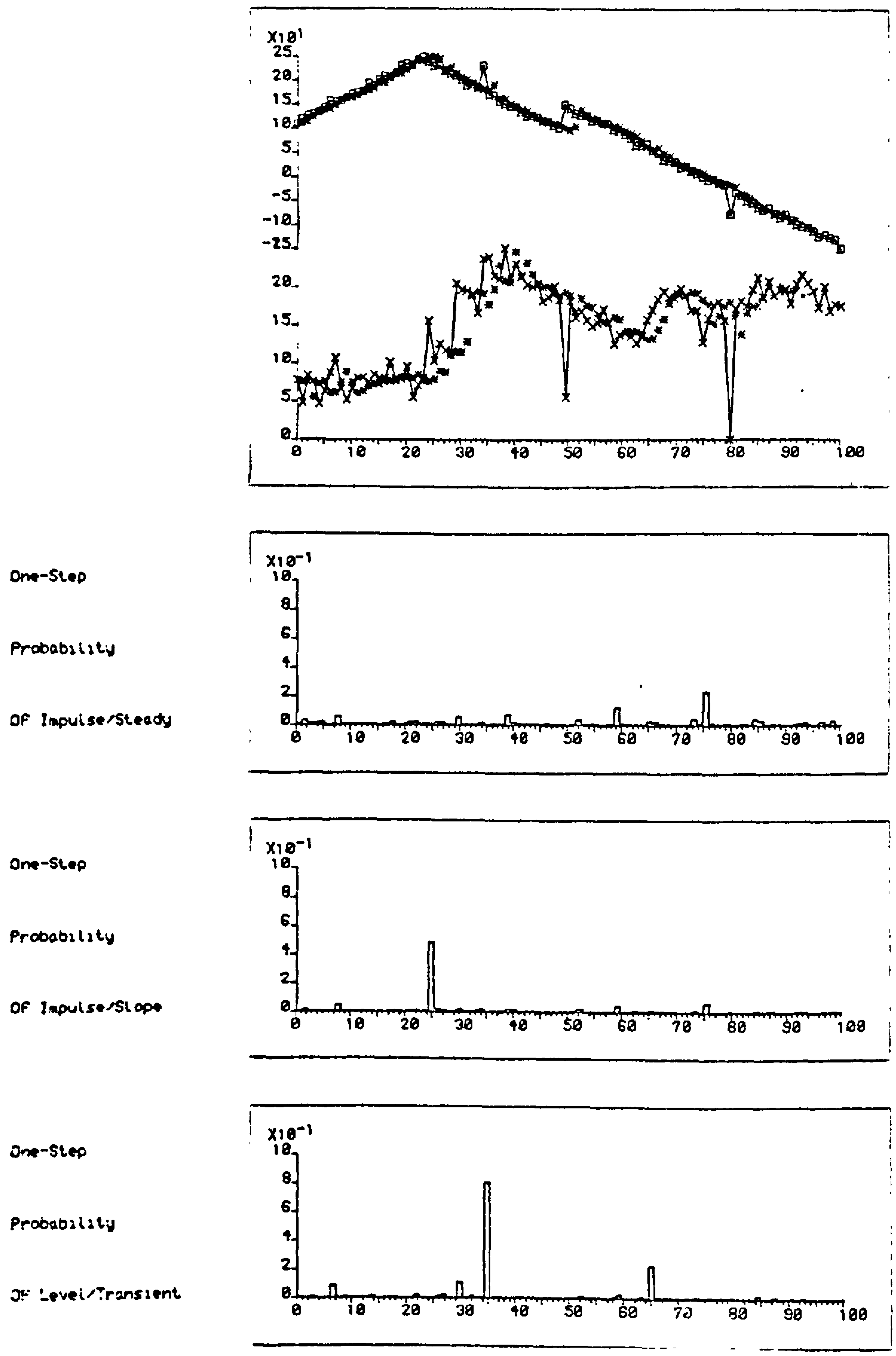
One-Step
Probability
Of Impulse/Steady
One-Step
Probability
Of Impulse/Slope
One-Step
Probability
Of Level/Transient

FIGURE 5.15



X10¹
X10⁻¹
One-Step
Probability
Of Impulse/Steady
One-Step
Probability
Of Impulse/Slope
One-Step
Probability
Of Level/Transient

FIGURE 5.16



X10¹
X10⁻¹
One-Step
Probability
Of Impulse/Steady
X10⁻¹
One-Step
Probability
Of Impulse/Slope
X10⁻¹
One-Step
Probability
Of Level/Transient

FIGURE 5.17

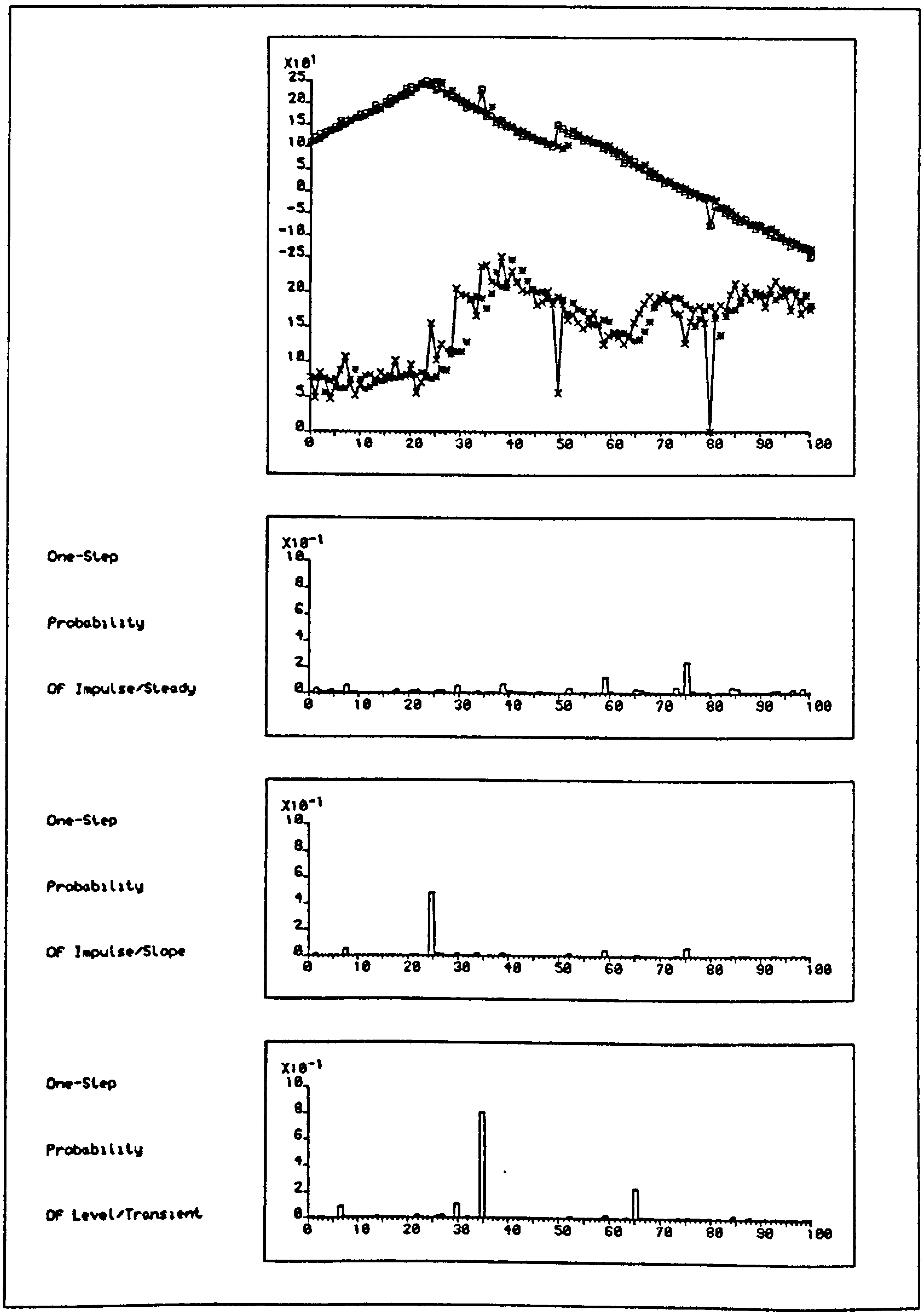

X10^1
25
20
15
10
5
0
-5
-10
-25
20
15
10
5
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Impulse/Steady
X10^-1
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Impulse/Slope
X10^-1
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100
One-Step
Probability
Of Level/Transient
X10^-1
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100

FIGURE 5.18

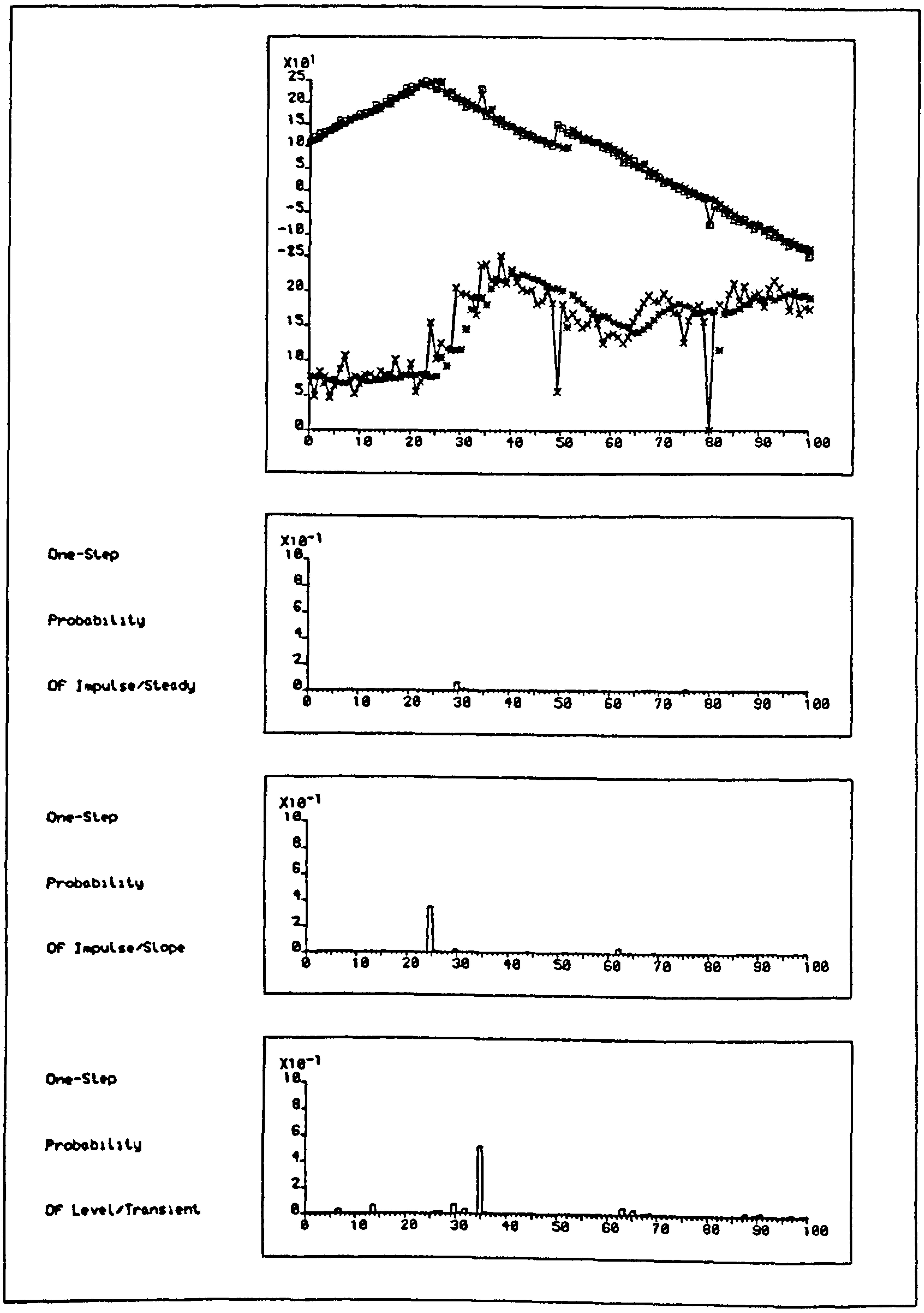

X10^1
X10^-1
One-Step
Probability
Of Impulse/Steady
One-Step
Probability
Of Impulse/Slope
One-Step
Probability
Of Level/Transient

FIGURE 5.19

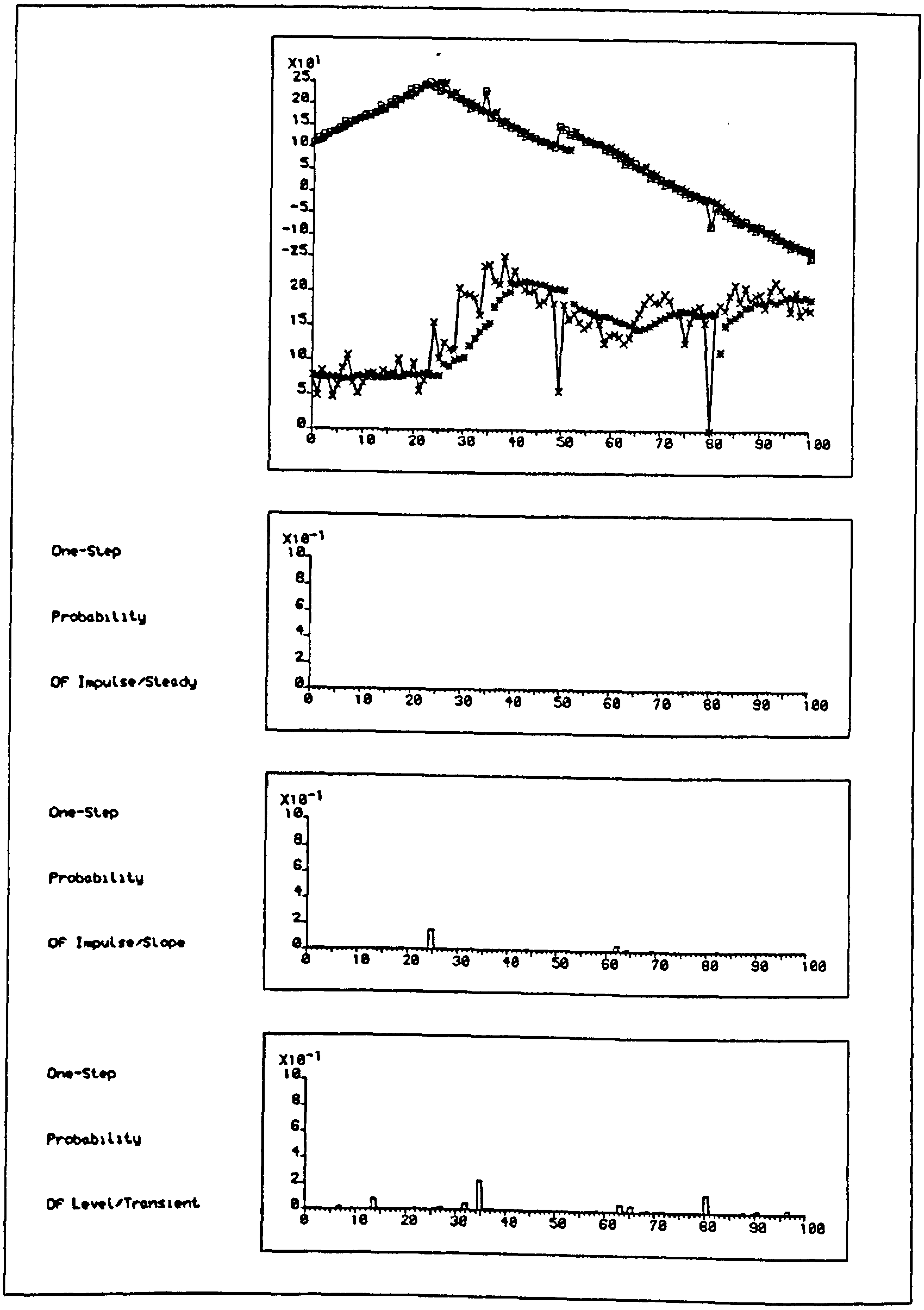

One-Step
Probability
Of Impulse/Steady
One-Step
Probability
Of Impulse/Slope
One-Step
Probability
Of Level/Transient

FIGURE 5.20

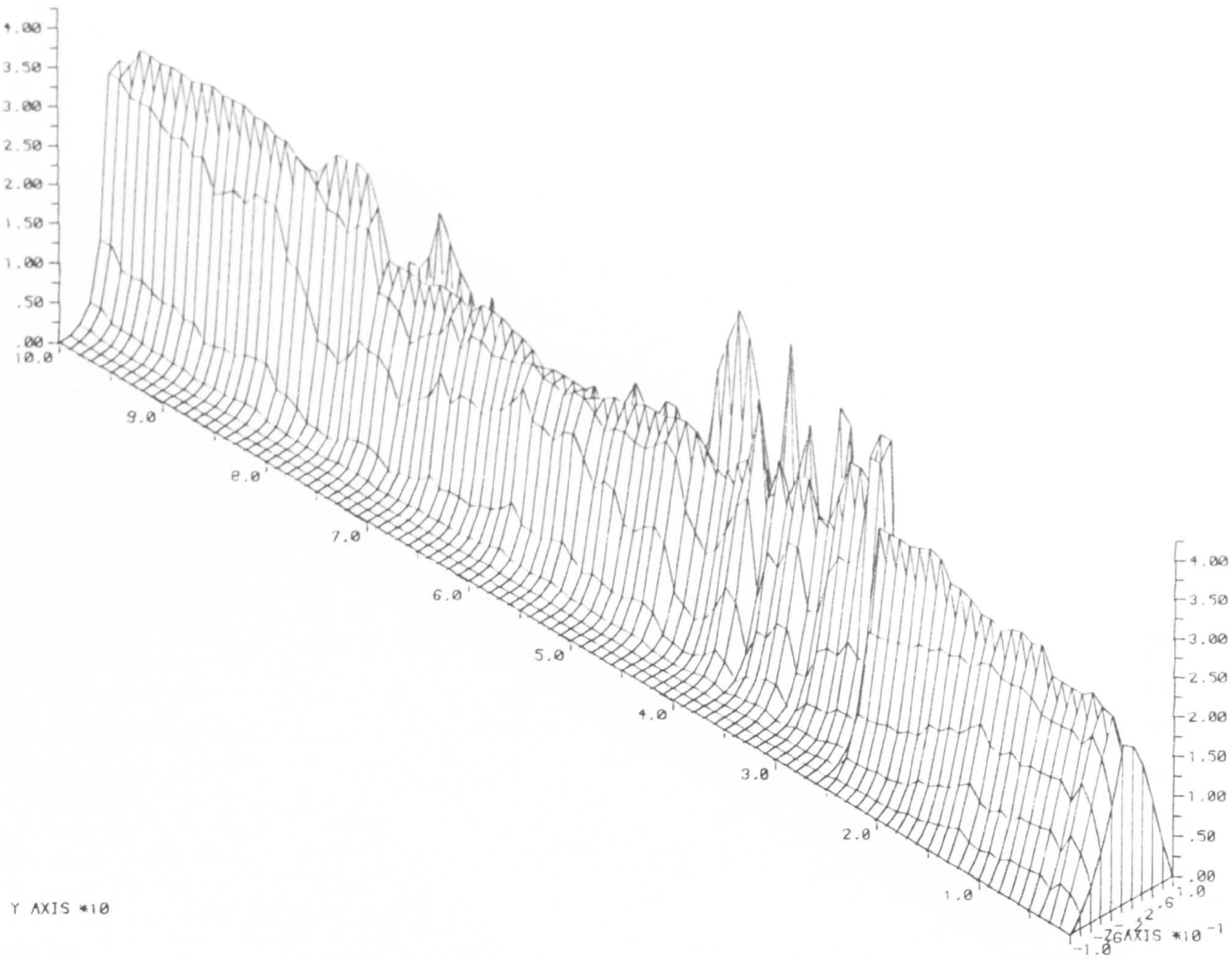

4.00
3.50
3.00
2.50
2.00
1.50
1.00
.50
.00
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
Z AXIS *10^-1

FIGURE 5.21

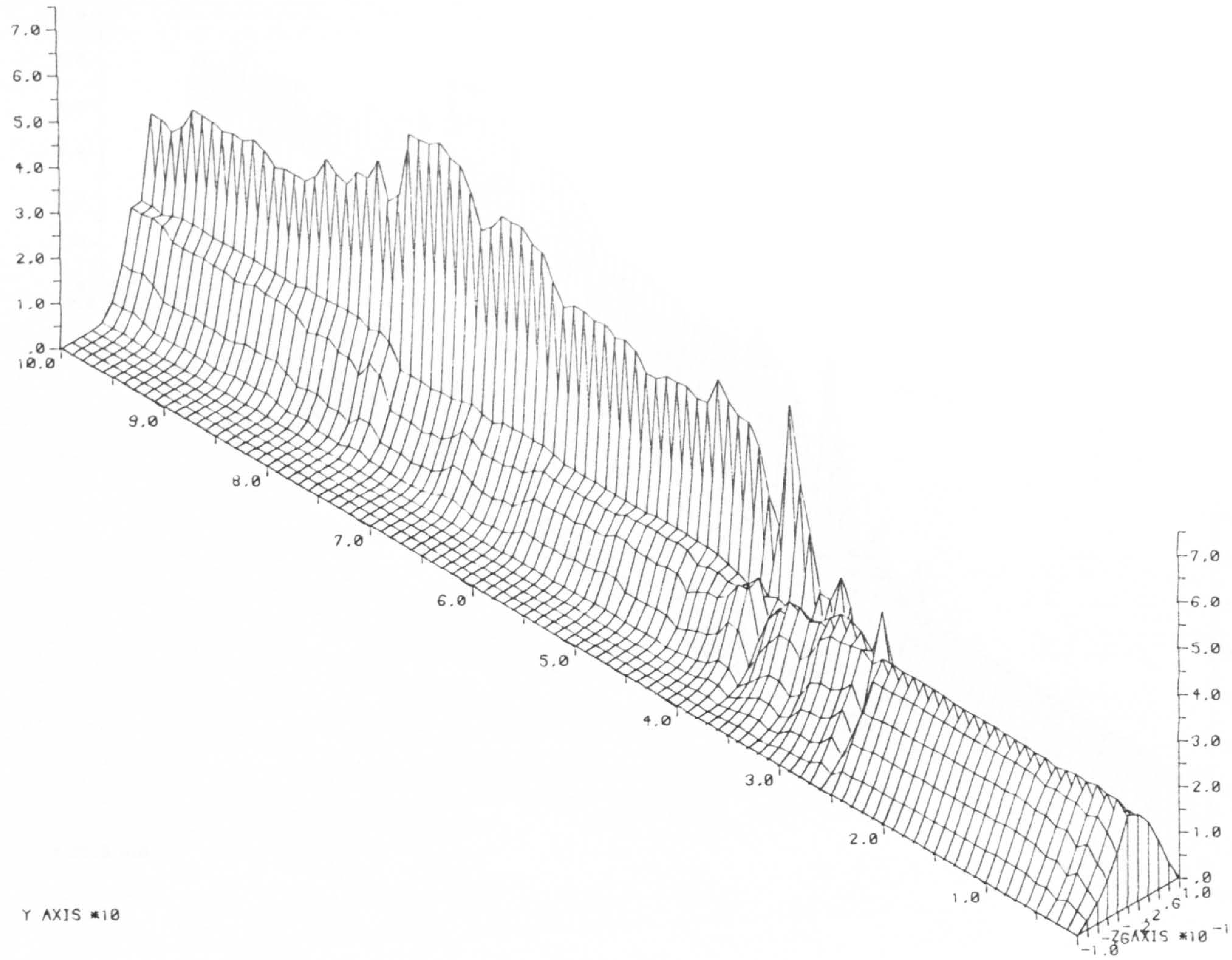
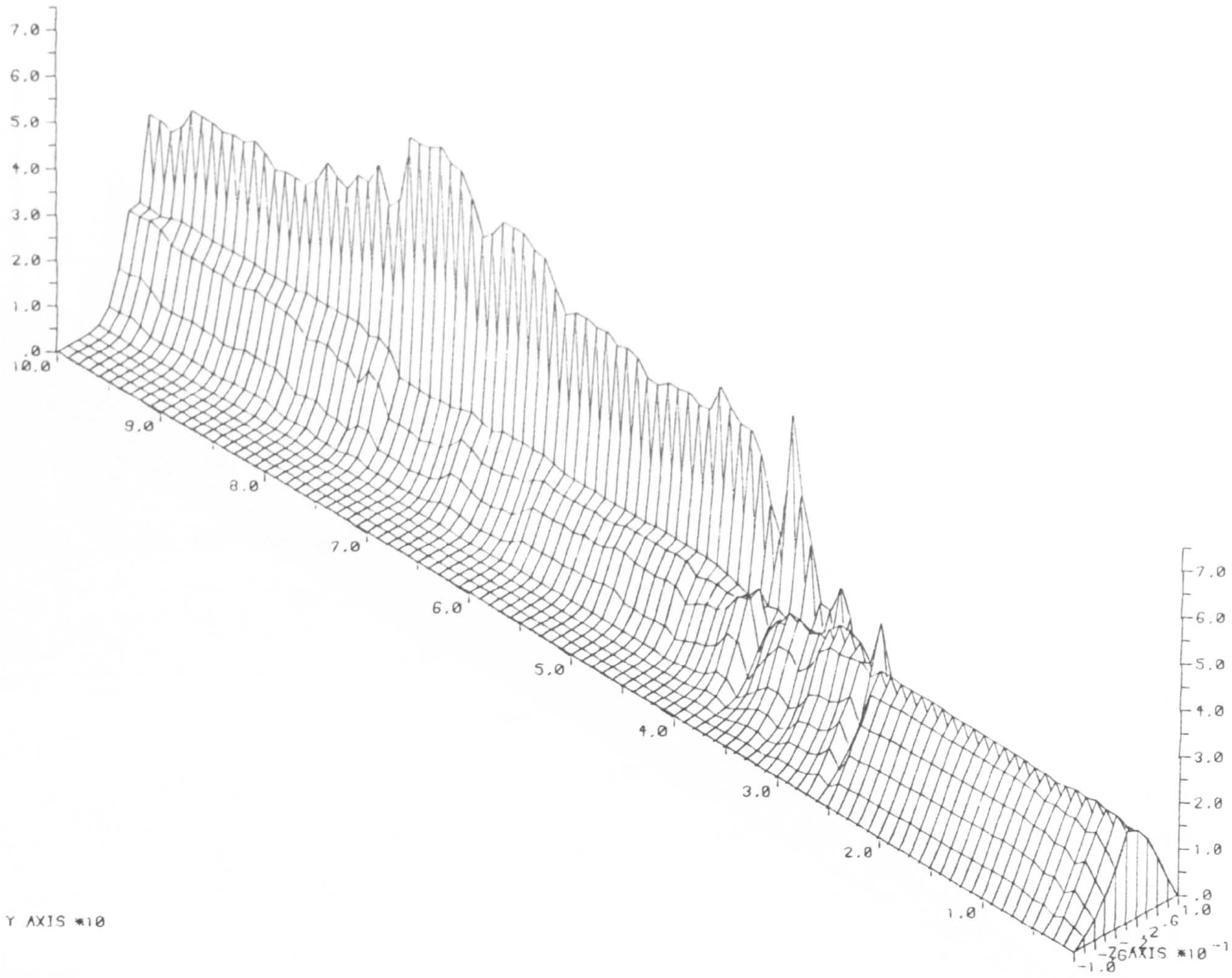

