APPLICATIONS OF LINEAR ESTIMATION THEORY TO CHEMICAL PROCESSES

A thesis submitted for
the Degree of Doctor of Philosophy
in the Faculty of Engineering of the University of London

by

Stephen Frederick Goldmann

Department of Chemical Engineering
and Chemical Technology
Imperial College
London S.W. 7.

October, 1969.

To my parents and Joyce.

Abstract

This thesis investigates the feasibility of applying linear statistical filtering theory to the solution of estimation problems encountered in the operation of chemical plants. The study is divided into two sections.

In the first part, the development of the theory is considered, and a unified derivation is presented which yields both growing and limited memory filtering algorithms.

In the second section, the application of filtering techniques to typical chemical engineering systems is considered. The sensitivity of the filter performance to errors in the statistical assumptions made about the system is investigated using data from a binary distillation column. The effects of system modelling errors is studied using a simulated fixed-bed catalytic reactor. For this example, a realistic industrial simulation is used, and the ability of the filter to estimate non-observable system parameters is demonstrated.

Acknowledgements

I am sincerely grateful to Professor R.W.H. Sargent for his continuing encouragement and assistance during the course of this work.

Thanks is also due to the Dyestuffs Division of Imperial Chemical Industries Ltd. for generously allowing me access to a description of one of their plants, and to my colleagues in the computing section at Imperial College for their assistance and suggestions.

I am also indebted to the Marshall Aid Commemoration Commission for their generous support, both moral and financial throughout my stay in England.

Stephen F. Goldmann

October, 1969.

CONTENTS

CONTENTS

(continued)

CONTENTS

(continued)

CHAPTER 1

INTRODUCTION

Though there have been impressive advances over the last decade in the theories of optimization and control, few of the new results have been applied within the chemical process industries. This gap between theory and application is caused by the fact that optimal control techniques require an exact mathematical model of the system being considered, and also require the observation of the true state of the system so that, using the model, its future behaviour can be predicted, and any corrective action determined.

Even when reliable deterministic models are available to the chemical engineer they are often too complex for repetitive optimization calculations. In practical situations, this problem is compounded by the fact that both the measuring devices and the system itself are subject to random disturbances which prevent the error free observation of the system state; indeed, the future evolution of the plant often depends upon a variable such as catalyst activity which cannot be directly observed, and must be deduced from other observations. Thus the problem becomes one of state estimation as well as stochastic control.

These difficulties have appeared to be so serious that many chemical engineers have dismissed optimal control theory as an academic exercise of little practical use. This attitude is reflected in the operation of many large industrial processes which are controlled by on-line computers capable of optimization calculations, but programmed merely to log data and simulate conventional analogue control algorithms.

In spite of the fact that the complexity of typical chemical processes often precludes the application of modern control techniques it would seem that, when a plant is installed with an on-line computer, better use could be made of the information which is available about the system. In a conventional plant, most operating decisions are made on the basis of a small sub-set of the measured process variables. For example, in the operation of a tubular reactor, control decisions are usually based solely upon the outlet conditions of the process stream, even though temperature measurements are likely to have been made at several points along the length of the reactor. Furthermore, process measurements are usually accepted as being completely correct, and no allowance is made for instrument errors, or the past performance of the measuring devices. Though it would be impossible for an operator to absorb and utilize all of the available process information, it is not unreasonable to propose that an on-line computer could be programmed to analyze for trends in plant performance, or detect changes in the reliability of the measuring devices.

The more efficient use of process information might, for instance, allow the estimation of a stream composition using several temperature measurements, and save the cost of installing a composition analyzer. Even when a complete mathematical model of the process is not available, effective use of an on-line computer would allow utilization of any system constraints that might be known. Thus when a process is in steady operation, account could be taken of the fact that a total mass balance must be satisfied, and this knowledge considered in conjunction with actual flow measurements might permit the detection of instrument malfunctions or the detection of leaks in the system. By monitoring successive sets of observations, it is possible that the computer could

be used to estimate the relative reliability and accuracy of the various process measurements. In brief, the installation of an on-line computer should allow better use of all information that is available about a process, even though the knowledge of the system is inadequate for the formal application of optimal control techniques.

In order to determine how more effective use can be made of the various sources of information in a chemical plant, it is useful to view the process as an information system, consisting of the physical units and the measuring and control devices, as well as any known system constraints. The amount of information that can be obtained by observing the system depends upon the speed with which the characteristics of its components are changing in relation of the observation rate. The extraction of any information from a system at a given observation rate requires that the process must contain some elements whose characteristics remain constant in relation to the rate of observation, in much the same way as navigation of the open sea relies upon the seeming immobility of the reference stars. In many situations, only the statistics of the observation noise can be assumed to be constant, and in these cases effective utilization of system information requires an accurate prediction of the dynamic behaviour of the plant. However, many industrial chemical processes move very slowly in relation to the rate at which they are observed, and under these circumstances not only can the statistics of the instrument errors be considered constant over long periods, but the plant state as well. In these cases, the assumption that the plant moves through a succession of steady-states may be justified, and it is reasonable to base an estimate of the plant state upon several sets of observations rather than the most recent one.

There are several possible methods for analyzing the system information in order to provide an improved estimate of the plant state. In cases when steady-state operation can be assumed, the most direct approach is to minimize some weighted function of the observation errors subject to the known steady-state constraints, where the weights reflect the relative reliability of the measuring devices. The actual form of the function to be minimized would depend upon the anticipated probability distribution of the instrument errors, but is unlikely to be very critical; a quadratic function can be shown to be adequate in most situations. Though this approach has the advantage of conceptual simplicity and does utilize all available information, it has some limitations. The technique requires the storage of the relevant data sets, and the performance of a new minimization as each new set of observations is acquired; it is therefore likely to be impractical for on-line computation. Furthermore, the performance of a minimization with nonlinear constraints is in itself a difficult numerical problem.

When the rate of change of the plant state is of the same order as the observation rate, the minimization approach is not applicable, since the governing constraints relating the observations to the plant state vary between data sets. A possible method of analysis under these circumstances is to interpret the plant records as stochastic time series. Thus each instrument reading can be considered to be the sum of the true reading and additive random observation errors. The estimation of the true plant state requires the removal of the random element of the measurements, and hence becomes a problem in statistical filtering. This approach is attractive in the sense that it is possible to formulate sequential filters, and thus eliminate the necessity of storing large sets of data.

Filtering techniques have been developed for the solution of the following problem : Given a set of observations of a system over the interval $(0, T]$, it is desired to estimate the output of the system at some time t, such that the estimate is optimal according to some predetermined criterion. When $t < T$, the problem is one of smoothing; when $t = T$, it is one of filtering and when $t > T$, it is one of prediction. The selection of a criterion for optimality is subject to the same considerations as those mentioned for the constrained minimization.

But the application of statistical filtering techniques would seem to be thwarted by the same limitation that prevents the calculation of optimal controls for realistic chemical processes. In order to remove the random observation noise, a deterministic model for the true value of the observed variable must be postulated. In cases when the system changes significantly between observations, this means that a dynamic model must be proposed before the filtering techniques can be applied.

However, mathematical models suitable for filtering applications are not as difficult to formulate as control models. A model used for the calculation of optimal controls must be valid for the entire period for which a control policy is needed, while a filtering model need only be locally valid. It should be possible to evaluate instrument performance and estimate non-observable system parameters with a local model, so long as the model errors remain relatively constant and small in relation to the observation errors. If the local model approximates the true system closely enough, the deterministic element of the observations can be removed, and a constant frame of reference based upon the characteristics of the instrument errors can be established in order to allow valid analysis of the information system.

Several of the arguments presented in the preceding paragraphs must be tested before filtering techniques can be applied to operating systems. Specifically, the sensitivity of the filter performance to errors in the estimate of the observation statistics or to local model errors must be determined. In order to evaluate the performance in typical chemical engineering applications, the filtering techniques have been applied to some simulated systems. In one example data from a distillation column operating in a known steady-state is analyzed to evaluate the effect of erroneous statistical information being used in the filter. In another simulation, the techniques are applied to a fixed bed tubular reactor with catalyst decay in order to test the ability of the method to estimate the catalyst activity and its rate of decay from temperature measurements only. In this case, the complete reactor model is too complex for on-line computation, and a simplified, locally valid model is used for the filtering calculations.

A review of the development of the theory of statistical filtering and estimation is presented in Chapter 2. An effort is made to reconcile the many different approaches that can be taken to establish the results of the theory, and in the following chapter, a unified derivation of the results is presented, based upon the characteristics of least-squares estimators. Chapter 4 describes previous applications of on-line estimating techniques in the process industries, and discusses the particular problems faced in applying statistical filtering theory to chemical engineering problems. The final chapters are devoted to the analysis of the performance of the filtering algorithms in the simulated situations mentioned above. Though the simulations are presented merely as examples of possible applications in the process industries, it is felt that they are representative of typical industrial systems, and will point the way towards more effective utilization of process information.

STATISTICAL FILTERING THEORY

CHAPTER 2

THE DEVELOPMENT OF STATISTICAL FILTERING TECHNIQUES

2.1 Introduction

The modern theory of statistical filtering was originated by Wiener[1], who combined the techniques of time series analysis with the classical theory of communication in order to provide solutions to the pressing problems of communication signal analysis which arose during World War II. Though Wiener's methods provided theoretical solutions to many problems involving the removal of random noise from radio signals, practical implementation of the new results was impossible in all but the most simple cases. Extensive research provided methods to solve a wider class of problems, but the Wiener theory was not widely applied until the advent of the digital computer. This powerful new tool for information analysis allowed the reformulation of the Wiener techniques into a more convenient form. Kalman[2] presented a technique which does not require a closed solution to the filtering problem, but rather allows its recursive solution with the aid of a computer. Though the original Kalman solution was developed for linear systems with gaussian disturbances, several modifications have been developed to allow the solution of a wide class of problems.

Before these techniques and their modifications are described, the question of a definition of optimality for filtering problems is considered. The relationship between filtering theory and its progenitors, the theories of estimation and communication, is also discussed.

2.2 A Criterion for Optimality

In all problems of filtering and estimation, it is desired to produce a solution that is "best" in some way. The form of the solution depends intimately upon the definition of what is considered to be "best". In selecting a criterion for optimality, account must be taken of the use to which the estimate will be put, and the types of errors that are assumed to corrupt the system. The engineer has traditionally turned to the criterion of least-squares to define optimality, and it is reasonable to consider first the desirability of this well known criterion.

The characteristics of the least-squares estimator were first described in detail by Gauss[3], and the publication of his theory is considered to have established estimation theory as a mathematical technique[4]. Gauss described the characteristics of an estimator for a linear system observed by measuring devices with uncorrelated errors, and defined as optimal the estimate of the system that minimized the sum of squared deviations of the true observations from the ones predicted by the fitted model. The principal advantage of the least-squares estimator is that its characteristics are independent of the probability distribution of the observation errors, as long as the errors are uncorrelated.

Later developments in estimation theory do not share this advantage, and this fact is reflected in the limited application of the techniques to practical engineering problems. For example, a large body of theory has been developed in order to rationalize the comparison of alternative estimators of the parameters in probability distribution functions. The qualities such as bias, consistency,

sufficiency, and efficiency have been defined for this purpose[5]. Thus, in estimating the variance of a normal population on the basis of a random sample; $\{x_i\}$, i = 1, .. n, the estimator;

$$s_1 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 \qquad (2.1)$$

is used in preference to ,

$$s_2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2 \qquad (2.2)$$

where $\bar{x}$ is the sample mean. s_1 is selected because it can be shown that its expected value is σ^2, the true population variance, and hence the estimator is unbiased by definition. Since the use of these comparative techniques for the evaluation of estimators requires a knowledge of the governing distribution, the criteria have been applied primarily in the life sciences where probabalistic models of error phenomena can be reliably postulated.

The same disadvantage is encountered in applying maximum likelihood techniques[5]. In this case the joint probability distribution of the observed events is calculated as a function of the unknown parameters in the governing probability distribution, and the optimal parameters are defined as those which maximize the joint distribution function. The use of this optimality criterion is therefore restricted to those situations for which the joint probability distribution can be defined analytically.

Though it has been suggested that the least-squares criterion is applied to practical problems out of ignorance rather than insight[6], a lack of probabilistic information is not the only justification for

its application to engineering problems. First of all, the estimator which produces a least-squares estimate can be shown to be identical to the maximum-likelihood estimator for linear systems subject to normal random disturbances. More significantly, the least-squares estimator can be considered to be optimal in a "wide sense" since it minimizes the first two moments of the observation errors about the predicted observations[4]. The neglecting of higher order moments is fortunately reasonable in many practical situations.

A further justification for choosing a least-squares criterion arises from the specification of electronic filters in communication theory. Using the classical theory, the filter is designed to have particular frequency response characteristics in order to remove the undesired noise components from the signal. The filter is synthesised so as to minimize the ratio of the noise signal power to the message power. Since power in electrical circuits depends quadratically upon current, minimizing the sum of squared errors caused by the signal fluctuations will minimize the power of the error component.

But the major advantage in employing a least-squares criterion is that it provides an estimator whose characteristics are independent of the stochastic characteristics of the system being observed. Though it might be possible to propose a more desirable optimality criterion for the solution of a particular problem, a filter which results in a least-squares estimate is likely to be applicable to a variety of practical problems. For this reason, any future reference to a best or optimal estimate should be taken to imply optimality in a least-squares sense.

2.3 Wiener Filter Theory

Wiener[1] considers the solution in the time domain of the problem of linear filter design in communication engineering. Using a minimum mean error power optimality criterion, the specification of the filter is made in terms of the first two moments of the signal statistics. The main result of the solution is the Wiener-Hopf equation which is an integral equation relating the cross-correlation function of the filter input with its output, to the auto-correlation function of the input. The solution of the equation provides the impulse response of the optimal linear filter.

Though the Wiener filter theory represents a major conceptual advance in information analysis, it involves several practical difficulties that limit its utility. A major limitation lies in the fact that the optimal filter is specified by its impulse response. Even when the Wiener-Hopf equation is soluble, the construction of the filter on the basis of its desired impulse response characteristics is a difficult problem. Furthermore, Wiener considers only simple systems with stationary statistics. Though multivariate systems are dealt with briefly, their solution requires the specification of auto- and cross-correlation functions relating all inputs to all outputs.

Nevertheless, a great deal of research has been inspired by the work of Wiener. Because the complexity of chemical engineering problems prevents the direct application of the theory, its developments and extensions will not be explored. Extensive references to studies based on the original Wiener theory are available in several sources[7,8].

The most significant aspect of the work of Wiener is the way in which the combining of statistical time series analysis with the classical techniques of communication engineering allows the consideration of filtering problems in the time domain. Time domain analysis is far more useful in the solution of chemical engineering problems where frequency domain analysis is unlikely to be physically meaningful, or is invalid due to system nonlinearities.

2.4 Kalman Filter Theory

The difficulties associated with the solution of the Wiener-Hopf equation were circumvented by Kalman[2]. He considers the estimation of the state of a linear dynamics system subject to gaussian disturbances and observed by measurements which are linearly related to the state. Using state-vector notation to describe the evolution of the system as a discrete function of time, and considering the geometric properties of optimal estimators, a recursive formulation of the filter was derived in a form ideally suited to digital computation.

2.4.1 State-Vector Notation

State vector notation allows the reduction of the matrix differential equation describing a linear multivariate dynamic system, to a set of recursive equations relating the current state of the system to the state at some interval of time later, where the state is defined as the minimum amount of information required to describe the future evolution of the system given the describing equations, and assuming no further inputs are applied. The salient features of the notation are presented below[9].

Consider a fixed homogeneous linear dynamic system described by,

$$\dot{\underline{x}} = A\,\underline{x} \qquad (2.3)$$

where $\underline{x}$ is a n-dimensional state vector, A is a (n x n) matrix of constant coefficients, and $\dot{\underline{x}}$ denotes the time derivative of the state. The solution of this equation is analogous to the scalar case.

$$\underline{x}(t_n) = \exp\left[A\,(t_n - t_0)\right]\,\underline{x}(t_0) \qquad (2.4)$$

for $t_n \geq t_0$.

The exponentiation of a matrix is defined by the infinite series:

$$\exp\left[At\right] = I + At + A^2t^2/2! + A^3t^3/3! + \ldots\ldots \qquad (2.5)$$

For the stationary system considered, the matrix exponential is defined as the state transition matrix relating the state at time t_n to the state at some earlier time t_0. Since the elements of the matrix A are time invariant, the state transition matrix depends only upon the interval $(t_n - t_0)$ and not upon either time explicitly. Thus equation (2.3) may be transformed into a recursive relationship,

$$\underline{x}(t_n) = \underline{\Phi}\,(t_n - t_0)\,\underline{x}(t_0) \qquad (2.6)$$

where $\underline{\Phi}\,(t_n - t_0)$ is defined from equation (2.4).

For the time-varying, homogeneous case, the elements of matrix A in equation (2.1) depend upon time, and the state transition matrix cannot in general be described in closed form. However, the relationship between $\underline{x}(t_n)$ and $\underline{x}(t_0)$ remains conceptually the same as described in (2.6), though in this case the transition matrix will be written as $\underline{\Phi}\,(t_n, t_0)$ to indicate the time dependence.

The addition of a forcing function to the system further complicates the analysis. In the non-homogeneous case, the general linear system becomes,

$$\dot{\underline{x}}(t) = A(t)\,\underline{x}(t) + B(t)\,\underline{u}(t) \qquad (2.7)$$

where both A and B may be time dependent, and $\underline{u}$ is an arbitrary forcing function. The solution to (2.7) consists of a homogeneous portion, and a contribution of the forcing term. Thus,

$$\underline{x}(t) = \Phi(t, t_0)\,\underline{x}(t_0) + \int_{t_0}^{t} \Phi(t, \lambda) B(t_0) \underline{u}(\lambda)\, d\lambda \qquad \ldots\ldots (2.8)$$

assuming the interval $(t - t_0)$ is small enough to allow A and B to be considered effectively constant. The integration in (2.8) may be avoided if it can be assumed that the short time interval permits the interpretation of $\underline{u}(t)$ as an impulse input just prior to time t. Under these conditions, the solution to (2.7) may be modified to describe the stagewise evolution of $\underline{x}$ through time.

$$\underline{x}(t) = \Phi(t, t-\Delta)\,\underline{x}(t-\Delta) + B(t)\,\underline{u}(t) \qquad (2.9)$$

The state transition matrix has several properties that will be utilized in the sequel.

1. $\Phi(t, t) = I$
2. $\Phi(t_2, t_1)\Phi(t_1, t_0) = \Phi(t_2, t_0)$
3. $\Phi(t_2, t_1) = \Phi^{-1}(t_1, t_2)$

These properties may be easily verified.

2.4.2 The Kalman Solution

Kalman considers the estimation of the linear dynamic system described in the continuous form of equation (2.7) or discretely as in (2.9), and assumes that B(t) = I for all t without loss of

generality. To the dynamic system is added an observation system described in the continuous case by,

$$\underline{y}(t) = G(t)\,\underline{x}(t) + \underline{v}(t) \tag{2.10}$$

where $\underline{x}$ is an n-dimensional state vector, $\underline{y}$ is an m-dimensional observation vector, and $\underline{v}$ is an m-dimensional vector of random observation errors. If the observations are perfect, and thus $\underline{v}(t)$ is zero for all t, the Kalman form of the solution of the filtering problem requires that $m \leq n$. The matrices A and G are known functions of time, and are of dimension (n x n) and (m x n) respectively. The dynamic input vector, $\underline{u}$, can be considered to be either normal random noise, or the output of a linear system fed by independent gaussian noise. The following statistical characteristics are assumed for the noise inputs:

$$E\{\underline{u}(t)\} = E\{\underline{v}(t)\} = \underline{0} \quad \text{for all } t$$

$$E\{\underline{u}(t)\,\underline{u}^T(t)\} = Q(t)$$

$$E\{\underline{v}(t)\,\underline{v}^T(t)\} = R(t)$$

$$E\{\underline{u}(t)\,\underline{u}^T(s)\} = E\{\underline{v}(t)\,\underline{v}^T(s)\} = 0 \quad \text{for all } t \neq s$$

$$E\{\underline{v}(t)\,\underline{u}^T(s)\} = 0 \text{ for all } t \text{ and } s.$$

The estimation problem is to determine the estimate $\hat{\underline{x}}(t_1)$ of $\underline{x}(t_1)$ such that the expected loss function is a minimum. The loss function is the least-squares criterion,

$$L = E\left\{\left[\hat{\underline{x}}(t_1) - \underline{x}(t_1)\right]^T\left[\hat{\underline{x}}(t_1) - \underline{x}(t_1)\right]\right\} \tag{2.11}$$

Kalman shows that the theory applies for any convex loss function symmetric about its mean[2].

It will be recalled that the general estimation problem can be subdivided into ones of smoothing, filtering and prediction. The Kalman theory applies only to the solution of the latter two; that is, the state at time t_1 can only be estimated at time t if $t_1 \geq t$. Fortunately, filtering and prediction are of particular interest in chemical engineering applications. The smoothing problem will not be pursued further; it is considered in detail by Meditch[10].

For convenience, the following notation convention is adopted. The estimate $\hat{\underline{x}}(t)$ at time t_n based upon observations from time t_0 to t_m will be denoted as $\hat{\underline{x}}_{n/m}$. Conditioning upon observations in the interval t_p to t_m with $p < m$ will be indicated by $\hat{\underline{x}}_{n/m-p}$. The state transition matrix $\underline{\Phi}(t_n, t_m)$ will be written $\underline{\Phi}_{n,m}$. Any time dependent matrix such as A(t) defined at time t_n will be written as A_n.

The analysis by Kalman yields the following algorithm which defines the filter for the discrete case. The analogous continuous solution[11] will not be considered.

Kalman algorithm

1. $$\hat{\underline{x}}_{n/n-1} = \underline{\Phi}_{n,n-1}\, \hat{\underline{x}}_{n-1/n-1} + \underline{e}_{n,n-1} \qquad (2.12)$$

2. $$P_{n/n-1} = \underline{\Phi}^T_{n,n-1}\, P_{n-1/n-1}\, \underline{\Phi}_{n,n-1} + Q_{n-1} \qquad (2.13)$$

3. $$P_{n/n} = P_{n/n-1} - P_{n/n-1}\, G_n^T \left[G_n\, P_{n/n-1}\, G_n^T + R_n \right]^{-1} G_n\, P_{n/n-1} \qquad (2.14)$$

4. $$\hat{x}_{n/n} = \hat{x}_{n/n-1} + P_{n/n-1}\, G_n^T \left[G_n P_{n/n-1}\, G_n^T + R_n \right]^{-1} \cdot \left[\underline{y}_n - G_n\, \hat{\underline{x}}_{n/n-1} \right] \qquad (2.15)$$

In the particular problem first considered by Kalman, the known non-random dynamic input (or control), $\underline{e}_{n,n-1}$, over the interval t_{n-1} to t_n is not considered; furthermore, perfect observations are assumed so that $R_n = 0$. The information required to initialize the filter is $\hat{\underline{x}}_{0/0}$ and $P_{0/0}$; that is, an estimate of the initial state and the covariance matrix of the error in that initial estimate. The assumption of knowledge of $\hat{\underline{x}}_{0/0}$ and $P_{0/0}$ is shown[11] to remove the requirement of stationarity of the dynamic system, which is inherent in the Wiener solution. All other elements of the algorithm are assumed to be known for all $t \geq t_0$. Note that as the estimator evolves through time, the information from all past observations starting from t_0 is contained in the current estimate $\hat{\underline{x}}_{n/n}$ and the error covariance matrix $P_{n/n}$.

The work of Kalman represents a major advance in statistical filtering theory. Besides offering a practical solution to the problem solved only in theory by Wiener, it demonstrates several significant points. Kalman shows that a linear filter is optimal for linear systems with gaussian inputs, and that the optimal estimate consists of a linear function of the current state and observation vectors. Furthermore, it is demonstrated that the results obtainable by a linear filter can be improved upon by a nonlinear estimator only if non-gaussian inputs are considered, and even then only if third order probability distributions are postulated to describe the error statistics. The Kalman theory is also applicable to non-stationary dynamic systems, and as such is more general than the Wiener techniques.

The analysis of continuous linear systems by Kalman and Bucy[11] contains a rigorous mathematical analysis of the properties of the estimator, which is interpreted as an orthogonal projection in Hilbert

space. Indeed both of the Kalman papers are elegant mathematical expositions, but there is little to be gained by reviewing in detail the Kalman derivation of the algorithm. The geometric approach taken is likely to be unfamiliar to the engineer; fortunately, the results can be obtained more easily by several alternative methods.

2.5 Alternative Derivations of the Kalman Filter

The filtering problem considered by Kalman is open to several interpretations, and each one allows a different method of solution. Kalman views the problem as one of determining the conditional expectation of the system state, where the estimate is conditional on all past observations, and he exploits the geometric properties of conditional estimates to derive the filter equations. However, the interpretation of the filter output as a conditional expectation also suggests the possibility of using the Bayesian theory of statistics to generate the solution to the problem. Another approach is suggested by the fact that the estimate of the system state is a function of time which minimizes a quadratic cost related to the observations and the predicted dynamic errors. Thus the problem can be couched in terms of a minimization in function space, and analyzed through the use of classical variational techniques. Alternatively, since the cost function considered is a least-squares criterion, the solution could be sought using standard least-squares theory. Indeed, all of these approaches have been considered by various authors, and each viewpoint has its advantages and disadvantages.

Ho and Lee[12] have derived the Kalman solution directly from Bayes Rule. Their technique provides a method for calculating the evolution in time of the probability distribution of the system state conditional on the observations of the system. The estimate is defined in terms of the parameters of the distribution, and in the case of linear systems with gaussian inputs the recursive equations defining the evolution of the mean and variance of the conditional distribution are the same as those derived by Kalman. Though the theory allows the calculation of the

distribution for any nonlinear system subject to inputs with known statistics, there will in general be no way of defining the estimate recursively in terms of the characteristics of the conditional probability distribution. Thus while the Bayesian approach is satisfying to the theorist, it is only useful to the pragmatist in the simplest situations.

A variational solution to the problem was first suggested by Bellman et al.[13], and is developed using the concepts of dynamic programming. The general nonlinear estimation problem is considered and a Hamilton-Jacobi equation is derived whose solution defines the trajectory of the optimal estimate of the state from the time of initialization to the current time. A recursive solution to the equation is found using the technique of invariant imbedding, which reduces the two-point boundary value problem to an initial value one. Once again, this formulation reduces to the Kalman solution for linear systems. Bellman and his co-workers were among the first to show that unknown parameters in the system model could be estimated along with the state without altering the form of the filtering equations.

The solution of the problem from a least-squares viewpoint is perhaps the most obvious approach to the engineer. In 1950, Plackett[14] proposed a method for combining least-squares estimates from two blocks of data, and his equations can be shown to be equivalent to a Kalman filter. Swerling[15] considered the generation of sequential least-squares estimates of the parameters of satellite orbits, and arrived at the same special case of the Kalman solution. Rosenbrock[16] has recently demonstrated that Kalman's results can be derived directly from a theorem by Gauss. While the least-squares derivations are not as elegant as some others that have been mentioned earlier, they are especially useful to engineers accustomed to the techniques of least-squares analysis.

2.6 Extensions to Kalman Theory

The class of problems soluble using linear Kalman theory can be easily extended within the framework of the original formulation. The application of a non-random forcing function to the dynamic system is allowed for in the algorithm presented above, and can be seen to enter only into equation (2.12) which describes the transition of $\hat{\underline{x}}_{n-1/n-1}$ to $\hat{\underline{x}}_{n/n-1}$. Sorenson[17] has shown that systems with correlated dynamic inputs and measurement biases can be analyzed by increasing the dimensionality of the filter without altering its form. Mayne[18] considers the estimation of the elements of the state transition matrix of a linear system with perfect measurements and shows that the problem can be transformed to a standard linear problem of dimension n^2.

However, the most serious limitation of the original Kalman theory is its restriction to the analysis of linear dynamic systems. Denham and Pines[19] have shown the detrimental effect small nonlinearities can have upon the performance of a linear filter. Nevertheless, Smith et al.[20] have used the linear theory to estimate the orbital parameters of satellites, whose behaviour is governed by highly nonlinear celestial mechanical equations of motion. They assume that the perturbations of the satellite from a given reference trajectory are governed by a set of linear dynamic equations, and hence estimate the parameters using the standard linear theory. This technique has been applied successfully to many estimation problems encountered in aerospace applications, where theoretical reference states can be accurately predicted.

The applicability of the Bayesian and variational approaches to the synthesis of nonlinear filters has already been mentioned. Sorenson and Stubberud[21] derive a filter for the analysis of slightly nonlinear

systems using Bayesian techniques, relying upon the assumption that the probability distribution of the state conditioned upon the measurements remains gaussian in spite of the nonlinearities. The method can be applied successfully if the nonlinear system is essentially quadratic. Pearson and Shridar[22] derive identical results from a variational viewpoint.

Though the Bayesian and variational approaches yield identical results, there is a distinct philosophical difference between the methods. A Bayesian analysis requires definite assumptions concerning the probability distributions of the system inputs, while as Pearson[23] observes, the variational approach does not necessitate the synthesis of stochastic input models, and as such is more likely to be useful for the analysis of practical problems. It should be stressed that though a completely general solution to the nonlinear, non-gaussian filtering problem can be formulated in theory[24], the practical specification of the filter can only be determined for, at best, quadratic systems with normal inputs.

A further extension of the linear Kalman theory of special interest in chemical engineering applications has been suggested by Seinfeld[25]. He considers the synthesis of filters for systems described by linear partial differential equations. The results follow from an extension of the variational analysis, and it is interesting to note that the modified filter differs in form from the standard continuous Kalman filter[11] only in the evolution of the state estimate, while the equation describing the estimate covariance matrix remains unchanged. In this case, of course, the difficult part of the estimation procedure lies in the solution of the partial differential equations of the model, and not in the actual updating of the filter.

Further extensions of the Kalman theory have been made to improve the performance of the filter in the analysis of real physical systems. Schlee and his co-workers[26] have demonstrated the adverse effects of rounding errors and model deficiencies on the estimator when applied to simulated aerospace problems. Smith[27] has made a similar analysis which shows the effect of unknown instrument biases on the estimator performance. In general, it seems that even in the unlikely situation when the system dynamic equations are perfectly known, computational rounding errors can still cause a degradation in the behaviour of the filter.

Several attempts have been made to modify the Kalman algorithm to counteract these effects, primarily by limiting the memory of the filter, thus allowing the estimates to be based upon the most recent observations of the system. An early attempt to synthesize a limited memory filter was made by Blum[28], but his technique requires the storage of all of the observations within the memory span. Jazwinski[29] has deduced a filtering algorithm whose memory oscillates between n and 2n sets, and requires twice the storage of the standard growing memory filter. The choice of the memory parameter n depends upon the nature of the system being observed. The filter output is characterized by a discrete shift in the estimate as the memory basis shifts from 2n to n records, and the detection of a pattern in successive shifts can often be used to suggest the form of the model deficiency. An alternative technique for reducing the influence of early data sets has been suggested by Fagin[30], who applies an exponentially decaying weighting factor to the memory, and specifies a filter requiring no more storage than the Kalman algorithm. A practical analysis of the performance of these limited memory techniques has not yet been reported.

2.7 Concluding Remarks

No effort has been made to include all of the contributions that have been made to the development of the linear filtering theories of Wiener and Kalman. Rather, an attempt has been made to reconcile the various approaches that have been taken to derive the results of the theory. The particular viewpoint taken in analyzing the filtering problem is clearly a matter of personal preference, and in practice, all methods yield identical algorithms.

In order to provide a sound basis for the application of the filtering techniques to chemical engineering problems, a unified derivation of both the growing memory (Kalman) and limited memory filters is presented in the following chapter. The least-squares approach is taken, since in most practical engineering applications, stochastic models for input and measurement disturbances are not available.

CHAPTER 3

DERIVATION OF THE FILTER EQUATIONS

3.1 Introduction

The discussion in Chapter 2 has indicated the variety of techniques which can be used to derive the recursive equations of the optimal filter. The approach taken here is inspired by Rosenbrock[16], who has shown that the Kalman filtering equations can be deduced directly from a theorem by Gauss on the characteristics of least-squares estimators. However, the present derivation differs from earlier efforts because it yields the recursive relationships describing not only the standard growing memory filter of Kalman[2], but also the limited memory filters derived using different methods by Jazwinski[29] and Fagin[30]. In order to allow this more general derivation, the dynamic input vector is restricted to be a deterministic function of time; hence the emphasis of the estimation problem is shifted to the determination of the system state from a set of noisy instrument readings, rather than estimation in the face of random input disturbances. It will become evident in the course of the analysis that random dynamic inputs preclude the use of certain types of limited memory filters unless every data set within the memory is stored.

3.2 The General Non-linear Problem

The filtering theory derived by Kalman[2] applies to a class of systems described by linear dynamic equations and observed by error free measurements. The linear theory can be applied to a more general class of problems by linearizing the system about the current best estimate of its state. Consider the following dynamic model:

$$\dot{\underline{x}}^{*}(t) = \underline{f}\left[\underline{x}^{*}(t), \underline{\alpha}(t), \underline{u}(t), t\right] \tag{3.1}$$

$$\underline{y}^{*}(t) = \underline{g}\left[\underline{x}^{*}(t), \underline{\alpha}(t), \underline{v}(t), t\right] \tag{3.2}$$

$$\dot{\underline{\alpha}}(t) = \underline{0}. \tag{3.3}$$

where $\underline{x}^{*}$ is an n-dimensional state vector, $\underline{y}^{*}$ is an m-vector of observations, $\underline{u}$ and $\underline{v}$ are random disturbance vectors, $\underline{\alpha}$ is a vector of unknown, constant parameters, and $\underline{f}$ and $\underline{g}$ are known functions. Any deterministic inputs to (3.1) and (3.2) can be accounted for by the time dependency.

For this analysis, the general problem is modified by assuming that the observation errors are additive, and by considering only deterministic dynamic inputs. The significance of excluding random inputs to the dynamic system will be discussed later. The problem is further simplified by adjoining the parameter vector to the state vector, so that $\underline{x}^{*}(t)$ is redefined to allow estimation of any unknown parameters in the system. The modified system description is :

$$\dot{\underline{x}}^{*}(t) = \underline{f}\left[\underline{x}^{*}(t), t\right] \tag{3.4}$$

$$\underline{y}^{*}(t) = \underline{g}\left[\underline{x}^{*}(t), t\right] + \underline{v}(t) \tag{3.5}$$

where the random observation error, $\underline{v}(t)$, has zero mean and known covariance matrix R(t).

The equations can be expanded in a Taylor series expansion about the current optimal estimate $\hat{\underline{x}}^{*}(t)$ to yield, to the first order :

$$\dot{\underline{x}}^{*}(t) = \underline{f}\left[\hat{\underline{x}}^{*}(t),\ t\right] + \left.\frac{\partial \underline{f}\left[\underline{x}^{*}(t),\ t\right]}{\partial \underline{x}^{*}(t)}\right|_{\hat{\underline{x}}^{*}(t)} \left[\underline{x}^{*}(t) - \hat{\underline{x}}^{*}(t)\right] \qquad (3.6)$$

$$\underline{y}^{*}(t) = \underline{g}\left[\hat{\underline{x}}^{*}(t),\ t\right] + \left.\frac{\partial \underline{g}\left[\underline{x}^{*}(t),\ t\right]}{\partial \underline{x}^{*}(t)}\right|_{\hat{\underline{x}}^{*}(t)} \left[\underline{x}^{*}(t) - \hat{\underline{x}}^{*}(t)\right] + \underline{v}(t) \qquad \ldots\ldots (3.7)$$

The nonlinear system of equations describing the state and observations can therefore be converted into a linear set describing the evolution of the errors in the estimation of the reference state. Defining

$$\left.\begin{aligned} \underline{x}(t) &= \underline{x}^{*}(t) - \hat{\underline{x}}^{*}(t) \\ \underline{y}(t) &= \underline{y}^{*}(t) - \underline{g}\left[\hat{\underline{x}}^{*}(t),\ t\right] \end{aligned}\right\} \qquad (3.8)$$

the converted problem can be written,

$$\dot{\underline{x}}(t) = F(t)\,\underline{x}(t) \qquad (3.9)$$

$$\underline{y}(t) = G(t)\,\underline{x}(t) + \underline{v}(t) \qquad (3.10)$$

where F and G are the Jacobian matrices in the series expansions (3.6) and (3.7) respectively.

The equation describing the error propagation with respect to time can now be converted to discrete state-vector notation according to the assumptions of section 2.4.1. Recalling the notation convention described in that section, it follows that :

$$\underline{x}_k = \Phi_{k,j}\,\underline{x}_j \tag{3.11}$$

$$\underline{y}_k = G_k\,\underline{x}_k + \underline{v}_k \tag{3.12}$$

where $\Phi_{k,j}$ is the appropriate state-transition matrix.

It should be stressed that this method of analysis of a nonlinear system assumes that the deviations from the reference state can be described by a set of linear dynamic equations. However, the reference state itself remains nonlinear.

3.3 The Filtering Equations

3.3.1 The basis in Least-Squares Theory

The derivation is based on the following theorem on the properties of least-squares estimators, which is shown by Plackett[14] and by Kendall and Stuart[5] to follow directly from work by Gauss.

Least-squares theorem

Given a set of observations $\underline{y}$ related to a state vector $\underline{x}$ by the equations :

$$\underline{y} = C\underline{x} + \underline{v} \tag{3.13}$$

where $$R = E\left\{\underline{v}\,\underline{v}^T\right\} \text{ and } E\left\{\underline{v}\right\} = \underline{0}. \tag{3.14}$$

then the estimate $$\hat{\underline{x}} = PC^T R^{-1}\,\underline{y} \tag{3.15}$$

where $$P = E\left\{(\hat{\underline{x}} - \underline{x})(\hat{\underline{x}} - \underline{x})^T\right\} = \left[C^T R^{-1} C\right]^{-1} \tag{3.16}$$

has the following properties :

(a) It minimizes $(\underline{y} - C\underline{x})^T R^{-1}(\underline{y} - C\underline{x})$.

(b) It minimizes $E\left\{(\hat{\underline{x}} - \underline{x})^T Q(\hat{\underline{x}} - \underline{x})\right\}$ for any positive semi-definite Q.

(c) It minimizes all elements of P.

3.3.2 The Derivation of the Algorithm

The state of the dynamic system described in (3.11) and (3.12) can be estimated using the theorem, by writing $\underline{y}_k$ in terms of $\underline{x}_j$, the unknown state at time t_j. Thus,

$$\underline{y}_k = G_k \Phi_{k,j}\,\underline{x}_j + \underline{v}_k \tag{3.17}$$

In general, an entire set of measurements can be related to this unknown state. That is,

$$\begin{bmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_k \\ \vdots \\ \underline{y}_N \end{bmatrix} = \begin{bmatrix} G_1 & & & \\ & \ddots & & \\ & & G_k & \\ & & & \ddots \\ & & & & G_N \end{bmatrix} \cdot \begin{bmatrix} \Phi_{1,j} \\ \vdots \\ \Phi_{k,j} \\ \vdots \\ \Phi_{N,j} \end{bmatrix} \underline{x}_j + \begin{bmatrix} \underline{v}_1 \\ \vdots \\ \underline{v}_k \\ \vdots \\ \underline{v}_N \end{bmatrix} \tag{3.18}$$

$$\text{or} \quad \underline{Y} = G \Phi_j \underline{x}_j + \underline{V} \tag{3.18a}$$

Note that the observation vector $\underline{y}_i$ at time t_i can have an arbitrary number of components, as long as G_i is appropriately dimensioned.

Equation (3.18) is in the form of (3.13) and the best estimate $\hat{\underline{x}}_{j/N}$ of $\underline{x}_j$ based upon N sets of data is thus given by equation (3.15).

$$\hat{\underline{x}}_{j/N} = P_{j/N} \Phi_j^T G^T R^{-1} \underline{Y} \tag{3.19}$$

where

$$P_{j/N} = E\left\{(\hat{\underline{x}}_{j/N} - \underline{x}_j)(\hat{\underline{x}}_{j/N} - \underline{x}_j)^T\right\}$$

$$= \left[\Phi_j^T G^T R^{-1} G \Phi_j\right]^{-1} \tag{3.20}$$

and unsubscripted R is the covariance matrix of $\underline{V}$. The corresponding best estimate $\hat{\underline{x}}_{k/N}$ of $\underline{x}_k$ is predicted by the system model (3.11),

$$\hat{\underline{x}}_{k/N} = \Phi_{k,j} \hat{\underline{x}}_{j/N} \tag{3.21}$$

Using the definition of P from (3.16) and the prediction of (3.21), it follows that :

$$P_{k/N} = \Phi_{k,j} P_{j/N} \Phi_{k,j}^T \tag{3.22}$$

If the observation errors, $\underline{v}_k$, are uncorrelated between samples, R is block diagonal, with blocks :

$$R_k = E\left\{ \underline{v}_k \underline{v}_k^T \right\} \tag{3.23}$$

Under these circumstances, (3.19) can be rewritten :

$$\hat{\underline{x}}_{j/N} = P_{j/N} \begin{bmatrix} \Phi_{1,j} \\ \vdots \\ \Phi_{N,j} \end{bmatrix}^T \cdot \begin{bmatrix} G_1 & & \\ & \ddots & \\ & & G_N \end{bmatrix}^T \cdot \begin{bmatrix} R_1 & & \\ & \ddots & \\ & & R_N \end{bmatrix}^{-1} \begin{bmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_N \end{bmatrix} \tag{3.24}$$

$$\text{or} \quad \hat{\underline{x}}_{j/N} = P_{j/N} \sum_{k=1}^{N} \Phi_{k,j}^T G_k^T R_k^{-1} \underline{y}_k \tag{3.24a}$$

Also, from equation (3.20) :

$$P_{j/N}^{-1} = \sum_{k=1}^{N} \Phi_{k,j}^T G_k^T R_k^{-1} G_k \Phi_{k,j} \tag{3.25}$$

Note that no restriction has been placed upon the relationship between t_k and t_j in the above equations, so the results apply for smoothing, filtering, and prediction. However, if there had been a random dynamic input to the system, this would not be the case. Under those circumstances, both $\underline{y}_k$ and $\underline{x}_k$ would be random variables, and in equation (3.18), the vectors $\underline{y}_i$ for $i \geq j$ would be correlated with $\underline{x}_j$. This means that for $i \geq j$, the observation error, $\underline{v}_i$, has extra components which complicate the analysis. Though it is possible to generalize this analysis to allow for random dynamic inputs, this will not be done, since this study is intended to provide techniques for interpreting noisy process observations, and is not directly concerned with the effects of random input disturbances.

A recursive relationship for combining two blocks of data sets can be deduced from (3.24) and (3.25). The summations can be segmented so that :

$$P^{-1}_{j/N}\,\hat{\underline{x}}_{j/N} = P^{-1}_{j/M}\,\hat{\underline{x}}_{j/M} + P^{-1}_{j/N-M}\,\hat{\underline{x}}_{j/N-M} \tag{3.26}$$

$$\text{and} \qquad P^{-1}_{j/N} = P^{-1}_{j/M} + P^{-1}_{j/N-M} \tag{3.27}$$

where N is the union of sets M and N-M.

Note that the set M can be any subset of the set N, and then the subset N-M will be its complement. In general, however, the formulae will be used for a limited memory filter, and will involve consecutive sequences within each block.

When the subset M is a single observation vector, say $\underline{y}_r$, then the equations (3.24a) and (3.25) reduce to the following :

$$P^{-1}_{j/1}\,\hat{\underline{x}}_{j/1} = \Phi^{T}_{r,j}\,G^{T}_{r}\,R^{-1}_{r}\,\underline{y}_r \tag{3.28}$$

$$\text{and} \qquad P^{-1}_{j/1} = \Phi^{T}_{r,j}\,G^{T}_{r}\,R^{-1}_{r}\,G_r\,\Phi_{r,j} \tag{3.29}$$

where the dependency j/1 implies estimation on the basis of one data set. These results can be used in conjunction with (3.26) and (3.27) for adding or dropping a single data set, or as an initial estimate. However, the removal of data from the filter memory one set at a time requires the storage of the right-hand sides of (3.28) and (3.29), or equivalently, the storage of the appropriate data sets for the duration of their retention in the filter memory.

Equations (3.26) and (3.27) form the basis of the limited memory filtering scheme derived by Jazwinski[29] using Bayesian techniques. $P^{-1}_{j/M}$, $P^{-1}_{j/M}\,\hat{\underline{x}}_{j/M}$, $P^{-1}_{j/N}$, and $P^{-1}_{j/N}\,\hat{\underline{x}}_{j/N}$ are generated by recursion, adding one set at a time through N sample intervals, and then the information in the first M sets is dropped by obtaining $P^{-1}_{j/N-M}$ and $P^{-1}_{j/N-M}\,\hat{\underline{x}}_{j/N-M}$ from (3.26) and (3.27). In practice N is chosen as a multiple of M to make the storage requirements reasonable, for if $N = qM$, it is necessary to store q sets of P^{-1}_{j} and $P^{-1}_{j}\,\hat{\underline{x}}_{j}$ at any one time.

In most applications, the current state is likely to be of more interest than some fixed previous state, and the relevant formulae may be obtained from the above. From (3.24a), letting $j = N$:

$$\hat{\underline{x}}_{N/N} = P_{N/N} \sum_{k=1}^{N} \Phi^{T}_{k,N}\, G_k^T\, R_k^{-1}\, \underline{y}_k \tag{3.30}$$

and from (3.25),

$$P^{-1}_{N/N} = \sum_{k=1}^{N} \Phi^{T}_{k,N}\, G_k^T\, R_k^{-1}\, G_k\, \Phi_{k,N} \tag{3.31}$$

Equation (3.22) can be modified using property (3) of the state transition matrix (see Section 2.4.1) which states that $\Phi_{k,j} = \Phi^{-1}_{j,k}$. Thus :

$$P^{-1}_{k/N} = \Phi^{T}_{j,k}\, P^{-1}_{j/N}\, \Phi_{j,k} \tag{3.32}$$

By combining (3.32) and (3.27) a method is provided to drop the earliest M sets of data from the filter memory, assuming that $P^{-1}_{M/M}$ and $P^{-1}_{M/M}\,\hat{\underline{x}}_{M/M}$ have been stored, and that the data sets have been acquired sequentially. Hence,

$$P^{-1}_{N/N} = \Phi^{T}_{M,N}\, P^{-1}_{M/M}\, \Phi_{M,N} + P^{-1}_{N/N-M} \tag{3.33}$$

Similarly, using (3.26) with (3.32) :

$$P^{-1}_{N/N}\,\hat{\underline{x}}_{N/N} = \Phi^{T}_{M,N}\,P^{-1}_{M/M}\,\Phi_{M,N}\,\hat{\underline{x}}_{N/M} + P^{-1}_{N/N-M}\,\hat{\underline{x}}_{N/N-M}$$

$$= \Phi^{T}_{M,N}\,P^{-1}_{M/M}\,\hat{\underline{x}}_{M/M} + P^{-1}_{N/N-M}\,\hat{\underline{x}}_{N/N-M} \qquad (3.34)$$

When the filter is being operated in real time, a single data set is more likely to be added than a block of sets. Earlier results can be simplified by setting j = r = N and M = N-1, noting that $\Phi_{N,N} = I$ from (3.11).

From (3.29) and (3.33), it follows that :

$$P^{-1}_{N/N} = \Phi^{T}_{N-1,N}\,P^{-1}_{N-1/N-1}\,\Phi_{N-1,N} + G^{T}_{N}\,R^{-1}_{N}\,G_{N} \qquad (3.35)$$

Also, from (3.26) and (3.27),

$$P^{-1}_{N/N}\,\hat{\underline{x}}_{N/N} = (P^{-1}_{N/N} - P^{-1}_{N/1})\,\hat{\underline{x}}_{N/N-1} + P^{-1}_{N/1}\,\hat{\underline{x}}_{N/1} \qquad (3.36)$$

The desired recursion is found by substituting the definitions of $P^{-1}_{N/1}$ and $P^{-1}_{N/1}\,\hat{\underline{x}}_{N/1}$ in (3.36). Thus,

$$P^{-1}_{N/N}\,\hat{\underline{x}}_{N/N} = P^{-1}_{N/N}\,\hat{\underline{x}}_{N/N-1} + G^{T}_{N}\,R^{-1}_{N}(\underline{y}_{N} - G_{N}\,\hat{\underline{x}}_{N/N-1})$$

$$\text{or} \qquad \hat{\underline{x}}_{N/N} = \hat{\underline{x}}_{N/N-1} + P_{N/N}\,G^{T}_{N}\,R^{-1}_{N}\,(\underline{y}_{N} - G_{N}\,\hat{\underline{x}}_{N/N-1}) \qquad (3.37)$$

It is sometimes reasonable to assume that current data sets are more relevant for state estimation than sets acquired early in the operation of the system, especially in cases when the reference model is only locally valid. In these situations it is possible to apply exponentially decreasing weights to the old data sets as new ones are

acquired. The weighting factor can be interpreted as an increase in the uncertainty in the measurement vector as it becomes older. If at time t_k, the covariance matrix of $\underline{y}_k$ is R_k, at time t_{k+1}, the elements of the matrix can be assumed to have increased so that every element, r_{ij}, obeys the relationship :

$$\left\{r_{ij}\right\}_{t_{k+1}} = \left\{r_{ij}\right\}_{t_k} \cdot \frac{1}{\sigma} \qquad (3.38)$$

where $0 \leq \sigma \leq 1$.