7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
Z AXIS *10^-1

FIGURE 5.22



7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
ZGAXIS *10^-1

FIGURE 5.23

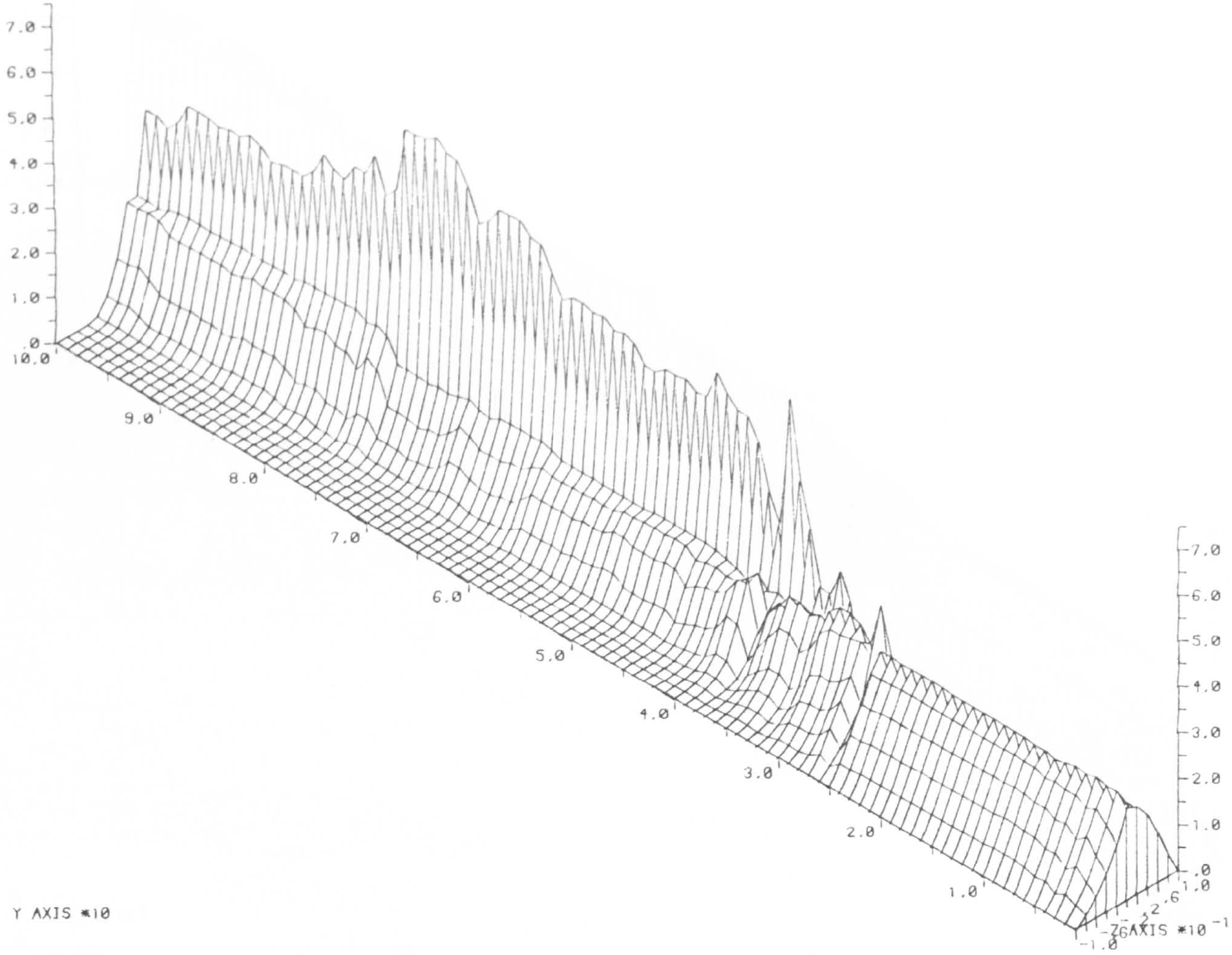

7.0
6.0
5.0
4.0
3.0
2.0
1.0
.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
Z AXIS *10 −1

FIGURE 5.24

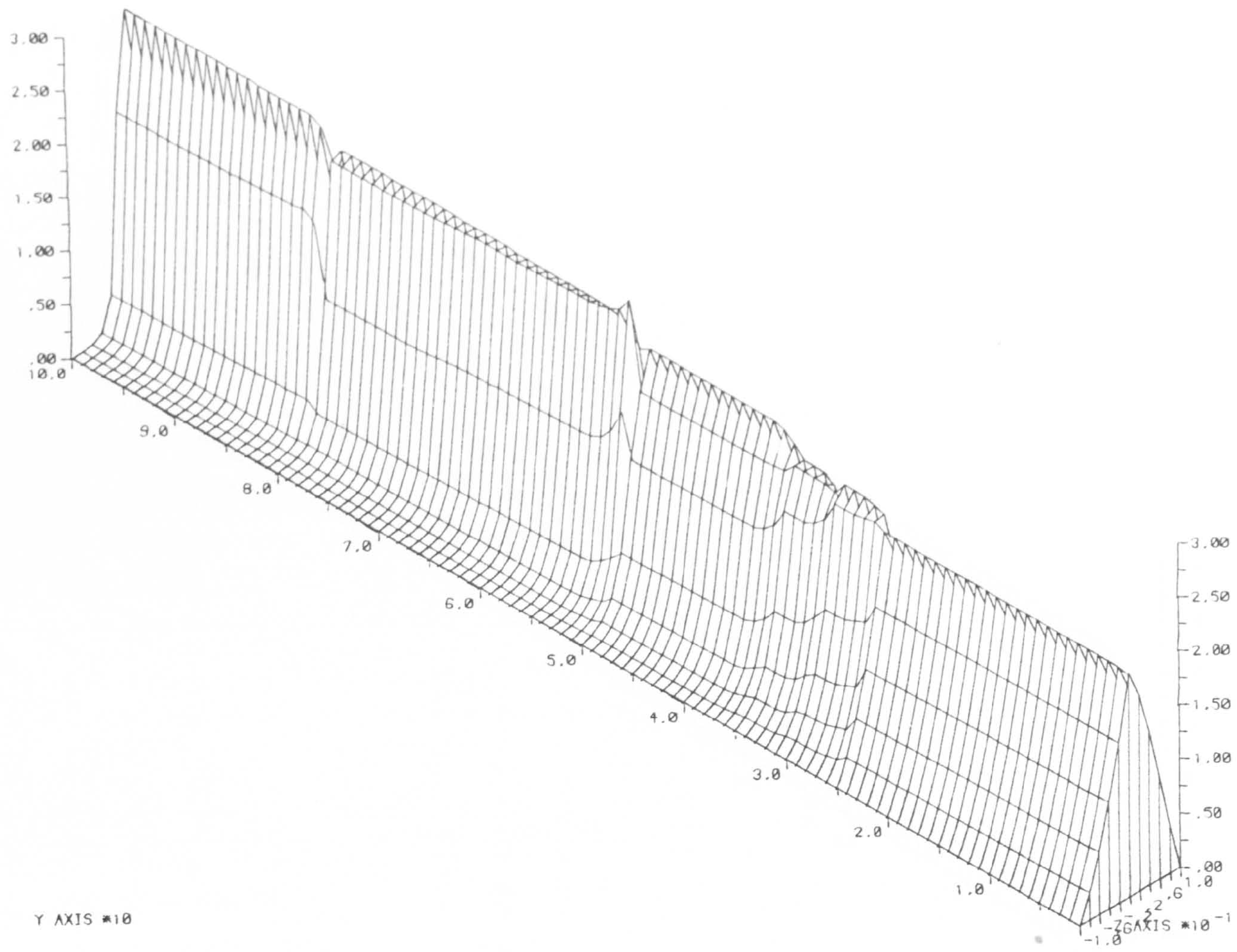

3.00
2.50
2.00
1.50
1.00
.50
.00
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
ZGAXIS *10^-1

FIGURE 5.25

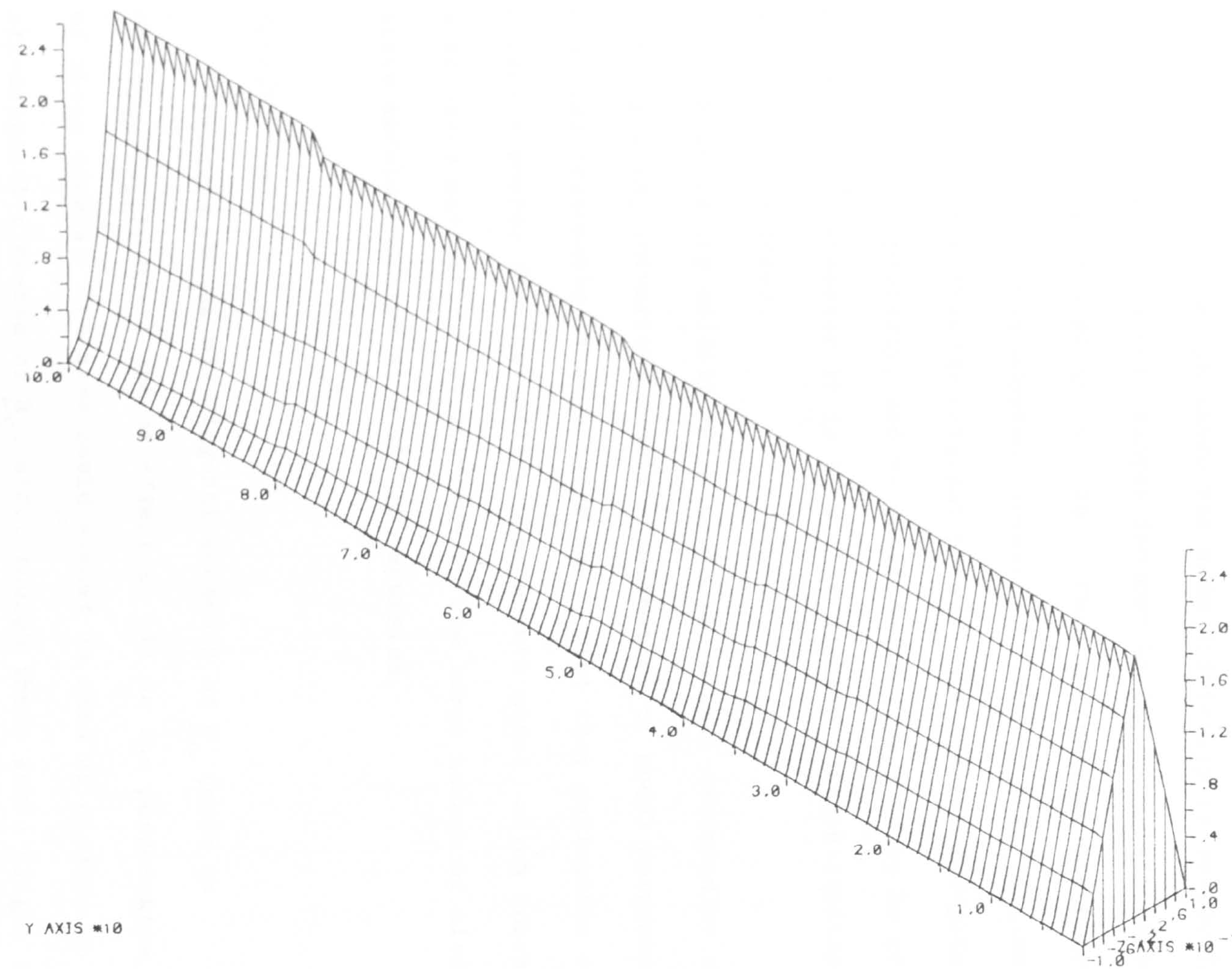

2.4
2.0
1.6
1.2
.8
.4
.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
Y AXIS *10
Z AXIS *10^-1
-1.0
.6
.2
1.0

FIGURE 5.26

5.4.3 CONCLUSIONS

Performance

Discrimination between alternative changepoint types was excellent, especially taking into account the number of competing models involved (further exacerbated when the model has a nuisance parameter, such as in the case of the AR(1)/linear growth model - see Section 5.2.1), although there was some difficulty in correctly identifying the adjacent changes induced in the bivariate linear growth model at t = 24 and t = 25. When a Markovian state-transition matrix, $\{p_o\}$, was adopted, however, involving only marginally more information than the original prior, this changepoint-pairing was identified properly, and we conclude that $\{p_o\}$ is to be preferred to $p_o^{(j)}$ whenever it is felt that adjacent discontinuities are likely to occur.

Forecasting ability was comparable to that achieved by the corresponding univariate models; estimation of model parameters was also reasonable, though there was a hint that estimation was slightly poorer for the AR(1)/linear growth model, which contains a nuisance parameter, possibly due to the large number of alternative models involved at a single recursion.

Sensitivity

Changes to the off-diagonal elements of $\underset{\sim}{C}_o$ (such as $cov(\delta\mu_{1t},\delta\mu_{2t})$, etc.) had no effect at all on the performance of these models. The same could almost be said of changes to the off-diagonal elements of $\underset{\sim}{R}_\varepsilon$, etc., though there was a slight indication that, by allowing higher correlations between concurrent changepoints, we were better at detecting the induced changes in the series. This is probably due to the fact that there was substantial changepoint-correlation in the series actually used, in that, not only were many of the discontinuities concurrent but, approximately 95% of the time, the series were in the steady-state together. However, the improved event detection may have been at the cost of poorer parameter estimation (see, for example, Figure 5.21 which shows how poorly the autoregressive parameter was estimated when these correlations were set too high) and, perhaps, a greater number of false positives.

It is very apparent, however, that the correct specification of the variance ratio, $var(\varepsilon_{1t})/var(\varepsilon_{2t})$, is vital if the models are to perform well. Incorrect specification of this ratio led to poor event discrimination, including more false positives and a number of events being missed altogether (false negatives), and poor estimation/forecasting. This was especially true for the AR(1)/linear growth model, particularly with regard to the estimation of the autoregressive parameter, ϕ. Figures 5.25 and 5.26 show how very poorly ϕ has been estimated when the incorrect variance ratio was specified; these patterns are reminiscent of the type of pattern one would expect to see, if one had specified the wrong univariate model for the y_{1t} series in the first place, i.e. if the AR(1) model was inappropriate (see Appendix A4.2).

Overall Conclusions

We might summarize the results of this chapter by saying that the models discussed perform reasonably well on generally unequally-spaced bivariate time series, with regard to changepoint detection/discrimination, especially when one considers the size of the models and the number of alternative states involved. This seems to suggest that the Kullback-Leibler-based collapsing procedures adopted are more than adequate in this context.

In terms of the amount of computing time necessary, we note that a single recursion (i.e. the recalculation of all model parameters, and all relevant multistate probabilities, upon the receipt of a single observation vector) takes approximately 0.5 s of CPU time (on a University Mainframe computer) for a bivariate model without a nuisance parameter (such as the bivariate linear growth model), approximately 4 s for a model with one nuisance parameter (such as the AR(1)/linear growth model), approximately 15 s for a model with two nuisance parameters (such as the bivariate AR(1) model, in which both series are first-order autoregressive in nature), etc. Therefore, even if observations were arriving every minute, the models should still be of some practical use.

The correct specification of the between-series variance-ratio is very important if the analyses, using these models, are to be believed. As it may often be very difficult to specify this ratio beforehand, it is suggested that alternative methods, for incorporating completely unknown variances into the bivariate, multistate dynamic linear model framework, should be investigated.

oOo

APPENDIX FIVE

A5.1 COROLLARY 5.2.2

Suppose that $\underset{\sim}{y}^T = [y_1 \quad y_2]$ is a single sample vector from a bivariate normal distribution for which the mean vector, $\underset{\sim}{\mu}^T = [\mu_1 \quad \mu_2]$, has an unknown value and the covariance matrix is of the form $\lambda^{-1}\underset{\sim}{\varepsilon}$, where $\underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{12} & \varepsilon_{22} \end{pmatrix}$ is a specified, positive definite matrix and λ is unknown. Suppose also that the prior joint distribution of $\underset{\sim}{\mu}$ and λ is given by:

$$p(\underset{\sim}{\mu}|\lambda) \sim N(\underset{\sim}{m},\lambda^{-1}\underset{\sim}{C}) \tag{A5.1}$$

and

$$p(\lambda) \sim G(\alpha,\beta) \tag{A5.2}$$

where

$$\underset{\sim}{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \qquad \text{and} \qquad \underset{\sim}{C} = \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix} \tag{A5.3}$$

Then the posterior joint distribution of $\underset{\sim}{\mu}$ and λ, if only y_1 is actually observed, is given by:

$$p(\underset{\sim}{\mu}|\lambda, y_1) \sim N(\underset{\sim}{m}^*, \lambda^{-1}\underset{\sim}{C}^*) \tag{A5.4}$$

$$p(\lambda|y_1) \sim G(\alpha^*, \beta^*) \tag{A5.5}$$

where

$$\underset{\sim}{m}^* = \underset{\sim}{C}^*(\underset{\sim}{C}^{-1}\underset{\sim}{m} + \varepsilon_{11}^{-1}\begin{pmatrix}1\\0\end{pmatrix}y_1) \tag{A5.6}$$

$$\underset{\sim}{C}^* = \underset{\sim}{C} - \underset{\sim}{C}\begin{pmatrix}1\\0\end{pmatrix}(C_{11} + \varepsilon_{11})^{-1}[1 \quad 0]\underset{\sim}{C} \tag{A5.7}$$

i.e.

$$(\underset{\sim}{C}^*)^{-1} = \underset{\sim}{C}^{-1} + \begin{pmatrix}1\\0\end{pmatrix}\varepsilon_{11}^{-1}[1 \quad 0] \tag{A5.8}$$

$$\alpha^* = \alpha + \tfrac{1}{2} \tag{A5.9}$$

and

$$\beta^* = \beta + \tfrac{1}{2}\,\frac{(y_1 - m_1)^2}{C_{11} + \varepsilon_{11}} \tag{A5.10}$$

Proof

In terms of a DLM, we may write

$$\begin{pmatrix}y_1\\y_2\end{pmatrix} = \begin{pmatrix}\mu_1\\\mu_2\end{pmatrix} + \begin{pmatrix}e_1\\e_2\end{pmatrix} \tag{A5.11}$$

where

$$\begin{pmatrix}e_1\\e_2\end{pmatrix} \sim N(\underset{\sim}{0}, \lambda^{-1}\begin{pmatrix}\varepsilon_{11} & \varepsilon_{12}\\ \varepsilon_{12} & \varepsilon_{22}\end{pmatrix})$$

Then, with the usual notation

$$\underset{\sim}{H} = \begin{pmatrix}h_1 & 0\\ 0 & h_2\end{pmatrix} = \underset{\sim}{I}, \quad \underset{\sim}{G} = \underset{\sim}{I} \quad \text{and} \quad \underset{\sim\omega}{R} = \underset{\sim}{0}. \tag{A5.12}$$

(A5.7), (A5.9) and (A5.10), therefore, follow directly from Theorem 5.2.2, since

$$f_1 = m_1, \quad \underset{\sim}{P} = \underset{\sim}{C}, \quad F_{11} = C_{11} + \varepsilon_{11} \quad \text{and} \quad \underset{\sim}{S} = \underset{\sim}{C}\begin{pmatrix}1\\0\end{pmatrix}(C_{11} + \varepsilon_{11})^{-1}$$

(A5.8) follows from the relationship given by (5.41), also contained in the theorem.

In order to demonstrate the truth of (A5.6), we note from Theorem 5.2.2 that:

$$\underset{\sim}{m}^* = \underset{\sim}{m} + \underset{\sim}{C}\begin{pmatrix}1\\0\end{pmatrix}(C_{11} + \varepsilon_{11})^{-1}(y_1 - m_1) \qquad \text{(using (A5.12))}$$

$$\Rightarrow (\underset{\sim}{C}^*)^{-1}\underset{\sim}{m}^* = (\underset{\sim}{C}^{-1} + \begin{pmatrix}1\\0\end{pmatrix}\varepsilon_{11}^{-1}[1 \quad 0])(\underset{\sim}{m} + \underset{\sim}{C}\begin{pmatrix}1\\0\end{pmatrix}(C_{11} + \varepsilon_{11})^{-1} \times (y_1 - m_1)) \qquad \text{(using (A5.8))}$$

$$= \underset{\sim}{C}^{-1}\underset{\sim}{m} + \varepsilon_{11}^{-1}\begin{pmatrix}1\\0\end{pmatrix}\{m_1 + \varepsilon_{11}(C_{11} + \varepsilon_{11})^{-1}(y_1 - m_1) + C_{11}(C_{11} + \varepsilon_{11})^{-1}(y_1 - m_1)\}$$

$$= \underset{\sim}{C}^{-1}\underset{\sim}{m} + \varepsilon_{11}^{-1}\begin{pmatrix}1\\0\end{pmatrix}y_1$$

i.e.

$$\underset{\sim}{m}^* = \underset{\sim}{C}^*(\underset{\sim}{C}^{-1}\underset{\sim}{m} + \varepsilon_{11}^{-1}\begin{pmatrix}1\\0\end{pmatrix}y_1).$$

NOTE: For the case when $\underset{\sim}{y} = \begin{pmatrix}y_1\\y_2\end{pmatrix}$ is observed, standard results (see, for example, DeGroot (1970)) yield:

$$\underset{\sim}{m}^* = \underset{\sim}{C}^*(\underset{\sim}{C}^{-1}\underset{\sim}{m} + \underset{\sim}{\varepsilon}^{-1}\underset{\sim}{y}) \qquad \text{(c.f. (A5.6))}$$

$$\underset{\sim}{C}^* = \underset{\sim}{C} - \underset{\sim}{C}(\underset{\sim}{C} + \underset{\sim}{\varepsilon})^{-1}\underset{\sim}{C} \qquad \text{(c.f. (A5.7))}$$

i.e.

$$(\underset{\sim}{C}^*)^{-1} = \underset{\sim}{C}^{-1} + \underset{\sim}{\varepsilon}^{-1} \qquad \text{(c.f. (A5.8))}$$

$$\alpha^* = \alpha + 1 \qquad \text{(c.f. (A5.9))}$$

and

$$\beta^* = \beta + \tfrac{1}{2}(\underset{\sim}{y} - \underset{\sim}{m})^T(\underset{\sim}{C} + \underset{\sim}{\varepsilon})^{-1}(\underset{\sim}{y} - \underset{\sim}{m}) \qquad \text{(c.f. (A5.10))}$$

A5.2 NON-CONJUGACY PROBLEM FOR THE UNKNOWN VARIANCE CASE

Theorem A5.2

Suppose that $\underset{\sim}{y}^T = [y_1 \quad y_2]$ is a single sample vector from a bivariate normal distribution for which the mean vector $\underset{\sim}{\mu} = \begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}$, has an unknown value and the covariance matrix,

$$\underset{\sim}{\Lambda} = \begin{pmatrix} \lambda_1^{-1} & \lambda_{12}^{-1} \\ \lambda_{12}^{-1} & \lambda_2^{-1} \end{pmatrix},$$

is an unknown positive definite matrix. Suppose also that the prior joint distribution of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Lambda}$ is given by:

$$p(\underset{\sim}{\mu}|\underset{\sim}{\Lambda}) \sim N(\underset{\sim}{m},\varepsilon\underset{\sim}{\Lambda}) \quad \text{(A5.13)}$$

$$p(\underset{\sim}{\Lambda}^{-1}) \sim W(\alpha,\underset{\sim}{\tau}) \quad \text{(A5.14)}$$

where $\underset{\sim}{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$, ε is a scalar and where $u \sim W(\alpha,\underset{\sim}{\tau})$ denotes the Wishart distribution with α degrees of freedom and precision matrix $\underset{\sim}{\tau}$. Then the posterior joint distribution of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Lambda}$, if only y_1 is actually observed, does not have the form of a joint Normal-Wishart distribution.