Equations (3.24a) and (3.25) become :

$$P^{-1}_{j/N}\,\hat{\underline{x}}_{j/N} = \sum_{k=1}^{N} \sigma^{N-k} \left[\Phi^{T}_{k,j}\, G^{T}_{k}\, R^{-1}_{k}\, \underline{y}_k\right] \qquad (3.39)$$

and

$$P^{-1}_{j/N} = \sum_{k=1}^{N} \sigma^{N-k} \left[\Phi^{T}_{k,j}\, G^{T}_{k}\, R^{-1}_{k}\, G_k\, \Phi_{k,j}\right] \qquad (3.40)$$

The derivation of the recursive relationships for the exponentially weighted filter follows as before from (3.39) and (3.40).

An algorithm can thus be constructed to update the estimate of the system state given the current best estimate, $\hat{\underline{x}}_{N-1/N-1}$, the current estimate of the inverse of the estimation error covariance matrix, $P^{-1}_{N-1/N-1}$, and the new observation $\underline{y}^{*}_{N}$, assuming the dynamic model of the system is known.

Least-squares algorithm

1) $\hat{\underline{x}}_{N/N-1} = \Phi_{N,N-1}\, \hat{\underline{x}}_{N-1/N-1}$ from (3.11)

2) $P^{-1}_{N/N-1} = \Phi^{T}_{N-1,N}\, P^{-1}_{N-1/N-1}\, \Phi_{N-1,N}$ from (3.32)

3) $P_{N/N}^{-1} = \alpha P_{N/N-1}^{-1} + G_N^T R_N^{-1} G_N$ from (3.32) and (3.35)

4) $\hat{\underline{x}}_{N/N} = \hat{\underline{x}}_{N/N-1} + P_{N/N} G_N^T R_N^{-1} (\underline{y}_N - G_N \hat{\underline{x}}_{N/N-1})$ from (3.37)

An a priori estimate of the state error and its covariance matrix, $\hat{\underline{x}}_{0/0}$ and $P_{0/0}$, is required to start the filtering procedure, and can, for example, be obtained from (3.28) and (3.29) when the initial measurement vector, $\underline{y}_1$, provides at least a minimal data set.

Note that step 3 includes the exponential weighting factor. Thus, this algorithm is suitable for both growing memory, and exponentially weighted memory filters.

The same algorithm is used as the basis of the oscillating memory filter proposed by Jazwinski[29]. The following procedure is used for the case when N = 2M.

Oscillating memory algorithm

1) Operate the least squares algorithm normally during the acquisition of the first M data sets, and store $P_{i/M}^{-1}$ and $\hat{\underline{x}}_{i/M}$, where $\hat{\underline{x}}$ at time t_i is being estimated.

2) Update the estimate using the least-squares algorithm for M additional data sets to generate $P_{i/2M}^{-1}$ and $\hat{\underline{x}}_{i/2M}$, which are $P_{i/N}^{-1}$ and $\hat{\underline{x}}_{i/N}$ by definition.

3) Using (3.33) calculate $P_{i/N-M}^{-1}$ and invert to obtain $P_{i/N-M}$.

4) Using (3.34), calculate $\hat{\underline{x}}_{N/N-M}$.

5) Since the estimates are now conditioned on the last M data sets (N-M=M), redefine $P_{i/M}^{-1}$ and $\hat{\underline{x}}_{i/M}$ from the results of steps 3 and 4, and return to step 2.

Note that these algorithms provide optimal estimates for the error in the state, which must be added to the reference trajectory to determine the actual state at any time. The method of calculation of the reference trajectory depends upon the system being considered, but usually consists of numerical integration of a set of nonlinear differential equations with initial conditions and unknown parameters provided by the estimator. The effects of any deterministic inputs, such as control, are accounted for in the calculation of the reference trajectory.

3.3.3 The Relationship with the Kalman Solution

The least-squares algorithm presented above reduces to the one derived by Deutsch[4] for the sequential estimation of unknown constants, where $\Gamma_{N,N-1} = I$. The equivalence is not surprising, since his method was derived originally by Swerling[15] using the characteristics of the stationary point of a quadratic cost criterion. However, the relationship between the Kalman[2] algorithm and the least-squares one is not as obvious.

The two algorithms can indeed be shown to be equivalent through the use of a matrix inversion lemma cited by Sorenson[17], and generalized here for use when exponential weighting is considered.

Matrix inversion lemma

If $A^{-1} = c\,B^{-1} + H^T R^{-1} H$

where A and B are (n x n) matrices, R is (m x m) and H is (m x n) with c scalar,

then $A = \frac{1}{c} B - \frac{1}{c^2} BH^T \left[\frac{1}{c} HBH^T + R\right]^{-1} HB$

Using the lemma, step 3 of the least-squares algorithm becomes :

$$P_{N/N} = \frac{1}{c} P_{N/N-1} - \frac{1}{c^2} P_{N/N-1} G_N^T \left[\frac{1}{c} G_N P_{N/N-1} G_N^T + R_N \right]^{-1} \cdot G_N P_{N/N-1} \quad (3.41)$$

Therefore it follows that the third steps of the two algorithms are equivalent (see equation 2.14), for c = 1.0. The algorithms are identical if the equivalence of their final steps can be demonstrated. Thus it must be shown that :

$$P_{N/N} G_N^T R_N^{-1} = P_{N/N-1} G_N^T \left[G_N P_{N/N-1} G_N^T + R_N \right]^{-1} \quad (3.42)$$

Denoting the right hand side of the preceding equation as K_N,

$$K_N = P_{N/N} P_{N/N}^{-1} P_{N/N-1} G_N^T R_N^{-1} \left[G_N P_{N/N-1} G_N^T R_N^{-1} + I \right]^{-1} \quad \ldots\ (3.43)$$

Using the definition of $P_{N/N}^{-1}$ from the least-squares algorithm, with c = 1.0,

$$K_N = P_{N/N} \left[P_{N/N-1}^{-1} + G_N^T R_N^{-1} G_N \right] P_{N/N-1} G_N^T R_N^{-1} \cdot \left[G_N P_{N/N-1} G_N^T R_N^{-1} + I \right]^{-1}$$

$$= P_{N/N} \left[I + G_N^T R_N^{-1} G_N P_{N/N-1} \right] G_N^T R_N^{-1} \left[G_N P_{N/N-1} G_N^T R_N^{-1} + I \right]^{-1}$$

$$= P_{N/N} G_N^T R_N^{-1} . \quad (3.44)$$

Q.E.D.

Note that random dynamic inputs have not been considered, so Q_{N-1} of step 2 in the Kalman algorithm is a null matrix, and the algorithms are identical.

3.3.4 Convergence of the Least-Squares Filter

Kalman and Bucy[11] present an elegant proof of the stability and convergence of the Kalman filter in its continuous form. They demonstrate that the optimal estimate generated is unbiased, and furthermore that the filter is uniformly asymptotically stable[31] and that $P_{N/N}$ approaches $\bar{P}$ as $N \to \infty$ for any non-negative $P_{0/0}$ as long as $\hat{x}_{0/0} = x_0$, the true state at time t_0. $\bar{P}$ is the covariance matrix of the optimal error based on an arbitrarily long record of past measurements, and generated assuming perfect knowledge of the initial state. The proof requires that the linear system is controllable and observable, and that the norms of Q, R, and $\Phi_{N/N-1}$ are bounded. Controllability implies that all of the states can be excited by permissible dynamic inputs while observability means that all states can be estimated by a finite number of perfect measurements. Since the only difference between the Kalman and least-squares algorithms lies in the fact that in the latter case Q is a null matrix (and therefore bounded), the proof still applies if all other conditions are met.

Deutsch[4] demonstrates that corrections to the estimate made by the unweighted least-squares filter will approach zero as the number of observations increase. However, this is not so for the exponentially weighted case, even in the simplest situation when R_N and G_N are constant for all N, and $\Phi_{N,N-1} = I$. In that case,

$$P_{N/N}^{-1} = c\, P_{N-1/N-1}^{-1} + G_N^T R_N^{-1} G_N \qquad (3.45)$$

from steps 2 and 3 of the least-squares algorithm. Rewriting this as :

$$P_N^{-1} = c\, P_{N-1}^{-1} + A^T A \qquad (3.46)$$

where $A = R_N^{-\frac{1}{2}} G_N$ and R_N is assumed positive definite, it follows that,

by premultiplying (3.46) by P_N and post-multiplying by $P_{N-1} A^T$,

$$P_N A^T = P_{N-1} A^T \left[A P_{N-1} A^T + cI \right]^{-1} \tag{3.47}$$

Successive substitutions for $P_{N-i} A^T$ in terms of $P_{N-i-1} A^T$ yields

$$P_N A^T = P_1 A^T \left[c^{N-1} I + (1 + c + c^2 + \ldots c^{N-2}) A P_1 A^T \right]^{-1} \tag{3.48}$$

Since $\lim_{i \to \infty} c^i = 0$ and $\sum_{i=0}^{\infty} c^i = \frac{1}{1-c}$ for $0 \leq c < 1$, it follows that

$$\lim_{N \to \infty} P_N A^T = (1-c) P_1 A^T \left[A P_1 A^T \right]^{-1} \tag{3.49}$$

$$\text{for } 0 \leq c < 1.$$

It can be seen from step 4 of the least-squares algorithm that,

$$\left| \hat{x}_{N/N} - \hat{x}_{N/N-1} \right| \propto P_N A^T \tag{3.50}$$

The correction to the estimate becomes zero in the limit only if $c = 1$. Therefore, the exponentially weighted filter will not converge to a constant estimate, and care must be exercised in its use to assure that the chosen weighting factor c does not cause instability.

The preceding discussion is not intended to be a complete analysis of the convergence properties of the least-squares filter. Indeed, even the rigorous analysis by Kalman and Bucy[11] applies only to linear systems, and cannot be extended to apply to linearized systems. The performance of the filter in the analysis of practical nonlinear problems cannot be predicted in theory, but must be evaluated by experimentation.

3.4 Comments on the Filtering Algorithms

A recursive filter has been derived for application to linear dynamic systems, and a technique has been suggested which, under certain circumstances, will permit the analysis of nonlinear systems. Before the practical application of the filtering techniques can be investigated, the assumptions which have been made in deriving the equations should be analysed to determine any limitations of the theory. Any practical problem can then be analysed in the light of these limitations.

The algorithms derived apply only to the analysis of linear dynamic systems. The linearization of the general problem can be justified if it is reasonable to assume that the perturbations from a given reference trajectory are small, and can be described by a set of linear dynamic equations. Of course, it is assumed that it is possible to postulate a reasonable reference state suitable for on-line computation.

The theory presented requires that the random observations errors have zero mean. This is not restrictive, since the system can have deterministic elements, and therefore instrument bias or drift can also be postulated. These effects, in addition to the random errors, should be adequate to describe any anticipated measurement errors.

The restriction to non-random dynamic inputs has been made to permit specification of limited memory filters. It can be seen, by comparing the Kalman and least-squares algorithm, that the random inputs only affect the calculation of $P_{N/N-1}$ from $P_{N-1/N-1}$. The extra term, however, prevents the expression of $P^{-1}_{j/N}$ as a simple summation in equation (3.25), and thus prevents the specification of an oscillating memory filter. The exponential weighting technique could still be applied, but the interpretation of the weighting factor would be modified. If there were

random dynamic inputs the uncertainty in old inputs as well as of old observations would increase in time.

Limited memory filtering is desirable in situations when the exact mathematical model of the system being observed is not perfectly known, and the model used can only be considered accurate for part of the operating period. In those cases, estimation on the basis of a recent subset of observations is preferred, and either the oscillating memory filter (OMF) or the exponentially weighted filter (EWF) could be used. Both forms require the specification of memory parameters, in the case of the OMF, the parameter M defines the record length, while c, the weighting factor, determines the rate of memory decay of the EWF. The values chosen for M and c depend upon the accuracy of the reference model, and should reflect the number of sample intervals for which the model can be assumed valid. The EWF is easier to apply than the OMF, since the latter requires an extra matrix inversion after every M^{th} observation, as shown in step 3 of the OMF algorithm. However, the OMF is characterized by discrete changes in the estimate after modification of the memory basis, and thus can sometimes be used to indicate model deficiencies. The relative desirability of the form chosen depends upon the application.

A final comment should be made on the equivalence of the Kalman and least-squares filters. Though the algorithms have been shown to be identical in theory, one form may be more convenient for use in a particular situation. This can be seen by comparing the final steps in the algorithms, which update $\hat{\underline{x}}_{N/N-1}$ to $\hat{\underline{x}}_{N/N}$. The Kalman form requires the inversion of a matrix of the same dimension as the observation vector, while the least-squares algorithm requires an n-dimensional inversion of $P^{-1}_{N/N}$ during each cycle. It follows that the least-squares

formulation is more convenient if there are more measuring devices than state variables. However, in most cases the instrument errors are uncorrelated within a set and R_k is diagonal for all k. In those situations the Kalman form should be used, since each element of the observation vector can be incorporated independently into the estimate, and the inversion in (2.15) is scalar.

It should be noted that the least-squares formulation does not allow any perfect observations, for then R_k would be singular. However, in that situation, the Kalman derivation is still valid, if all observations are linearly independent and thus G_N is full rank. This assures that $G_N^T P_{N/N-1} G_N$ remains non-singular.

No matter what form of the estimator is used, a certain minimum amount of statistical information is required to initialize the filter, and update the estimate as new observations are made. This information includes not only the initial estimate of the state system, which is likely to be known, but also an estimate of the covariance matrix of the error in that state, which is often difficult to specify. In addition, a covariance matrix must be postulated for the instrument errors at every sampling instant. The sensitivity of the filter performance to errors in this input information as well as to modelling errors must be determined in order to assess its usefulness in practice.

CHEMICAL ENGINEERING APPLICATIONS

CHAPTER 4

INTRODUCTORY COMMENTS ON CHEMICAL ENGINEERING APPLICATIONS

4.1 Systems Modelling and Parameter Estimation

The estimation techniques presented in Part I have been applied successfully to several aerospace problems[32]. The systems dealt with have certain characteristics which make them ideal for the application of the theory. The dynamic equations governing the motion of space vehicles are perfectly known, and the effects of any control inputs to the system can be exactly calculated. Furthermore, the vehicles are not subject to random dynamic disturbances, and are observed by tracking stations whose statistical characteristics are well known. Thus all of the information required for the operation of the optimal filter is available.

The systems encountered in chemical engineering are less well defined, and the application of estimation theory is more difficult. Dynamic models of large industrial processes are rarely known to any degree of accuracy; even when theoretical models can be proposed, they often consist of sets of partial differential equations which are impractical for repetitive on-line calculations. Moreover, these distributed parameter systems can only be described in infinite dimensional state-space. The situation is further complicated by the fact that the input disturbances to the plant are rarely known, either in stochastic or deterministic terms.

However, the advantages of modern control techniques have been appreciated, and a great deal of effort has been directed towards the determination of reliable models for process elements. In several

papers, Eykhoff[33,34,35] has reviewed recent advances in modelling techniques. Cuneod and Sage[36] and Balakrishnan and Peterka[37] have also described several alternative methods for determining process characteristics. There are two major approaches to the problem, one based upon the physical laws governing the process, and the other seeking to fit local models of specific mathematical form which may be unrelated to the actual physics of the system. Into this second class fall the well known techniques of linear analysis requiring the input of known distrubances to the process being identified. No attempt will be made to evaluate the relative desirability of the approaches, except to note that it is generally advisable to take advantage of any physical information that is available about the process.

A distinction must be made between the problems of determining the form of a model to describe a plant, and that of actually estimating the parameters of the model once its form is specified. It is the on-line solution of the latter problem that is considered here; the form of the mathematical model describing the plant will be assumed known. The theory developed in Part I can be used to determine parameters for any type of system description of known form whether it is physically based or merely a regression model.

4.2 Previous Applications of Estimation Theory in Chemical Engineering

When the describing equations of a chemical process are perfectly known, the theory developed by Kalman is directly applicable to chemical engineering problems. Seinfeld[25] has applied the theory to a very simple heat conduction problem, where the system is governed by a known linear partial differential equation, and the initial temperature profile is being estimated. More recently, he has shown with Gavalas[38] that the filtering technique can be applied to a reactor with catalyst decay to monitor the change in catalyst activity. A concurrent study of a similar application to a more realistic problem is reported in Chapter 6.

Coggan and Noton[39] have recently applied Kalman filtering methods to a simulated mixing system and a simulated furnace to demonstrate their applicability to typical chemical engineering processes. The applications are based directly upon the ordinary differential equations describing the system. In order to generate a state transition matrix from the linearized estimation error equations, the Jacobian matrix, F, as in equation (3.9) is assumed constant over the sample interval; thus the transition matrix is obtained from the exponentiation of F. Though this is quite reasonable in theory, the practical problem of calculating repetitively a matrix exponential in a realistic situation is likely to be prohibitive.

Off-line estimation techniques have been applied to many chemical engineering problems, primarily to discriminate between alternative kinetic models, and to fit parameters for those models. The extensive body of research that has emanated from the University of Wisconsin due to Box, Hunter, Draper, Mezaki and many others will not be reviewed.

The techniques deal in general with the design of sequential laboratory experiments to determine the best least-squares parameters with a minimum of effort.

An off-line technique more closely related mathematically to the filtering methods has recently been proposed by Lee[40] in the chemical engineering literature. The technique is called quasilinearization, and is based upon the theory of dynamic programming and invariant imbedding due to Bellman. It is a very powerful method for fitting parameters in ordinary or partial differential equations.

Because of the difficulty of modelling chemical plants, there have been few applications of on-line estimation techniques. Bray and his co-workers[41] have used a sequential least-squares filter to update the regression model of a water-gas shift reactor. The filter used can be shown to be identical to the exponentially weighted filter developed in Part I. Their results seem quite encouraging, though not enough information was presented to fully evaluate the filter performance. Furthermore, as presented, the filter seems to be restricted to use in regression type models, though this has been shown in Part I not to be the case.

Several other investigators have considered the use of linear regression models for on-line control, though the model parameters are usually not adjusted after initial selection. Astrom and Bohlin[42] present a method for the off-line fitting of parameters for general linear models, though it is intimated that the least-squares minimization proposed could be performed sequentially. Their technique has been applied successfully to the description of a paper-making plant[43]. An alternative approach is suggested by Dahlin[44], who determines the poles and zeros of the linear transfer functions of the same type of process.

He uses a standard filter, as well as one operating on the integrated process output. The integration has the effect of removing high frequency noise, and for linear processes could be effected by prefiltering of input and output data. The approach of Astrom and Bohlin is in fact equivalent in theory to the one of Dahlin, since Dahlin deals with the frequency domain representation of the linear difference model.

The same type of linear model is proposed by Box and Jenkins[45], and has been used in the optimization of a water-gas shift reactor by Price and Rippin[46], as well as by Wise[47]. The model remained fixed after off-line specification, and was used in conjunction with an extremum seeking controller.

An industrial use of on-line estimation techniques has been reported by Noton and Choquette[48]. They have applied the Kalman theory directly to an operating plant described in terms of a twelve dimensional state-vector. Their report is unfortunately quite sketchy, since most of the process details are priority information.

4.3 Estimation using Steady-state Models

Though it is seldom possible to derive reliable dynamic models for industrial processes, very good steady-state models are often available. It is proposed that these static models can be used for on-line estimation, since the characteristics of many industrial processes vary so slowly that the plant can be considered to move through a succession of steady-states. In these situations, the process measurements are related at any time by the known steady-state constraints of the system, and can be used to monitor the slowly varying process elements, whether they are model parameters, or characteristics of the measuring devices.

In terms of the theory presented in Part I, the steady-state model provides the functionality, g, in (3.2), which relates the observation vector to the state vector. The Jacobian matrix G can be computed either numerically or analytically by perturbing or differentiating the steady-state model with respect to the state vector elements. In typical situations, the steady state model is a set of differential equations, as for a plug flow reactor, whose solution provides the temperature and composition profile of any time. The state in this case will consist of the set of initial conditions at any time, as well as the values for slowly varying parameters such as catalyst activity.

The dynamic model describing the evolution of the state will often be easy to postulate. In the absence of random dynamic input disturbances, most problems can be converted so that a set of constant initial conditions for the differential equations describing the evolution of the state can be estimated, and $\Phi_{N,N-1} = I$ for all N. The true state at any time is obtained by numerical integration. If there are random inputs the state vector must be updated one step at a time, and this

simplification is not possible. In these cases, the matrix exponentiation used, for instance, by Coggan and Noton[48] is required.

Whether random dynamic inputs present a serious problem in chemical process systems is a debatable point which can only be resolved by on-line application of the filtering techniques. However, it is felt that any small random distrubances would be damped out by the system, and any large disturbances, such as the effects of changes in ambient temperature, could probably be accounted for by a deterministic model. Indeed, the most serious plant disturbances, which could not be accounted for by a stochastic model, are likely to be due to operator errors or equipment malfunctions. For these reasons, random dynamic inputs are not considered in the examples presented, and the estimation problems can be solved by the determination of constant state variables.

The type of sequential steady-state analysis suggested above offers distinct advantages over regression modelling for multi-variate systems, especially when it is desired to make full use of process measurements. Since every instrument reading is a system output, a regression model requires a term relating every input to every measurement, and is therefore very difficult to apply to heavily instrumented systems. On the other hand, a steady-state model will provide all likely observations of the system, and changes in the measurement scheme require only the modification of the observation matrix, G, in the formulation of the filter.

4.4 The Filter Program

A Fortran subroutine has been written to implement the algorithms described in Part I. The routine requires two problem dependent routines which must be supplied by the user. These two routines predict the next set of observations on the basis of the current state vector, and calculate the current value of G, the Jacobian matrix of the observations with respect to the state variables.

The prediction routine actually performs two tasks in the examples presented. Since the systems dealt with are not subject to random dynamic inputs, the state vector consists of the system state at time zero. Thus the routine, PREDCT, first calculates the state at the current time; if x_1 is a flowrate, and x_2 is the drift in that rate, the predicted value of the flow rate at time t_n is $x_1 + x_2 \cdot t_n$. In more complex situations, this calculation is performed by numerical integration. Given the state at the current time, the routine next calculates a set of predicted observations, using the steady-state model.

The gradient routine, CALCG, must define G using the current state-vector calculated in PREDCT. This requires the differentiation of the steady-state equations with respect to the current state. In many cases, the differentiation can be performed analytically, but numerical differentiation can be used if necessary.

Note that no particular form of system description has been specified. The user must merely provide a predicted observation vector, and a gradient matrix. An alternative form of the routine can be used if there are random dynamic inputs to the system. In these cases, the state transition matrix would have to be defined at each step, but the updating of the state vector to current time would be done internally, and not in PREDCT.

In the simulations the filtering subprogram is called from a main routine, which increments the time and supplies new observation vectors. An initial estimate of the plant state and the covariance matrix of the instrument errors at the time of initialization must be provided before the first call of the sub-routine. The initial estimation error covariance matrix, $P_{o/o}$ is calculated within the routine using a modification of equation (3.29), which allows a measure of the uncertainty of the initial state to be incorporated in the initial coveriance matrix. The selection and significance of this extra term is discussed in Chapter 5.