Proof

In terms of a DLM, we may write

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \quad \text{(A5.15)}$$

where

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \sim N(\underset{\sim}{0}, \begin{pmatrix} \lambda_1^{-1} & \lambda_{12}^{-1} \\ \lambda_{12}^{-1} & \lambda_2^{-1} \end{pmatrix})$$

i.e.

$$p(y_1|\underset{\sim}{\mu},\underset{\sim}{\Lambda}) \sim N(\mu_1,\lambda_1^{-1}) \quad \text{(A5.16)}$$

Now

$$p(\underset{\sim}{\mu},\underset{\sim}{\Lambda}|y_1) \propto p(y_1|\underset{\sim}{\mu},\underset{\sim}{\Lambda})p(\underset{\sim}{\mu}|\underset{\sim}{\Lambda})p(\underset{\sim}{\Lambda}^{-1}) \qquad \text{(using Bayes theorem)}$$

$$\propto \lambda_1^{\frac{1}{2}}\exp\{-\frac{\lambda_1}{2}(y_1 - \mu_1)^2\}.|\underset{\sim}{\Lambda}|^{\frac{1}{2}}\exp\{-\frac{1}{2\varepsilon}(\underset{\sim}{\mu} - \underset{\sim}{m})^T\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\mu} - \underset{\sim}{m})\}$$

$$\times |\underset{\sim}{\Lambda}|^{(\alpha-3)/2}\exp\{-\tfrac{1}{2}tr(\underset{\sim}{\tau}\underset{\sim}{\Lambda}^{-1})\}$$

$$= \lambda_1^{\frac{1}{2}}|\underset{\sim}{\Lambda}|^{(\alpha-2)/2}\exp\{-\tfrac{1}{2}[\lambda_1(y_1 - \mu_1)^2 + \frac{1}{\varepsilon}(\underset{\sim}{\mu} - \underset{\sim}{m})^T\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\mu} - \underset{\sim}{m}) + tr(\underset{\sim}{\tau}\underset{\sim}{\Lambda}^{-1})]\} \qquad (A5.17)$$

Now suppose that the posterior joint distribution of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Lambda}$ is Normal-Wishart, i.e.

$$p(\underset{\sim}{\mu}|\underset{\sim}{\Lambda},y_1) \sim N(\underset{\sim}{m}^*,\varepsilon^*\underset{\sim}{\Lambda}) \qquad (A5.18)$$

$$p(\underset{\sim}{\Lambda}^{-1}|y_1) \sim W(\alpha^*,\underset{\sim}{\tau}^*) \qquad (A5.19)$$

Then

$$p(\underset{\sim}{\mu},\underset{\sim}{\Lambda}|y_1) \propto |\underset{\sim}{\Lambda}|^{(\alpha^*-2)/2}\exp\{-\tfrac{1}{2}[\frac{1}{\varepsilon^*}(\underset{\sim}{\mu} - \underset{\sim}{m}^*)^T\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\mu} - \underset{\sim}{m}^*) + tr(\underset{\sim}{\tau}^*\underset{\sim}{\Lambda}^{-1})]\} \qquad (A5.20)$$

Upon equating (A5.17) and (A5.20), we find:

$$|\underset{\sim}{\Lambda}|^{(\alpha^*-2)/2} = \lambda_1^{\frac{1}{2}}|\underset{\sim}{\Lambda}|^{(\alpha-2)/2}$$

i.e.

$$\alpha^* = \alpha + \frac{\ln(\lambda_1)}{\ln(|\underset{\sim}{\Lambda}|)} \qquad (A5.21)$$

Also,

$$\frac{1}{\varepsilon^*}(\underset{\sim}{\mu} - \underset{\sim}{m}^*)^T\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\mu} - \underset{\sim}{m}^*) + tr(\underset{\sim}{\tau}^*\underset{\sim}{\Lambda}^{-1}) = \lambda_1(y_1 - \mu_1)^2 + \frac{1}{\varepsilon}(\underset{\sim}{\mu} - \underset{\sim}{m})^T\underset{\sim}{\Lambda}^{-1}(\underset{\sim}{\mu} - \underset{\sim}{m}) + tr(\underset{\sim}{\tau}\underset{\sim}{\Lambda}^{-1})$$

i.e.

$$A = B,$$

say, where

$$A = \frac{1}{\varepsilon^*}[\mu_1 - m_1^* \quad \mu_2 - m_2^*]\underset{\sim}{\Lambda}^{-1}\begin{pmatrix}\mu_1 - m_1^* \\ \mu_2 - m_2^*\end{pmatrix} + tr(\underset{\sim}{\tau}^*\underset{\sim}{\Lambda}^{-1})$$

$$= \frac{1}{\varepsilon^*|\underset{\sim}{\Lambda}|}\{\frac{(\mu_1 - m_1^*)^2}{\lambda_2} + \frac{(\mu_2 - m_2^*)^2}{\lambda_1} - \frac{2(\mu_1 - m_1^*)(\mu_2 - m^*_2)}{\lambda_{12}}\} + tr(\underset{\sim}{\tau}^*\underset{\sim}{\Lambda}^{-1})$$

$$= \frac{1}{\varepsilon^*|\underset{\sim}{\Lambda}|}\{\frac{\mu_1^2}{\lambda_2} + \frac{\mu_2^2}{\lambda_1} - \frac{2\mu_1\mu_2}{\lambda_{12}} + 2(\frac{m_2^*}{\lambda_{12}} - \frac{m_1^*}{\lambda_2})\mu_1 + 2(\frac{m_1^*}{\lambda_{12}} - \frac{m_2^*}{\lambda_1})\mu_2\}$$

$$+ \{\text{terms not involving } \underset{\sim}{\mu}\} \qquad \text{(A5.22)}$$

Similarly,

$$B = (\frac{1}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_2} + \lambda_1)\mu_1^2 + \frac{1}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_1}\mu_2^2 - \frac{2}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_{12}}\mu_1\mu_2$$

$$+ 2(\frac{m_2}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_{12}} - \frac{m_1}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_{12}} - \frac{\lambda_1 y_1}{\varepsilon})\mu_1 + 2(\frac{m_1}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_{12}} - \frac{m_2}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_1})\mu_2$$

$$+ \{\text{ terms not involving } \underset{\sim}{\mu}\} \qquad \text{(A5.23)}$$

Since A must equal B, (A5.22) and (A5.23) imply that, for instance,

$$\frac{1}{\varepsilon^*|\underset{\sim}{\Lambda}|\lambda_2} = \frac{1}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_2} + \lambda_1 \qquad \text{(from the terms in } \mu_1^2\text{)} \qquad \text{(A5.24)}$$

and

$$\frac{1}{\varepsilon^*|\underset{\sim}{\Lambda}|\lambda_1} = \frac{1}{\varepsilon|\underset{\sim}{\Lambda}|\lambda_1} \qquad \text{(from the terms in } \mu_2^2\text{)} \qquad \text{(A5.25)}$$

and, because $\underset{\sim}{\Lambda}$ is positive definite, it is clear that this leads to a contradiction since, from (A5.24)

$$\varepsilon^* = \frac{\lambda_{12}^2}{(\varepsilon + 1)\lambda_{12}^2 - \varepsilon\lambda_1\lambda_2}\varepsilon \qquad \text{(A5.26)}$$

whereas, from (A5.25)

$$\varepsilon^* = \varepsilon \tag{A5.27}$$

Therefore the posterior joint distribution for $\underset{\sim}{\mu}$ and $\underset{\sim}{\Lambda}$ does not have a Normal-Wishart form.

A5.3 ADDITIONAL DATA SET

Using the starting values of $\underset{\sim}{\theta}_o = \begin{pmatrix} 500 \\ 4 \end{pmatrix}$, 100 observations were generated according to the linear growth model of Section 2.3.1, with the errors, ε_t, simulated from $\varepsilon_t \sim N(0,30)$. At $t = 24$, a level change was simulated by setting $R\mu_{24} = 20$, i.e. $\delta\mu_{24} \sim N(0,600)$. At $t = 75$, a slope change was simulated by setting $R\beta_{75} = 10$, i.e. $\delta\beta_{75} \sim N(0,300)$. At $t = 50$ and $t = 80$, transients were simulated by setting $R\varepsilon_{50} = R\varepsilon_{80} = 30$, i.e. $\varepsilon_{50} \sim N(0,900)$ and $\varepsilon_{80} \sim N(0,900)$. At $t = 100$, the true value of $\underset{\sim}{\theta}$ was $\underset{\sim}{\theta}_{100} = \begin{pmatrix} 3527.6 \\ 104.8 \end{pmatrix}$. The following data set obtained:

TIME	OBSERVATION	TIME	OBSERVATION
1.0	512.2	23.0	568.1
2.0	475.9	24.0	641.0
3.0	535.1	25.0	616.7
4.0	488.8	26.0	609.4
5.0	507.1	27.0	606.7
6.0	516.8	28.0	553.7
7.0	524.5	29.0	604.2
8.0	557.2	30.0	635.7
9.0	526.4	31.0	628.4
10.0	579.1	32.0	625.5
11.0	551.3	33.0	630.5
12.0	561.4	34.0	626.5
13.0	520.7	35.0	642.3
14.0	577.4	36.0	638.9
15.0	511.2	37.0	645.9
16.0	584.6	38.0	686.8
17.0	585.3	39.0	702.5
18.0	580.0	40.0	676.3
19.0	580.1	41.0	684.3
20.0	632.1	42.0	735.7
21.0	567.3	43.0	697.7
22.0	549.8	44.0	688.6

TIME	OBSERVATION	TIME	OBSERVATION
45.0	691.1	73.0	801.6
46.0	682.8	74.0	809.6
47.0	675.9	75.0	866.0
48.0	715.8	76.0	1001.6
49.0	706.9	77.0	1123.1
50.0	557.8	78.0	1205.9
51.0	703.5	79.0	1264.3
52.0	729.4	80.0	1808.0
53.0	709.1	81.0	1520.7
54.0	722.1	82.0	1616.8
55.0	738.5	83.0	1703.3
56.0	728.1	84.0	1844.4
57.0	742.5	85.0	1962.2
58.0	766.3	86.0	2047.1
59.0	739.4	87.0	2203.8
60.0	746.7	88.0	2215.3
61.0	713.1	89.0	2348.9
62.0	799.7	90.0	2425.0
63.0	753.1	91.0	2535.2
64.0	748.3	92.0	2673.3
65.0	783.1	93.0	2755.1
66.0	776.1	94.0	2891.2
67.0	826.9	95.0	3005.0
68.0	778.4	96.0	3100.5
69.0	820.1	96.0	3211.9
70.0	803.1	98.0	3311.9
71.0	835.5	99.0	3405.1
72.0	781.0	100.0	3588.7

o0o

CHAPTER SIX

APPLICATIONS

Introduction

In order to illustrate the methodology outlined in previous chapters, we now present a selection of applications of these models to medical time series. In particular, we shall concentrate on the univariate models which were examined in Chapters 3 and 4 (namely, the linear growth, sinusoidal and AR(1) models) and the bivariate models which were examined in Chapter 5 (namely, the bivariate linear growth model and the AR(1)/linear growth model).

Although model identification is crucial to the successful implementation of this methodology , we shall not discuss this aspect in detail in what follows and we do not claim that the particular model chosen for a specific application is necessarily the 'correct' or 'best' model in some general 'scientific' sense. The examples have been chosen to illustrate the potential application of this methodology to a variety of medical monitoring problems.

6.1 UNIVARIATE EXAMPLES

6.1.1 LINEAR GROWTH MODEL

6.1.1.1 Kidney Transplantation

When a patient receives a transplanted kidney it becomes important to monitor the state of kidney function, since rejection of the new kidney is a common phenomenon. In order to assess how well the kidneys are performing (in the sense of clearing 'poisonous' substances from the body), one would ideally like to monitor the progress of the Glomerular Filtration Rate (GFR) which is a direct measure of this rate of clearance. The GFR, however, is unobservable and other biochemical indicators related to kidney function serve as proxy measures. The blood concentration of creatinine, for example, is widely used as such an indicator. Several investigators (for instance Knapp et al. 1977, West 1982) have shown that body-weight adjusted reciprocal serum creatinine concentrations provide a time series which is well-suited to analysis by the linear growth model, and the corresponding multistate dynamic linear model has been shown to be very useful in detecting kidney rejection episodes (Trimble et al. 1983). However, the patients involved were part of a study to determine the usefulness of the statistical methodology and, therefore, great care was taken in order to obtain equally-spaced creatinine measurements. Here, we consider cases where the data have been collected during routine clinical monitoring and therefore contain a number of gaps. In order to apply our extended methodology to these unequally-spaced series we have used the same baseline prior information as was used by Smith and West (1983).

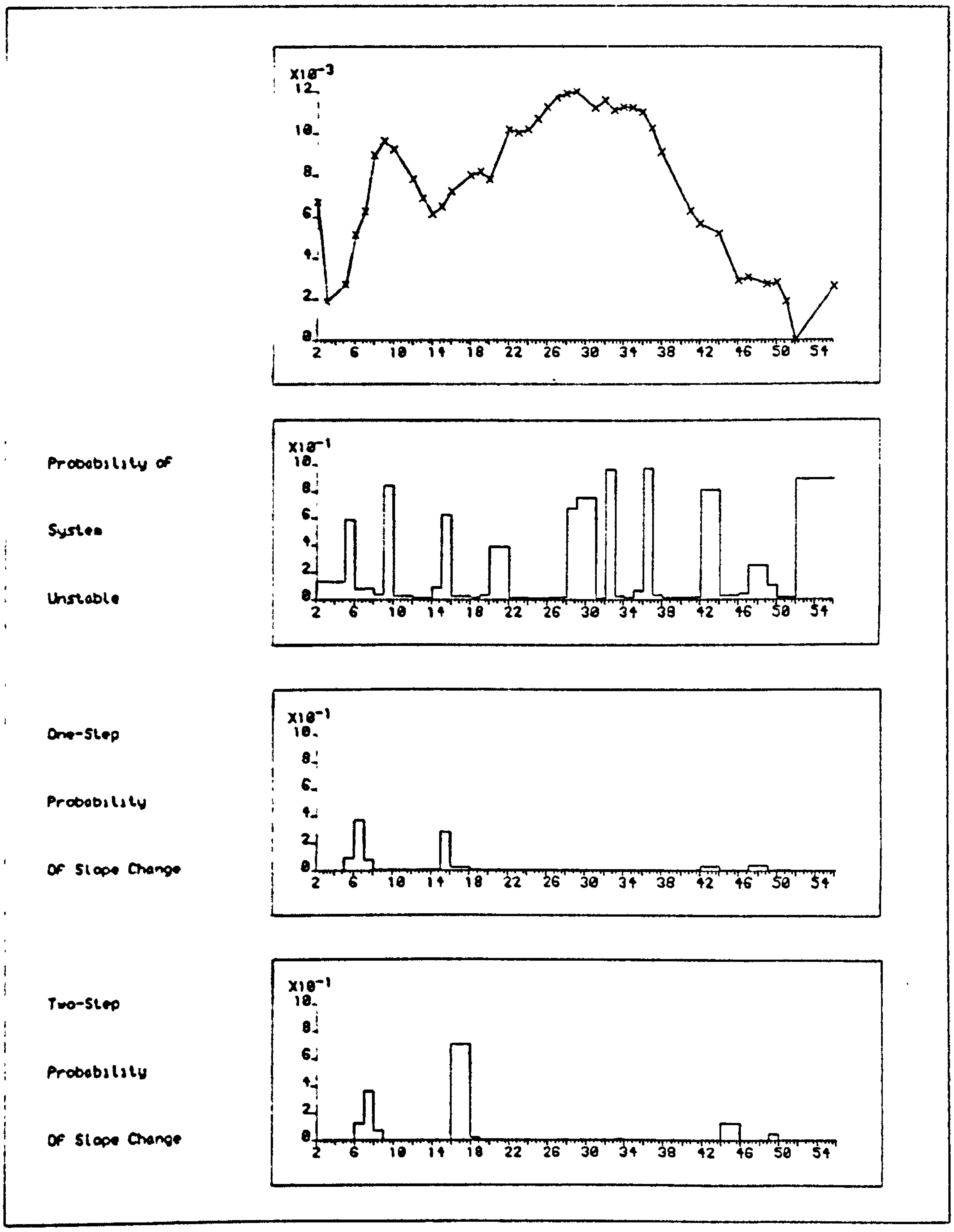
The first two series presented are each from patients who were transplanted by the Renal Transplant Unit Team at the Cardiff

Royal Infirmary. For the first of these series (Figure 6.1), we see that positive signals ($O^{(3)}_{T_k} > 0.2$) arise on two occasions: $T_k = 7$ and $T_k = 16$, and we therefore suggest that rejection episodes are associated with these events. A retrospective look at the clinical record showed that the clinicians concerned initiated rejection therapy at precisely the same two timepoints.

In the second series (Figure 6.2), we again notice two positive signals at $T_k = 9$ and at $T_k = 111$. Reference to the clinical record showed that rejection therapy was initiated at $T_k = 9$ and again at $T_k = 112$, i.e. the second of the signals was one day earlier than the clinician's reaction. This is hardly surprising, considering both the sparseness of the data leading up to this event (implying very little patient/clinician interaction) and the fact that chronic renal failure had already begun (depicted by the slow upward trend). Note, too, that at $T_k = 19$ a 'diminished' signal ($O^{(3)}_{19} = 0.125$) is produced by the analysis; it is of interest that the clinical record reports that the patient had left the ward, and returned home, just one day beforehand.

These two time series were part of a larger collection of cases from Cardiff whose detailed analyses produced results very similar to those reported by Trimble et al. (1983), who studied a group of patients from Nottingham. The statistical analysis detected 24 out of the 25 rejection episodes presented by the Cardiff group, with median difference (statistical signal-clinical 'reaction') of zero days.

The final series in this section was obtained from a patient who was involved in a pilot study initiated by the United Kingdom Transplant Services (UKTS) Management Committee. One of the aims



Probability of
System
Unstable
One-Step
Probability
Of Slope Change
Two-Step
Probability
Of Slope Change

FIGURE 6.1

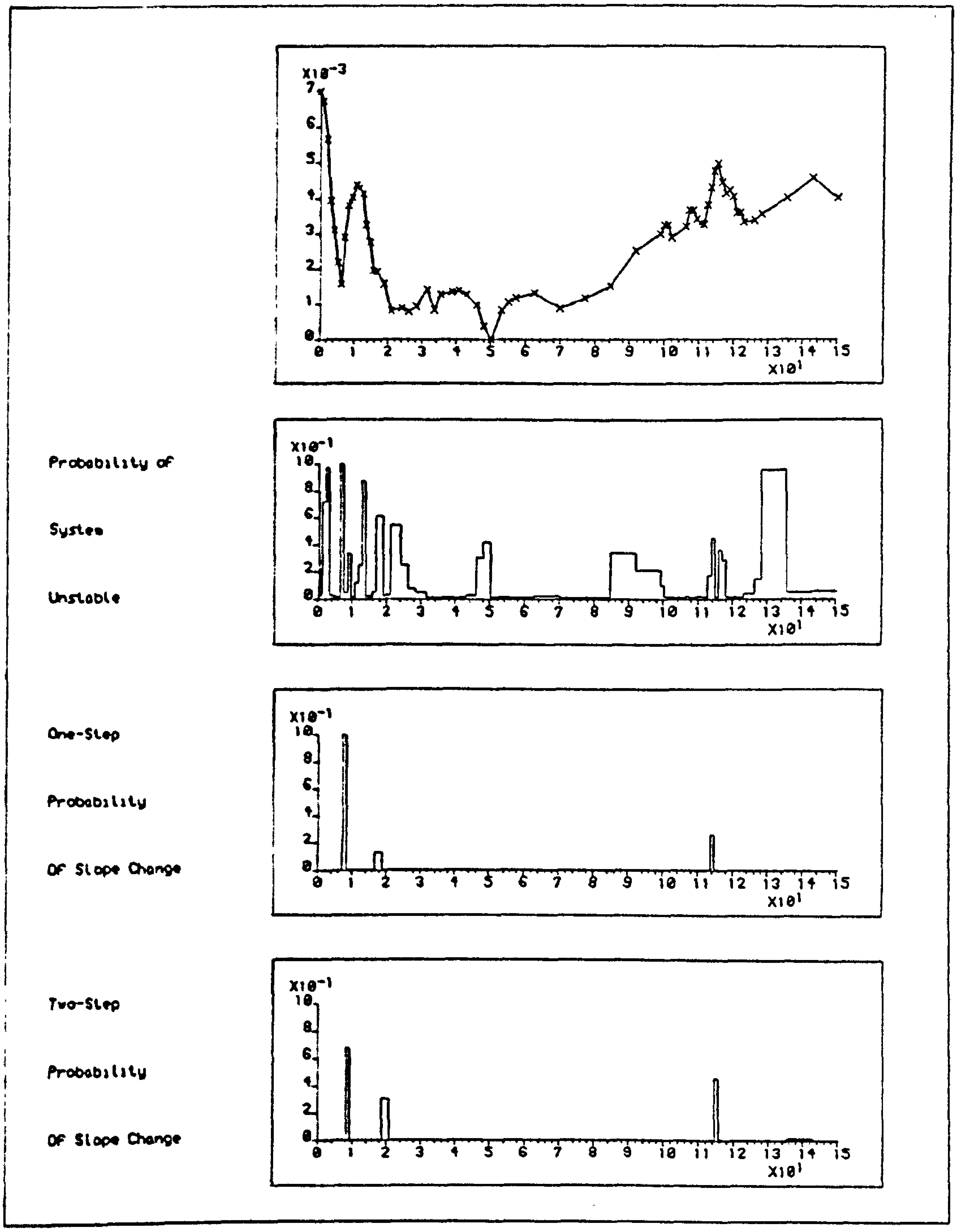

Probability of
System
Unstable
One-Step
Probability
Of Slope Change
Two-Step
Probability
Of Slope Change

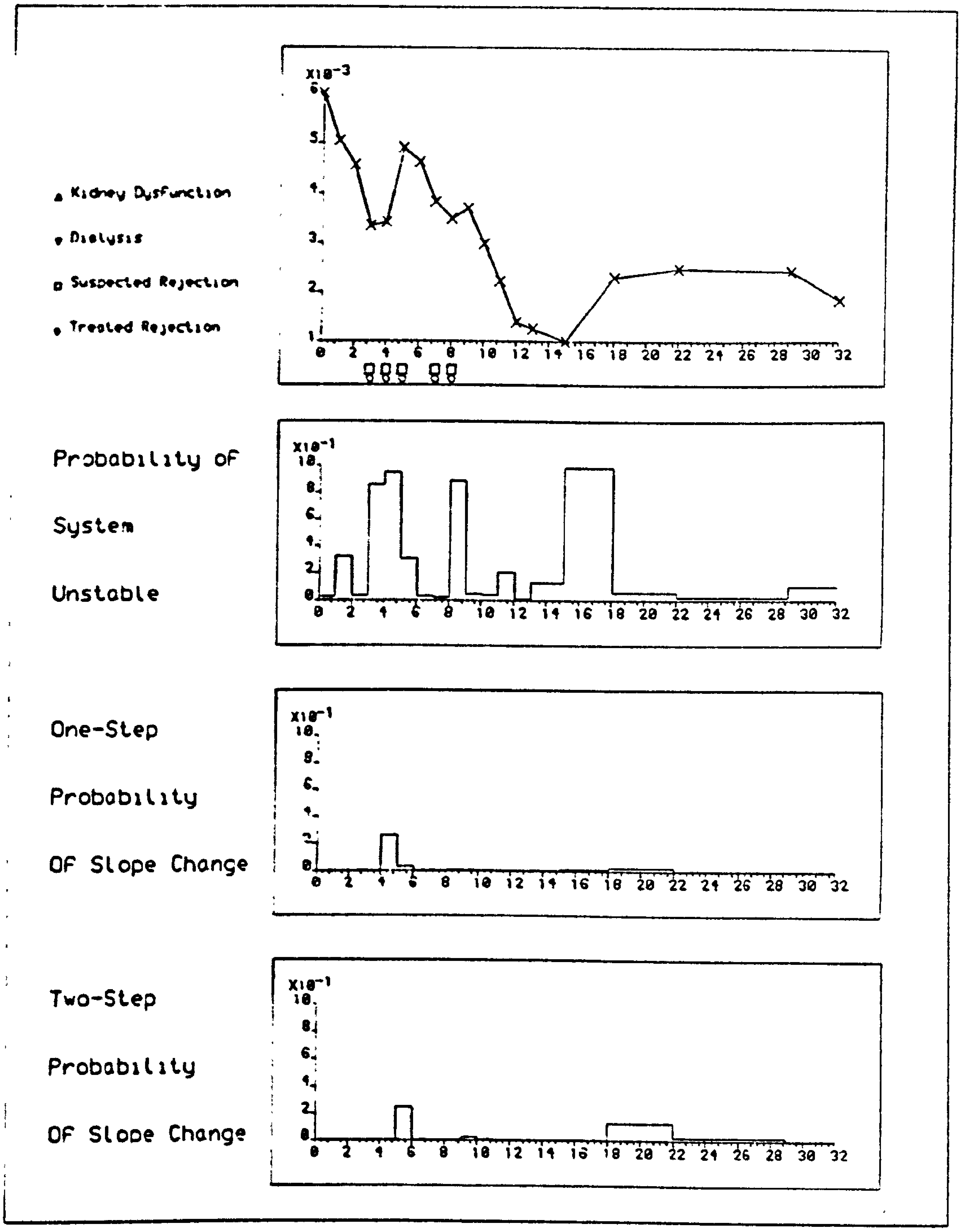
FIGURE 6.2

of this study was to determine whether or not so-called 'highly-sensitized' patients exhibited a different post-transplant pattern to patients who were not 'highly-sensitized'. There was a suggestion that the patterns did indeed differ, in that highly sensitized patients were more inclined to have a 'grumbling start', in the sense that it tended to take longer for the new kidney to start to function properly (sometimes reflected by initially high concentrations of creatinine). In these situations the analysis presented above, using the baseline prior information of Smith and West (1983), often failed to detect early rejection episodes (see Figure 6.3). We attempted, however, to modify this information by adjusting our initial estimate of the slope parameter, which was seen (in Section 3.4) to be rather sensitive. We found that by changing our initial estimate from $b_o = 0.0$ to $b_o = 0.003$ (i.e. reflecting a belief that the reciprocal of creatinine was initially increasing) we were able to identify correctly many of these very early events (see Figure 6.4).

6.1.1.2 Bone Marrow Transplantation

Bone marrow transplantation is one form of treatment for patients who have the blood disorder leukemia and, post-transplant, such patients are often monitored in terms of time series of blood cell counts or concentrations. Jones (1984) presented an example in which three such indicators were utilized: white blood cell count (WBC), platelet count and haematocrit. In this section, and in a number of subsequent sections, we shall re-examine these data sets.

Jones (1984) suggested that, having first taken the logarithms of the WBC counts and the platelet counts, each of these



Kidney Dysfunction
Dialysis
Suspected Rejection
Treated Rejection
Probability of System Unstable
One-Step Probability Of Slope Change
Two-Step Probability Of Slope Change
x10^-3
x10^-1

FIGURE 6.3

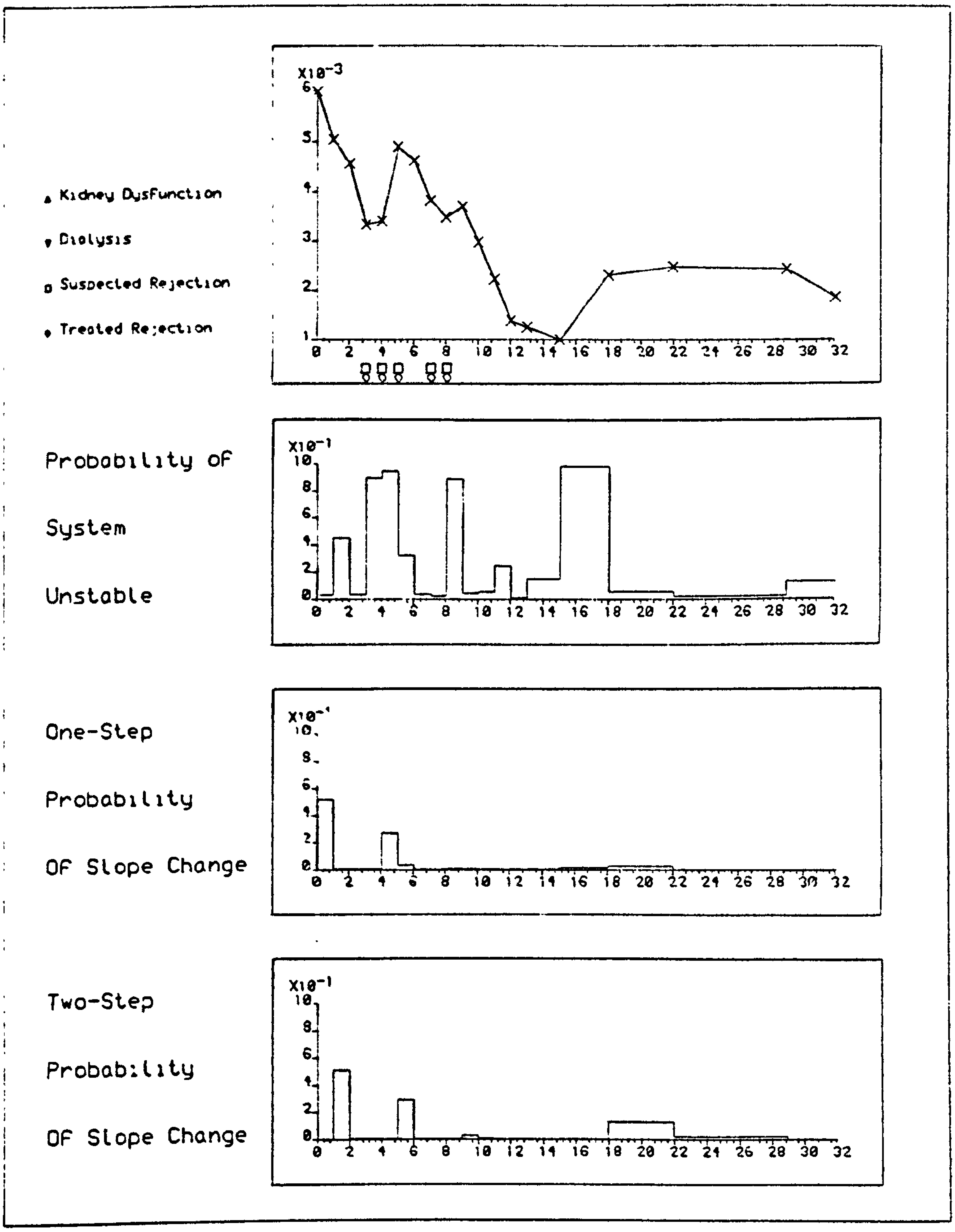

Kidney Dysfunction
Dialysis
Suspected Rejection
Treated Rejection
Probability of
System
Unstable
One-Step
Probability
Of Slope Change
Two-Step
Probability
Of Slope Change

FIGURE 6.4

series could be reasonably modelled by a first order autoregressive model. Previous experience with series of WBC counts, however, has led us to believe that the untransformed counts are likely to exhibit AR(1)-type behaviour (see Section 6.1.3.1). It is then not unnatural to suppose that the logarithm of these values will display a linear structure, since (ignoring error terms) if

$$X_t = \phi X_{t-1} \tag{6.1}$$

then

$$\log(X_t) = \log(\phi) + \log(X_{t-1}) \tag{6.2}$$

and, since ϕ is assumed constant, it is clear that (6.2) has the (recursive) form of a straight line. It is, of course, true that the setting of Section 6.1.3.1 (that of renal dialysis) is very different from that of bone-marrow transplantation, so that there is no reason why the steady-state patterns of WBC counts should be identical to those encountered in that case. However, inspection of the graph of log(WBC) against time (see Figure 6.5) seems to suggest a straight line form (bearing in mind that the graph depicts post-transplantation observations and may well contain changepoints), and so we shall examine the WBC series using the linear growth model.

Before we do so, we note that, whereas Jones (1984) used the $\log_{10}$ transformation for WBC and platelet counts, we have preferred to use natural logarithms; the scales of measurement therefore differ from those presented by Jones (1984). We have assumed that the raw WBC count was about 5000 (the lower end of the 'normal range'), with a negligible slope, initially. On the log scale we thus have prior parameters:

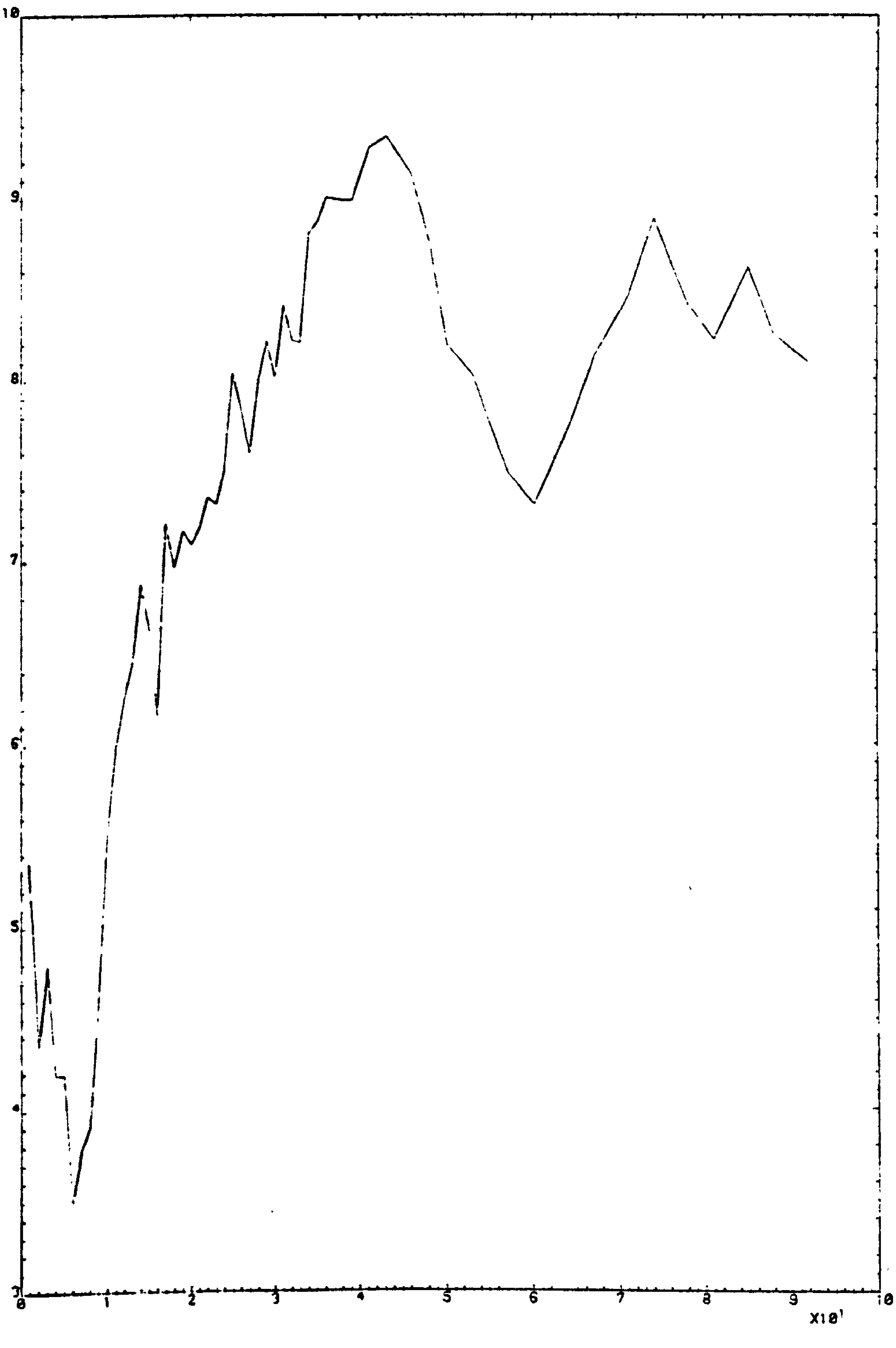


FIGURE 6.5

$$\underset{\sim}{m}_o = \begin{pmatrix} 8.5 \\ 0.0 \end{pmatrix}, \qquad \underset{\sim}{C}_o = \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}$$

where the elements of $\underset{\sim}{C}_o$ are chosen to give a fairly diffuse prior distribution. Results from the analysis are presented in Figure 6.6, in which the transformed data along with one-step-ahead forecasts can also be seen.

It is interesting to note that changepoint signals ($O_{T_k} > 0.2$) are obtained on four occasions: slope changes were detected at $T_k = 10$, 16 and 50, and a level change was signalled on day 67. The latter illustrates the fact that changepoint-types may be confused when the data is very sparse (see Section 4.4.4).

What do these changepoints represent? Clearly, the first slope change is an abrupt reversal of slope from negative to positive (i.e. from deterioration to improvement) and therefore seems to indicate that the treatment became effective at this point. The second change can be seen to be a (downward) deflection of the slope, possibly implying that a less vigorous course of therapy was initiated at this point. At $T_k = 50$, however, a change in slope from positive to negative was signalled, and we tentatively suggest that this may be due to a transplant rejection episode.

As a further guide to the suitability of the model, Figure 6.7 shows on-line estimation of the ϕ-grid, from an analysis in which it was assumed that an AR(1), with autoregressive parameter ϕ, was the correct model for the transformed data. On comparison of this display with that found in Appendix A4.2, it seems likely that the AR(1) is not the most suitable model. This is also confirmed by a drop in forecasting ability (linear growth: SSFE = 23.3, MAD = 0.45; AR(1): SSFE = 30.4, MAD = 0.65).

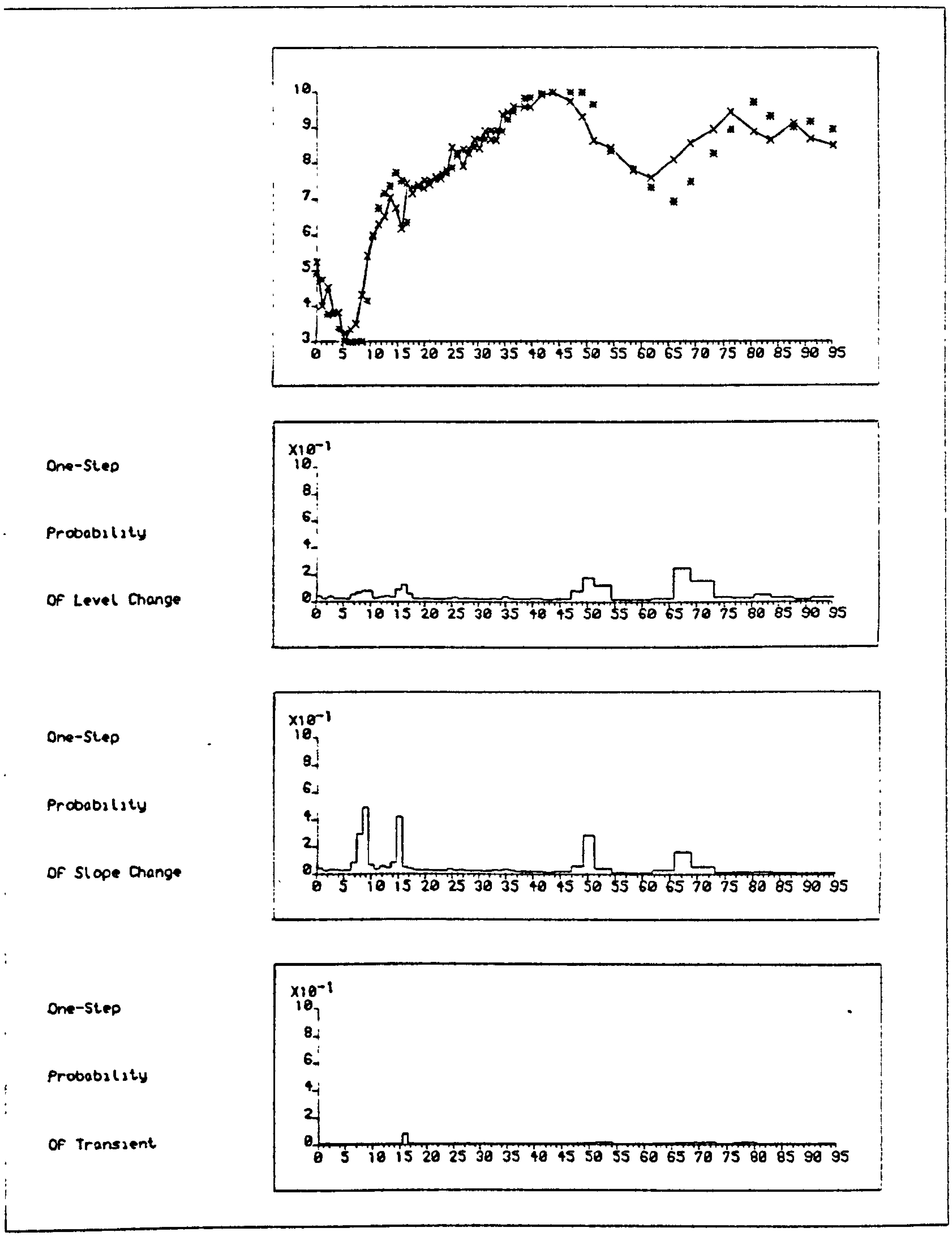

One-Step
Probability
Of Level Change
X10⁻¹
One-Step
Probability
Of Slope Change
X10⁻¹
One-Step
Probability
Of Transient
X10⁻¹
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95

FIGURE 6.6

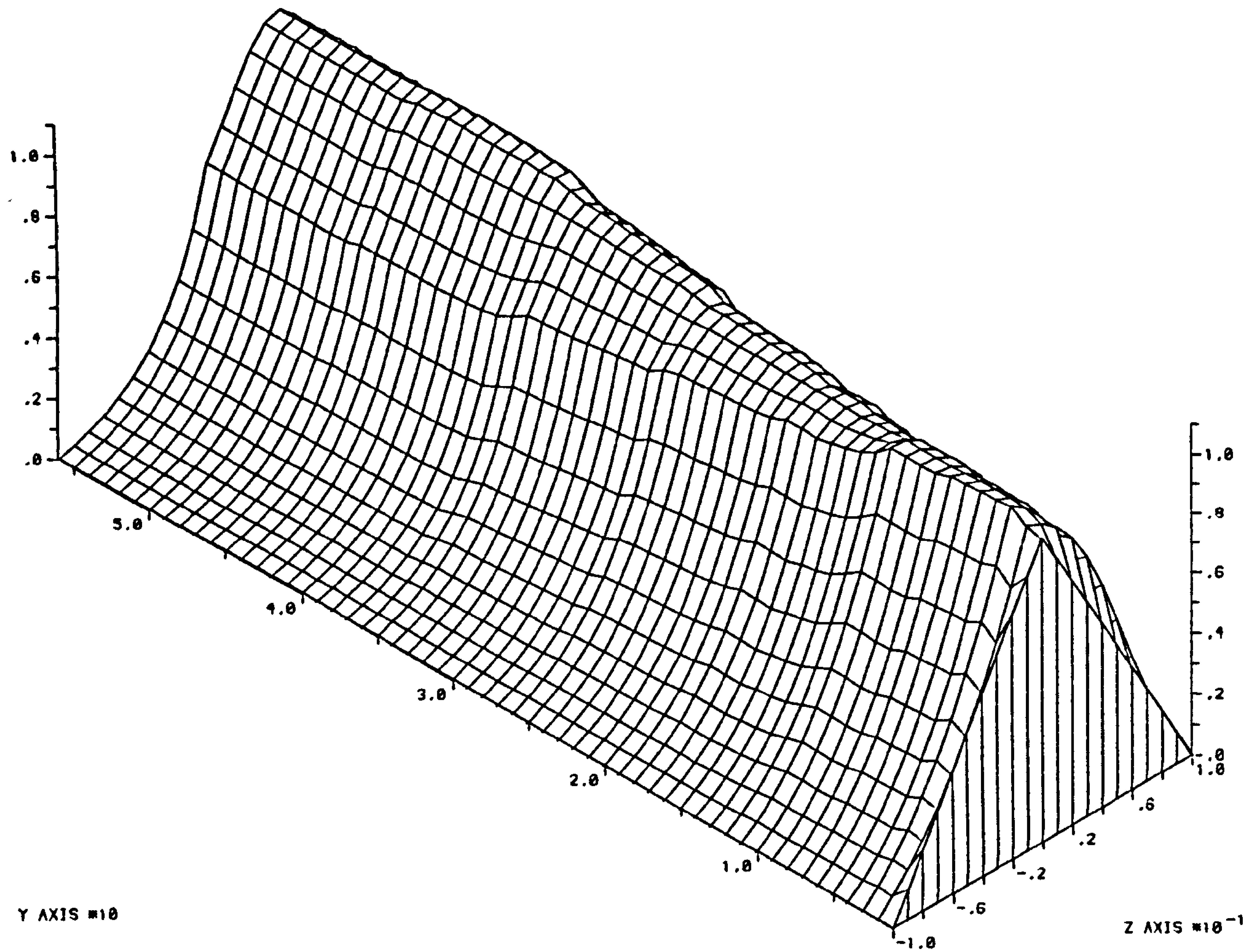


FIGURE 6.7

6.1.2 SINUSOIDAL MODEL

Many time series of clinical indicators measured on human beings exhibit approximately twenty-four-hour rhythmicity (see, for instance, Minors and Waterhouse 1981). Although the study of these 'circadian rhythms' is a growing area of research, most practising clinicians are not yet convinced of the practical advantage that might be gained by taking such information into account, even though some studies have shown that the timing of drug intake (within a 24 hour time-span) may have substantial effects on disease prognosis and drug toxicity (see, for instance, Kowanko et al. 1980, Hrushesky et al. 1982). There is clearly an additional cost involved in obtaining more than one measurement per day (both in terms of resource allocation and, perhaps, patient inconvenience), and it is therefore very rare for series with many measurements per day to be collected in routine clinical care at the present time.

In order to illustrate the sinusoidal model, we shall examine series arising from two very different clinical research studies (one involving animals, the other involving humans) but we emphasize that this type of model may be suited to more and more diverse applications as the clinical significance of human rhythmicity becomes more apparent.

6.1.2.1 Urinary Rhythms

Circadian rhythms associated with urine production and its constituents have been investigated by several authors (see, for instance, Buchsbaum and Harris 1971). Perhaps more importantly, however, a number of studies have suggested that certain drugs and/or diseases may induce abnormal rhythms or destroy some rhythms altogether (see, for instance, Hillier, Knapp and Cove-Smith 1980). The following example concerns a sample data set obtained from an experiment performed on a group of rats, in order to determine whether or not their urinary rhythms were disturbed by the introduction of steroidal drugs (Kowanko 1982).

During the experiment the rats were kept under a constant 12:12 light/dark cycle (i.e. 06.00 to 18.00 Light, 18.00 to 06.00 Dark) for a period of about twelve days. The rats were untreated for the first week of the experiment but, on the eighth day, the steroidal drug, dexamethasone, was introduced. Urine collections were made (mechanically) every four hours throughout the duration of the experiment, and a number of urinary variables were measured. For our illustration, we examine the resulting time series of urinary flow (i.e. volume/4 hours) obtained from one of the rats. Due to the expected rhythmicity of this variable, we adopted the sinusoidal model (described in Sections 2.3.3, 3.3.3 and 4.3.3) with a fixed periodicity of 24 hours. Figure 6.8 displays the series along with output from the analysis, for which 4 hours is equivalent to one time unit. A positive signal ($O_{T_k} > 0.2$) occurs at one timepoint only, $T_k = 44$, and at this point the system is unsure as to whether there has been a change in level ($O_{44}^{(2)} = 0.316$) or a change in amplitude ($O_{44}^{(3)} = 0.415$). In other words, a discontinuity starting at $T_k = 43$ (the first value on the eighth day) is suspected, suggesting that the introduction of dexamethasone had an immediate disturbing effect on the urinary flow rhythm for this particular rat.