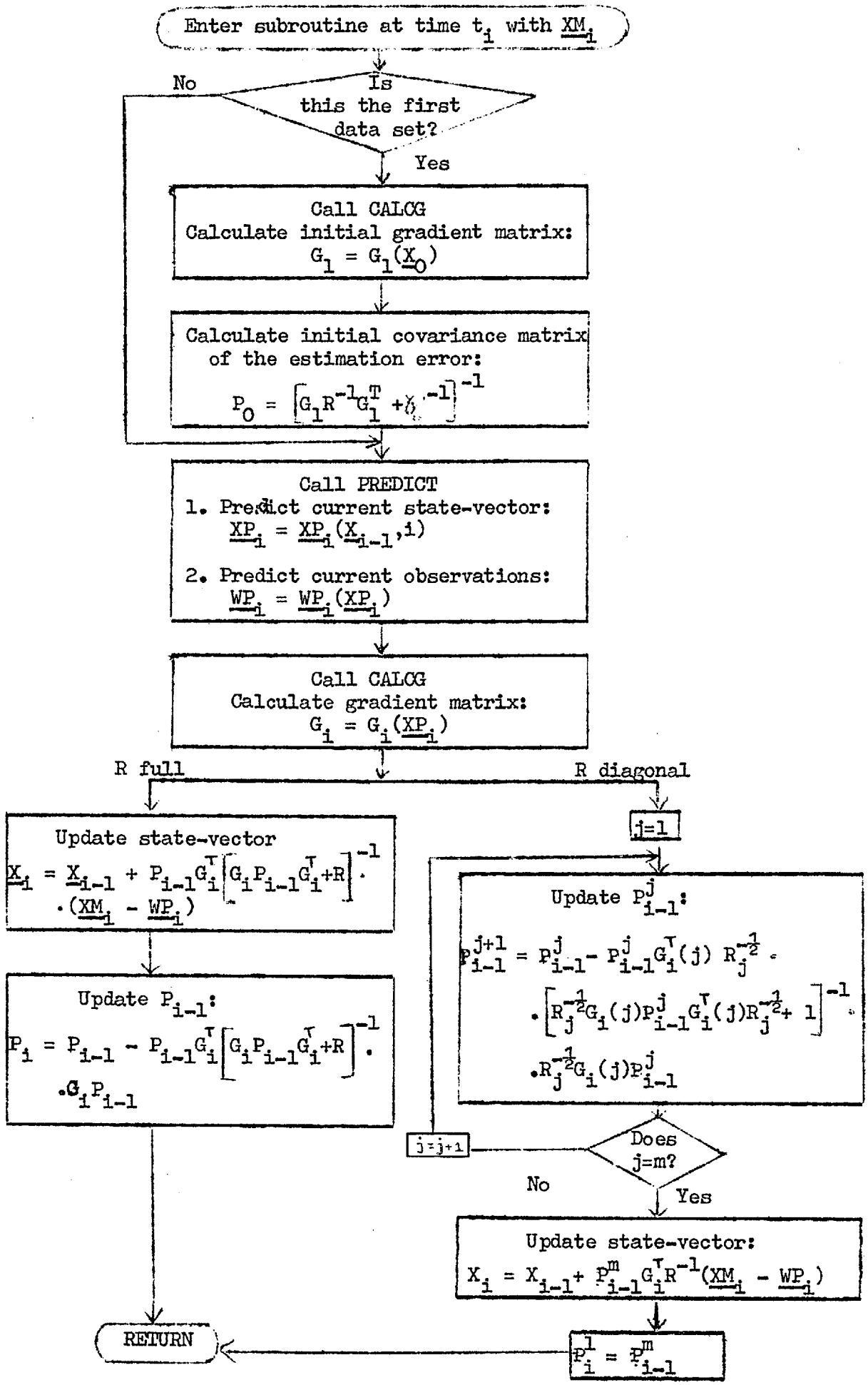
Note that the filter is initialized with an estimate of the actual state, rather than the estimate of a deviation from a reference state. This causes the filter output to be the actual state at all times, and is permitted because the corrections to the initial estimate are proportional to the difference between the actual and predicted observations. This difference is unaffected by the inclusion of the reference state in the estimate.

A flow chart of the filter is presented in Figure 4.1; a table of nomenclature is given in Table 4.1. For convenience, the instrument statistics are assumed constant in time, and R(t) = R for all t. The flow chart branches to allow the consideration of both uncorrelated (R diagonal) and correlated instruments (R full). Since no random dynamics are considered, all examples are converted to a form in which $\Phi_{N,N-1} = I$ for all N, so $P_{N/N-1} = P_{N-1/N-1}$. Therfore the notation in the flow chart has been simplified accordingly.

Table 4.1 Nomenclature for Filter Flow chart

Symbol	Description	Dimension
G_i	Jacobian matrix of observations with respect to state variables at time t_i	m×n
$G_i^T(j)$	jth column of G_i^T	nxl
P_i	Covariance matrix of estimation error at time t_i	nxn
R	Covariance matrix of instrument errors	mxm
R_j	Diagonal element r_{jj} of R	scalar
$\underline{WP}_i$	Predicted observation vector at time t_i	mxl
$\underline{X}_i$	Best estimate at time ti of the system state at time zero	nxl
$\underline{XM}_i$	Observation vector at time t_i	mxl
$\underline{XP}_i$	Best estimate at time t_i of the system state at current time	nxl
γ_o	Diagonal matrix which is a measure of the uncertainty in $\underline{x}_o$	nxn
Superscript T	Matrix transpose	-
Superscript -1	Matrix inverse	-

Figure 4.1: Flowchart for Filtering Subroutine



CHAPTER 5

FILTERING APPLIED TO A BINARY DISTILLATION COLUMN

The monitoring of an ammonia-water distillation column operating in steady-state is the first chemical engineering application of filtering techniques which is considered. The system simulated is representative of process elements which, after attaining a given operating condition, are maintained in that state for an extended period. In these situations, the filter can be used to monitor the performance of plant instrumentation as well as detect any unplanned changes in the system state.

Two series of experiments have been performed using a digital simulation of the column. In the first, the affects of errors in the statistical information input to the filter are examined by determining the average performance of the estimator over an ensemble of one hundred simulated plant records containing fifty sets of observations per record. In the second, the ability of the estimator to detect instrument biases and drifts in system inputs is studied using individual plant records of fifty observations.

5.1 System Description

5.1.1 Column Model

The steady-state operation of the column is described by three equations: an overall material balance, a component balance, and an enthalpy balance. There are nine unknown variables in the system, consisting of the flowrates, compositions, and enthalpies of the feed, bottoms and product streams. Thus;

$$F_F = F_B + F_P \tag{5.1}$$

$$F_F H_F = F_B H_B + F_P H_P \tag{5.2}$$

$$F_F c_F = F_B c_B + F_P c_P \tag{5.3}$$

where F, H, and c refer to melar flowrate, enthalpy, and composition respectively. The subscripts F, B, and P denote the three access streams. Using the steady-state balance equations, three of the unknown variables can be calculated in terms of the other six. Solving for the product stream components;

$$F_P = F_F - F_B \tag{5.4}$$

$$H_P = \frac{F_F H_F - F_B H_B}{F_F - F_B} \tag{5.5}$$

$$c_P = \frac{F_F c_F - F_B c_B}{F_F - F_B} \tag{5.6}$$

The stream enthaplies cannot be directly observed, but it is assumed that they can be deduced from temperature measurements.

The column is operated with a total condenser, and the bottoms stream is taken from the reboiler, so the bottoms and product streams

are saturated liquids. Using data from an ammonia-water enthalpy-concentration diagram at 10 atm .[49], a third order polynomial has been fitted to describe the saturated liquid enthalpy as a function of temperature, yielding;

$$H_S(T) = 3504.31 - 17799.62\ T' - 22169.67\ T'^2 - 16919.18\ T'^3 \qquad (5.7)$$

where T' is a normalized temperature, such that;

$$T' = T/182.22$$

with T in degrees centigrade.

The feed to the column is assumed to be a subcooled liquid whose enthalpy is specified by;

$$H_F\ (T_F) = H_S(TB) - 32.2575\ (TB - T_F) \qquad (5.8)$$

where TB is the bubble point of the feed stread with composition c_F. The constant in (5.8) is determined from the enthalpy-concentration diagram cited above, and implies constant specific heat of the liquid feed over the range considered.

The bubble point of an ammonia-water mixture at 10 atm. can be determined if the equilibrium constants of ammonia and water are known functions of temperature. Third order polynomials in temperature have been fitted using published data[50]. Hence;

$$K_A\ (T) = 0.427799 + 4.420266\ T' - 2.144682\ T'^2 + 3.400508\ T'^3 \qquad (5.9)$$

$$K_W\ (T) = -0.046666 + 0.223197\ T' - 0.692144\ T'^2 + 1.344717\ T'^3 \qquad (5.10)$$

where T' = T/170 with T in degrees centigrade. These definitions apply for $30 \leq T \leq 170$ in the case of K_A and $70 \leq T \leq 170$ for Kw.

Simple exponential functions define the constants for values outside of these ranges.

$$K_A(T) = \exp\left[(T - 25.3)/31.7\right] \quad \text{for } T \leq 30$$

$$Kw(T) = 0.22 \exp\left[(T - 70.0)/25.0\right] \quad \text{for } T \leq 70 \qquad (5.11)$$

$$K_A(T) = 5.5875 \exp(T - 170.0) \quad \text{for } T > 170$$

$$Kw(T) = \exp\left[(T - 182.22)/62.8\right] \quad \text{for } T > 170$$

The exponential extensions at the ends of the range are required to allow use of a Newton-Raphson search in the bubble point calculation. This calculation determines the temperature such that;

$$c_{NH_3} K_A(T) + c_{H_2O} K_w(T) = 1.0 \qquad (5.12)$$

where the compositions refer to the overall mole fractions of water and ammonia in the liquid.

Note that at this stage a relationship has been provided which relates T_B to c_B and T_p to c_p, and the dimensionality of the problem could be reduced. However, in practice equilibrium data may not be reliable, and it is assumed for this problem that the bubble point relationship cannot be used. The feed enthalpy is thus considered to be a known function of T_F and c_F. The following algorithm is used to specify the product stream components given values for the components of the other streams.

Steady-State Column Model

1. Given T_B, calculate the enthalpy of the bottoms stream from equation (5.7).
2. Given c_F and T_F, calculate the feed enthalpy. (In practice, this was performed using equations (5.8) to (5.12)).

3. Calculate F_p, H_p and c_p using (5.4) to (5.6).
4. Calculate T_p by solving for the root of (5.7) using a Newton-Raphson search.

5.1.2 Simulated Data

Plant data was simulated by perturbing the outputs calculated using the steady-state model with various types of artificially generated instrument noise. All three components of the access streams were assumed to be observed every five minutes during the operation of the plant. Though this measurement scheme is probably unrealistic for a simple binary distillation column, the filter will in practice be applied to key process elements that are likely to be heavily instrumented, and it is desired to evalute the filter preformance with systems having redundant measurements.

The instrument noise was generated using package routines for the calculation of random normal and rectangular number sequences on an IBM 7094[51]. In some cases, the random noise was modified to simulate correlation between sample intervals. In all cases, the instrument noise was independent within an observation set, so R(t) was a diagonal matrix. The observation errors were scaled in all but one example so that the 95% confidence interval of the normal noise, and the range of the rectangular noise corresponded to $\pm$ 2% accuracy for temperature and flowrate measurements, and $\pm$ 3% for composition measurements. In one example all accuracies were set to be $\pm$ 5%.

Three general situations were simulated: constant steady-state, a slowly drifting steady-state, and a system with instrument biases. In all cases, the unperturbed measurements satisfy the steady-state

equations instantaneously. For the constant state simulation, the system is subject to random instrument errors only, and all inputs and outputs are theoretically constant. For the drifting system, a slow linearly time dependent drift in one of the access stream flowrates is superimposed. In the third situation, constant zero-point off-sets are present in some of the instruments.

5.1.3 Filtering Model

Since the system is not subject to random dynamic inputs, the filtering problem can be converted to one of estimating constants. The state variables estimated depend upon the situation considered. For the constant steady-state column, a six dimensional state-vector is required, since the bubble point relationship is not used. Hence;

$$\underline{x}^T = \left[F_B, T_B, c_B, F_F, T_F, c_F\right]^T \qquad (5.13)$$

Since the elements are unknown constants, $\underline{\Phi}_{N,N-1} = I$ for all N. When there is a drift in one of the independent variables, with constant unknown time derivative α, a seventh state variable is adjoined to $\underline{x}$ to estimate the constant α. Further variables can be added to allow the estimation of biases. $\underline{\Phi}_{N,N-1}$ remains an identify matrix of appropriate dimension. Thus the state vector itself provides the best estimate of the constant steady-state, and contains all the information necessary for the prediction of the state and measurements in the other cases. For instance, if there is a drift in the feed flowrate, the value of the flowrate at time t is predicted by;

$$F_F(t) = F_F(t_o) + \alpha \cdot (t - t_o) \qquad (5.14)$$

where $F_F(t_o)$ is element four of the state vector. Of course, a more complex time dependency can be proposed, and the constants in

the drift model estimated in the same way.

Routines for the prediction of the current instrument readings and the calculation of the Jacobian matrix G must be provided for the filter program. Consider the case when there is a bias in the measurement of the feed flowrate as well as a drift in the flowrate. The relevant elements in the state vector are F_F, the flowrate at time zero; α, the drift rate; and β, the measurement bias. The best estimate of the true flowrate at time t_{n+1} based upon the state vector at time t_n is determined from (5.14). The predicted measurement at t_{n+1} includes the bias term. Thus;

$$F'_F = F_F + \alpha . (t_{n+1} - t_o) + \beta \qquad (5.15)$$

where the prime indicates a predicted measurement.

For this example, most of the elements of the Jacobian matrix, G, are constants. Figure 5.1 shows the matrix for the case with nine measurements when the state vector has eight elements, consisting of the six independent stream components, a drift rate for the feed flowrate and a bias for the feed flowmeter. The only variable elements of the matrix are the gradients of the dependent stream components with respect to the state variables. All of these elements can be determined analytically, except for the fifth and sixth ones in the eighth column, which required the knowledge of $\partial H_F/\partial T_F$ and $\partial H_F/\partial c_F$. These derivatives are determined numerically using a quadratic approximator[52].

Figure 5.1 Elements of the Jacobian Matrix : G^T

	F_B	T_B	c_B	F_F	T_F	c_F	F_P	T_P	c_P
F_B	1	0	0	0	0	0	-1	$\frac{\partial T_P}{\partial F_B}$	$\frac{\partial c_P}{\partial F_B}$
T_B	0	1	0	0	0	0	0	$\frac{\partial T_P}{\partial T_B}$	0
c_B	0	0	1	0	0	0	0	0	$\frac{\partial c_P}{\partial c_B}$
F_F	0	0	0	1	0	0	1	$\frac{\partial T_P}{\partial F_F}$	$\frac{\partial c_P}{\partial F_F}$
T_F	0	0	0	0	1	0	0	$\frac{\partial T_P}{\partial T_F}$	0
c_F	0	0	0	0	0	1	0	$\frac{\partial T_P}{\partial c_F}$	$\frac{\partial c_P}{\partial c_F}$
α	0	0	0	t	0	0	t	$\frac{\partial T_P . t}{\partial F_F}$	0
β	0	0	0	1	0	0	0	0	0

5.2 Studies of Filter Performance

5.2.1 Ensemble Results

The algorithm derived in Chapter 3 requires knowledge of several statistical characteristics of the observed system; the first two moments of the initial estimate error must be provided, as well as the covariance matrix of the instrument errors at every sample interval. It is assumed, in effect, that the initial estimate error, $\underline{x}_{o/o}$, is described by a multivariate normal probability distribution with zero mean and covariance matrix γ_o, and that the observation errors are governed by a distribution with zero mean and covariance matrix R(t). It is further assumed that the instrument errors are uncorrelated between data sets.

The sensitivity of the filter performance to errors in these assumptions has been examined by analysing ensembles of one hundred runs rather than single simulated plant records. The data used for the ensemble results is based on a drifting system with a $\frac{1}{2}\%$ per hour drift in the feed flowrate, which is assumed initially to be zero. The initial covariance matrix of the estimation error is determined from a modification of equation (3.35) which permits the specification of the uncertainty in the initial state. Thus;

$$P_{o/o} = \left[G_1^T R_1^{-1} G_1 + \gamma_o^{-1} \right]^{-1} \qquad (5.16)$$

Note that when the initial estimate, $\underline{x}_{o/o}$, is very uncertain, its variance will be very large, and γ_o^{-1} becomes a null matrix. In that case, equation (5.16) reverts to the original form of (3.29).

Four matrices have been defined for use in the ensemble studies: R^T, R^F, γ_o^T and γ_o^F, where the superscript T denotes the true variance matrix used to scale the random numbers generated as instrument or

initial state errors, and F denotes the matrix used in the filter algorithm. In all of the studies with $\pm$ 2% and $\pm$ 3% instrument accuracies, R^T is a diagonal matrix, with diagonal elements;

$$\{r_{ii}\} = \{0.457,\ 3.14,\ 0.39 \times 10^{-6},\ 1.05,\ 0.0276,\ 0.277 \times 10^{-4},\ 0.116$$
$$0.0755,\ 0.238 \times 10^{-3}\} \qquad (5.17)$$

In this example, the first six elements of the state vector are observed directly, and thus the first six diagonal elements of γ_o are defined in (5.17). The seventh diagonal element, defining the uncertainty in the initial value chosen for the drift is set to be 0.25×10^{-4}. This small number implies that the expected drift will be very slight. The covariance matrices are chosen to be diagonal, which indicates that there is no correlation between instrument errors within a data set. This is a reasonable assumption for chemical process systems; correlation between successive readings on a given instrument is more likely, and its effect is considered below.

The following program scheme was used to generate the ensemble results.

1. Read in R^T, γ_o^T, R^F and γ_o^F.
2. Set NTRIAL to zero.
3. Increment NTRIAL by 1.
4. Set NSET to zero.
5. Select $\underline{x}_{o/o}$ using random numbers scaled according to γ_o^T.
6. Increment NSET by 1.
7. Select a set of observations by perturbing the true measurements for the current time with random numbers scaled according to R^T.
8. If NSET is 1, calculate $P_{o/o}$ using (5.17) with γ_o^F.

9. Update the filtering algorithm using R^F.
10. Update the average filter output for the current value of NSET.
11. If less than 50 sets have been analysed, go to 6.
12. If less than 100 trials have been made, go to 3.
13. Print out the ensemble results for 50 observation sets based on 100 trials.

Table 5.2 presents a summary of the ensemble tests performed. The situations considered are the effects of errors in the magnitude of R^F and χ_o^F in relation to the true values, as well as the effects of non-normal and correlated noise. In all cases but the last, R^T is described by (5.17); in the final test, all readings are generated to be $\pm$ 5% accurate, so that;

$$\{r_{ii}\} = \{2.856,\ 19.625,\ 0.108 \times 10^{-5},\ 6.562,\ 0.172,\ 0.769 \times 10^{-4},\ 0.725,\ 0.472,\ 0.661 \times 10^{-3}\} \qquad (5.18)$$

Table 5.2 Ensemble Results

RUN	$Q^{*}_{\gamma_o}$	Q_R^{*}	Comments
E1	1	1	Normal random noise
E2	10^4	1	Normal random noise
E3	10^{20}	1	Normal random noise
E4	0.1	1	Normal random noise
E5	2.0	1	Normal random noise
E6	1	4	Normal random noise
E7	1	0.25	Normal random noise
E8	1	0.04	Normal random noise
E9	1	1	Rectangular random noise
E10	10^4	1	Rectangular random noise
E11	1	1	Normal correlated noise **
E12	10^4	1	Normal correlated noise **
E13	10^4	1***	Normal random noise

$$* \; Q_K = \frac{k^F_{ii}}{k^T_{ii}}$$

** Correlation coeff. = 0.2

*** R^T defined by (5.18).

Figures 5.2 to 5.6 illustrate the effect of variations of γ_o^F on the output of the filter. It can be seen that the assumption of high uncertainty in the value of the initial state improves the convergence of the algorithm, especially in the case of the dependent product stream components. However, increasing the magnitude of γ_o^F

will not improve the performance after a certain point. This is to be expected, for it is the point at which the second term in the brackets in (5.16) becomes insignificant in relation to the first. No instability due to a large value of γ_o^F has occurred in this example, and it is safe to assume that an infinite variance of the initial state can be used in most applications. Note, however, that the variance of the initial estimate error, $P_{o/o}$, remains finite.

The estimation of the dependent, product stream variables can be seen to be most indicative of the filter performance. The ensemble results for the independent variables are essentially those which could have been obtained by simple averaging over the 100 trials. Thus, for convenience, the remaining ensemble results will be illustrated by observing the estimation of the product temperature and the drift parameter only.

The effect of varying R^F in relation to the true instrument error variance is shown in Figure 5.7; the filter performance improves as smaller and smaller variances are assumed. This result was not expected, since the small variance implies a high degree of certainty in the instrument readings, and would be expected to cause oscillations in the filter output if the values selected for the elements of R^F are unrealistically optimistic. However, it can be seen by inspecting the LS algorithm that a decrease in the magnitude of R^F has two effects; it will cause $P_{N/N}$ to be artificially small and R_N^{-1} to be artificially large. The filter gain is a function of the product of these two matrices, and it is evident that in this example, the decrease in $P_{N/N}$ overshadows the increase in R_N^{-1}. Note that the variation in $P_{N/N}$ is also a function of the Jacobian matrix which is problem dependent; thus the effect of

(text continues on Page 87)

FIGURE 5.2

ENSEMBLE RESULTS: THE EFFECT OF THE ASSUMED VARIANCE OF THE INITIAL STATE

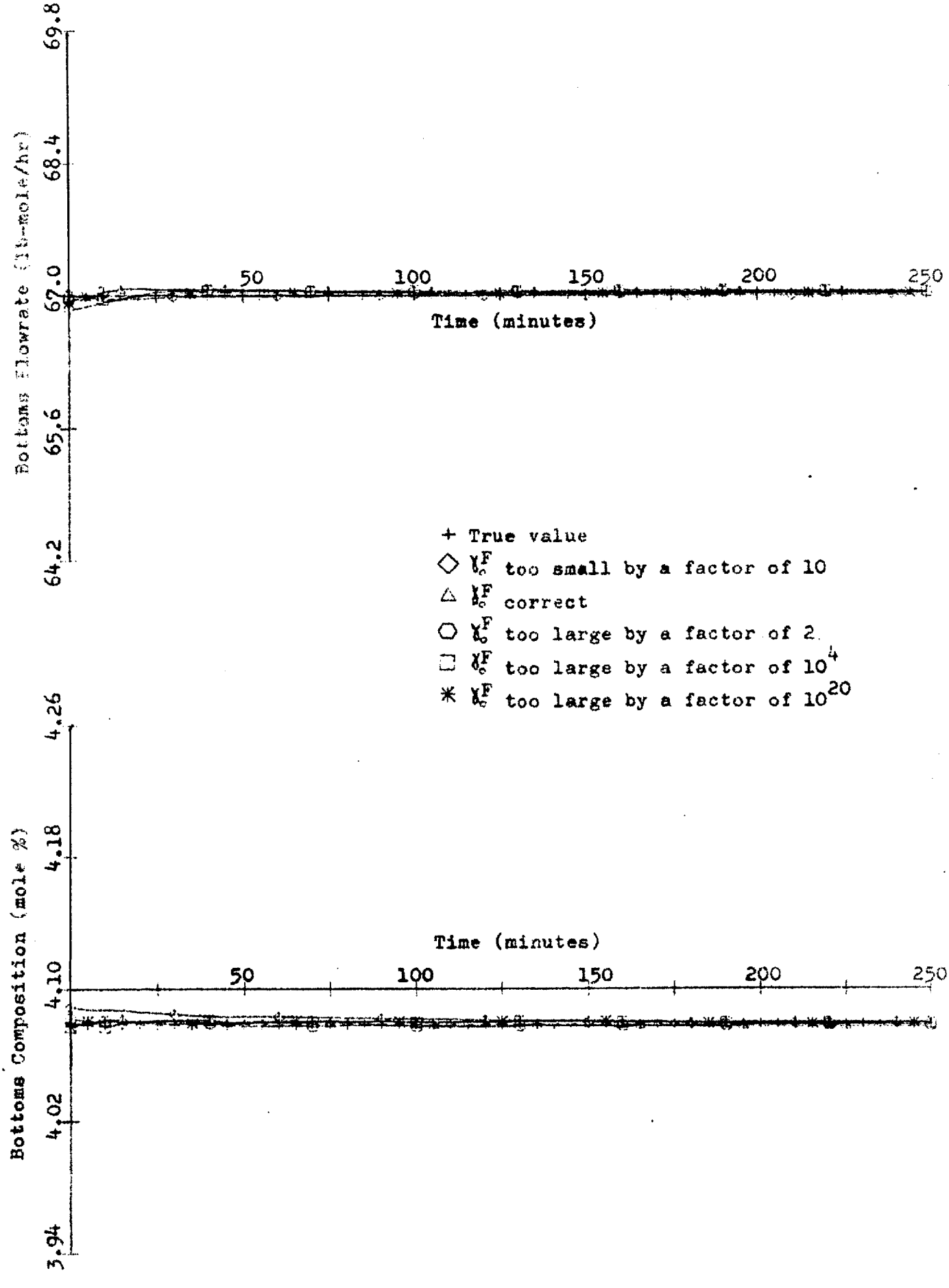


FIGURE 5.3

ENSEMBLE RESULTS: THE EFFECT OF THE ASSUMED VARIANCE OF THE INITIAL STATE

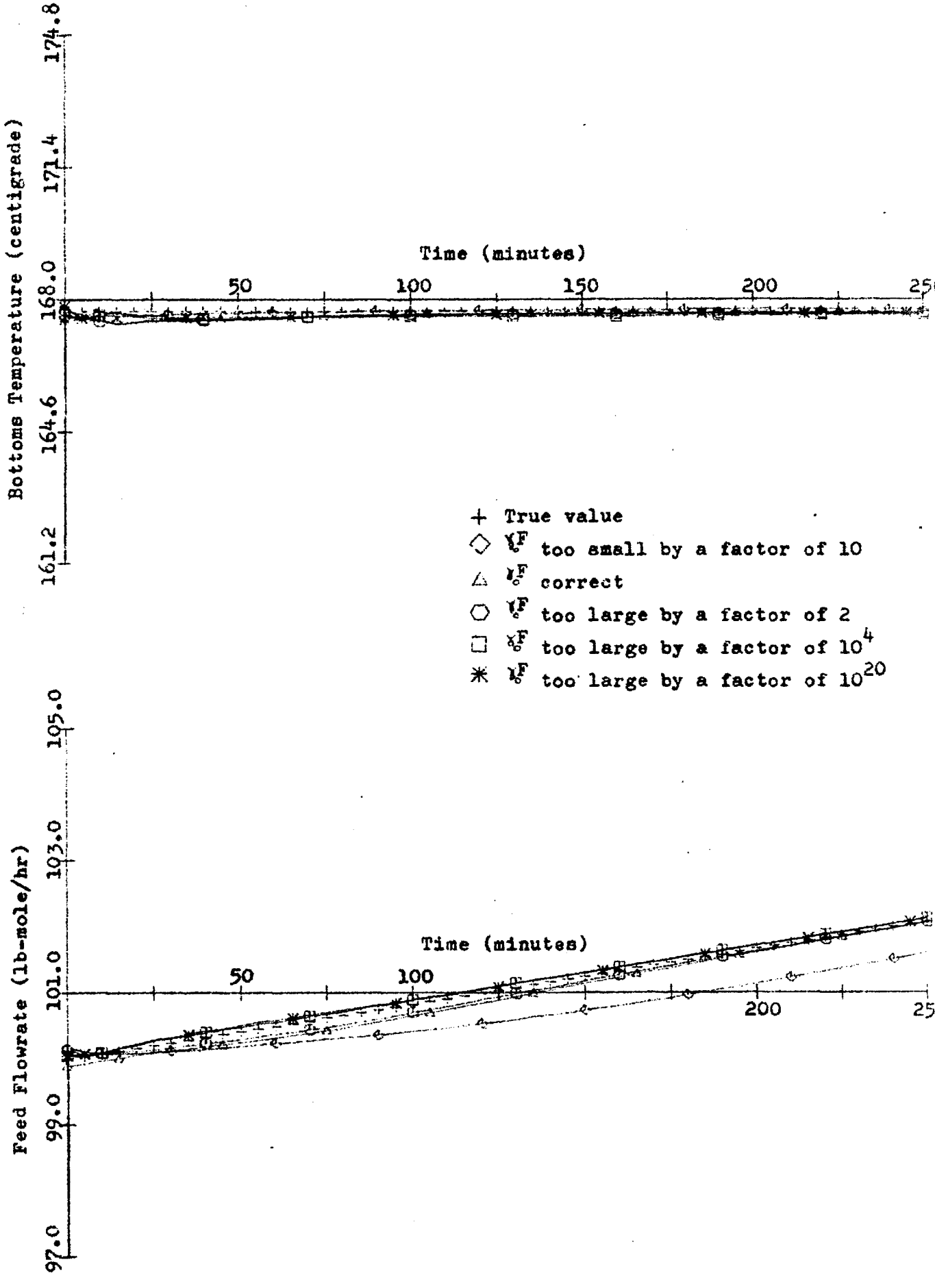


FIGURE 5.4

ENSEMBLE RESULTS: THE EFFECT OF THE ASSUMED VARIANCE OF THE INITIAL STATE

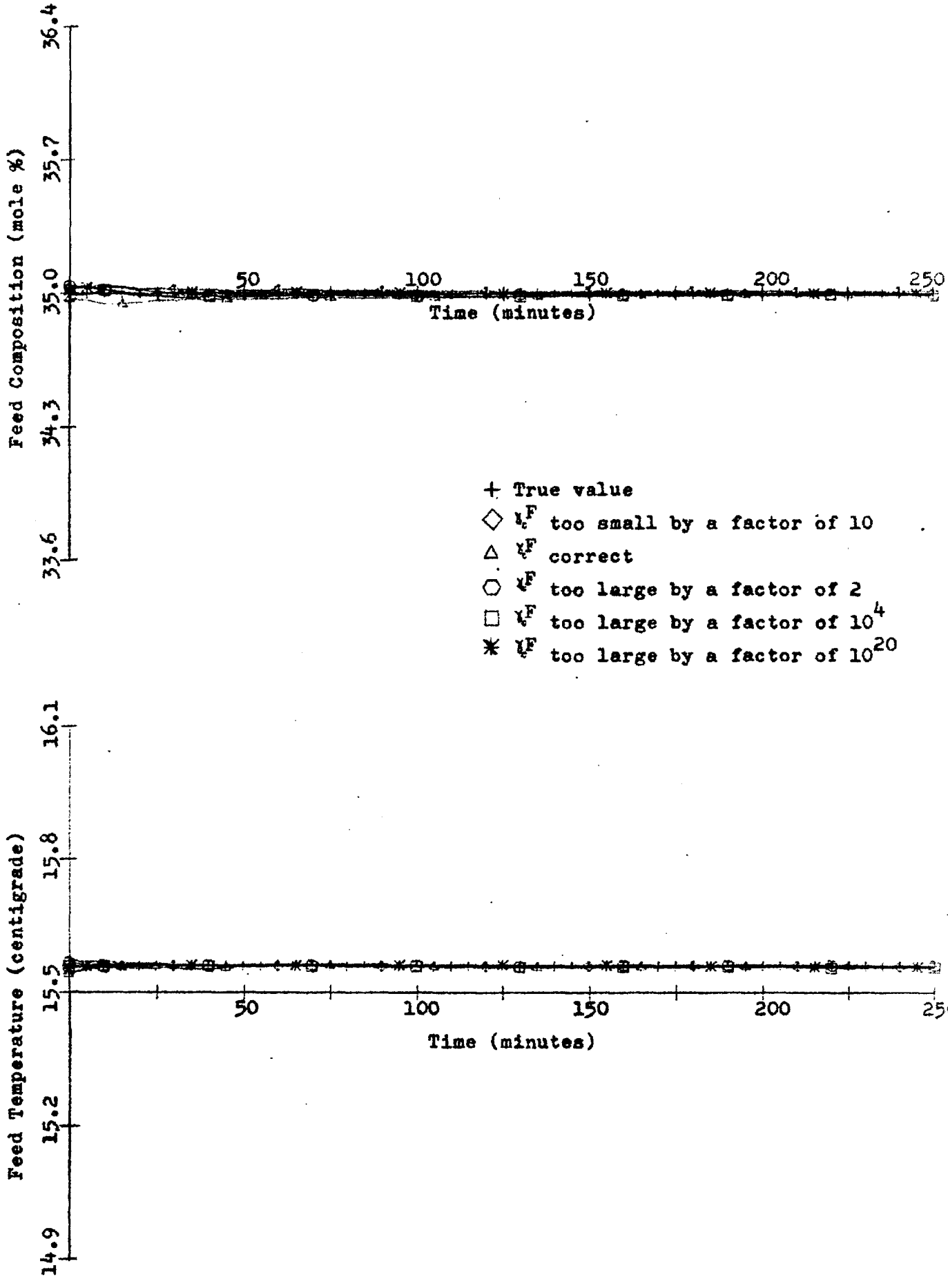


FIGURE 5.5

ENSEMBLE RESULTS: THE EFFECT OF THE ASSUMED VARIANCE OF THE INITIAL STATE

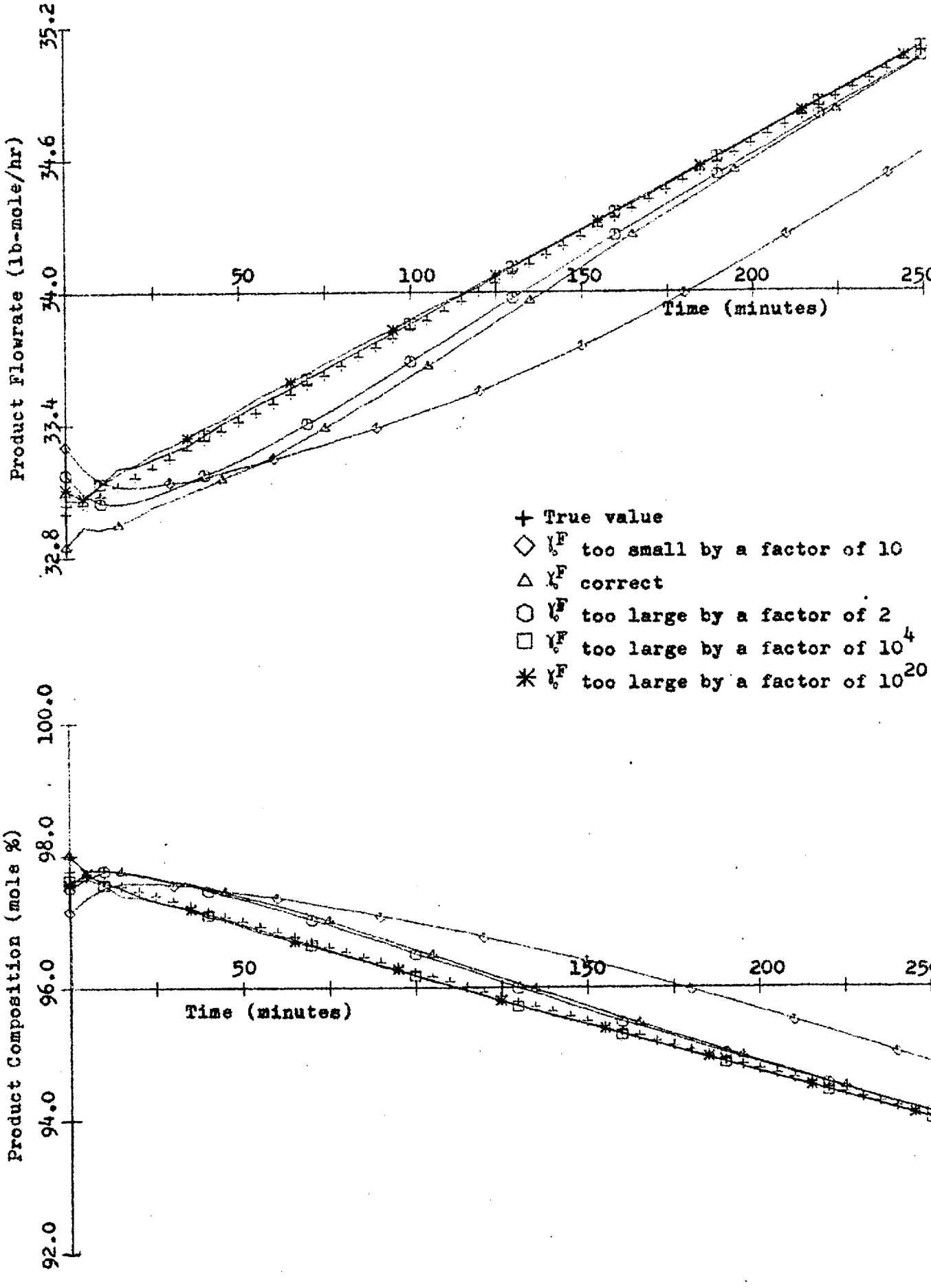


FIGURE 5.6

ENSEMBLE RESULTS: THE EFFECT OF THE ASSUMED VARIANCE OF THE INITIAL STATE

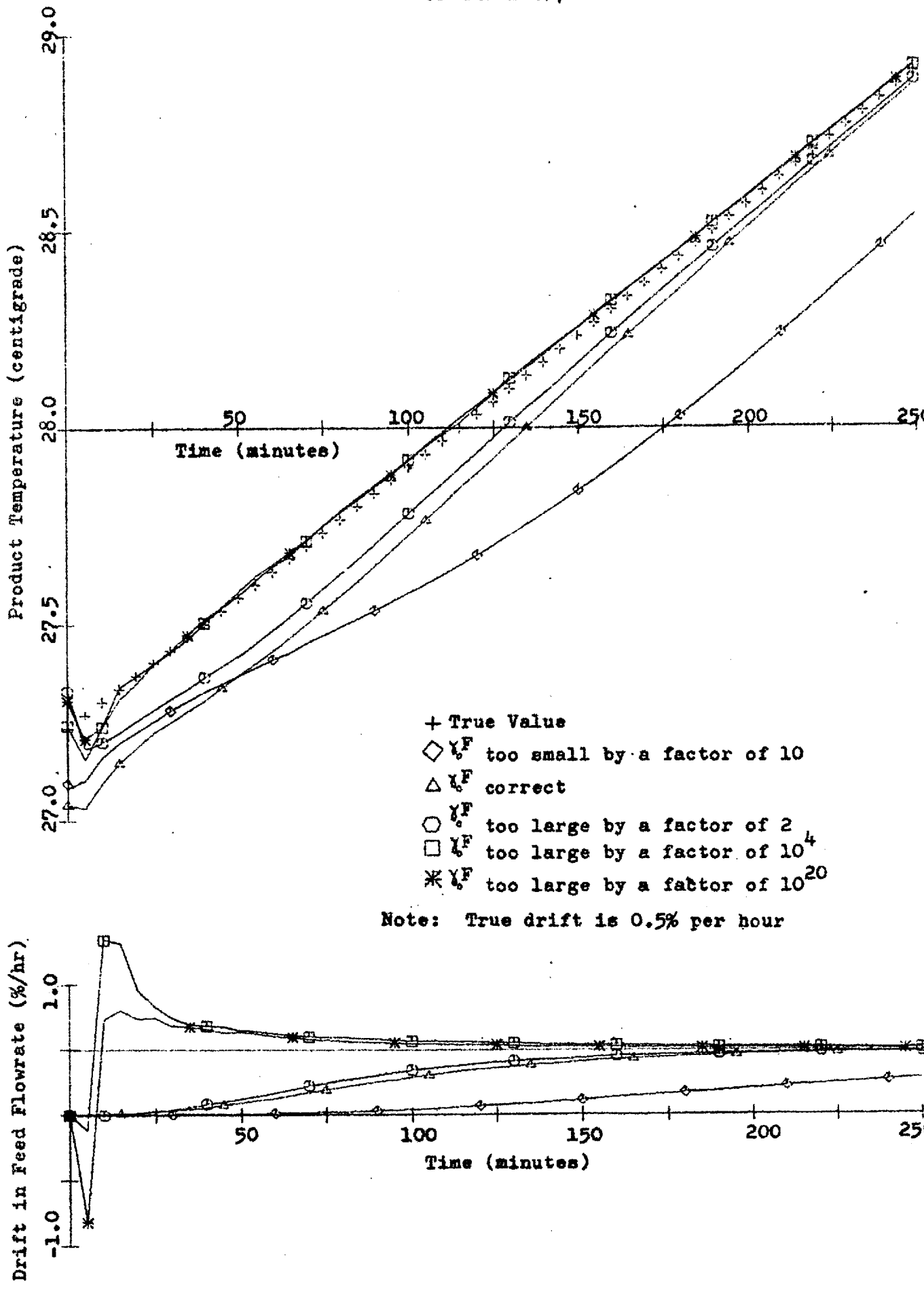


FIGURE 5.7

ENSEMBLE RESULTS: THE EFFECT OF THE ASSUMED VARIANCE OF THE MEASURING DEVICES

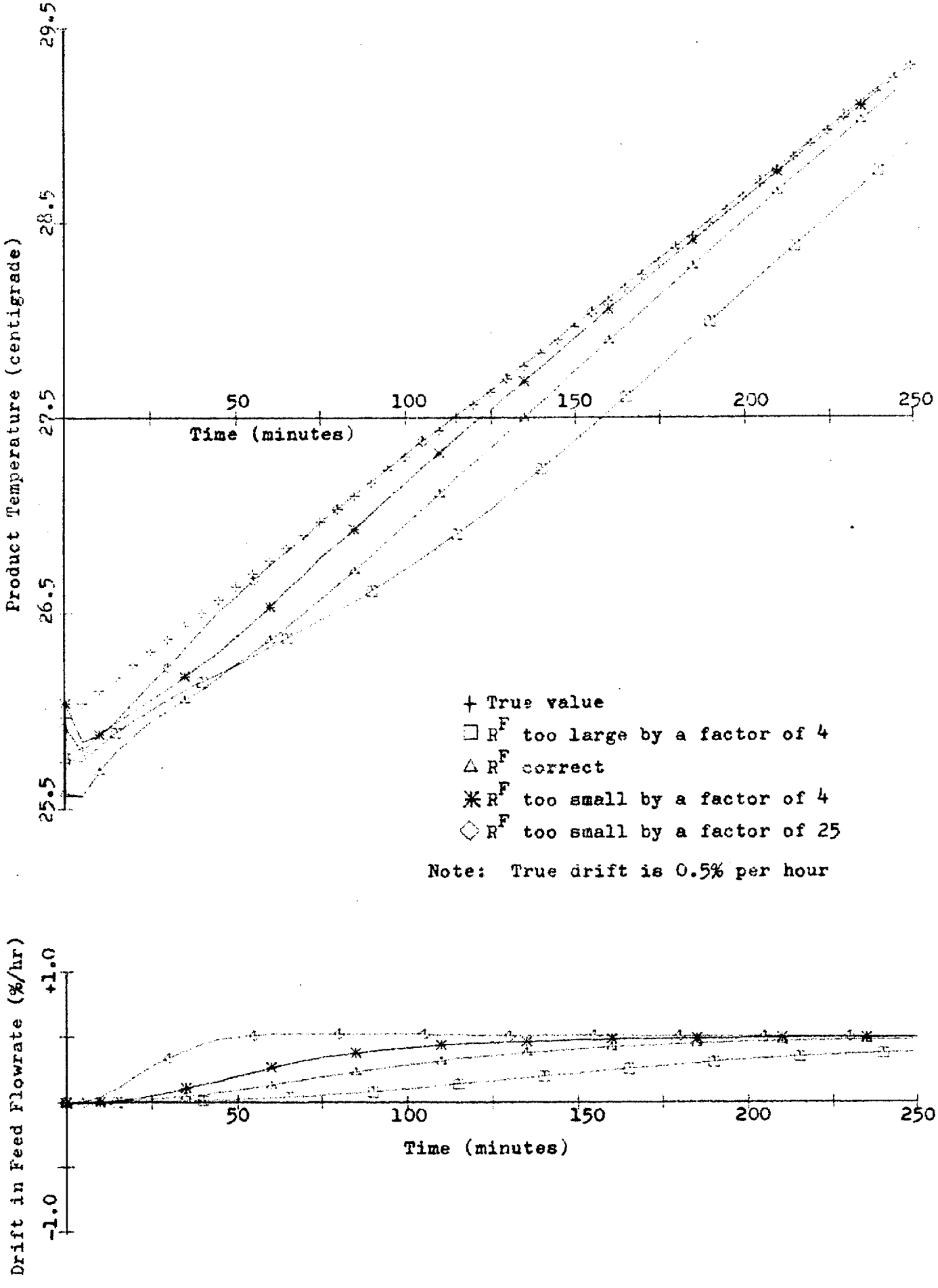


FIGURE 5.8

ENSEMBLE RESULTS: THE EFFECT OF NON-NORMAL NOISE

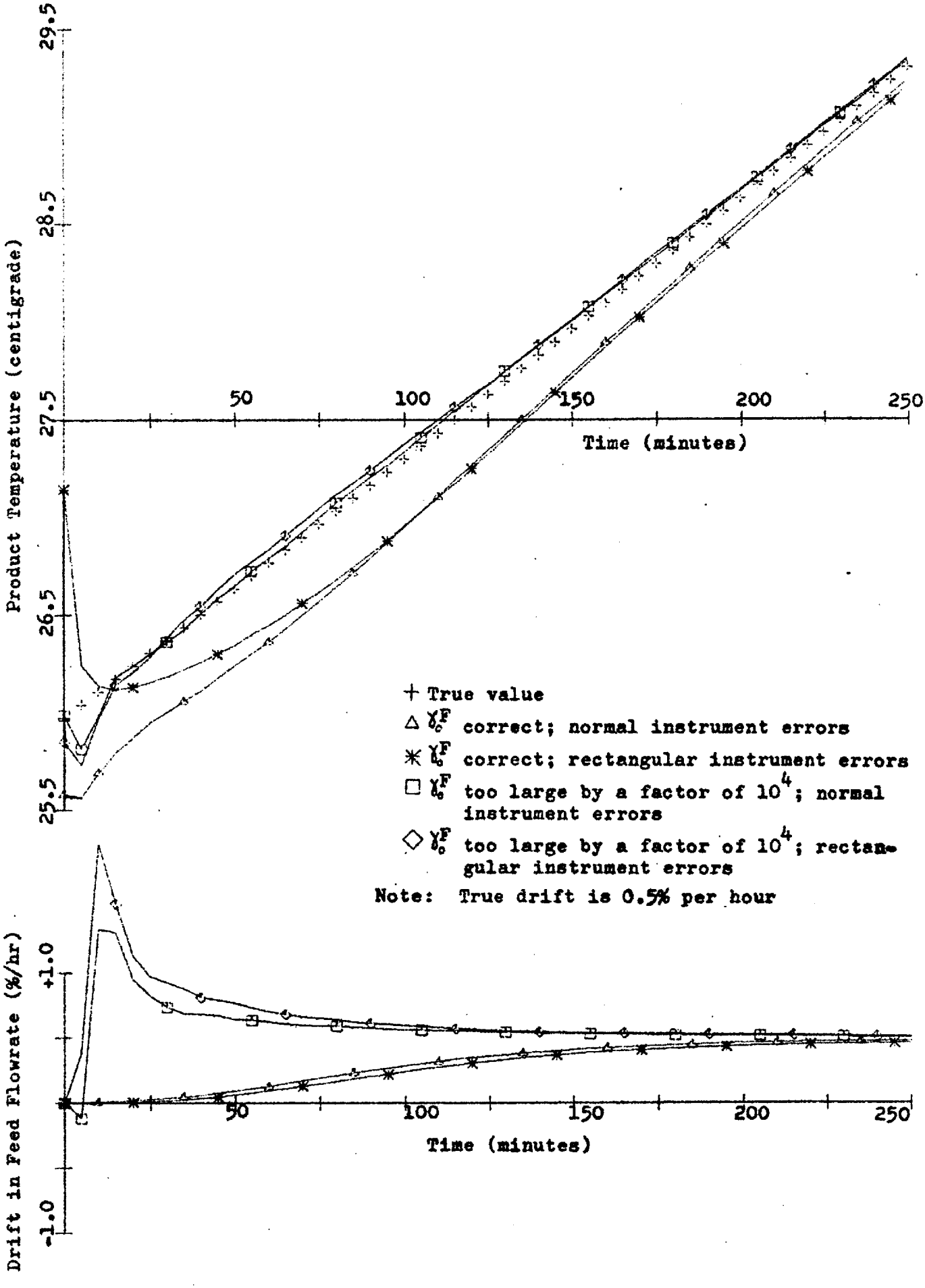


FIGURE 5.9

ENSEMBLE RESULTS: THE EFFECT OF CORRELATED NOISE

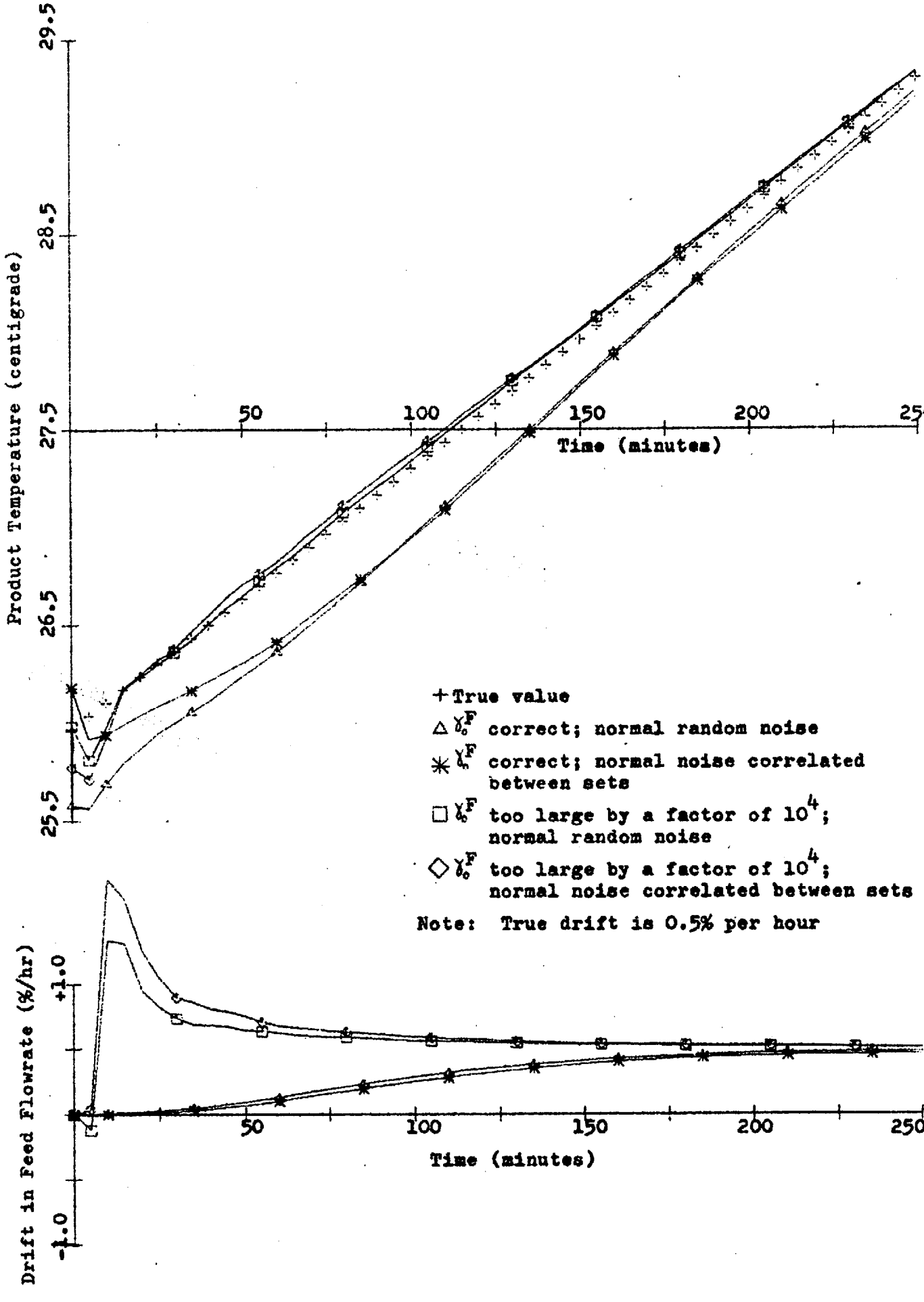
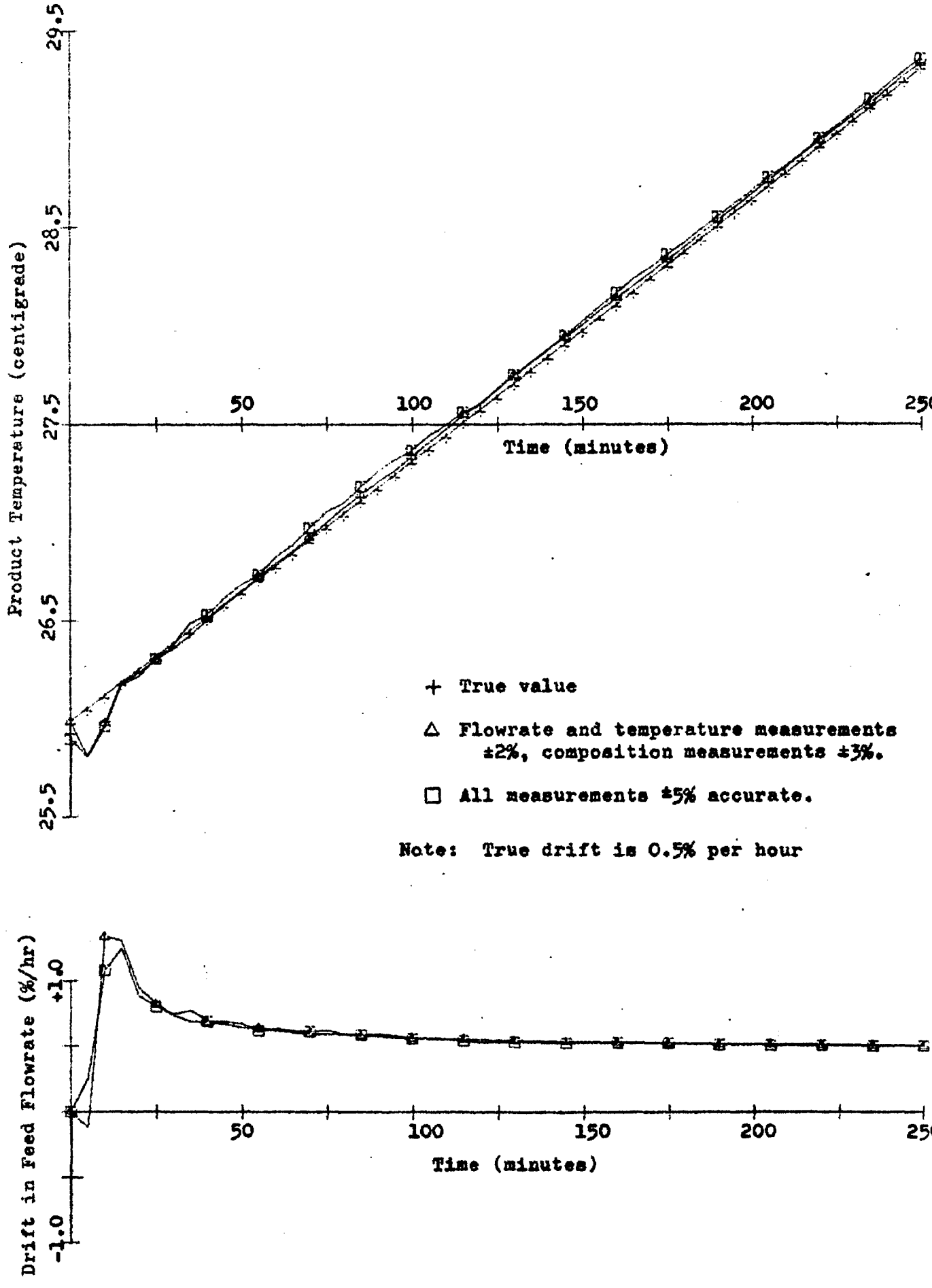


FIGURE 5.10

ENSEMBLE RESULTS: THE EFFECT OF LARGE INSTRUMENT ERRORS



using erroneous values for the elements of R^F must be examined for each particular case.

Figure 5.8 illustrates the effect of the presence of rectangular rather than normal instrument noise. It can be seen that there is essentially no change in the filter performance. A similar result is obtained when the noise is correlated between successive sample intervals, as is shown in Figure 5.9. The correlated errors were generated in the following manner:

$$ERR_N = 0.8\ RN_N + 0.2\ ERR_{N-1} \tag{5.19}$$

where ERR_N is the instrument error at time t_N and RN_N is the random number generated by the computer at time t_N.

The effect of larger instrument noise is illustrated in Figure 5.10. Though the estimate bias is increased slightly due to the increased instrument errors, the effect can be seen to be negligible.

5.2.2 Studies of Individual Plant Records

The ensemble results presented above are based on a column in a drifting steady-state. The filter used to analyse the effects of errors in statistical inputs was designed to estimate the drift in the feedrate; thus the form of the system model was specified, and only the parameters in the model were not known. In a practical situation, it is unlikely that all of the types of system disturbances will be known. The following studies explore the effects of model deficiencies which ignore some of the characteristics of the observed system.

The examples presented are based on the analysis of individual plant records, usually consisting of a set of fifty observations made at five minute intervals. Since the independent variables in the model are all directly observed, the initial state vector is defined from the first data set. Any biases or drifts which are estimated are initially set to zero.

It is assumed that the instrument error variances can be deduced from past operating records or can be based on the manufacturer's specification, and R(t) is known. R(t) is defined in this example by equation (5.17) for all t. Biases and drifts are assumed to be within 1% or 1% per hour of the nominal instrument reading, and their variances are defined accordingly. In all of the studies, the nine measurements of the access stream components are made every five minutes, but the dimension of the state vector is varied depending upon which biases and drifts are estimated.

Data sets for constant, drifting and biased steady-state systems were generated on a digital computer and punched on to data cards. The sets were analysed by filters assuming various types of models, using both growing memory and oscillating memory filters. For instance, constant steady-state data was analysed with a filter designed to detect a bias in the product flowmeter to determine if a zero bias can be accurately estimated. In several tests a general 18 dimensional model was used to detect biases in all instruments as well as drifts in the feed stream components. Table 5.3 lists the symbols used to describe the situations considered; Table 5.4 describes the various studies performed on the simulated data.

Table 5.3 List of Symbols

SYMBOL	MEANING
CSS	Constant steady-state
BSS	Biased steady-state
DSS	Drifting steady-state
BDSS	Drifting steady-state with bias
LS	Least-squares filter
OMF	Oscillating memory filter

For a system in a constant steady-state, the instrument readings should in theory be constant in time. If the instrument errors are independent and normally distributed, it can be shown that the best estimate of the true instrument reading based upon a series of measurements is merely the mean of the series. Under the assumptions of normality and independence, it is furthermore possible to show that the sample mean is the minimum variance unbiased estimator of the true reading. Thus the optimal estimate of the constant steady-state model could be determined by simple recursive averaging. Figure 5.11 shows the results of test S1 and demonstrates that the filter and sequential averaging provide essentially identical results for the constant system with independent noise. One independent variable and one dependent variable are taken to be representative of the results of the test. The large oscillation in the filter estimate of the product temperature at sample 2 is due to the fact that the dependent variables are forced to satisfy the system constraints; the averaged results are only consistent with the constraints in the limit.