Examination of the recursive estimates of rhythm level and amplitude, provided by the analysis, showed that the mean level of the series rose from 1.16 to 1.42 on the introduction of the

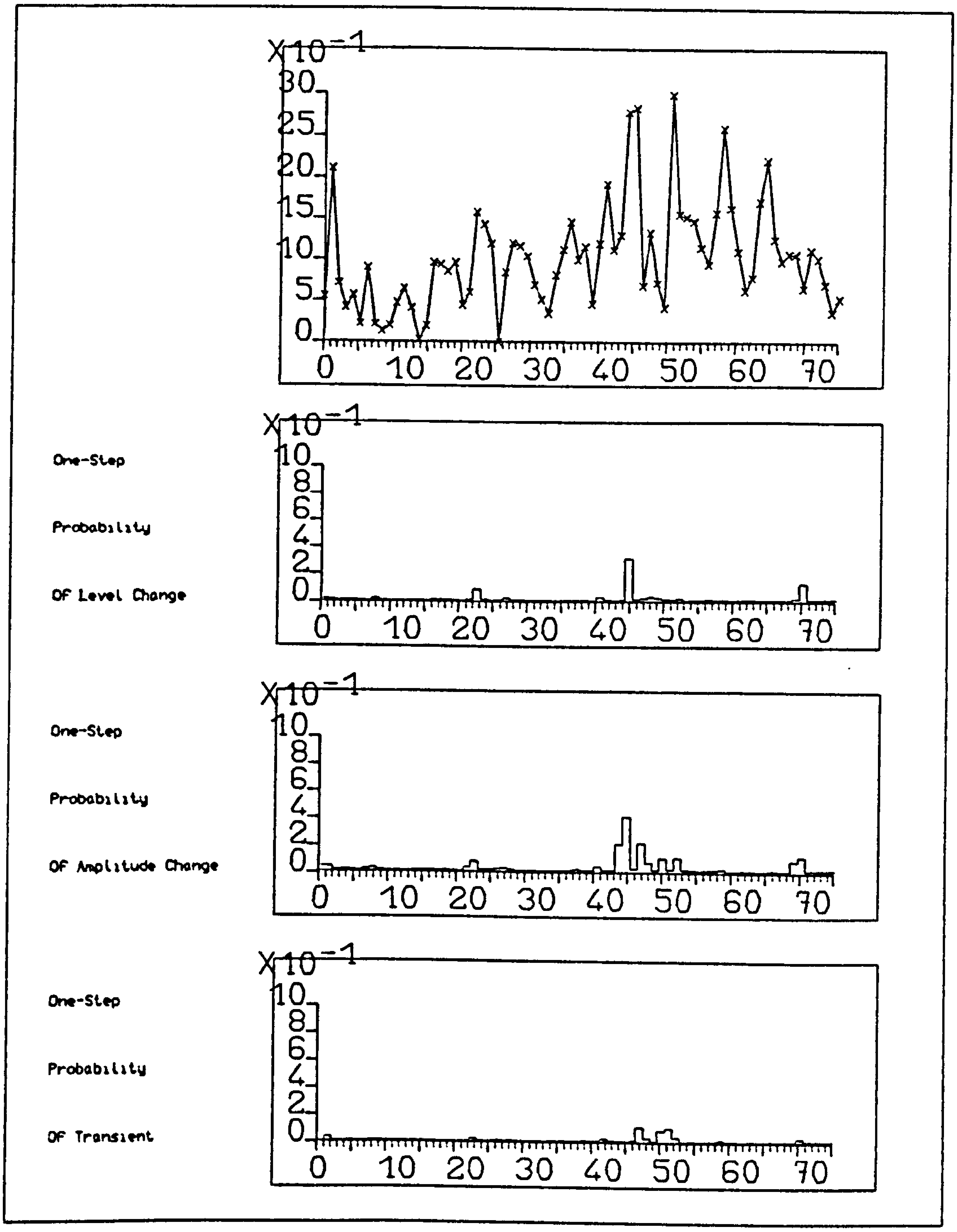

X10^-1
30
25
20
15
10
5
0
0 10 20 30 40 50 60 70
One-Step
Probability
OF Level Change
X10^-1
10
8
6
4
2
0
0 10 20 30 40 50 60 70
One-Step
Probability
OF Amplitude Change
X10^-1
10
8
6
4
2
0
0 10 20 30 40 50 60 70
One-Step
Probability
OF Transient
X10^-1
10
8
6
4
2
0
0 10 20 30 40 50 60 70

FIGURE 6.8

steroid, and that the rhythm amplitude, in fact, almost doubled from 0.36 to 0.68. This analysis has proved to be useful, therefore, in not only quantifying the magnitude of changes to the rhythm characteristics, but also in identifying the existence of a discontinuity in rhythm when it was very difficult to detect this by eye.

6.1.2.2 Respiration Studies

The next data set consists of 4-hourly measurements (06.00, 10.00, etc.) of peak expiratory flow rate (PEFR, a measure of airway capacity) made on an asthmatic patient who was part of a trial to examine the effects of various asthmatic drugs and their administration. As the trial subjects were human, it was decided to minimize patient inconvenience by omitting to take measurements at 02.00 hours, thereby producing a cyclic sampling pattern. The patients were monitored for several weeks and the underlying drug regimen was changed on a number of occasions during this period.

It has been shown that many respiratory patterns in asthmatics, including that of PEFR, exhibit circadian rhythmicity, and that the airway capacity is often lowest at night, at which time there is the greatest risk of respiratory difficulties (Hetzel and Clark 1979). For this reason we chose once more, the sinusoidal model in an attempt to reflect the rhythmic characteristics of the data. The series, along with one-step-ahead forecasts, and the results from the multistate analysis can be seen in Figure 6.9 (1 unit = 1 hour). Changes in the level of the series have been signalled at T_k = 150, 226 and 330, and although these signals correlated very closely with changes to the drug regime, it is

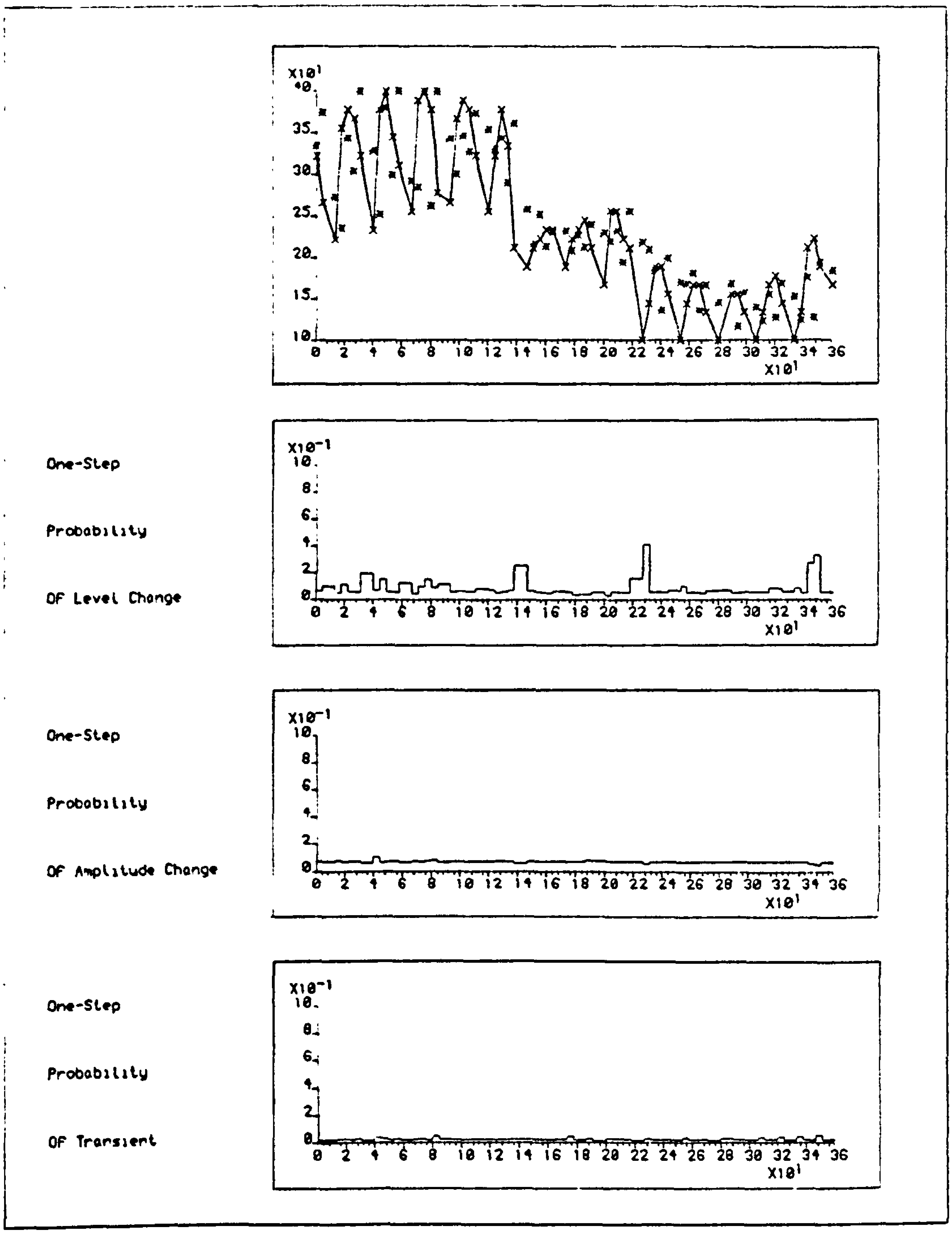

X10¹
40.
35
30
25
20
15
10
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
X10¹
One-Step
Probability
Of Level Change
X10⁻¹
10.
8
6
4
2
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
X10¹
One-Step
Probability
Of Amplitude Change
X10⁻¹
10
8
6
4
2
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
X10¹
One-Step
Probability
Of Transient
X10⁻¹
10.
8
6
4
2
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
X10¹

FIGURE 6.9

readily apparent that no amplitude change has been signalled (even though a change in the amplitude seems perfectly obvious from an 'eyeball' of the time series) and that, in general, the forecasts are poor. We might explain this by suggesting that there were simultaneous changes in level and amplitude and that our simple multistate structure was unable to handle this type of phenomenon. By allowing for an additional 'combined' state, perhaps, we might have been able to model the situation more closely (see Section 3.3.1).

On the other hand, examination of the recursively calculated estimate of the rhythm amplitude reveals that the estimate dropped from 50.6 before the first changepoint to 25.9 afterwards (suggesting that there had been a change in the amplitude of the PEFR rhythm). For the second changepoint, however, this estimate changed from 24.1 to 27.1, suggesting that this alteration in the drug regimen had little effect on the rhythm amplitude.

6.1.3 AR(1) MODEL

6.1.3.1 Long-Term Dialysis

Renal dialysis is a common form of treatment for patients who have serious kidney dysfunction, and is sometimes chosen as an alternative to kidney transplantation. Conversely, some patients who have received a transplanted kidney are not only initially supported by dialysis but, if their kidney eventually fails, they may need to revert to full-time dialysis therapy. It is therefore not uncommon for a patient to undergo long-term treatment by dialysis in which case, since the condition is chronic rather than acute, the patient will be investigated clinically rather infrequently. Successive clinic appointments may be several months apart unless a deterioration in condition occurs, in which case the sampling frequency might be increased. Sequences of data, arising from these clinic visits, will therefore tend to be very unequally-spaced.

The example considered here consists of a sequence of WBC counts, obtained over a period of just over a year, from a patient on long-term dialysis. From the point of view of clinical care, one is anxious to ensure that the WBC count does not fall to too low a level, since this would result in a weakening of the body's defence mechanisms with regard to infections. One is often in the position where downward trends in WBC count need to be watched very closely, but these trends may be 'confused' by the introduction of drugs (especially steroids) which are sometimes required to treat other symptoms associated with poor renal function. These drugs often have the effect of producing transient increases in the WBC count and, although these effects are anticipated, they still need to be properly considered since they might otherwise give a false impression of the underlying trend.

Retrospective examination of twelve renal patients who were considered, clinically, to have stable WBC counts indicated that half of the WBC series were indistinguishable from white noise, whereas the remaining series contained non-negligible first-order autocorrelations. For this reasons, we chose to adopt the AR(1) model (as described in Sections 2.3.5.1, 3.3.5.1 and 4.3.5.1) for the multistate analysis, for which we set

$$\underset{\sim}{m}_o = \begin{pmatrix} 5000 \\ 5000 \end{pmatrix} \qquad \text{and} \qquad \underset{\sim}{C}_o = \begin{pmatrix} 25000 & 0 \\ 0 & 25000 \end{pmatrix}$$

The example data series, along with results from the multi-state analysis, can be seen in Figure 6.10, where the timescale is given in days. Notice that three impulses were detected, each coinciding with the introduction of steroids. These events are, perhaps, obvious even by eye, but it is emphasized that:

(i) the detection of these changes was on-line, so that quick, automatic signals were available before the full data set was 'uncovered', and

(ii) the analysis has in each case distinguished these clinically-meaningful impulses from error-based transients, even though the data is, at times, exceedingly scarce and that the impulses themselves differ in length of duration.

Figure 6.11 shows on-line estimation of the grid for ϕ, the autoregressive parameter, which (after a hesitant beginning) soon displays reasonable confidence (considering the sparseness of the data) in a positive autocorrelation. Comparison with Figure 6.7 in Section 6.1.1.2 (roughly the same number of observations, but over a much condensed timescale) confirms our belief that the AR(1) model is more suited to a series of raw WBC counts than to a series of their logarithms.

6.1.3.2 Foetal Heart Monitoring

With the recent advent of sophisticated computerized foetal monitors, the study of foetal heart rates and other foetal parameters is of growing importance. Before the invention of these machines it was very difficult to obtain accurate information

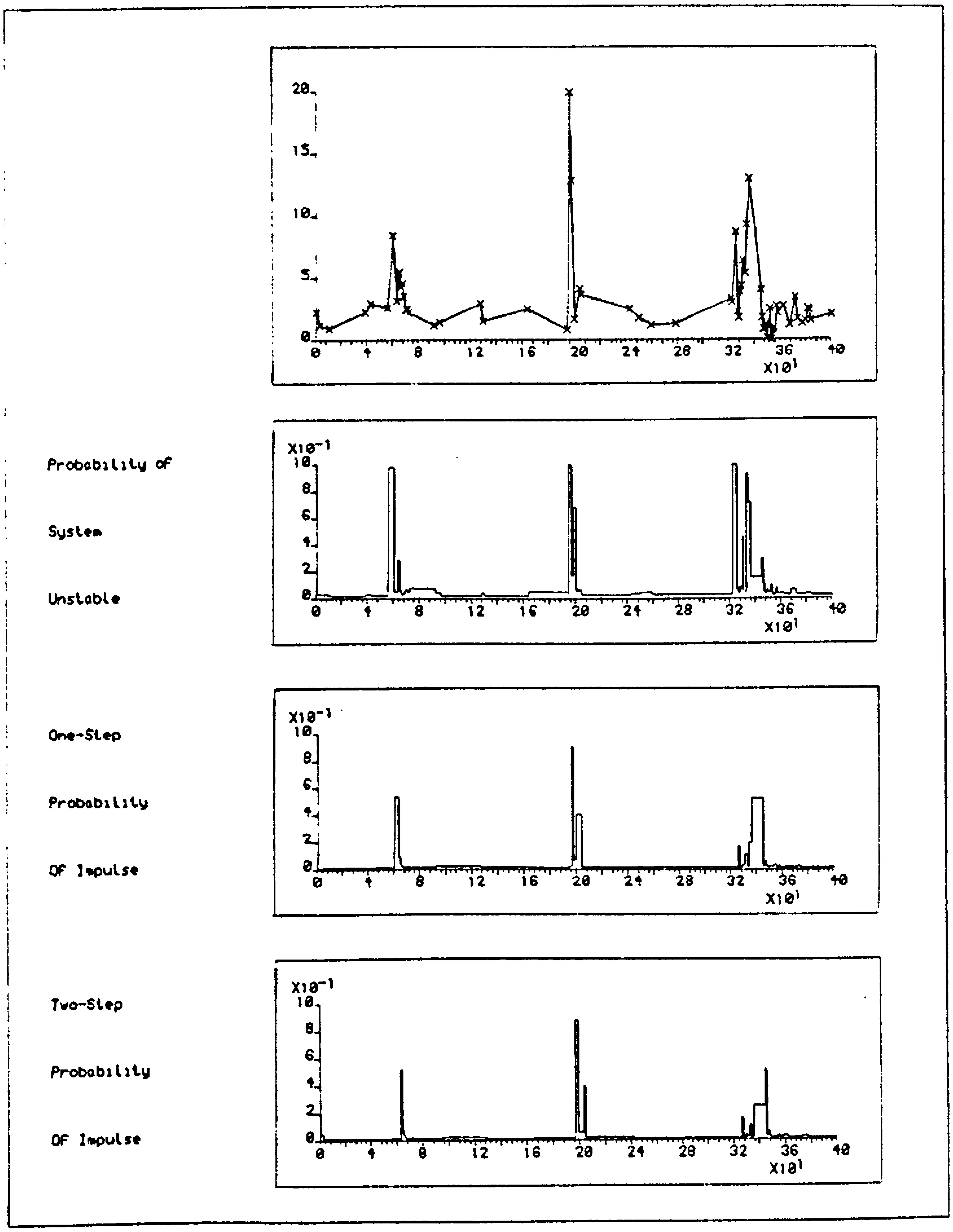

Probability of
System
Unstable
One-Step
Probability
Of Impulse
Two-Step
Probability
Of Impulse
X10^-1
X10^1

FIGURE 6.10

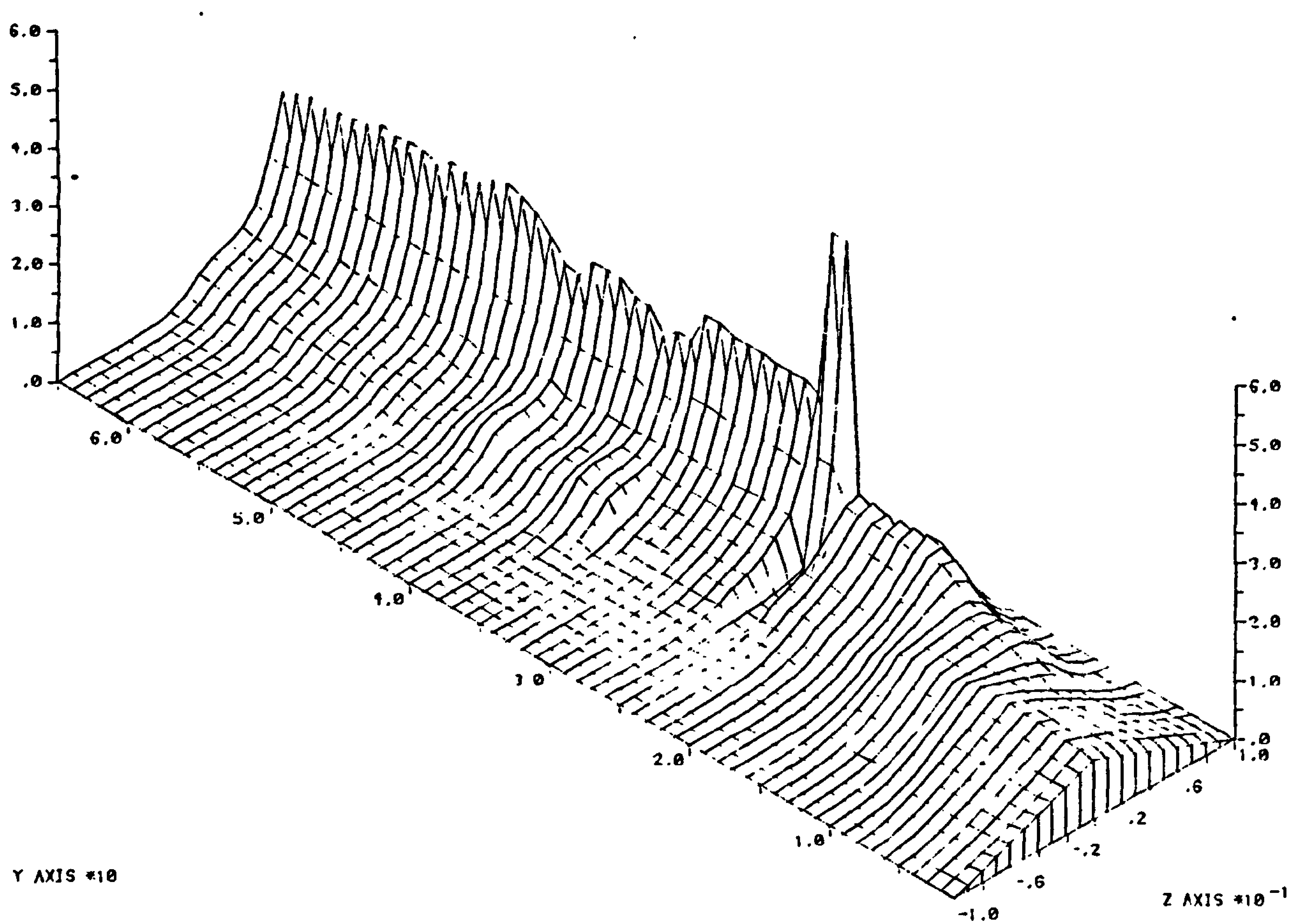


FIGURE 6.11

regarding the characteristics of the foetal heart, but the latest generation of monitors provide a tremendous quantity of accurate data. In fact, these multi-channelled devices can obtain data from, perhaps, a dozen alternative foetal parameters over a time scale of seconds, or even milliseconds. There is, of course, much debate as to which parameters are important, in terms of monitoring the well-being of the foetus and, because it has only recently become possible to collect such data in a routine manner, it is uncertain at this stage just what information is useful, and what is irrelevant.

What has become clear over the years, however, is that one beat-to-beat cycle of the (foetal) heart produces a heart-wave which has a consistent, but very complicated, 'shape' in the healthy foetus. The statistical or mathematical modelling of such a waveform is an horrendously complicated task, owing to the complexity of the waveform and the fact that it is not clear how this waveform changes in times of foetal distress; we make no attempt to approach this modelling problem here.

Obstetricians (and others involved in the examination of heart rates) tend to split up this overall waveform into components, or sub-waves, generally referred to as the P-wave, the Q-R-S complex and the T-wave. For the example given in this section, the data set was provided by Mr. H. Jenkins (formerly of the University Hospital, Nottingham) and consists of measurements of the duration of the S-T segment (the section of the wave starting at the end of the Q-R-S complex and ending at the mid-point of the T-wave) measured about a baseline of 1000 (the isoelectric line). These measurements were recorded every minute for about six hours (throughout the period of labour), and a graph of the resulting series can be seen in Figure 6.12.

In terms of choosing a model for this series, we have very little prior information to go on, since not enough is yet known about even the stable-state behaviour of such parameters. As in the previous section, however, similar data sets suggested that there may be some autocorrelation present and so, for simplicity, we attempted to use the AR(1) model in the multistate analysis. Figure 6.13 shows output from this analysis, along with one-step-ahead forecasts for the original series, and there are several points worth making.

Firstly, it should be apparent that in terms of tracking, as depicted by the one-step-ahead forecasts, the model seems to perform admirably even towards the end of the series when the pattern changes completely. There are also a number of change-point signals, sometimes indicating an impulse (e.g. $T_k = 180$) though, more commonly, transients are indicated ($T_k = 127, 181, 263$ and 275). These signals are associated with sharp downward dips (of differing severity and duration) which are each due to a single <u>contraction</u>.

The 'event' which begins at around $T_k = 249$ (as signalled by a level change), and which totally alters the pattern of the time series, seems to throw the event-detection analysis into total confusion. This 'crisis' is, in fact, a normal, and anticipated, event: the successful birth of the infant!

Figure 6.14 displays the on-line estimation of the ϕ-grid (autoregressive parameter). Notice that the impulse (at $T_k = 180$) sharpens the confidence in the estimate for ϕ. It is clear,

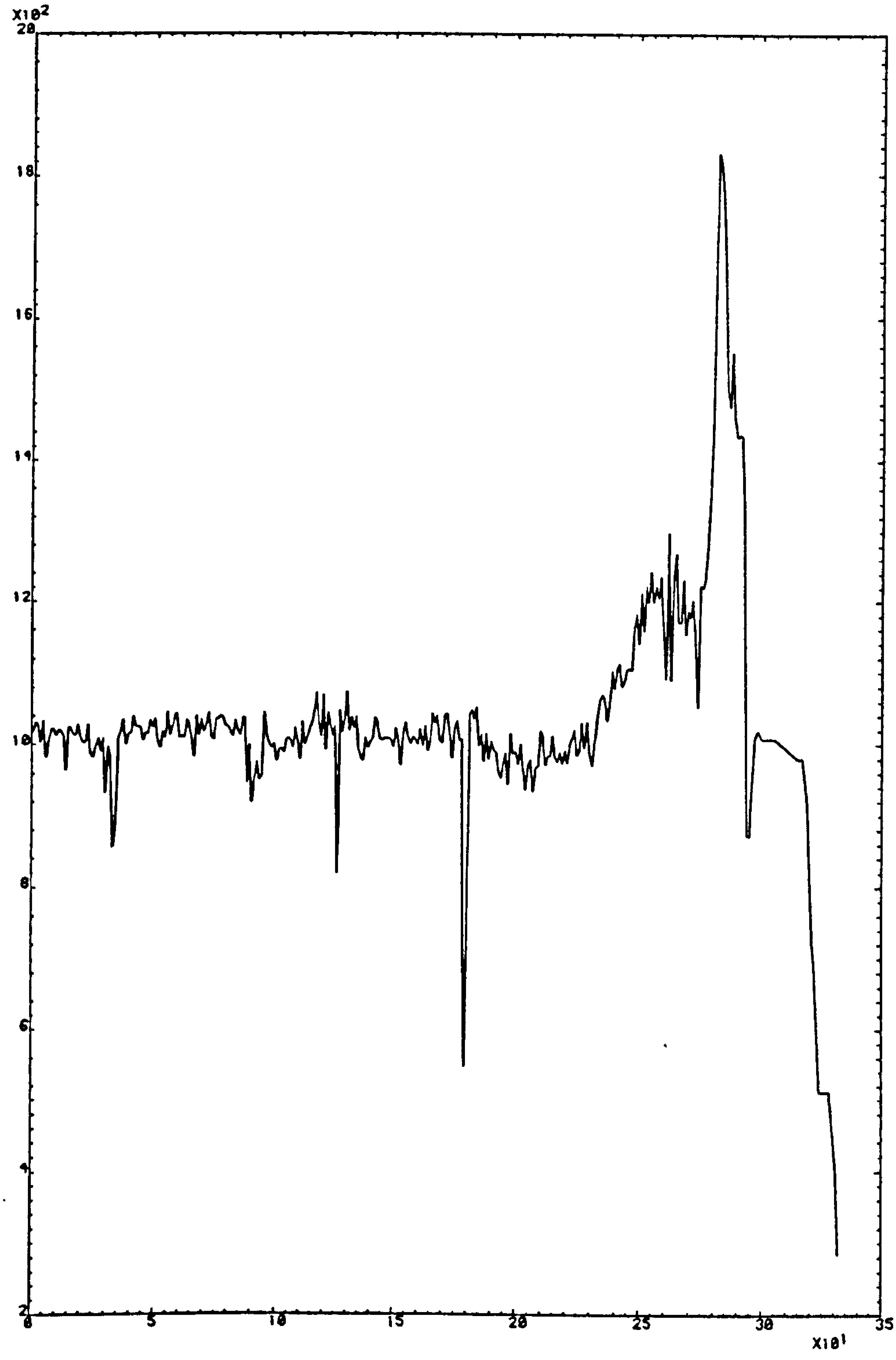
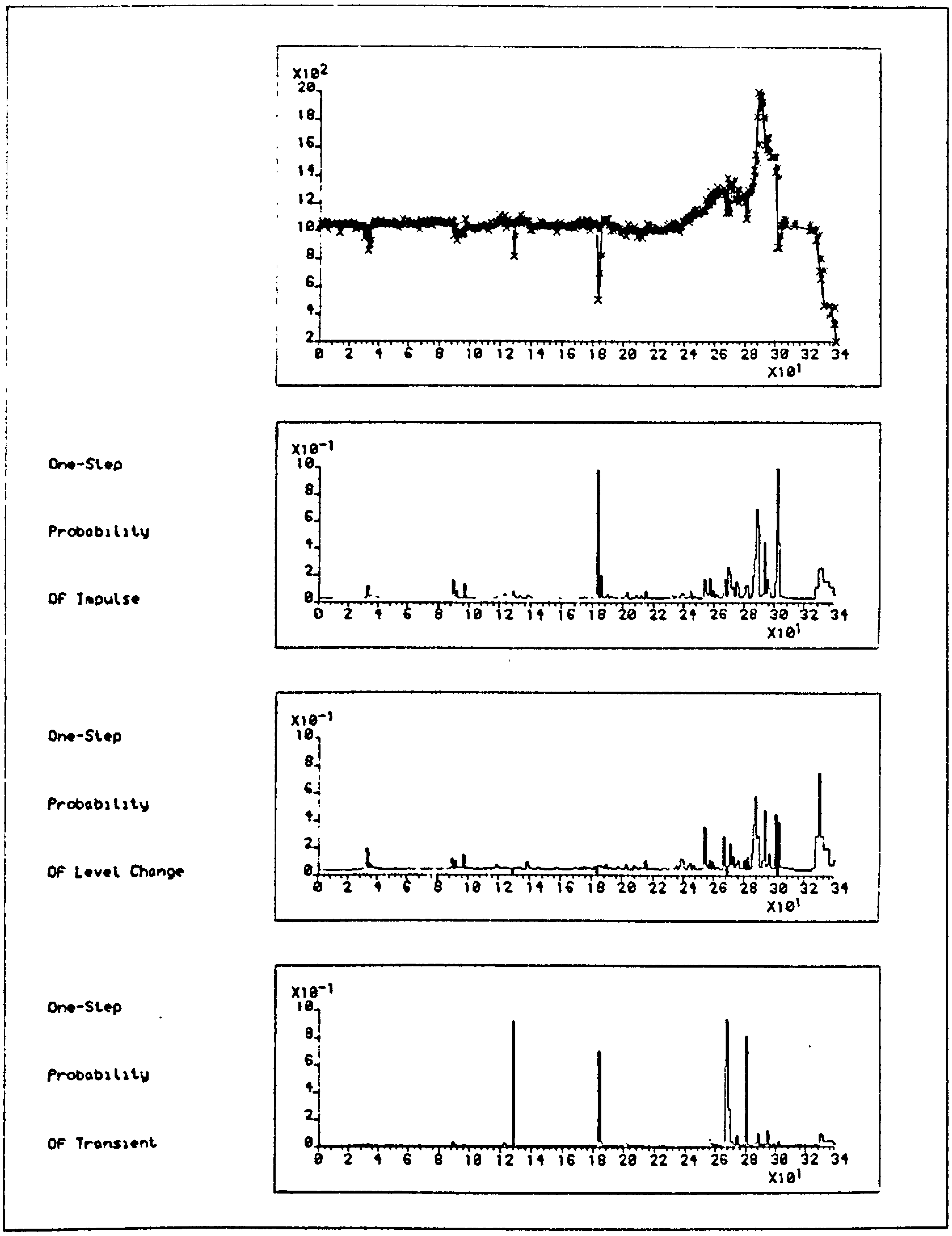


FIGURE 6.12



X10²
X10¹
One-Step
Probability
Of Impulse
X10⁻¹
One-Step
Probability
Of Level Change
X10⁻¹
One-Step
Probability
Of Transient
X10⁻¹

FIGURE 6.13

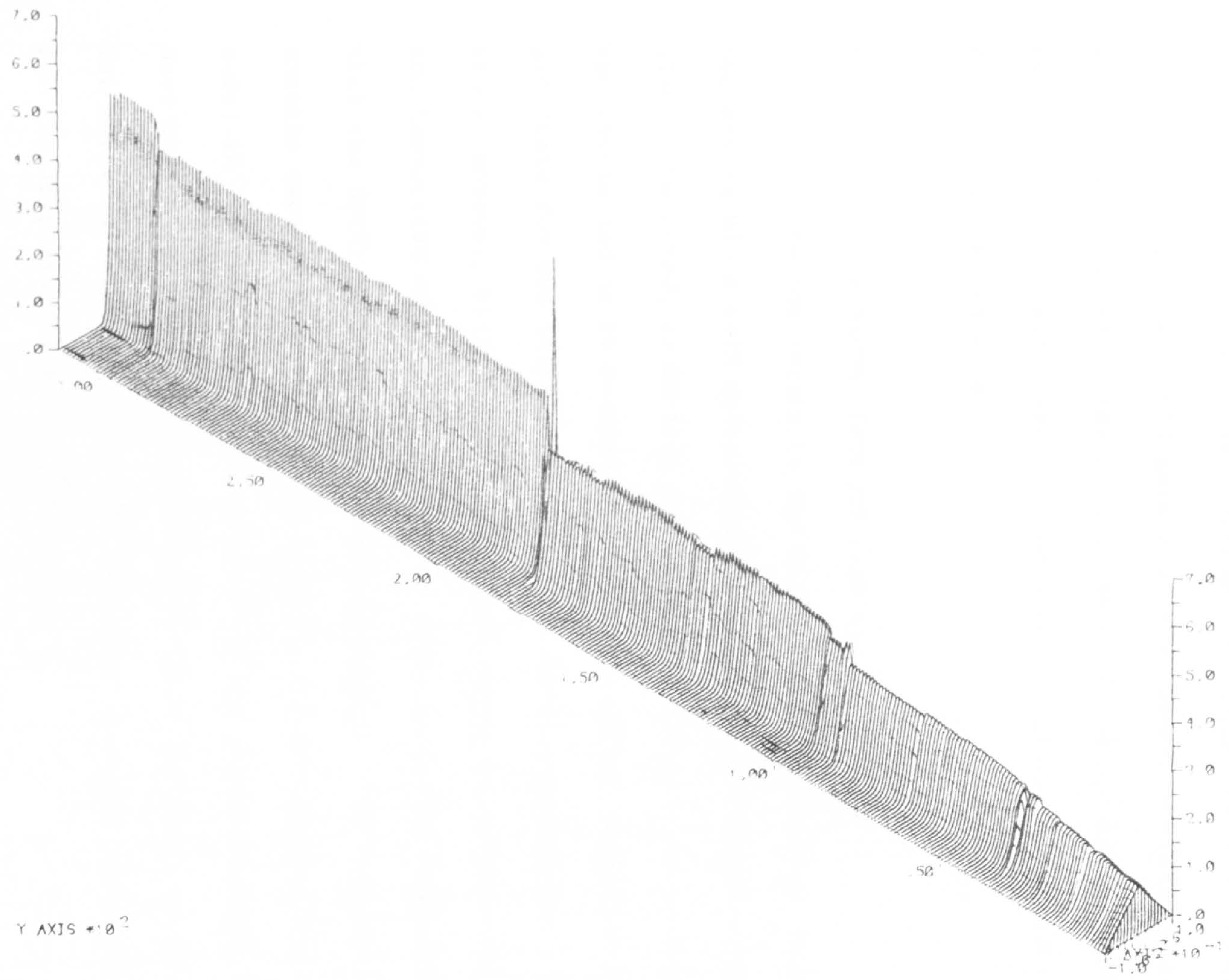


FIGURE 6.14

however, that a great deal of data was necessary to inspire any confidence in this parameter estimate, indicating that the AR(1) process is, perhaps, not the most suitable choice of model, and that we may be able to do better with some alternative model. We feel, though, that, given the accuracy of the forecasts and the ability to 'detect' contractions, the signs are very encouraging and that if we could gain a greater understanding of the mechanisms involved in the production of these complex waveforms, this methodology will have considerable potential in the area of foetal heart monitoring.

6.1.3.3 Bone Marrow Transplantation

We now return to the data sets given in Jones (1984) concerning blood cell measurements following a bone-marrow transplant. We noted, in Section 6.1.1.2, that our experience with WBC counts led us to believe that a linear growth model was appropriate for the ln(WBC) data set. We have no additional experience, however, with series of platelet counts or haematocrit, and inspection of the time series of these variables suggests that the AR(1) model may not be inappropriate. We therefore examine each of these series by adopting the multistate AR(1) model in the analysis. Recall that, for the platelet count, we have chosen to take the natural logarithm of the data, so that the scale of measurement differs from that given by Jones (1984). For the ln(platelet) series, we choose

$$\underset{\sim}{m}_o = \begin{pmatrix} 12.0 \\ 12.0 \end{pmatrix}, \qquad \underset{\sim}{C}_o = \begin{pmatrix} 10.0 & 0 \\ 0 & 10.0 \end{pmatrix},$$

while for the haematocrit series we set

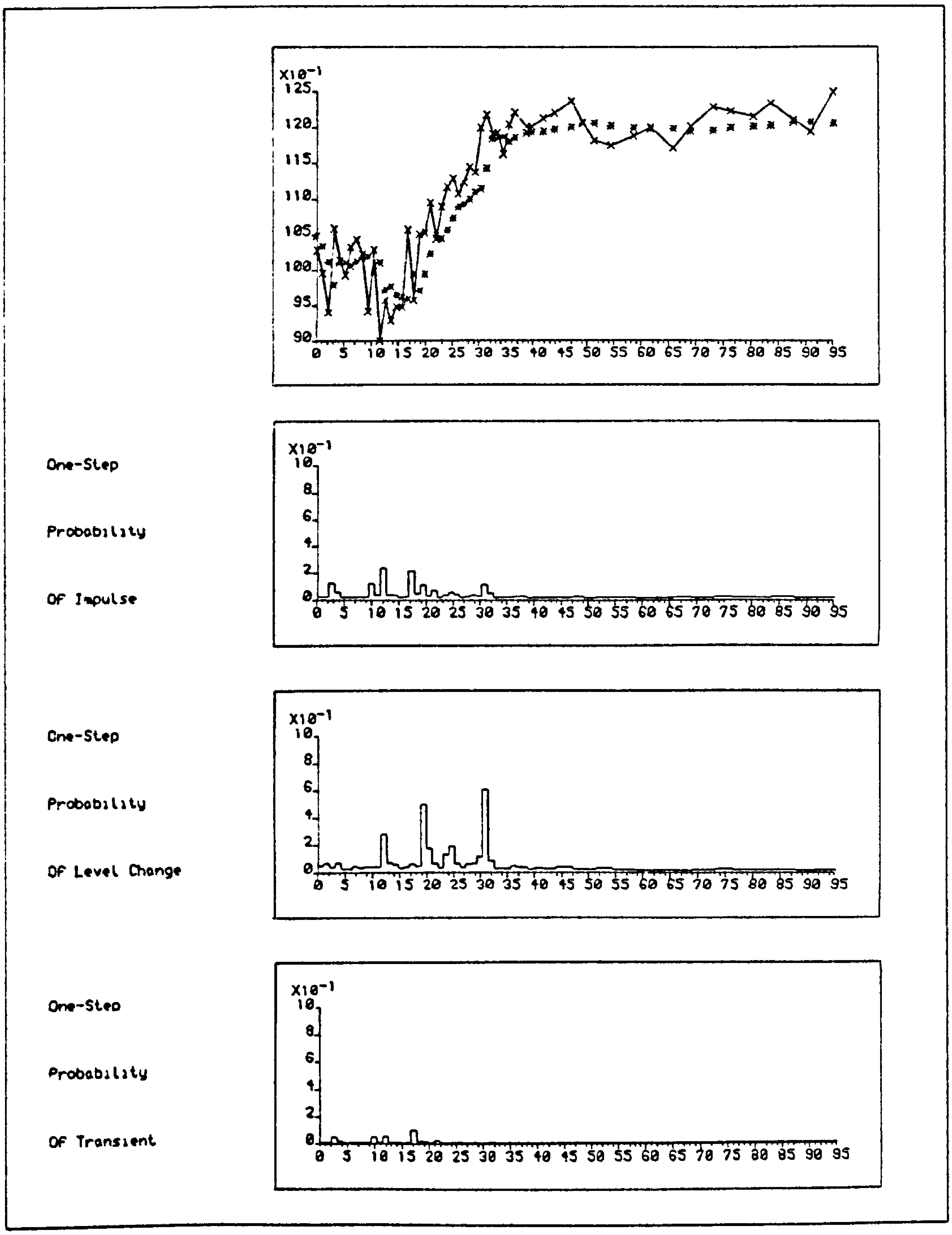
$$\underset{\sim}{m}_o = \begin{pmatrix} 30 \\ 30 \end{pmatrix}, \qquad \underset{\sim}{C}_o = \begin{pmatrix} 250 & 0 \\ 0 & 250 \end{pmatrix} .$$

The results from the univariate multistate analyses for each of these series, along with one-step-ahead forecasts, can be seen in Figures 6.15 and 6.16. Figure 6.15, showing the platelet series, indicates that level changes ($0_{T_k}^{(3)} > 0.2$) were suspected at $T_k = 13$, 20 and 31. Figure 6.16, showing the haematocrit series, indicates an impulse at $T_k = 27$, with level changes at $T_k = 43$ and 67. The remaining signals ($T_k = 78$ and 85) seem to be a feature of the sparseness of the data.

The drop in the level of haematocrit at $T_k = 43$ confirms our belief that an untoward event, possibly rejection, occurred at around this time (see Section 6.1.1.2), although the haematocrit series does not reflect the initial delay in the onset of treatment effectiveness. This delay is, perhaps, reflected by the ln(platelet) series since a level change was indicated at $T_k = 13$; the second level change signalled in this series ($T_k = 20$) may well be associated with the decrease in slope signalled in the ln(WBC) series at $T_k = 16$.

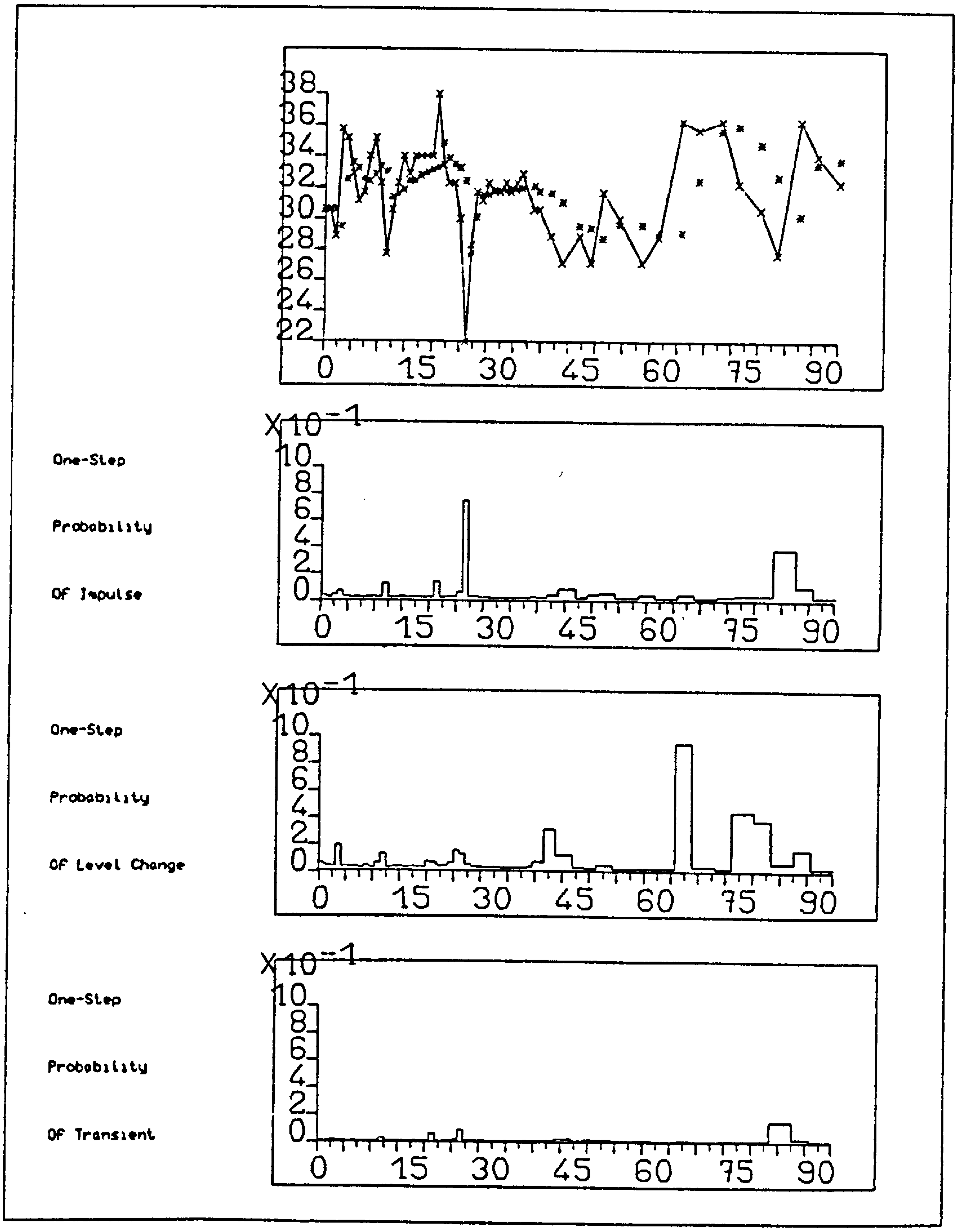
If the changepoints that have been signalled by these analyses are genuine, we conclude that, of the indicators examined, ln(WBC) provides the clearest view of the post-transplant course, although haematocrit may be just as useful in the early detection of acute deterioration.

N.B. For the ln(platelet) series we obtained SSFE = 12.0, MAD = 0.38 and for the haematocrit series we obtained SSFE = 479.6, MAD = 2.26.



X10^-1
125
120
115
110
105
100
95
90
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
One-Step
Probability
Of Impulse
X10^-1
10
8
6
4
2
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
One-Step
Probability
Of Level Change
X10^-1
10
8
6
4
2
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
One-Step
Probability
Of Transient
X10^-1
10
8
6
4
2
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95

FIGURE 6.15



38
36
34
32
30
28
26
24
22
0
15
30
45
60
75
90
X10^-1
One-Step
Probability
Of Impulse
10
8
6
4
2
0
One-Step
Probability
Of Level Change
One-Step
Probability
Of Transient

FIGURE 6.16

6.2 BIVARIATE EXAMPLES

6.2.1 BIVARIATE LINEAR GROWTH MODEL

In order to illustrate the use of the bivariate linear growth model, we return to the setting of kidney transplantation. Besides creatinine, there are many other indicators of kidney function, some of them involving alternative blood measurements, others involving the execution of certain clinical tests (e.g. a biopsy). One of the other most common blood chemicals that is used as a guide to the state of a patient's kidneys is the substance serum urea, and it was felt that by attempting to model the urea and creatinine series in a bivariate manner we might be able to further discriminate between those changes which were clinically significant and those changes which were not.

In the example which follows, the creatinine series is the first of those presented in Section 6.1.1.1 (see Figure 6.1), and the corresponding urea series (from the same patient over the same period of time) has been utilized. We note that, from earlier studies, a time series of the reciprocal of urea (corrected for weight) may also be modelled satisfactorily by the linear growth model in the setting of kidney transplantation (Knapp et al. 1977).

In terms of the interrelationships between the two series, we know very little. We feel (from physiological considerations) that there is unlikely to be steady-state causality (see Section 5.3.1.2) between the two variables, since a rise (say) in one of these chemicals does not directly produce a change in the other. However, since the two variables are both indicators of the same function (i.e. kidney function) it is likely that the two series will be highly correlated. Moreover, if a change in kidney function occurs, this will probably be reflected by pattern changes in each of the series, i.e. we need to incorporate the possibility of changepoint-causality (see Section 5.3.1.3). Using the notation from Section 5.3.2.1, we shall set $R_{\mu\mu} = R_{\beta\beta} = 0.1$ and we shall also set the variance ratio $var(\varepsilon_{1t})/var(\varepsilon_{2t}) = 0.5$ (where ε_{1t} is associated with the creatinine series, ε_{2t} the urea series), since this corresponds, approximately, to the adjusted ratio of coefficients of variation for the two chemicals, provided by the laboratory.

Using this prior information, Figure 6.17 shows the two series (creatinine the lower of the two), along with the probabilities of a dual positive slope change, i.e. a concurrent slope change in each of the series, which we shall (as in the univariate case) use as a guide to whether or not the kidney has rejected.

Using our previous criterion for a positive signal, we see that $0^{(11)}_{T_k} > 0.2$ on two occasions; when $T_k = 6$ and when $T_k = 16$. Notice that the first of these signals is one day earlier than the corresponding signal obtained from the univariate analysis of creatinine alone and, therefore, one day earlier than the clinician's reaction. Notice, too, that there are several occasions when only one of the two indicators has been measured, so that this is an example of a generally unequally-spaced bivariate time series.

6.2.2 AR(1)/LINEAR GROWTH MODEL

In order to illustrate the use of the AR(1)/linear growth model, we return, once more, to the setting of bone marrow

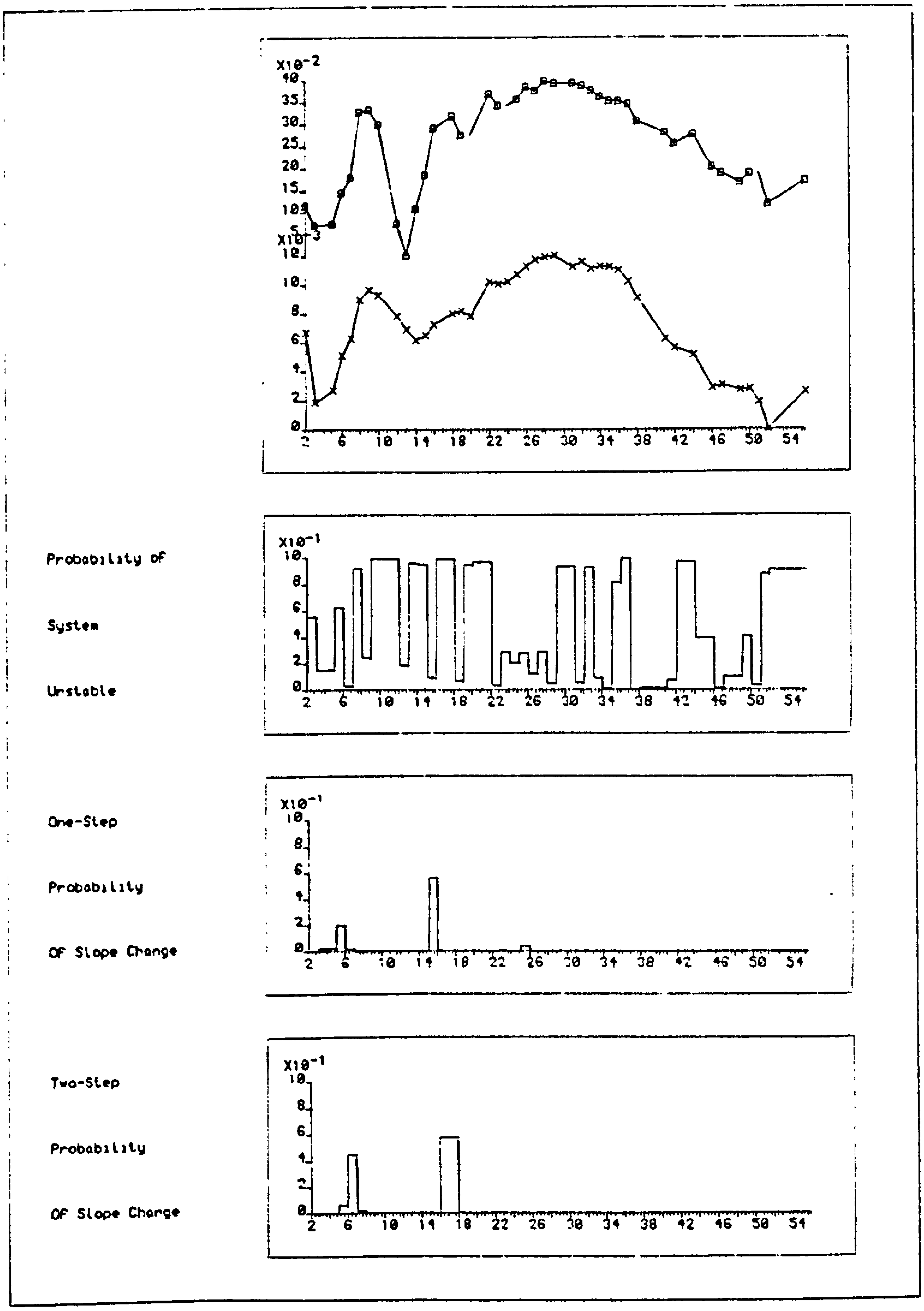

Probability of
System
Unstable
One-Step
Probability
Of Slope Change
Two-Step
Probability
Of Slope Change

FIGURE 6.17

transplantation, and the data sets given by Jones (1984). Recall (from Section 6.1.1.2) that we have assumed that the ln(WBC) can be modelled by the linear growth model and (from Section 6.1.3.3) that the ln(platelet) and haematocrit series can each be modelled by the AR(1) model. We shall therefore consider the analysis as applied to each of two bivariate time series, the first consisting of ln(platelet) and ln(WBC), and the second consisting of haematocrit and ln(WBC).