Table 5.4 Simulated Experiments Performed

RUN	FILTER TYPE	DIMEN-SION	MODEL	DATA	COMMENTS
S1	LS	6	CSS	CSS	
S2	LS	7	BSS	CSS	Model assumed bias in bottoms flowmeter
S3	LS	7	DSS	CSS	Model assumes drift in bottoms flowrate
S4	LS	7	BSS	BSS	Model assumes bias in bottoms flowmeter
S5	LS	6	CSS	DSS	Drift in bottoms flowrate
S6	OMF	6	CSS	DSS	Drift in bottoms flowrate
S7	LS	18	BDSS	BDSS	Drift in feed flowrate, bias in product flowmeter
S8	LS	18	BDSS	BDSS	Bias set to zero if less than 0.1%
S9	OMF	18	BDSS	BDSS	Memory oscillates between 20 and 40 sets

The ability of the filter to estimate zero biases and drifts is investigated in runs S2 and S3, and representative results are shown in Figures 5.12 and 5.13. It can be seen that the absence of biases and drifts in the constant data is clearly indicated.

The estimation of instrument bias is investigated in run S4. The bias is assumed to be in the bottoms flowmeter, and thus this test is similar to the ensemble studies in that the form of the system error is known, and only its magnitude must be determined. Figure 5.14 shows that the bias has been detected; however, note that the convergence to the true value of the bias is slower than the convergence experienced in estimating drifts. This is due to the fact that the bias is independent of all other measurements, so that the information provided by the steady-state model does not give a direct indication of the instrument error. Nevertheless, the presence of some bias is clearly indicated; this information could not have been determined by independent analysis of the bottoms flowmeter record.

The effects of model deficiencies on the filter performance is illustrated by the analysis of drifting data with a filter designed for a constant steady-state. Figures 5.15 to 5.19 show the complete results of runs S5 and S6, and the diagrams include the true states and simulated data as well as the output of the filters. It can be seen that serious estimation errors are caused by ignoring the drift in the bottoms flowrate, and the errors are especially pronounced in the estimation of the dependent variables. Note that the filter based on the steady-state model predicts the average value of the drifting variable rather than the true value. The discrete change in the prediction of the OMF provides a clear indication of the presence of a linear drift; this is particularly noticeable in the estimation of the product temperature.

It can be seen from the preceding results that model deficiencies can seriously degrade the performance of the estimator. The safest procedure to use which avoids these deficiencies would seem to be to allow for as many contingencies as possible. The eighteen-dimensional studies were made to test this procedure. Though the data generated contained only one bias and one drift, the filtering model was designed to estimate biases in all instruments and drifts in the feed stream components.

In run S7 a standard LS filter was used to estimate the biases and drifts in the system on the basis of 100 observations. The drifts were correctly estimated, but Figure 5.20 shows that the biases were less well determined. This is again due to the fact that the system contains no information about the biases which can be cross-checked using the steady-state model. The filter is unable to cope with so many independent instrument errors. In run S8, the same filter was used, but every 20 sets the biases were tested, and set to zero if they were less than 0.1%; the appropriate row of the Jacobian matrix was also made null. This in effect reduces the dimensionality of the system, and slightly improves the performance of the estimator by eliminating erroneous bias indications. However, in both cases, the convergence to the true value of 1% in the product flowmeter is slow. It is evident that the estimation of a large number of independent variables will tend to swamp the estimator.

An oscillating memory filter was used in S9 to analyse the same data. It was hoped that the discrete changes in memory length would produce random changes in the biases of the correct instruments, and produce some non-random indication of the true biases. However, it was found that the updating of the P matrix (c.f. equation (3.33)) resulted in a singular matrix. This is further indication that the observations of the plant did not contain enough information to estimate nine independent biases.

(text continues on Page 102)

FIGURE 5.11

INDIVIDUAL RESULTS: A COMPARISON OF AVERAGING AND FILTERING OF STEADY-STATE DATA

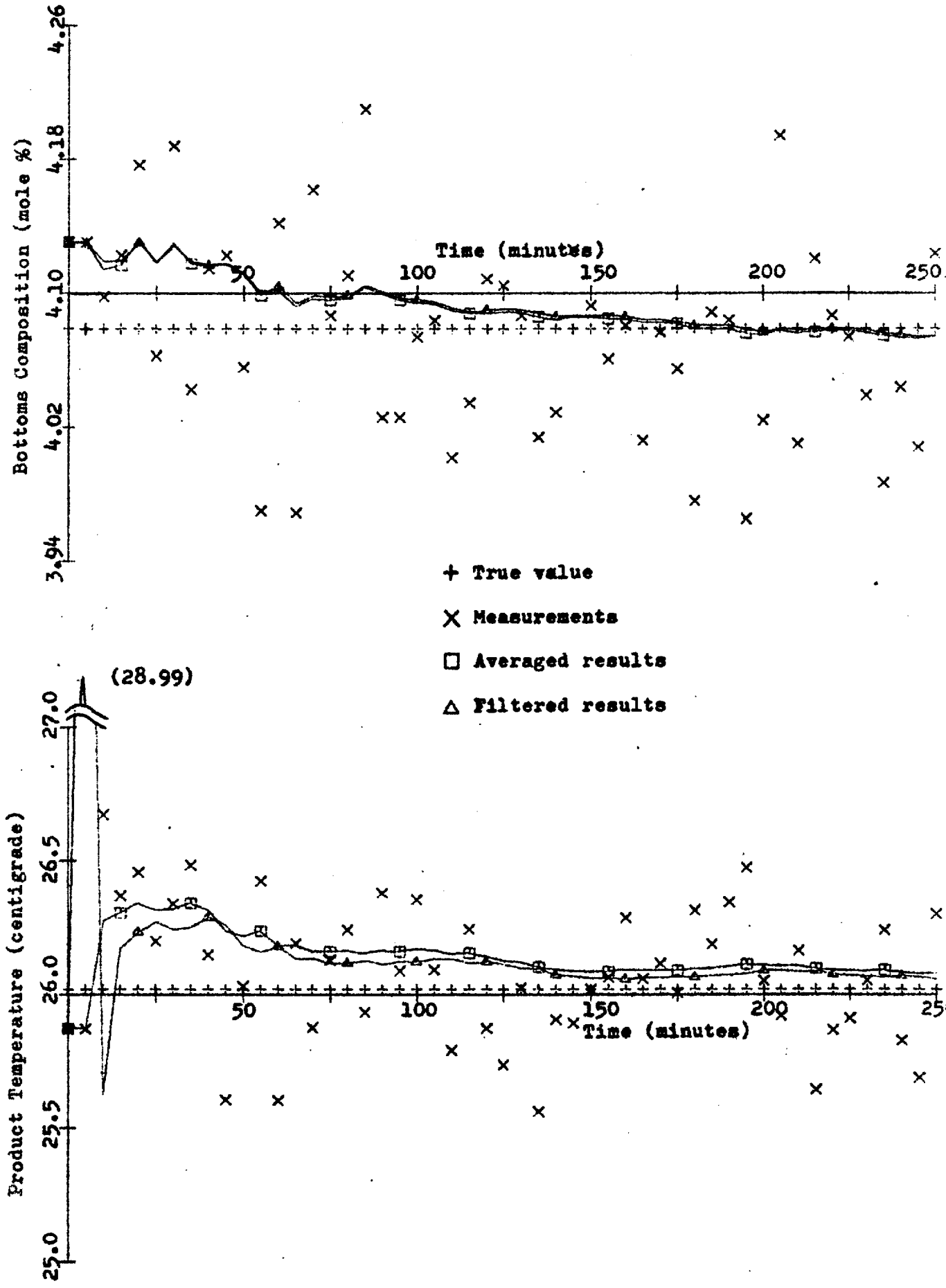


FIGURE 5.12

INDIVIDUAL RESULTS: THE ANALYSIS OF STEADY-STATE DATA FOR BIAS IN BOTTOMS FLOWMETER

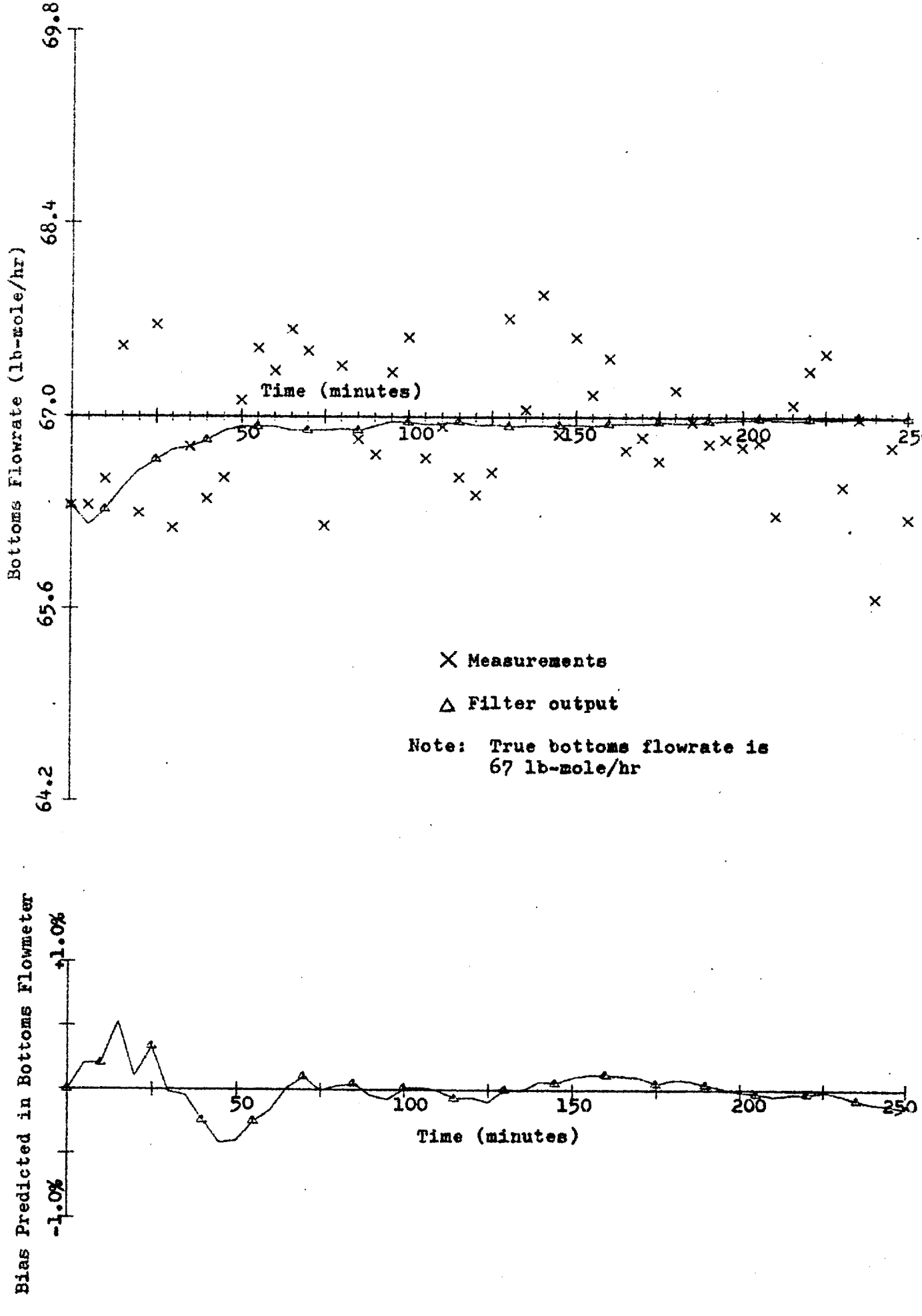
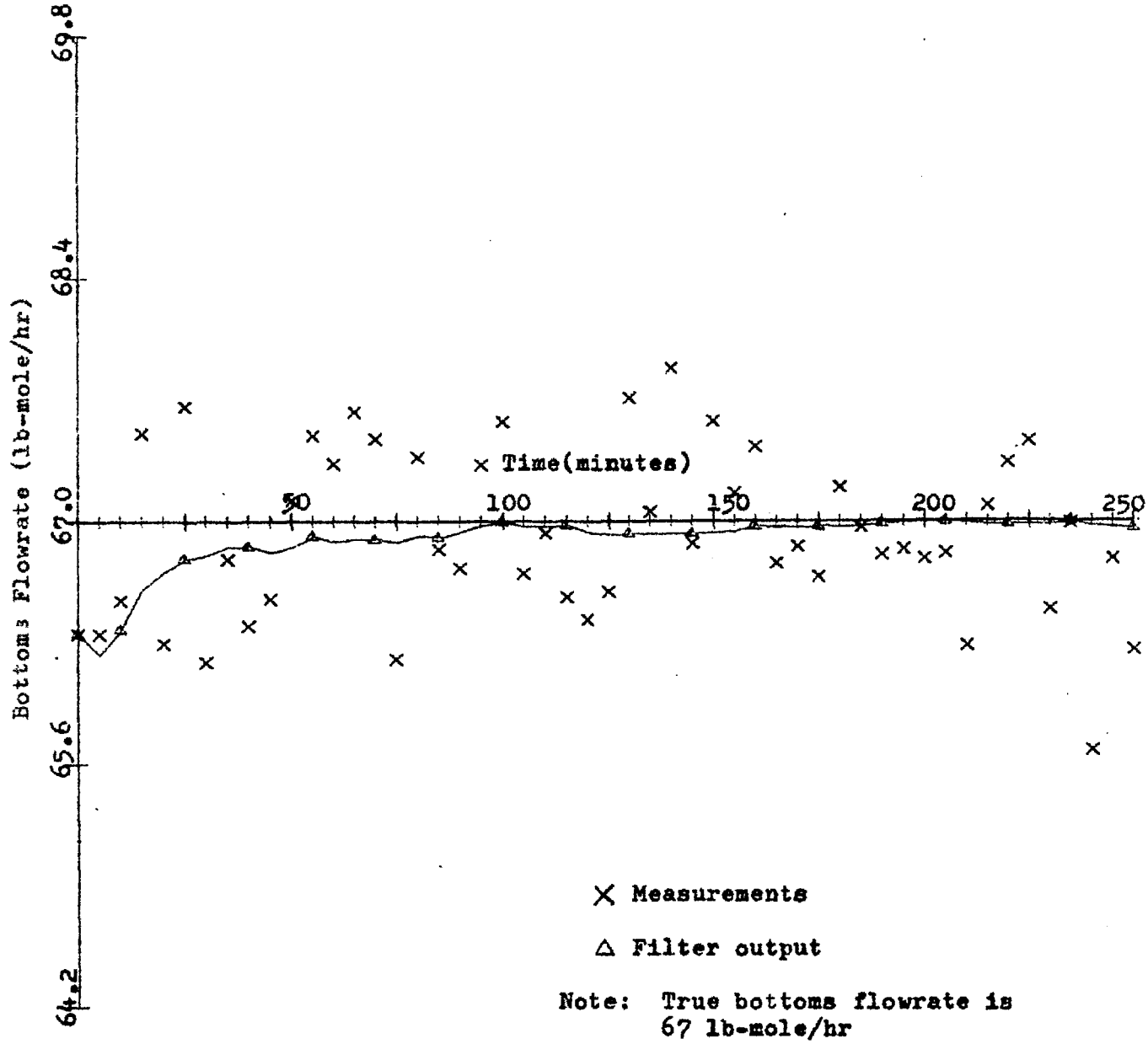