6.2.2.1 No Causality

Jones (1984) discovered that the 'best' trivariate model for the three indicators was one in which the WBC counts and platelet counts were not causally linked, either through the transition matrix or the system-perturbation covariance matrix. We have therefore assumed that a suitable bivariate model for ln(platelet) and ln(WBC) is the AR(1)/linear growth model without steady-state causality or changepoint-causality. This leads to a diagonalization of the various matrices involved (see Section 5.3.1), so that very little prior information is required in addition to that previously utilized in Sections 6.1.1.2 and 6.1.3.3. Because these two indicators are both affected by the effectiveness of the bone marrow transplant, we shall assume that coincident observations are correlated, by setting $R_{\varepsilon\varepsilon} = 0.1$ (see Section 5.3.2.2). In order to specify the ratio of observation variances, we note that Jones (1984) found that the variance associated with the log(platelet) series was nearly twice that associated with the log(WBC) series, and so we set the variance ratio accordingly.

Results from the bivariate analysis can be seen in Figure 6.18 (ln(platelet) the lower of the two series). Apart from the detection of the occasional transient in one or other of the series (but not both together - this 'event' could be of some importance), positive signals ($O_{T_k} > 0.2$) were obtained on several occasions. The level changes indicated by the univariate analysis for the ln(platelet) series (see Section 6.1.3.3) at $T_k = 20$ and $T_k = 31$ have each been signalled as the event 'Level change/Steady state', i.e. the ln(WBC) fails to confirm the suggested changes. Similarly, the slope change found in the ln(WBC) series (see Section 6.1.1.2) at $T_k = 16$ is not mirrored by a change in the ln(platelet) series, since at $T_k = 16$ the event signalled is 'Steady state/Slope change'.

There are two occasions, however, when the event 'Level change/Slope change' is signalled: at $T_k = 8$ and $T_k = 25$. The first of these seems to correspond to an early signal of the onset of transplant effectiveness, though the second signal appears to be a false positive. Notice that the (possible) event of most clinical interest (i.e. the deterioration at or around $T_k = 45$) has been missed altogether.

For this model we found SSFE = 12.0 and MAD = 0.37 for the ln(platelet) series, with SSFE = 18.7 and MAD = 0.43 for the ln(WBC) series. Examination of the corresponding values obtained in the univariate analyses (Sections 6.1.1.2 and 6.1.3.3) shows that the forecasting ability of the bivariate model is comparable to that attained by the univariate models, with perhaps slightly better performance achieved for the ln(WBC) series.

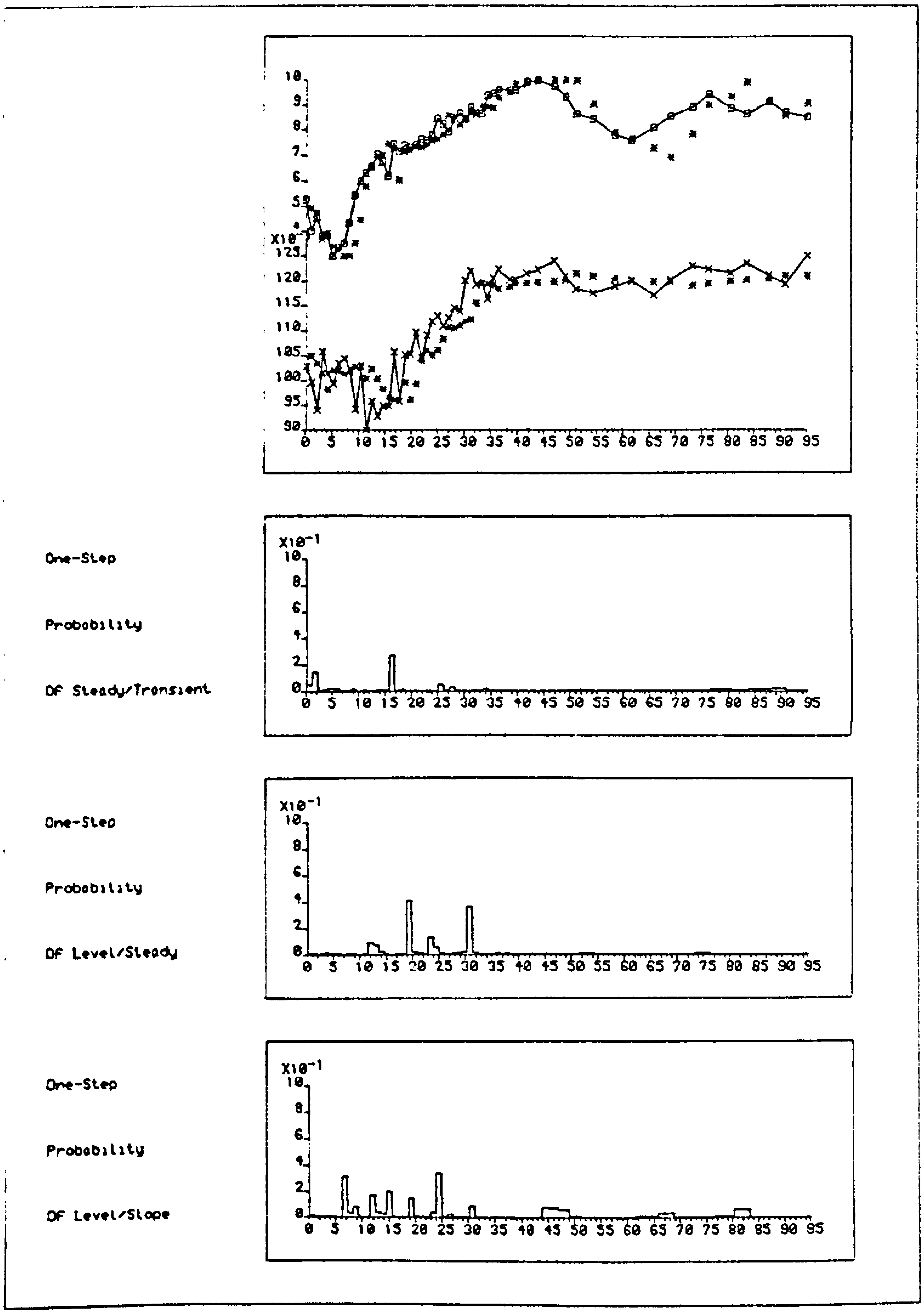
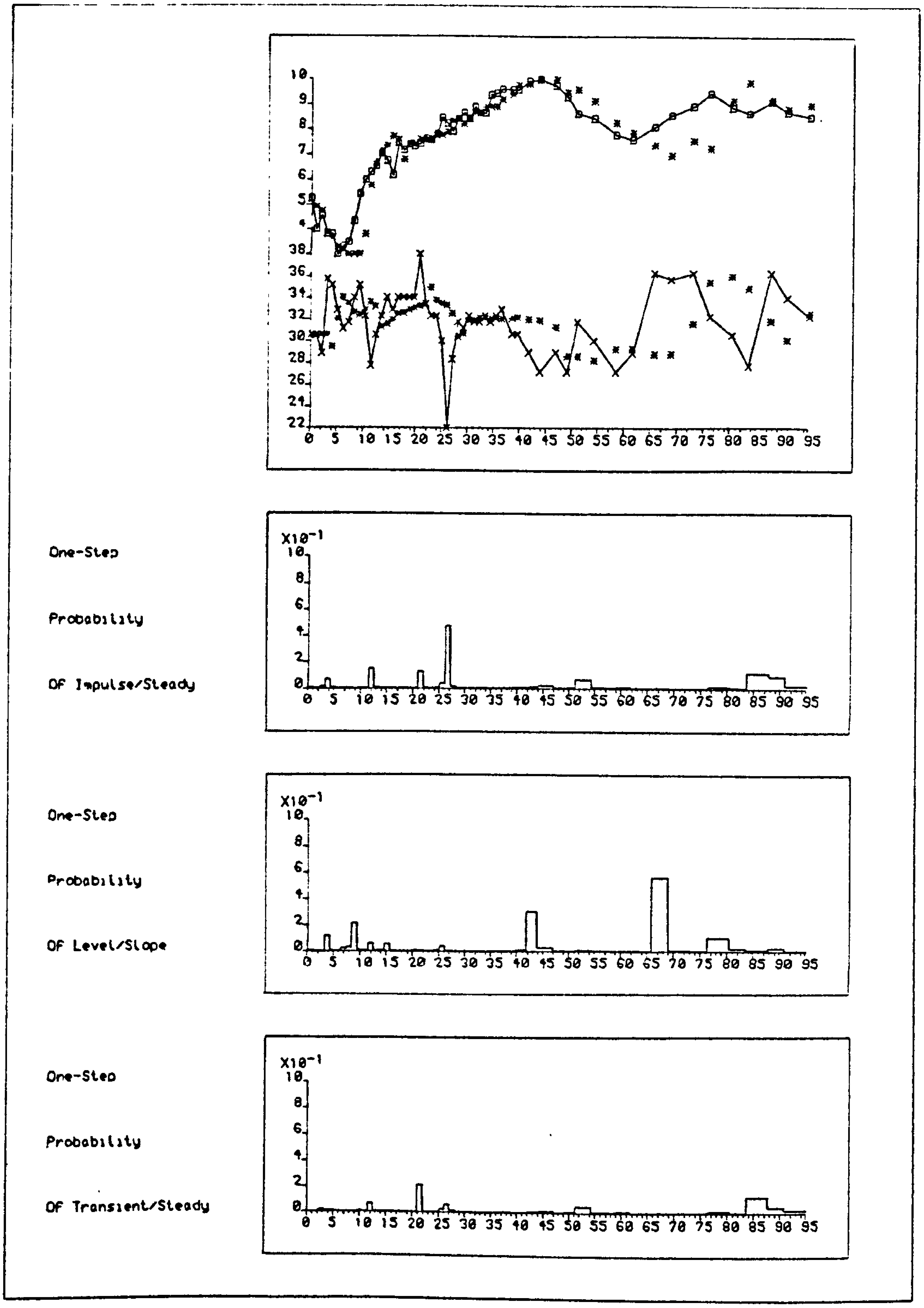

X10⁻¹
One-Step
Probability
Of Steady/Transient
One-Step
Probability
Of Level/Steady
One-Step
Probability
Of Level/Slope
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95

FIGURE 6.18

6.2.2.2 Unidirectional Causality

Here we examine a bivariate model for the sequences of haematocrit and ln(WBC) which, according to Jones (1984), showed some signs of causality in that it was suggested that a rise in the haematocrit level would be the direct cause of a rise in the ln(WBC) level, but not vice versa (so that we have a case of unidirectional causality). We tried to incorporate this type of behaviour into the model by setting $c = 0.025$ (using the notation of Section 5.3.2.2), thus allowing for steady-state unidirectional causality via the transition matrix, G. Although Jones (1984) found no causal links between the system perturbations, it was felt that we ought to allow for changepoint-causality by setting $R_{\varepsilon\varepsilon} = R_{\mu\mu} = 0.1$. A variance ratio of 1:10 was specified since Jones (1984) found that the observation error variance for the haematocrit series was negligible.

Results from the bivariate analysis are given in Figure 6.19 (the haematocrit series is the lower of the two) and we focus our attention, once more, upon the signals ($O_{T_k} > 0.2$) obtained. We see that we are again able to screen out transient observations in one or other of the series (e.g. $T_k = 22$), and that the impulse detected, at $T_k = 27$, by the univariate analysis for haematocrit (see Section 6.1.3.3) is now signalled as 'Impulse/Steady state', i.e. this event is not reflected by the ln(WBC) series. Of main interest is the fact that the event 'Level change/Slope change' (with which we are most concerned) has been signalled on three occasions: $T_k = 10$, 43 and 67. The first of these signals confirms our belief that the bone-marrow transplant needed around eight to nine days before it



One-Step
Probability
Of Impulse/Steady
X10-1
One-Step
Probability
Of Level/Slope
X10-1
One-Step
Probability
Of Transient/Steady
X10-1

FIGURE 6.19

began to take effect, whereas the second signal is further confirmation of an acute deterioration (possible rejection) at around day 41. The final signal (T_k = 67) seems to clarify those given by the univariate analyses, and may well represent the improvement due to 'treatment' of the earlier deterioration.

For this model, SSFE = 490.6, MAD = 2.24 for the haematocrit series and SSFE = 24.1, MAD = 0.44 for the ln(WBC) series, which are comparable to the values obtained from the univariate analyses (Sections 6.1.1.2 and 6.1.3.3).

Summary of Results from Bone Marrow Transplantation Example

From our own analyses of the data sets presented by Jones (1984) concerning blood cell counts following bone marrow transplantation, we tentatively conclude that:

(i) the transplant was ineffectual for the first week or so, at which point the treatment suddenly began to take effect;

(ii) there was a deterioration (possibly 'rejection') in the patient's condition beginning at around day 41, which seemed to be treated (or was self-corrected) around day 64;

(iii) the bivariate pairing of haematocrit and WBC appears to provide the earliest warning signs for both improvement and, more importantly, deterioration; the platelet count appears to be misleading in this latter respect.

Unfortunately, we do not have access to the original clinical record for this patient, in order to discover whether or not the changepoints we have detected relate to genuine clinical events.

Acknowledgements

It is convenient at this point to acknowledge those people who very kindly provided many of the time series that were presented in the previous chapter: the Cardiff Renal Transplant Team, the UKTS Management Committee, Dr. Inge Kowanko, Mr. Howard Jenkins and, last but not least, the patients themselves who, for obvious reasons, remain nameless.

oOo

APPENDIX SIX

DATA SETS

A6.1 RENAL TRANSPLANTATION

Cardiff Data

DAY	WEIGHT	CREATININE	UREA
2.0	49.8	221.0	4.0
3.0	50.0	126.0	3.5
5.0	52.0	127.0	3.3
6.0	52.7	161.0	4.0
7.0	52.9	186.0	4.5
8.0		303.0	10.5
9.0	54.3	343.0	10.5
10.0	54.2	312.0	8.0
12.0	55.0	224.0	3.0
13.0		192.0	2.5
14.0		171.0	3.3
15.0	50.3	209.0	5.0
16.0	48.8	248.0	9.0
18.0	49.0	281.0	11.0
19.0	49.2	288.0	8.0
20.0	51.3	251.0	
22.0	52.0	432.0	17.0
23.0	51.6	418.0	12.5
24.0	50.7	451.0	
25.0	51.9	519.0	14.5
26.0	53.9	623.0	21.0
27.0	55.0	788.0	17.5

DAY	WEIGHT	CREATININE	UREA
28.0	55.2	882.0	27.0
29.0	55.2	954.0	24.0
31.0	55.5	584.0	24.0
32.0	57.6	666.0	20.0
33.0	50.8	641.0	20.0
34.0	51.5	673.0	16.0
35.0	51.8	655.0	14.0
36.0	53.7	560.0	13.0
37.0	54.0	413.0	12.0
38.0		300.0	8.5
41.0	52.3	189.0	7.5
42.0	52.1	175.0	6.5
44.0	51.5	168.0	7.5
46.0	47.2	153.0	6.0
47.0	46.6	159.0	5.8
49.0	46.0	158.0	5.5
50.0	45.7	161.0	6.0
51.0	44.6	155.0	
52.0	44.3	135.0	5.0
56.0	42.4	183.0	6.5

DAY	WEIGHT	CREATININE
1.0	65.4	1231.0
2.0	67.7	908.0
3.0	68.4	484.0
4.0	68.1	284.0
5.0	66.7	244.0
6.0	67.5	202.0
7.0	68.9	176.0
8.0	70.0	216.0
9.0	68.5	272.0
10.0	68.6	287.0
11.0	69.4	308.0
12.0	69.8	299.0
13.0	69.0	292.0
14.0	68.6	240.0
15.0	69.1	215.0
16.0	68.5	189.0
17.0	69.0	186.0
19.0	68.7	177.0
21.0	67.7	162.0
24.0	68.1	162.0
26.0	68.1	160.0
28.0		163.0
31.0		175.0
33.0	68.6	159.0
35.0	69.4	166.0
38.0		168.0
40.0	69.4	169.0
42.0		166.0
45.0		159.0
47.0	69.0	148.0
49.0		141.0
52.0		157.0
54.0	69.6	160.0

DAY	WEIGHT	CREATININE
56.0		163.0
61.0	71.0	161.0
68.0	68.3	161.0
75.0	71.7	155.0
82.0		163.0
89.0	74.9	180.0
96.0	73.2	204.0
97.0		215.0
98.0		215.0
99.0		200.0
103.0	73.6	211.0
104.0	74.5	229.0
105.0	75.8	223.0
106.0	74.9	215.0
107.0	77.2	200.0
108.0	77.2	199.0
109.0	73.9	240.0
110.0	72.3	282.0
111.0	74.5	306.0
112.0	74.2	327.0
113.0	73.2	287.0
114.0	72.2	270.0
115.0	72.6	274.0
116.0	72.8	261.0
117.0	74.0	228.0
118.0	73.3	232.0
119.0	73.5	218.0
122.0		220.0
124.0		229.0
131.0	73.4	255.0
138.0	72.2	304.0
145.0	71.2	268.0

UKTS Pilot Study Data

DAY	WEIGHT	CREATININE
0.0		623.0
1.0	53.8	371.0
2.0	52.2	333.0
3.0	49.6	264.0
4.0	50.4	261.0
5.0	51.0	384.0
6.0	50.5	357.0
7.0	51.0	282.0
8.0		260.0
9.0		273.0
10.0		233.0
11.0		202.0
12.0		176.0
13.0	52.4	165.0
15.0		159.0
18.0	51.8	200.0
22.0	51.1	211.0
29.0		209.0
32.0		189.0

A6.2 URINARY FLOW

TIME	FLOW	TIME	FLOW
1.00	0.77	26.00	1.03
2.00	2.18	27.00	1.36
3.00	0.91	28.00	1.33
4.00	0.65	29.00	1.21
5.00	0.78	30.00	0.89
6.00	0.47	31.00	0.74
7.00	1.09	32.00	0.59
8.00	0.46	33.00	1.01
9.00	0.39	34.00	1.29
10.00	0.45	35.00	1.59
11.00	0.70	36.00	1.17
12.00	0.85	37.00	1.32
13.00	0.64	38.00	0.69
14.00	0.29	39.00	1.36
15.00	0.45	40.00	2.01
16.00	1.14	41.00	1.28
17.00	1.12	42.00	1.45
18.00	1.05	43.00	2.79
19.00	1.14	44.00	2.83
20.00	0.67	45.00	0.89
21.00	0.81	46.00	1.48
22.00	1.69	47.00	0.92
23.00	1.55	48.00	0.66
24.00	1.35	49.00	2.99
25.00	0.27	50.00	1.67

TIME	FLOW	TIME	FLOW
51.00	1.65	62.00	2.27
52.00	1.61	63.00	1.40
53.00	1.32	64.00	1.16
54.00	1.14	65.00	1.24
55.00	1.70	66.00	1.24
56.00	2.62	67.00	0.87
57.00	1.75	68.00	1.29
58.00	1.27	69.00	1.19
59.00	0.84	70.00	0.90
60.00	0.99	71.00	0.60
61.00	1.82	72.00	0.75

A6.3 RESPIRATION

TIME (HOURS)	PEFR	TIME (HOURS)	PEFR	TIME (HOURS)	PEFR
18.00	320.00	130.00	320.00	234.00	200.00
22.00	270.00	134.00	370.00	238.00	170.00
30.00	230.00	138.00	330.00	246.00	120.00
34.00	350.00	142.00	220.00	250.00	160.00
38.00	370.00	150.00	200.00	254.00	180.00
42.00	360.00	154.00	220.00	258.00	180.00
46.00	320.00	158.00	230.00	262.00	150.00
54.00	240.00	162.00	240.00	270.00	120.00
58.00	370.00	166.00	240.00	278.00	170.00
62.00	390.00	174.00	200.00	282.00	170.00
66.00	340.00	178.00	230.00	286.00	150.00
70.00	310.00	182.00	240.00	294.00	120.00
78.00	260.00	186.00	250.00	298.00	150.00
82.00	380.00	190.00	220.00	302.00	180.00
86.00	390.00	198.00	180.00	306.00	190.00
90.00	370.00	202.00	260.00	310.00	160.00
94.00	280.00	206.00	260.00	318.00	120.00
102.00	270.00	210.00	230.00	322.00	150.00
106.00	360.00	214.00	220.00	326.00	220.00
110.00	380.00	222.00	120.00	330.00	230.00
114.00	370.00	226.00	160.00	334.00	200.00
118.00	320.00	230.00	200.00	342.00	180.00
126.00	260.00				

A6.4 LONG-TERM DIALYSIS

DAY	WBC	DAY	WBC	DAY	WBC
1.00	3.70	56.00	4.00	66.00	5.50
4.00	2.80	60.00	9.00	67.00	5.70
11.00	2.60	63.00	4.50	68.00	4.80
39.00	3.70	64.00	6.40	70.00	3.90
43.00	4.30	65.00	6.50	71.00	3.90

DAY	WBC
72.00	3.70
91.00	2.80
95.00	3.00
126.00	4.30
128.00	3.10
161.00	3.90
191.00	2.50
193.00	18.80
194.00	12.70
196.00	3.20
200.00	5.30
201.00	4.90
238.00	3.90
245.00	3.30
254.00	2.80
273.00	2.90
315.00	4.60
316.00	4.40
319.00	9.20
320.00	3.70
321.00	3.30
322.00	5.10
323.00	5.50
324.00	7.20
326.00	6.40
327.00	9.70
329.00	12.80
338.00	5.20
339.00	3.30
340.00	2.50
341.00	2.50
342.00	2.80
343.00	1.90
345.00	3.90
346.00	1.80
348.00	2.60
349.00	2.30
350.00	4.10
351.00	3.60
355.00	4.10
360.00	2.80
364.00	4.70
366.00	3.20
370.00	2.90
372.00	3.10
374.00	3.90
375.00	3.80
376.00	3.10
392.00	3.50

A6.5 FOETAL HEART DATA

TIME (MINUTES)	ST LENGTH
1.00	1017.20
2.00	1030.70
3.00	1029.00
4.00	1002.90
5.00	1033.50
6.00	981.50
7.00	1005.40
8.00	1020.90
9.00	1021.40
10.00	1011.90
11.00	1020.30
12.00	1017.60
13.00	1012.70
14.00	963.20
15.00	1023.20
16.00	1022.20
17.00	1014.30
18.00	1011.90
19.00	1028.00
20.00	1012.70
21.00	1001.50
22.00	1003.10
23.00	1027.80
24.00	987.60
25.00	982.70
26.00	997.30
27.00	1009.40
28.00	989.60
29.00	1009.60
30.00	931.30
31.00	997.80
32.00	981.90
33.00	856.00
34.00	890.10
35.00	1008.50
36.00	1020.00
37.00	1034.70
38.00	999.90
39.00	1014.50
40.00	1016.20
41.00	1039.30
42.00	1025.40
43.00	1025.10
44.00	1022.30
45.00	1004.90
46.00	1016.40
47.00	1014.00
48.00	1033.80
49.00	1023.90
50.00	1035.70
51.00	1004.60
52.00	994.50
53.00	1018.50
54.00	1008.50
55.00	1045.20
56.00	1012.40
57.00	1024.70
58.00	1039.70
59.00	1042.60
60.00	1009.40

TIME (MINUTES)	ST LENGTH	TIME (MINUTES)	ST LENGTH
61.00	1011.00	113.00	1011.00
62.00	1008.90	114.00	1023.20
63.00	1033.00	115.00	1029.10
64.00	1027.50	116.00	1049.30
65.00	1006.90	117.00	1071.40
66.00	982.00	118.00	1029.70
67.00	1039.40	119.00	1010.20
68.00	1011.80	120.00	1068.90
69.00	1035.20	121.00	991.70
70.00	1016.40	122.00	1044.30
71.00	1029.30	123.00	1026.20
72.00	1043.70	124.00	1011.10
73.00	1011.10	125.00	1025.60
74.00	1006.30	126.00	817.40
75.00	1034.70	127.00	1046.80
76.00	1034.90	128.00	1016.70
77.00	1039.80	129.00	1035.10
78.00	1036.00	130.00	1072.40
79.00	1025.70	131.00	1017.60
80.00	1024.80	132.00	1037.50
81.00	1017.40	133.00	1021.90
82.00	1011.00	134.00	1037.50
83.00	1033.20	135.00	992.00
84.00	1019.40	136.00	979.40
85.00	1011.80	137.00	975.90
86.00	1036.10	138.00	1008.50
87.00	1035.80	139.00	994.00
88.00	944.80	140.00	1001.50
89.00	998.20	141.00	1007.90
90.00	917.30	142.00	1035.80
91.00	951.80	143.00	1030.00
92.00	976.30	144.00	1007.40
93.00	949.60	145.00	1003.50
94.00	956.50	146.00	1008.50
95.00	1043.90	147.00	1005.60
96.00	1014.20	148.00	1008.20
97.00	998.80	148.00	1003.50
98.00	994.20	150.00	994.10
99.00	1000.10	151.00	1020.90
100.00	976.00	152.00	1005.80
101.00	992.00	153.00	970.60
102.00	992.80	154.00	1015.60
103.00	987.50	155.00	1029.40
104.00	1005.40	156.00	1010.30
105.00	1009.30	157.00	998.50
106.00	1004.90	158.00	1009.60
107.00	994.00	159.00	1005.50
108.00	1022.40	160.00	997.90
109.00	1005.30	161.00	1020.50
110.00	978.60	162.00	995.80
111.00	1030.70	163.00	1024.60
112.00	998.20	164.00	989.20