FIGURE 5.13

INDIVIDUAL RESULTS: THE ANALYSIS OF STEADY-STATE DATA FOR DRIFT IN BOTTOMS FLOWRATE



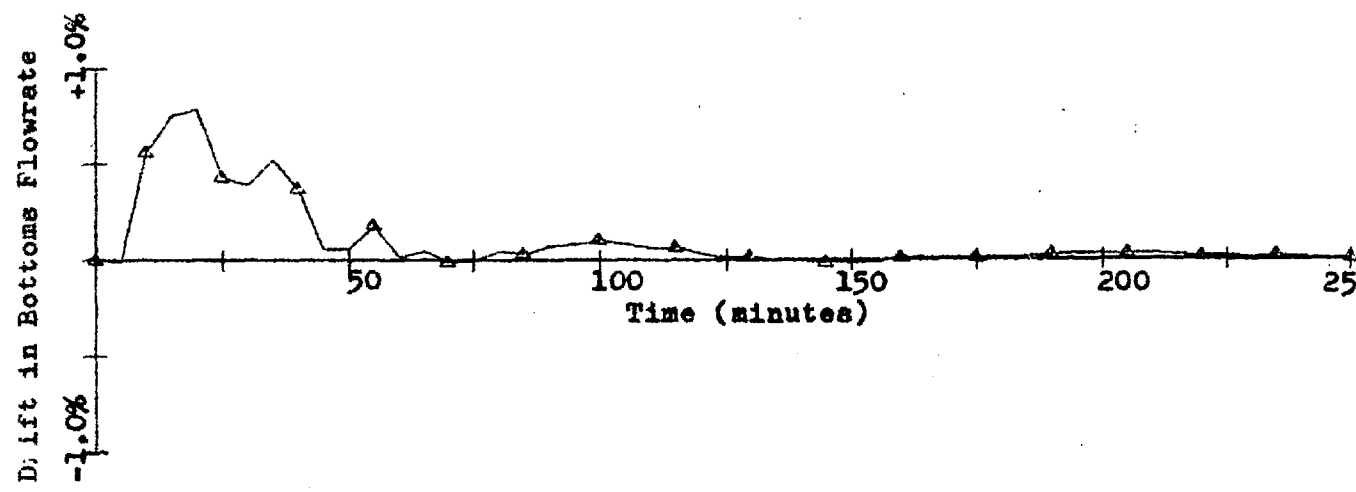


FIGURE 5.14

INDIVIDUAL RESULTS: THE ANALYSIS OF BIASED DATA FOR BIAS IN BOTTOMS FLOWMETER

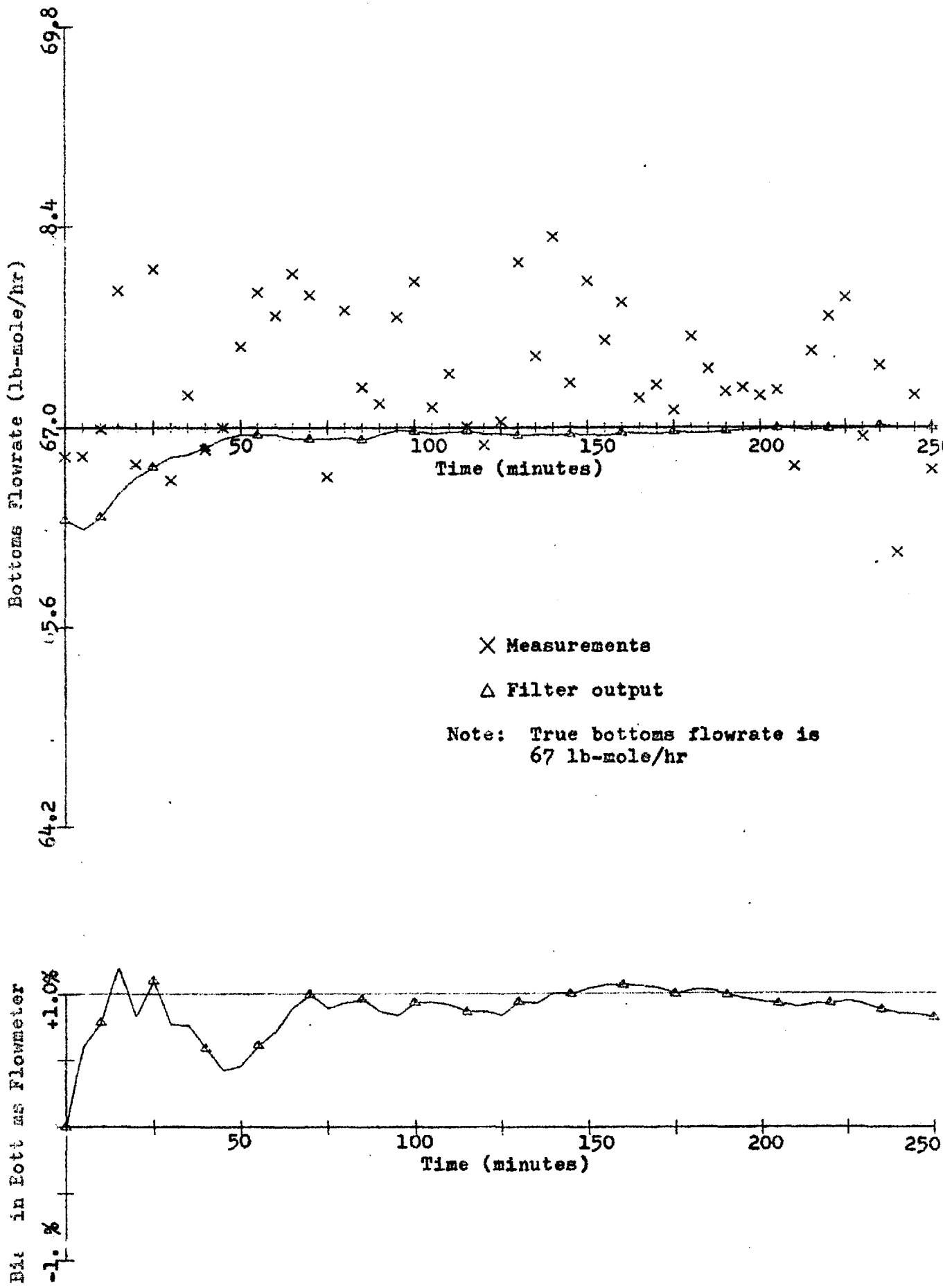


FIGURE 5.15

INDIVIDUAL RESULTS: THE ANALYSIS OF DATA WITH A DRIFT IN THE BOTTOMS FLOWRATE

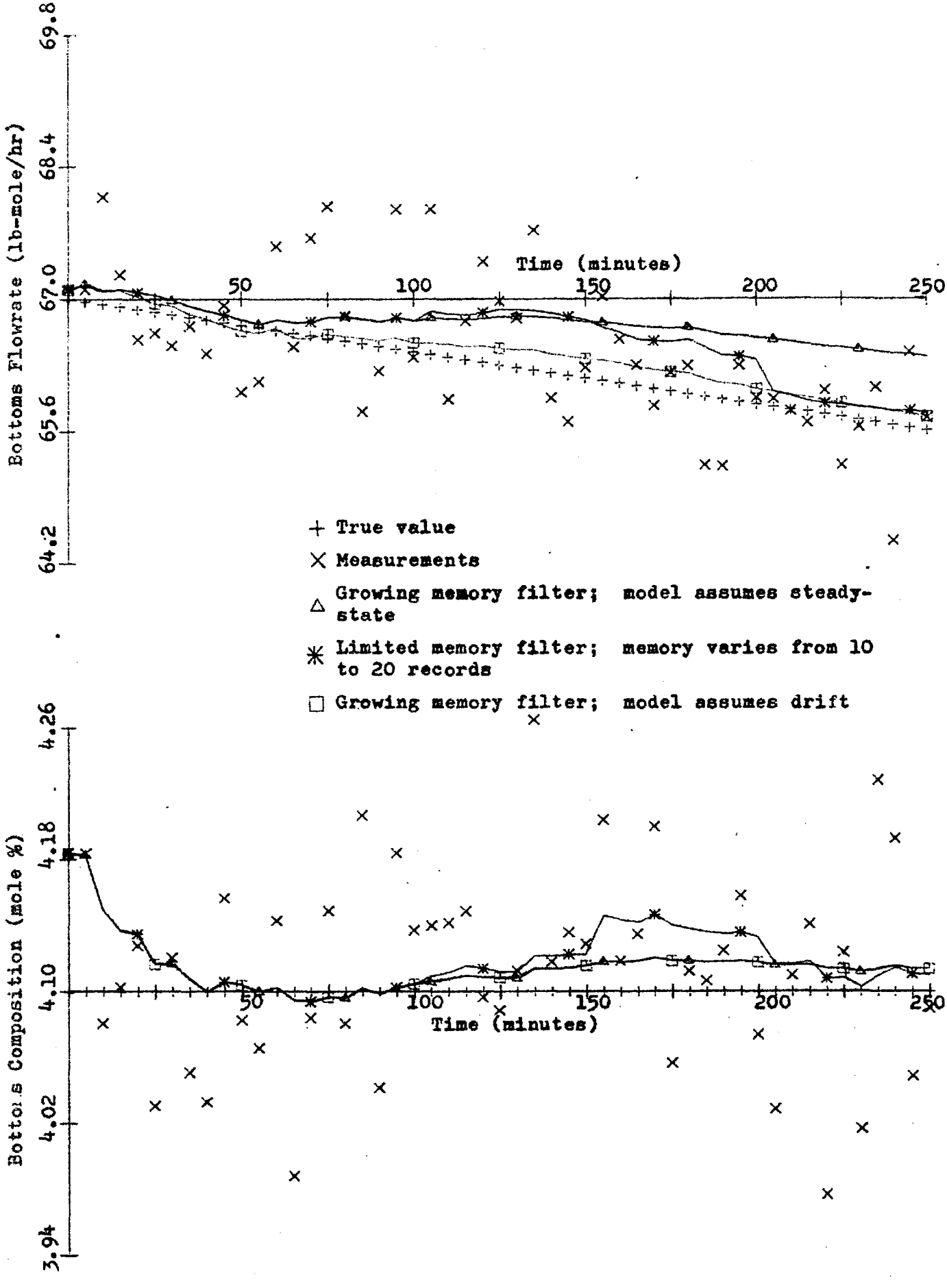


FIGURE 5.16

INDIVIDUAL RESULTS: THE ANALYSIS OF DATA WITH A DRIFT IN THE BOTTOMS FLOWRATE

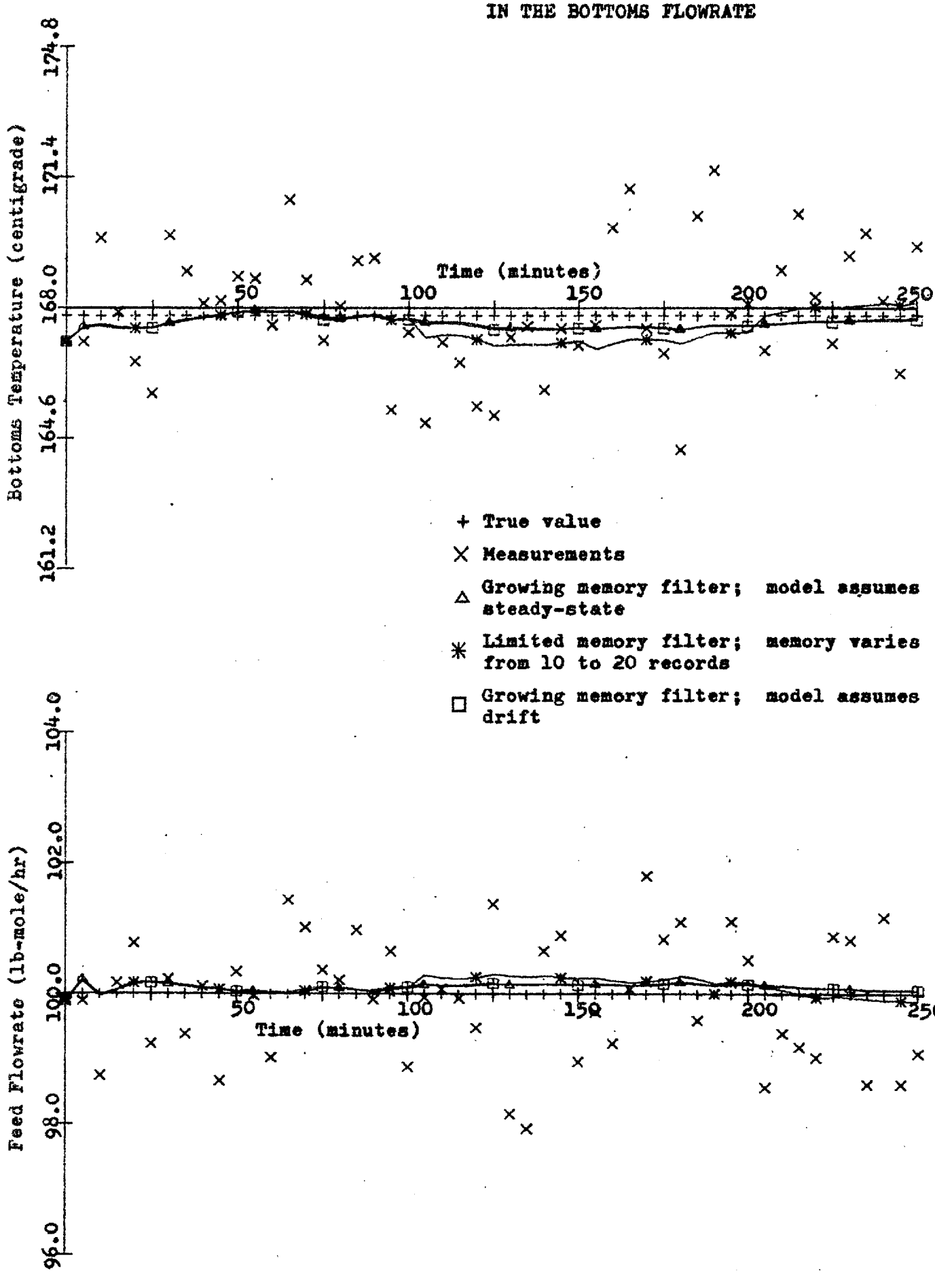


FIGURE 5.17

INDIVIDUAL RESULTS: THE ANALYSIS OF DATA WITH A DRIFT IN THE BOTTOMS FLOWRATE

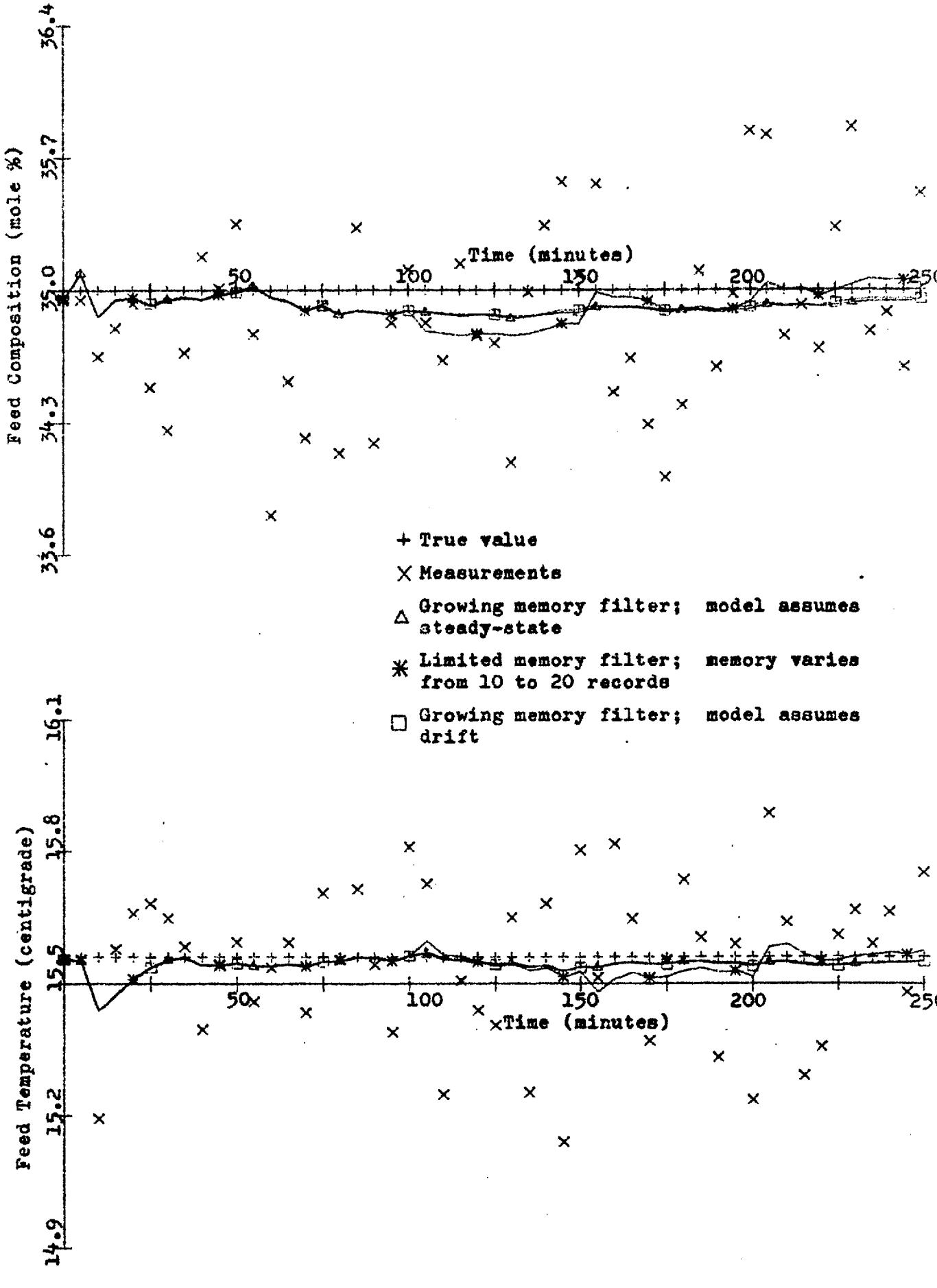


FIGURE 5.18

INDIVIDUAL RESULTS: THE ANALYSIS OF DATA WITH A DRIFT IN THE BOTTOMS FLOWRATE

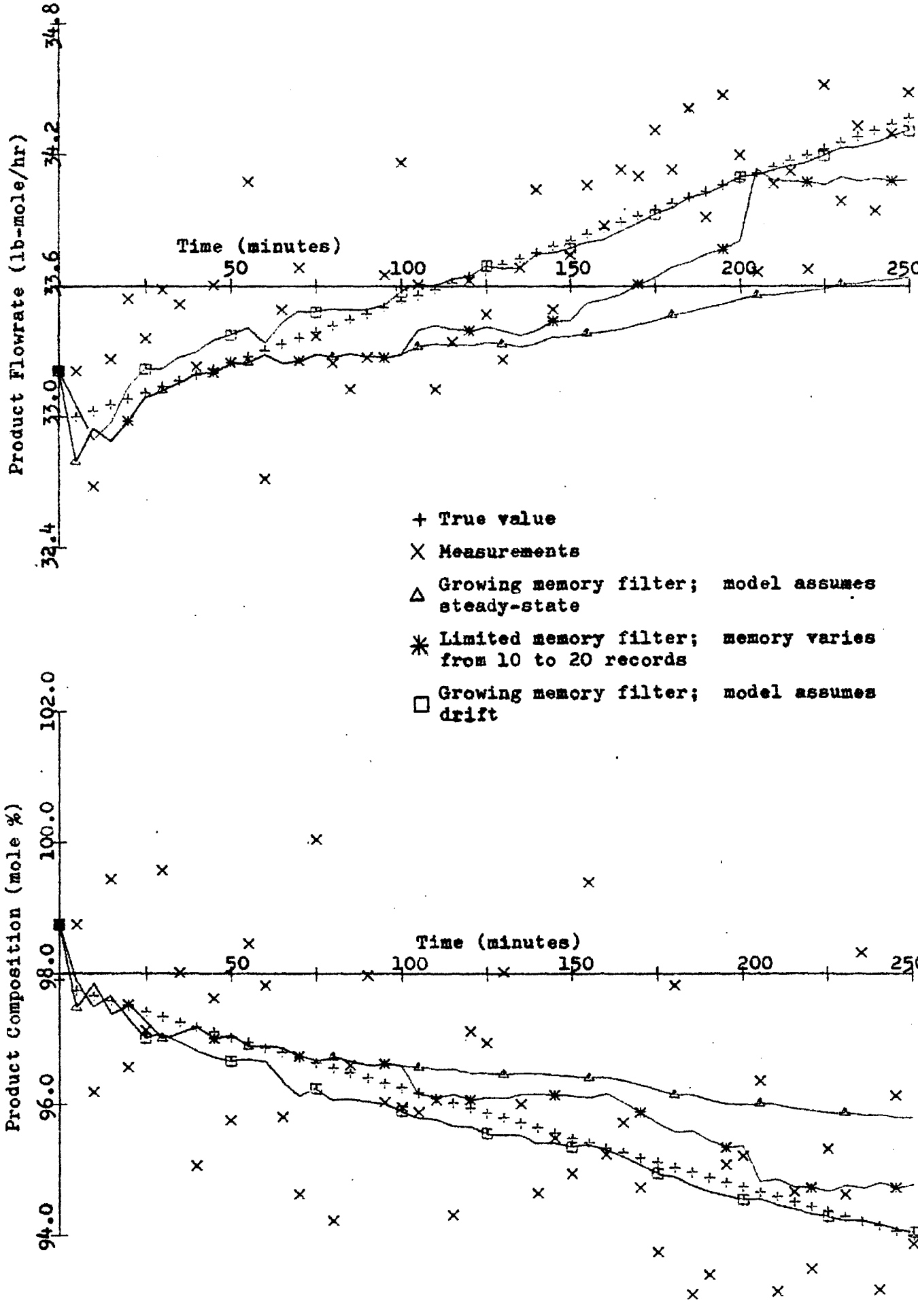


FIGURE 5.19

INDIVIDUAL RESULTS: THE ANALYSIS OF DATA WITH A DRIFT IN THE BOTTOMS FLOWRATE

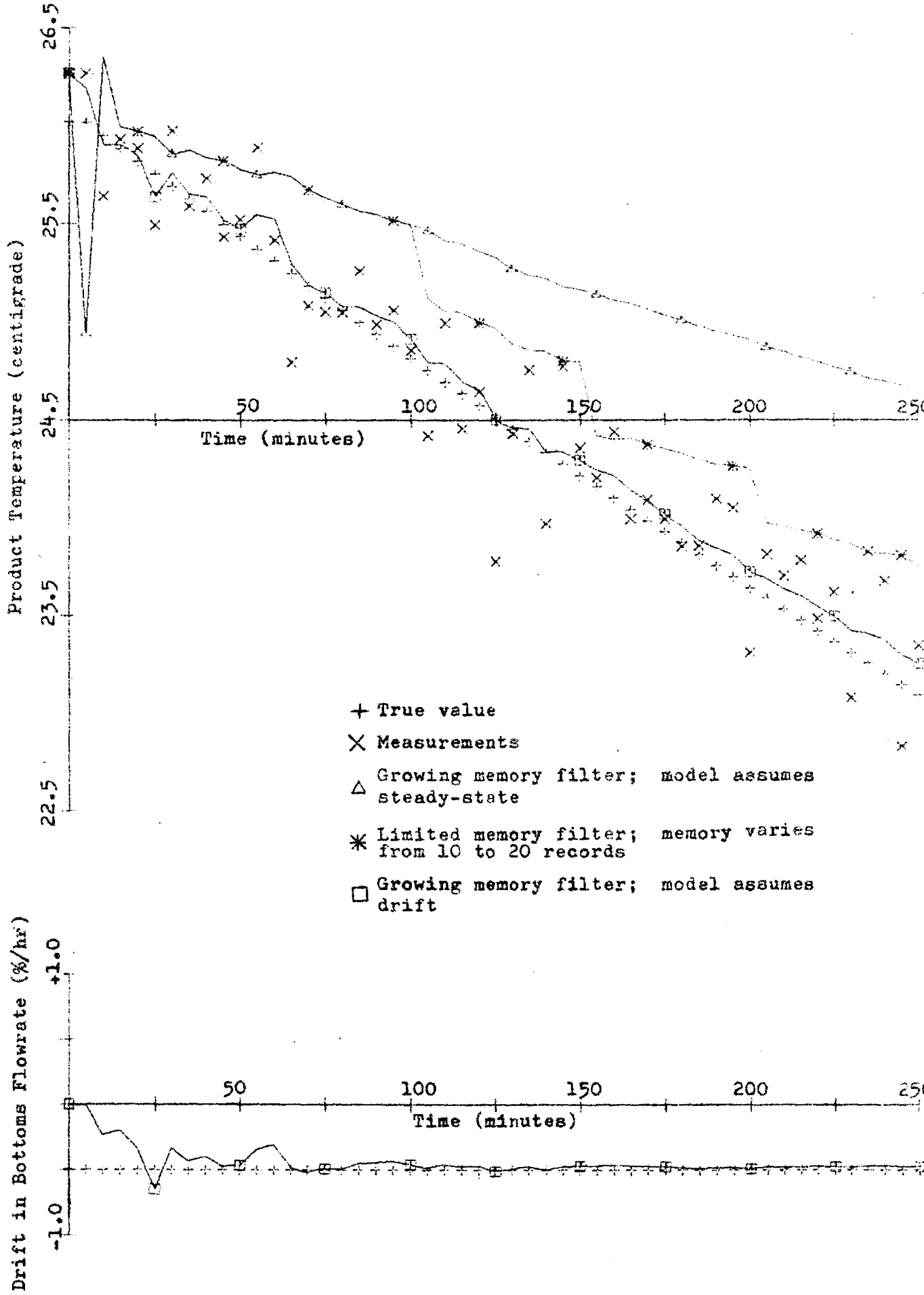
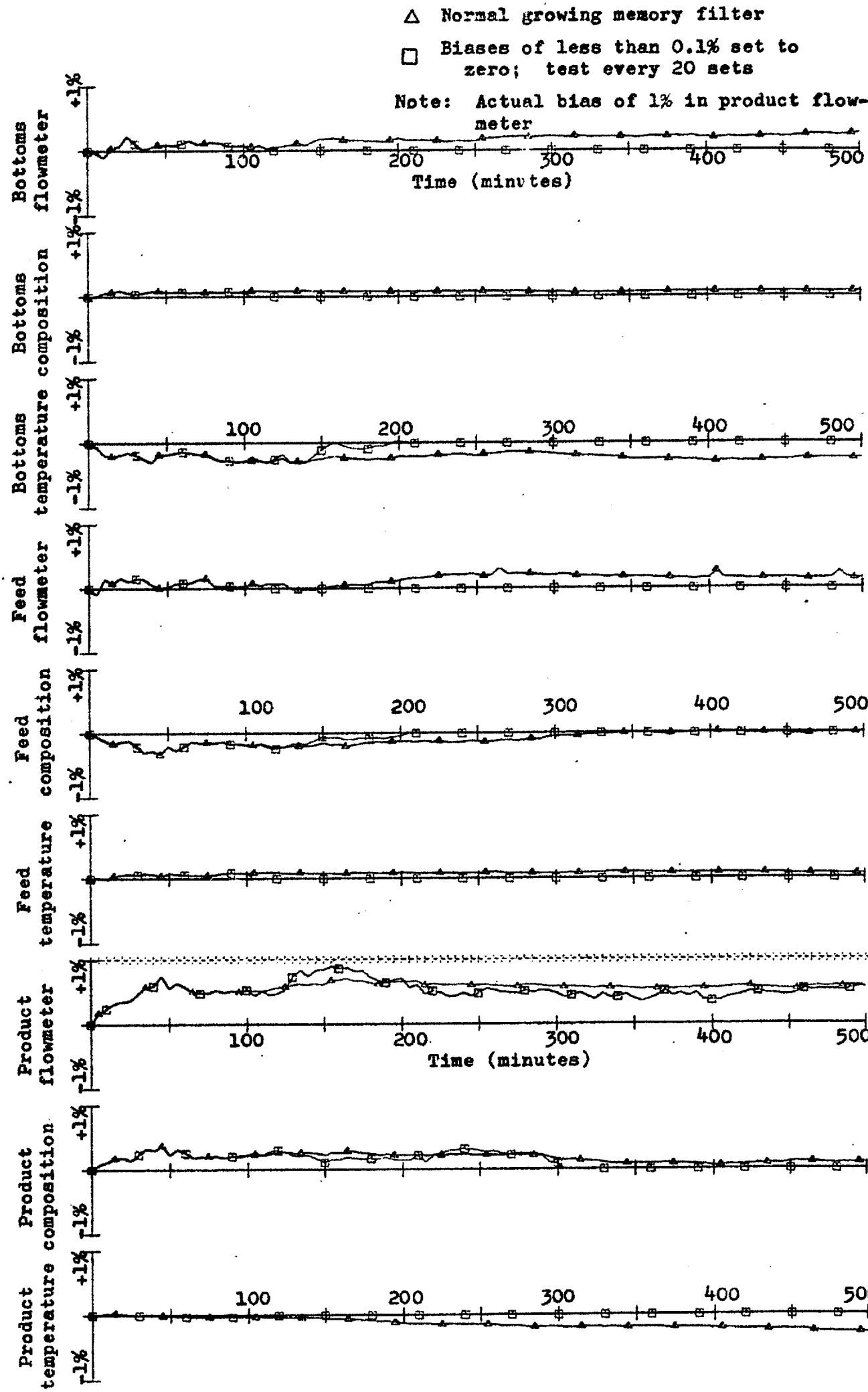


FIGURE 5.7

INSTRUMENT BIASES PREDICTED BY THE 18-DIMENSIONAL FILTER



5.3 Discussion of Results

This distillation column example has been chosen to be representative of typical chemical engineering systems. It should be noted that although the describing equations of the system seem to be nonlinear, the response to the disturbances considered is nearly linear. This can be seen by observing that a linear drift in the bottoms flowrate causes nearby linear responses in the produce temperature and composition. However, this has occurred because the perturbations are small; the linearity of the response is not a requirement of the application of the filter theory.

The sequential least-squares filter has been shown to be applicable to practical problems when accurate statistical data is not available. The ensemble results indicate that the convergence of the estimator is improved if an infinite variance is assumed for the initial error estimate. The effects of errors in the assumed instrument statistics are more difficult to predict, since they are problem dependent. However, some information concerning instrument reliability is usually available. In any event, erroneous statistical inputs merely alter the rate of convergence of the estimator, but do not affect its stability.

The presence of non-normal and correlated noise does not degrade the filter performance in this application. This is especially encouraging for chemical engineering applications where the nature of error statistics will rarely be known. Though it is possible to allow for correlation by altering the dimension of the filter, as is shown by Smith[27], it is preferable to be able to ignore all but the most pronounced relationships between instrument errors at successive readings. Most reasonable measurement errors can be allowed for by a bias with normal or rectangular fluctuations superimposed.

The estimation of independent biases has been shown to be possible, though they are more difficult to detect than effects which are coupled through the system model. There is unfortunately no set method for determining how many independent errors can be determined; however, the updating step with an OMF will yield a singular matrix if not enough information has been extracted by the measurements. It is clearly desirable to minimize the number of biases which must be detected, and to take advantage of any physical relationships between measurements.

The filter performs most poorly when there are errors in the model used to describe the plant. This is the most serious limitation of the theory in terms of practical chemical engineering applications, since exact process models are not often available. Though a limited memory filter has been shown to be useful for detecting model deficiencies in this example, the effects of model errors on the filter performance warrants further investigation. The problem is studied in Chapter 6 in an application of the filtering techniques to a chemical reaction system.

CHAPTER 6

FILTERING APPLIED TO A FIXED-BED CATALYTIC REACTOR.

In the preceding chapter, the application of the filtering algorithm to a system whose mathematical model is perfectly known has been considered. It has been demonstrated that the filter performance is not inordinately sensitive to errors in the statistical descriptions postulated for the process disturbances. However, the situation considered is rather unrealistic since the process model is rarely known to any degree of accuracy and even if an accurate simulation is available, it is likely to be too complex for on-line calculations. Thus errors in the system model, or simplifications made to allow on-line use of the model are likely to be a more serious source of errors in the filter output than poor assumptions of the statistical characteristics of the process disturbances.

This chapter describes the application of the filtering techniques to an industrial reactor for which a model has been provided by Imperial Chemical Industries, Ltd. The filter is used to estimate the parameters in a catalyst decay law which is required for a control optimization calculation. The techniques are tested using the complete ICI model, and these results are compared with those obtained using a greatly simplified model based upon the general physical characteristics of the system.

6.1 System Description.

6.1.1 The Complete Model

The reaction considered is multi-phase, and occurs at high pressure in a fixed-bed catalytic reactor with recycle. The reaction scheme is effectively A → B → C → D where the last product is formed by thermal degradation of C. The first two reactions are not first order, and proceed nearly to completion. There is a phase separator at the exit from the reactor, and the gas phase is recycled. The recycle composition is a function of the separation temperature, and the amount of unreacted A in the reactor outlet stream.

A mathematical model of the system has been developed by ICI, and consists of a series of adiabatic stirred tanks for which heat and material balances as well as equilibrium relationships are provided. The model is completely specified by the following information:

1. Feed flowrates
2. Reactor input temperature
3. Separator temperature
4. Recycle flowrate
5. Catalyst activity profile along the bed.

The model is solved by direct iteration on the amount of reactant A in the outlet stream, since knowledge of this concentration is required for the calculation of the recycle composition. However, it has been found that this composition is always nearly zero, and when the value is set to zero for calculation of the recycle composition the model is reported to predict the sixteen equidistant temperature measurements along the length of the reactor and the outlet compositions to within 2%.

When the catalyst decay law is specified, the model can be used in the calculation of optimal controls for the entire operating period. It is intended to test the feasibility of using statistical filtering techniques for the on-line estimation of the parameters in the decay law.

The specific details of the ICI model cannot be revealed here; however, the proprietary nature of the information is actually an advantage for this study. The complete model is used to generate "real" plant data by perturbing its output with random measurement noise, but for the purposes of filtering, it is assumed that the plant model is either unknown, or too complex for on-line computation. Therefore, a simplified model based upon the general physical characteristics of the reactor is postulated for use in the filtering algorithm. This is clearly a more realistic situation than one which assumes perfect knowledge of the mathematical model of the plant.

6.1.2 The Simplified Model

The simplified model chosen is an A $\rightarrow$ B $\rightarrow$ C reaction with first order kinetics occurring in an adiabatic plug flow reactor. It has been assumed by ICI that the small amount of D formed by degradation can be predicted from the outlet conditions, and this reaction is ignored in the sequel. Thus;

$$\frac{da}{dx} = - k_1 \lambda (x,t) \exp \left[-E_A/R(T + 273)\right] \cdot a \qquad (6.1)$$

$$\frac{db}{dx} = - k_2 \lambda (x,t) \exp \left[-E_B/R(T + 273)\right] \cdot b - \frac{da}{dx} \qquad (6.2)$$

$$\frac{dT}{dx} = - k_3 \frac{da}{dx} - k_4 \left[\frac{da}{dx} + \frac{db}{dx}\right] \qquad (6.3)$$

where λ (x,t) is a function describing the evolution of the catalyst profile, and a and b are extents of reaction; T is the temperature in degrees centigrade and R is the universal gas constant. E_A and E_B, the activation energies, are assumed known, while k_1 ... k_4 are determined by off-line minimization of the sum of squares deviations of the temperature profile predicted by the simplified model from the profile of the "real" reactor. A minimization which allowed E_A and E_B to vary as well as the k's failed to converge; the poor performance of the six parameter search is probably due to the sensitivity of the objective function of the exponential in the Arrenhius dependencies, and to the fact that the model becomes too general. Powell's method for minimizing a sum of squares without derivatives was used for the search[53].

Table 6.1 lists the deviations of the output predicted by the simplified model from that predicted by the complete one for various values of catalyst activity, assuming no variation of activity with distance along the reactor. Table 6.2 lists the optimal parameter values when both E_A and E_B are 17,000 cal/g.mole. The modelling errors indicated in Table 6.1 must be accounted for when using the simplified model to predict the reactor profile, and the values of the prediction biases are added to the output of the simplified model to correct for these errors. Methods of updating the model bias corrections are considered later.

Table 6.1 Deviation of the Best Fit Plug Flow Reactor from the True Profile

		λ=3000	λ=2893*	λ=2500	λ=2000
TEMPERATURE INDICATORS	1	-0.78C°	-0.76C°	-0.59C°	-0.52C°
	2	-1.30	-1.28	-1.05	-0.96
	3	-1.46	-1.45	-1.24	-1.23
	4	-1.31	-1.34	-1.23	-1.37
	5	-0.93	-0.99	-1.03	-1.38
	6	-0.36	-0.47	-0.68	-1.28
	7	+0.29	+0.17	-0.23	-1.08
	8	+0.91	+0.79	+0.28	-0.80
	9	+1.26	+1.22	+0.76	-0.47
	10	+1.18	+1.28	+1.12	-0.09
	11	+0.73	+0.92	+1.20	+0.27
	12	+0.20	+0.36	+1.03	+0.57
	13	-0.17	-0.11	+0.41	+0.72
	14	-0.36	-0.38	-0.18	+0.68
	15	-0.44	-0.50	-0.62	+0.44
	16	-0.46	-0.55	-0.89	+0.06
Output Conc. of A		+1.91%	+1.08%	+1.08%	+0.03%

* Activity after 8 hours assuming decay over 150 hours.

Note: Typical inlet temperature = 95°C.; typical outlet temperature = 130°C.

Table 6.2 Parameters for the Best Fit Plug Flow Reactor

λ	$\ln k_1'$	$\ln k_2'$	k_3	k_4
3000	23.91	24.09	-0.59	30.44
2893	23.88	24.06	-0.60	30.34
2500	23.72	23.90	+0.30	29.06
2000	23.45	23.63	+0.49	30.69