TIME (MINUTES)	ST LENGTH	TIME (MINUTES)	ST LENGTH
165.00	1005.00	217.00	973.70
166.00	1042.10	218.00	987.30
167.00	1028.90	219.00	970.60
168.00	1038.90	220.00	989.20
169.00	1004.70	221.00	971.90
170.00	1000.10	222.00	998.00
171.00	1039.30	223.00	1008.10
172.00	1041.80	224.00	1018.80
173.00	1017.60	225.00	982.00
174.00	978.80	226.00	989.60
175.00	1022.90	227.00	1028.80
176.00	1031.20	228.00	991.80
177.00	1003.90	229.00	1029.90
178.00	1006.00	230.00	984.30
179.00	546.80	231.00	968.10
180.00	714.00	232.00	1004.30
181.00	1041.20	233.00	1034.90
182.00	1046.90	234.00	1061.50
183.00	1034.10	235.00	1068.20
184.00	1050.70	236.00	1061.90
185.00	996.40	237.00	1031.40
186.00	1012.50	238.00	1060.30
187.00	974.60	239.00	1100.70
188.00	1014.60	240.00	1076.30
189.00	977.00	241.00	1102.80
190.00	1002.90	242.00	1111.70
191.00	990.60	243.00	1078.80
192.00	986.50	244.00	1086.70
193.00	960.50	245.00	1103.00
194.00	951.20	246.00	1103.30
195.00	974.30	247.00	1101.20
196.00	986.20	248.00	1162.30
197.00	941.40	249.00	1180.70
198.00	1014.10	250.00	1138.70
199.00	984.50	251.00	1211.50
200.00	988.30	252.00	1155.50
201.00	970.60	253.00	1221.40
202.00	998.20	254.00	1196.80
203.00	968.10	255.00	1240.70
204.00	934.60	256.00	1197.00
205.00	970.30	257.00	1220.90
206.00	976.20	258.00	1202.70
207.00	932.90	259.00	1234.00
208.00	965.10	260.00	1163.60
209.00	968.30	261.00	1089.10
210.00	1016.90	262.00	1295.90
211.00	1008.90	263.00	1087.20
212.00	969.20	264.00	1235.50
213.00	980.80	265.00	1266.80
214.00	981.00	266.00	1168.70
215.00	1010.20	267.00	1168.70
216.00	980.40	268.00	1229.50

TIME (MINUTES)	ST LENGTH	TIME (MINUTES)	ST LENGTH
269.00	1150.80	291.00	1434.90
270.00	1184.70	292.00	1430.60
271.00	1175.80	293.00	1335.90
272.00	1199.80	294.00	871.70
273.00	1148.60	295.00	869.80
274.00	1050.30	296.00	945.50
275.00	1220.40	297.00	1006.50
276.00	1217.50	298.00	1017.60
277.00	1230.80	299.00	1013.80
278.00	1277.80	300.00	1006.30
279.00	1345.60	305.00	1006.30
280.00	1443.30	315.00	978.60
281.00	1679.70	316.00	978.60
282.00	1830.70	317.00	978.60
283.00	1816.00	319.00	917.90
284.00	1774.50	321.00	724.30
285.00	1677.50	322.00	678.80
286.00	1499.60	324.00	510.80
287.00	1473.10	328.00	510.80
288.00	1550.40	331.00	398.20
289.00	1457.90	332.00	285.60
290.00	1431.10		

A6.6 BONE MARROW TRANSPLANTATION

DAY	ln(WBC)	ln(PLATELET)	HAEMATOCRIT
1.00	5.37	10.29	30.00
2.00	4.34	9.98	30.00
3.00	4.79	9.43	28.50
4.00	4.19	10.60	34.50
5.00	4.19	10.15	34.00
6.00	3.50	9.95	32.00
7.00	3.78	10.34	30.50
8.00	3.91	10.45	31.00
9.00	4.61	10.24	33.00
10.00	5.52	9.44	34.00
11.00	5.99	10.30	31.50
12.00	6.25	9.02	27.50
13.00	6.45	9.60	30.00
14.00	6.88	9.31	31.50
15.00	6.63	9.50	33.00
16.00	6.15	9.50	32.00
17.00	7.21	10.58	33.00
18.00	6.97	9.60	33.00
19.00	7.16	10.52	33.00
20.00	7.09	10.54	33.00
21.00	7.18	10.96	36.50
22.00	7.35	10.45	32.50
23.00	7.31	10.91	31.50
24.00	7.50	11.18	31.50

DAY	ln(WBC)	ln(PLATELET)	HAEMATOCRIT
25.00	8.04	11.29	29.50
26.00	7.84	11.08	22.50
27.00	7.59	11.25	28.00
28.00	8.01	11.45	31.00
29.00	8.22	11.39	30.50
30.00	8.02	12.00	31.50
31.00	8.41	12.18	31.00
32.00	8.22	11.91	31.00
33.00	8.20	11.94	31.50
34.00	8.81	11.63	31.00
35.00	8.87	12.04	31.50
36.00	9.00	12.21	32.00
38.00	8.99	12.01	30.00
39.00	8.99	12.01	30.00
41.00	9.27	12.13	28.50
43.00	9.33	12.21	27.00
46.00	9.12	12.38	28.50
48.00	8.76	12.07	27.00
50.00	8.19	11.82	31.00
53.00	8.02	11.74	29.50
57.00	7.48	11.88	27.00
60.00	7.31	12.00	28.50
64.00	7.74	11.71	35.00
67.00	8.13	12.01	34.50
71.00	8.46	12.29	35.00
74.00	8.88	12.23	31.50
78.00	8.41	12.16	30.00
81.00	8.22	12.35	27.50
85.00	8.61	12.11	35.00
88.00	8.24	11.94	33.00
92.00	8.09	12.50	31.50

o0o

REFERENCES

ABRAHAM, B. (1980). Intervention analysis and multiple time series, Biometrika, 67, 73-78.

ABRAHAM, B. and BOX, G. E. P. (1978). Linear models and spurious observations, Applied Statistics, 27, 131-138.

ABRAHAM, B. and BOX, G. E. P. (1979). Bayesian analysis of some outlier problems in time series, Biometrika, 66, 229-236.

AITCHISON, J. and DUNSMORE, I. R. (1975). Statistical Prediction Analysis, Cambridge University Press: Cambridge.

AKAIKE, H. (1974). A new look at the statistical model identification, I.E.E.E. Transactions Auto. Cont., AC19, 716-723.

AMEEN, J. R. M. and HARRISON, P. J. (1982). Normal discount Bayesian models. Warwick Statistics Research Report, No. 30.

AMEEN, J. R. M. and HARRISON, P. J. (1983). Discount Bayesian multiprocess modelling with cusums. Warwick Research Report, No. 34.

AMEEN, J. R. M. and HARRISON, P. J. (1984). Discount Weighted Estimation, Journal of Forecasting, 3, 285-296.

ANDERSON, B. D. O. and MOORE, J. B. (1979). Optimal Filtering, Prentice Hall: New Jersey.

ANDERSON, T. W. (1980). Maximum likelihood estimation for vector autoregressive moving average models, in Directions in Time Series (Eds. D. R. Brillinger and G. C. Tiao), 49-59.

ANSLEY, C. F. and KOHN, R. (1983). Exact likelihood of vector autoregressive-moving average process with missing or aggregated data, Biometrika, 70, 275-278.

BACON, D. W. and WATTS, D. G. (1971). Estimating the transition between two intersecting straight lines, Biometrika, 58, 525-534.

BAGSHAW, M. and JOHNSON, R. A. (1977). Sequential procedures for detecting parameter changes in a time-series model, J. Amer. Statist. Assoc., 72, 593-597.

BLOOMFIELD, P. (1970). Spectral analysis with randomly missing observations, J. R. Statist. Soc., B, 32, 369-380.

BOOTH, N. B. and SMITH, A. F. M. (1982). A Bayesian approach to retrospective identification of changepoints, J. of Econometrics, 19, 7-22.

BOX, G. E. P. and JENKINS, G. M. (1962). Some statistical aspects of adaptive optimization and control (with discussion), J. R. Statist. Soc., B, 24, 297-343.

BOX, G. E. P. and JENKINS, G. M. (1970). Time Series Analysis: Forecasting and Control, Holden-Day: San Francisco.

BOX, G. E. P. and TIAO, G. C. (1965). A change in level of a non-stationary time series, Biometrika, 52, 181-192.

BOX, G. E. P. and TIAO, G. C. (1968). A Bayesian approach to some outlier problems, Biometrika, 55, 119-129.

BOX, G. E. P. and TIAO, G. C. (1975). Intervention analysis with applications to economic and environmental problems, J. Amer. Statist. Assoc., 70, 70-79.

BROWN, R. G. (1963). Smoothing, Forecasting and Prediction of Discrete Time Series, Prentice Hall: New Jersey.

BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion), J. R. Statist. Soc., B, 37, 149-192.

BUCHSBAUM, M. and HARRIS, E. K. (1971). Diurnal variation in serum and urine electrolytes, J. Applied Physiology, 30(1), 27-35.

CANTARELIS, N. and JOHNSTON, F. R. (1983). On-line variance estimation for the steady state Bayesian forecasting model, J. Time Series Anal., 3, 225-234.

CARSON, E. R., COBELLI, C. and FINKELSTEIN, L. (1983), The Mathematical Modelling of Metabolic and Endocrine Systems, Wiley: New York.

CHIN CHOY, J. H. and BROEMELING, L. D. (1980). Some Bayesian inferences for a changing linear model, Technometrics, 22, 71-78.

CLINGER, W. and VAN NESS, J. W. (1976). On unequally spaced time points in time series, Ann. Statist., 4, 736-745.

COX, D. R. (1961). Prediction by exponentially weighted moving averages and related methods, J. R. Statist. Soc., B, 23, 414-422.

COX, D. R. and MILLER, H. D. (1965). The Theory of Stochastic Processes, Methuen: London.

CRANE, D. B. and CROTTY, J. R. (1967). A 2-stage forecasting model: exponential smoothing and multiple regression, Management Science, 13, 501-507.

DEGROOT, M. H. (1970). Optimal Statistical Decisions, McGraw-Hill: New York.

DIAZ, J. (1982). Bayesian detection of a change of scale parameter in sequences of independent gamma random variables, J. of Econometrics, 19, 23-29.

DIXON, W. J. (1953). Processing data for outliers, Biometrics, 9, 74-89.

van DOBBEN DE BRUYN, C. S. (1968). Cumulative Sum Tests: Theory and Practice, Griffin: London.

DORAN, H. E. (1974). Prediction of missing observations in the time series of an economic variable, J. Amer. Statist. Assoc., 69, 546-554.

DUNSMUIR, W. and ROBINSON, P. M. (1981). Estimation of time series models in the presence of missing data, J. Amer. Statist. Assoc., 76, 560-568.

EDWARDS, R. W. H. (1980). Internal analytical quality control using the CUSUM chart and truncated V-mask procedure, Annals of Clinical Biochemistry, 17, 205-211.

EZZET, F. L. (1985). Unpublished Ph.D. Thesis: University of Nottingham.

EZZET, F. L. and SMITH, A. F. M. (1985). Detecting parameter changes in time series models, Nottingham: Statistics Research Report, No. 04-85.

FERREIRA, P. E. (1975). A Bayesian analysis of a switching regression model: known number of regimes, J. Amer. Statist. Assoc., 70, 370-374.

FILIPPOV, A. F. (1962). On certain questions in the theory of optimal control, J. S.I.A.M., Control, A, 1, 76-84.

FOX, A. J. (1972). Outliers in time series, J. R. Statist. Soc., B, 34, 350-363.

GODOLPHIN, E. J. and HARRISON, P. J. (1975). Equivalence theorems for polynomial-projecting predictors, J. R. Statist. Soc., B, 37, 205-215.

HARRISON, P. J. and DAVIES, O. L. (1964). The use of cumulative sum (Cusum) techniques for the control of routine forecasts of product demand, Operations Research, 12, 325-333.

HARRISON, P. J. and JOHNSTON, F. R. (1983). A regression method with non-stationary parameters. Warwick Statistics Research Report, No. 35.

HARRISON, P. J. and STEVENS, C. F. (1971). A Bayesian approach to short-term forecasting, Operational Research Quarterly, 22, 341-362.

HARRISON, P. J. and STEVENS, C. F. (1975). Bayes forecasting in action: case studies. Warwick Statistics Research Report, No. 14.

HARRISON, P. J. and STEVENS, C. F. (1976). Bayesian forecasting (with discussion). J. R. Statist. Soc., B, 38, 205-247.

HARVEY, A. C. and PHILLIPS, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances, Biometrika, 66, 49-58.

HARVEY, A. C. and PIERSE, R. G. (1984). Estimating missing observations in economic time series, J. Amer. Statist. Assoc., 79, 125-131.

HAUGH, L. D. and BOX, G. E. P. (1977). Identification of dynamic regression (distributed lag) models connecting two time series, J. Amer. Statist. Assoc., 72, 121-130.

HETZEL, M. R. and CLARK, T. J. H. (1979). Does sleep cause nocturnal asthma?, Thorax, 34, 749-754.

HILLIER, P., KNAPP, M. S. and COVE-SMITH, J. R. (1980). Circadian variations in urine excretion in chronic renal failure, Quarterly J. of Medicine N. S., 49, 461-478.

HILLMER, S. C. and TIAO, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models, J. Amer. Statist. Assoc., 74, 652-660.

HINKLEY, D. V. (1971). Inference about the changepoint from Cumulative sum tests, Biometrika, 58, 509-523.

HOLBERT, D. (1982). A Bayesian analysis of a switching linear model, J. of Econometrics, 19, 77-87.

HRUSHESKY, W. J. M. (1984). The sine-o-graph pulse monitor - a potential quantifier of cardiovascular wellness, in Annual Review of Chronopharmacology, Volume 1 (eds: A. Reinberg, M. Smolensky and G. Labrecque), 325-328, Pergamon Press: Oxford.

HRUSHESKY, W. J. M., LEVI, F. A., HALBERG, F. and KENNEDY, B. J. (1982). Circadian stage dependence of cis-Diamminedichloroplatinum lethal toxicity in rats, Cancer Research, 42, 945-949.

HSU, D. A. (1979). Detecting shfits of parameter in gamma sequences with application to stock price and air traffic flow analysis, J. Amer. Statist. Assoc., 74, 31-40.

JOHNSON, R. A. and BAGSHAW, M. (1974). The effect of serial correlation on the performance of CUSUM tests, Technometrics, 16, 103-112.

JOHNSTON, F. R. and HARRISON, P. J. (1983). Discount weighted moving averages. Warwick Statistics Research Report, No. 42.

JONES, R. H. (1962). Spectral analysis with regularly missed observations, Ann. Math. Statist., 33, 455-461.

JONES, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations, Technometrics, 22, 389-395.

JONES, R. H. (1981). Fitting a continuous time autoregression to discrete data, in Applied Time Series II (ed: D. F. Findley), 651-682, Academic Press: New York.

JONES, R. H. (1984). Fitting multivariate models to unequally spaced data, in Time Series Analysis of Irregularly Observed Data (eds: D. Brillinger, S. Fienberg, J. Gani, J. Hartigan and K. Krickeberg), 276-289, Springer-Verlag: Berlin.

JONES, R. H. (1985). Time series analysis with unequally spaced data, in Handbook of Statistics Volume 5: Time Series in the Time Domain (eds: E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 157-177, Elsevier: Amsterdam.

JONES, R. H. and TRYON, P. V. (1986). Continuous time series models for unequally spaced data applied to modelling atomic clocks, S.I.A.M. J. of Sci. Stat. Comput., 7 (to appear).

JOSEPH, P. and TOU, J. T. (1961). On linear control theory, A.I.E.E. Trans. (Applications and Industry), 80, 193-196.

KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems, J. of Basic Eng., 82D, 35-45.

KALMAN, R. E. and BUCY, R. S. (1961). New results in linear filtering and prediction theory, J. of Basic. Eng., 83D, 95-108.

KIRBY, R. M. (1967). A comparison of short and medium range statistical forecasting methods, Management Science, 13, 202-210.

KITAGAWA, G. (1984). State space modelling of non-stationary time series and smoothing of unequally spaced data, in Time Series Analysis of Irregularly Observed Data (eds: D. Brillinger, S. Fienberg, J. Gani, J. Hartigan and K. Krickeberg), 189-210, Springer-Verlag: Berlin.

KNAPP, M. S., BLAMEY, R. W., COVE-SMITH, J. R. and HEATH, M. (1977). Monitoring the function of renal transplants, Lancet, 2, 1183.

KOWANKO, I. C. (1982). Unpublished Ph.D. Thesis: University of Nottingham.

KOWANKO, I.C., KNAPP, M. S., POWNALL, R. and SWANNELL, A. J. (1982). Domiciliary self-measurement in rheumatoid arthritis and the demonstration of circadian rhythmicity, Annals of the Rheumatic Diseases, 41, 453-455.

KOWANKO, I. C., SWANNELL, A. J., MAHONEY, P. G. C., POWNALL, R. and KNAPP, M. S. (1980). Preliminary observations on the therapeutic effectiveness of Froben at different times of day , British J. of Clinical Practice, Supplement 9, 10-14.

KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency, Ann. Math. Statist., 22, 79-86.

LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model (with discussion), J. R. Statist. Soc., B, 34, 1-41.

LOYNES, R. M. (1969). Aliassing and related questions in stationary processes, in Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress on Time Series and Stochastic Processes; Convexity and Combinatories, 89-98.

MARKUS, L. and LEE, E. B. (1962). On the existence of optimal controls, J. of Basic. Eng., 84, 13-20.

MATIS, J. H. and WEHRLY, T. E. (1979). Stochastic models of compartmental systems, Biometrics, 35, 199-220.

McCULLAGH, P. and NELDER, J. A. (1983). Generalised Linear Models, Chapman and Hall: London.

McGILL, R. (1965). Optimal control, inequality state constraints and the generalized Newton-Raphson algorithm, J. S.I.A.M., Series A, Control, 3, 291-298.

MEHTA, J. S. and SWAMY, P. A. V. B. (1974). Bayesian analysis of a bivariate normal distribution when some observations are missing, in Studies in Bayesian Econometrics and Statistics: In Honor of Leonard J. Savage (eds: S. E. Fienberg and A. Zellner), 289-309, North-Holland: Amsterdam.

MINORS, D. S. and WATERHOUSE, J. M. (1981). Circadian Rhythms and the Human, John Wright: Bristol.

MONTGOMERY, D. C. and JOHNSON, L. A. (1976). Forecasting and Time Series Analysis, McGraw-Hill: New York.

MORE-EDE, M. C., SULZMAN, F. M. and FULLER, C. A. (1982). The Clocks that Time Us, Harvard University Press: Cambridge and London.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models, J. R. Statist. Soc., A, 135, 370-384.

NEWBOLD, P. (1979). Bivariate time series model-building methods, in Forecasting and Planning (eds: R. Fildes and D. Wood), 104-118, Saxon House: Farnborough.

NICHOLLS, D. F. and HALL, A. D. (1979). The exact likelihood function of multivariate autoregressive-moving average models, Biometrika, 66, 259-264.

OSBORN, D. R. (1977). Exact and approximate maximum likelihood estimators for vector moving average processes, J. R. Statist. Soc., B, 39, 114-118.

PAGE, E. S. (1954). Continuous inspection schemes, Biometrika, 41, 100-114.

PARZEN, E. (1963). On spectral analysis with missing observations and amplitude modulation, Sankhya, A, 25, 383-392.

PARZEN, E. (1977). Multiple time series: determining the order of approximating autoregressive schemes, in Multivariate Analysis IV (ed: P. R. Krishnaiah), 283-295, North-Holland: Amsterdam.

PHADKE, M. S. and WU, S. M. (1974). Modelling of continuous stochastic processes from discrete observations with application to sunspots data, J. Amer. Statist. Assoc., 69, 325-329.

PRIESTLEY, M. B. (1980). System identification, Kalman filtering and stochastic control, in Directions in Time Series (eds: D. R. Brillinger and G. C. Tiao), 188-218.

QUENOUILLE, M. H. (1957). The Analysis of Multiple Time Series, Griffin: London.

QUINN, B. G. (1980). Order determination for a multivariate autoregression, J. R. Statist. Soc., B, 42, 182-185.

REBAR, R. W. and YEN, S. S. C. (1979). Endocrine rhythms in gonadotrophins and ovarian steroids with reference to reproductive processes, in Endocrine Rhythms (ed: D. T. Krieger), 259-298, Raven Press: New York.

ROBINSON, P. M. (1977). Estimation of a time series model from unequally spaced data, Stochastic Processes and their Applications, 6, 9-24.

ROBINSON, P. M. (1980). Continuous model fitting from discrete data, in Directions in Time Series (eds: D. R. Brillinger and G. C. Tiao), 263-278.

ROBINSON, P. M. (1984). Multiple time series analysis of irregularly spaced data, in Time Series Analysis of Irregularly Observed Data (eds: D. Brillinger, S. Fienberg, J. Gani, J. Hartigan and K. Krickeberg), 276-289, Springer-Verlag: Berlin.

SARGAN, J. D. and DRETTAKIS, E. G. (1974). Missing data in an autoregressive model, International Economic Review, 15, 39-58.

SCHEINOK, P. A. (1965). Spectral analysis with randomly missed observations: the binomial case, Ann. Math. Statist., 36, 971-977.

SCHMIDT, S. F. (1966). Application of state-space methods to navigation problems, in Advances in Control Systems, Theory and Applications, Volume 3 (ed: C. T. Leondes), 293-340, Academic Press: New York.

SEARLE, S. R. (1971). Linear Models, Wiley: New York.

SHAPIRO, H. S. and SILVERMAN, R. A. (1960). Alias-free sampling of random noise, J. S.I.A.M., 8, 225-248.

SMITH, A. F. M. (1977). A Bayesian analysis of some time-varying models, in Recent Developments in Statistics (eds: J. R. Barra et al.), North-Holland: Amsterdam.

SMITH, A. F. M. and COOK, D. G. (1980). Straight lines with a changepoint: a Bayesian analysis of some renal transplant data, Applied Statistics, 29, 180-189.

SMITH, A. F. M. and WEST, M. (1983). Monitoring renal transplants: an application of the multiprocess Kalman filter, Biometrics, 39, 867-878.

SORENSON, H. W. (1966). Kalman filtering techniques, in Advances in Control Systems, Theory and Applications, Volume 3 (ed: C. T. Leondes), 219-292, Academic Press: New York.

SORENSON, H. W. and SACKS, J. E. (1971). Recursive fading memory filtering, Information Sciences, 3, 101-119.

STOODLEY, K. D. C. and MIRNIA, M. (1979). The automatic detection of transients, step changes and slope changes in the monitoring of medical time series, The Statistician, 28, 163-170.

TIAO, G. C. and BOX, G. E. P. (1981). Modelling multiple time series with applications, J. Amer. Statist. Assoc., 76, 802-816.

TRIGG, D. W. (1964). Monitoring a forecasting system, Operational Research Quarterly, 15, 271-274.

TRIGG, D. W. and LEACH, A. G. (1967). Exponential smoothing with an adaptive response rate, Operational Research Quarterly, 18, 53-59.

TRIMBLE, I. M., WEST, M., KNAPP, M. S., POWNALL, R. and SMITH, A. F. M. (1983). Detection of renal allograft rejection by computer, British Medical Journal, 286, 1695-1699.

WEST, M. (1981). Robust sequential approximate Bayesian estimation, J. R. Statist. Soc., B, 43, 157-166.

WEST, M. (1982). Unpublished Ph.D. Thesis: University of Nottingham.

WEST, M., HARRISON, P. J. and MIGON, H. S. (1984). Dynamic generalised linear models and Bayesian forecasting. Warwick Statistics Research Report.

WHEELWRIGHT, S. C. and MAKRIDAKIS, S. (1973). Forecasting Methods for Management, Wiley: New York.

WHITTLE, P. (1953). The analysis of multiple stationary time series, J. R. Statist. Soc., B, 15, 125-139.

WHITTLE, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix, Biometrika, 50, 129-134.

WICHERN, D. W., MILLER, R. B. and HSU, D. A. (1976). Changes of variance in first-order autoregressive time series models - with an application, Applied Statistics, 25, 248-256.

WINTERS, P. R. (1960). Forecasting sales by exponentially weighted moving averages, Management Science, 6, 324-342.

YOUNG, P. C. (1974). Recursive approaches to time series analysis, Bull. Inst. Math. and its Applics., 10, 209-224.

ZADEH, L. A. and DESOER, C. A. (1963). Linear System Theory: The State-Space Approach, McGraw-Hill: New York.

oOo