$$k_1' = k_1 . \lambda$$

6.1.3 The Simulated Data

Plant data was generated by perturbing the measured variables generated by the complete model with normal random noise. The instrument errors had zero means, and were scaled so that the 95% confidence intervals of all temperature sensors is $\pm$ 1C^{o} and that of all flowmeters is $\pm$ 5%. Only temperature and flowrate measurements were generated, as it was assumed that compositions at the exit from the reactor could not be conveniently measured.

Various sets of data were generated; all had the same feed and recycle flowrates and input and separation temperatures. Several different catalyst decay laws were considered, as is shown in Table 6.3.

Table 6.3 Data Sets for Reactor Study

Set	Catalyst Activity	Measurements
S1	Constant; $\lambda(x,t) = 3000$ for all x and t	Feed flowrates, recycle and purge flowrate, input and separation temperatures, and 16 temperatures along the length of the reactor.
S2	Constant with respect to distance at any time. Decay over 150 hours proportional to the inlet temperature. $\lambda(o) = 3000$, $\lambda(150) = 1000$.	16 temperatures along reactor and input temperature. Made every 5 minutes during first eight hours.
S3	As S2, but decay dependent upon local temperature. $\lambda(x,o) = 3000$ for all x.	16 temperatures and inlet temperature measured every 5 minutes from hour 16 to hour 24.

The data for set S3 was generated by defining a separate catalyst activity, $\lambda(i)$, for each of the stirred tanks in the model. Thus at time t_{N+1}, the activity in tank i is calculated by:

$$\lambda(i)_{N+1} = \lambda(i)_N - \alpha T_N \qquad (6.4)$$

where T_N is the temperature in the tank at time t_N. The data in set S3 was taken from hour 16, by which time the catalyst profile varied from 2785 at the inlet to 2720 at the outlet.

6.1.4 The Filtering Models

The only dynamic effect considered in the system is the catalyst decay. The dimension of the state vector of the estimator varies depending upon the form of decay that is proposed, but in each case the problem is formulated so that only constants are being estimated, and $\Phi_{N/N-1}=I$ for all N. Table 6.4 describes the various models used in the studies.

Table 6.4 Models Used for Filtering of Reactor Data

Model	Catalyst Decay Law	State Variables
M1	$\dot{\lambda}(x,t) = 0$ $\lambda(o,o) = \lambda_o$	λ_o, feed flowrates, recycle flowrate, input and separation temperatures. (ICI Model)
M2	$\dot{\lambda}(x,t) = 0$ $\lambda(o,o) = \lambda_o$	λ_o
M3	$\dot{\lambda}(x,t) = -\alpha T_I$ $\lambda(x,o) = \lambda_o$ for all x. Thus $\lambda(x,t) = \lambda_o(x,o) - \alpha T_I t$ where T_I is input temperature	λ_o, input temperature, α (decay rate)
M4	$\dot{\lambda}(x,t) = -\alpha T(x)$ $\lambda(x,o) = \lambda_o$ for all x. Approximated by: $\lambda(x,t) = \lambda_o - \alpha T_I \cdot t - \beta x$ where x is the normalized reactor length.	λ_o, T_I, α, β

Note that model M4, the case where the catalyst activity is a function of local temperature, is in fact a distributed parameter system that requires an infinite dimensional state vector to describe the activity at every point along the length of the reactor. The situation is made tractable by postulating a low order polynomial in x (distance along the reactor) to describe the profile at any time. For this particular example, simulations showed that the catalyst profile could be accurately described by a linear function of distance. However, a higher order polynomial could easily be postulated, and would require only the estimation of a few more coefficients. Note that the first three variables in the state vector were truly constant during the data generation, or a known function of time in the case of T_I. However, the value of α and β in the simplified dynamic description would be expected to vary slowly in time.

It should also be noted that in models M2 to M4, the flowrates are assumed perfectly known, and only temperatures are measured. This simplification was made for convenience, since it will be shown that uncertain flow measurements do not alter the stability of the filter, but only slow its rate of convergence. Using this simplified measurement scheme it is possible to ignore the recycle stream, and interpret the system as a one-pass reactor.

Given the elements of the state vector, the catalyst activity as a function of distance can be calculated at any time. Once the catalyst profile is known, the steady-state model (either simplified or complete) can be used to predict the temperature profile and output compositions.

The Jacobian matrix, G, required for the filtering algorithm is calculated using a 0.1% perturbation of the state variable. Thus

the steady-state profile is generated using the current value of the state-vector, and then the profile is recalculated n times, perturbing a different element of the state-vector each time. Whenever possible, a solution of the reactor model is avoided by using the chain rule for differentiation. For instance, in the case where the catalyst decay depends upon inlet temperature (model M3);

$$\lambda(t) = \lambda_o - \alpha T_I t \qquad (6.5)$$

at any time t. $dy_i/d\lambda$ is calculated by perturbation for each measuring device y_i, and then;

$$\frac{dy_i}{d\lambda_o} = \frac{dy_i}{d\lambda} \cdot \frac{d\lambda}{d\lambda_o} \qquad i = 1 \ldots m \qquad (6.6)$$

where m is the number of instruments. The other rows of G follow similarly using (6.5) and (6.6). The size of the perturbation has been selected rather arbitrarily on the basis of off-line studies. In most cases it should be relatively easy to select a reasonable step length for each state variable. No difficulties were encountered in this example.

6.2 Studies of the Filter Performance

The ability of the filter to determine catalyst activity was tested in the simplest situation, where the activity is constant in both time and distance along the reactor. Data set S1 was analysed using the full recycle model (M1) in the filter, and in PREDCT, the prediction routine. Various initial estimates, λ_o, of the catalyst activity were chosen, and the convergence to the true value was observed. All other elements of the state vector were correctly specified initially, and the outlet composition of reactant A was fixed when calculating the recycle composition, so that the prediction of the reactor profile did not require iterative calculation of the model. As was noted, this simplification is also employed by ICI when using the complete recycle model for optimization, and does not degrade its performance.

The same data set was analysed in a similar study for which model M2 was used in the filter. In this case, all flowrates were assumed to be measured accurately, and only temperature measurements along the length of the bed were considered; the outlet composition of reactant A was fixed as before for use in calculation of the recycle composition. This test was performed to determine whether the characteristics of the estimator could be evaluated using a one-pass model which ignores the uncertainty in the system flowrates.

It can be seen from Table 6.5 that the one-dimensional model yields results that are comparable with those obtained from the full model. The convergence of the filter using the simpler model is faster because there is no uncertainty in the flowrate measurements, but the filter is stable in both cases. Since the one-dimensional model is faster computationally, it was used in the remainder of the study.

Table 6.5 Convergence of the Estimate to the True Catalyst Activity of 3000

Time (min)	Full recycle	Full one-pass	Simplified one-pass
0	2500	2500	2500
10	2721	2780	2783
20	2782	2852	2854
30	2801	2887	2890
40	2839	2914	2917
50	2864	2927	2930
60	2887	2936	2939
70	2895	2947	2950
80	2901	2954	2956
90	2907	2956	2959
100	2916	2960	2962
110	2917	2964	2966
120	2921	2967	2968

a) λ_o = 2500

Time (min)	Full recycle	Full one-pass	Simplified one-pass
0	2900	2900	2900
10	2960	2954	2954
20	2974	2972	2972
30	2958	2982	2983
40	2973	2992	2994
50	2985	2993	2995
60	3001	2994	2996
70	2996	2998	3000
80	2996	2998	3002
90	2996	2997	3000
100	3002	2999	2999
110	3000	2999	3001
120	2999	2999	3001

b) λ_o = 2900

Time (min)	Full recycle	Full one-pass	Simplified one-pass
0	3100	3100	3100
10	3026	3019	3019
20	3014	3011	3010
30	2984	3010	3011
40	2991	3015	3016
50	3001	3012	3013
60	3005	3010	3012
70	3006	3013	3014
80	3006	3013	3014
90	3005	3009	3010
100	3011	3008	3009
110	3007	3009	3010
120	3007	3008	3009

c) λ_o = 3100

Time (min)	Full recycle	Full one-pass	Simplified one-pass
0	3500	3500	3500
10	3121	3117	3118
20	3028	3061	3061
30	2975	3043	3045
40	2973	3040	3042
50	2983	3031	3033
60	2998	3026	3028
70	2988	3026	3029
80	2990	3024	3027
90	2990	3019	3022
100	2998	3018	3019
110	2995	3017	3019
120	2997	3015	3017

d) λ_o = 3500

Table 6.5 also illustrates the effect of using the simplified model in the predictor. The biases listed in Table 6.1 for an activity of 3000 were added to the predictions of the model. In practice, values for these biases would be determined by performing an off-line minimization based on operating plant data. The optimal parameters would be determined as before, and the biases chosen to compensate for the difference between the predicted and measured temperature profiles. As is noted in Table 6.2, the minimization calculates k_1' and k_2' which are overall reaction constants. The activity level of 3000 was defined arbitrarily for a fresh catalyst bed, and therefore $k_1 = k_1'/3000$ and $k_2 = k_2'/3000$. Of course, any other factor for the catalyst activity could be chosen.

A comparison of the second and third colums of the results shown in Table 6.5 indicates that the simplified and full models yield essentially identical results for the situation with constant catalyst activity. The table also shows that the convergence of the estimator is sensitive to the initial value assumed for the catalyst activity. A poor initial estimate, however, is clearly indicated by a large change in the estimate after the first observation. It should be further noted that the convergence is not symmetric; that is convergence curves from estimates initiating at points below the true value of the activity are not mirror images of those initiating the same amount above the true value. This is a reflection of the dependence of the Jacobian on both time and position in state-space.

It is ultimately intended to use the filter to provide an estimate of the catalyst decay law for an optimization calculation. Since the objective function of the optimization is likely to depend upon the output concentration of reactants from the system, the true test of the filter performance lies in its ability to predict reactor

(text continues on Page 124)

FIGURE 6.1

PREDICTION OF OUTPUT CONCENTRATION USING AN INCORRECT FILTERING MODEL

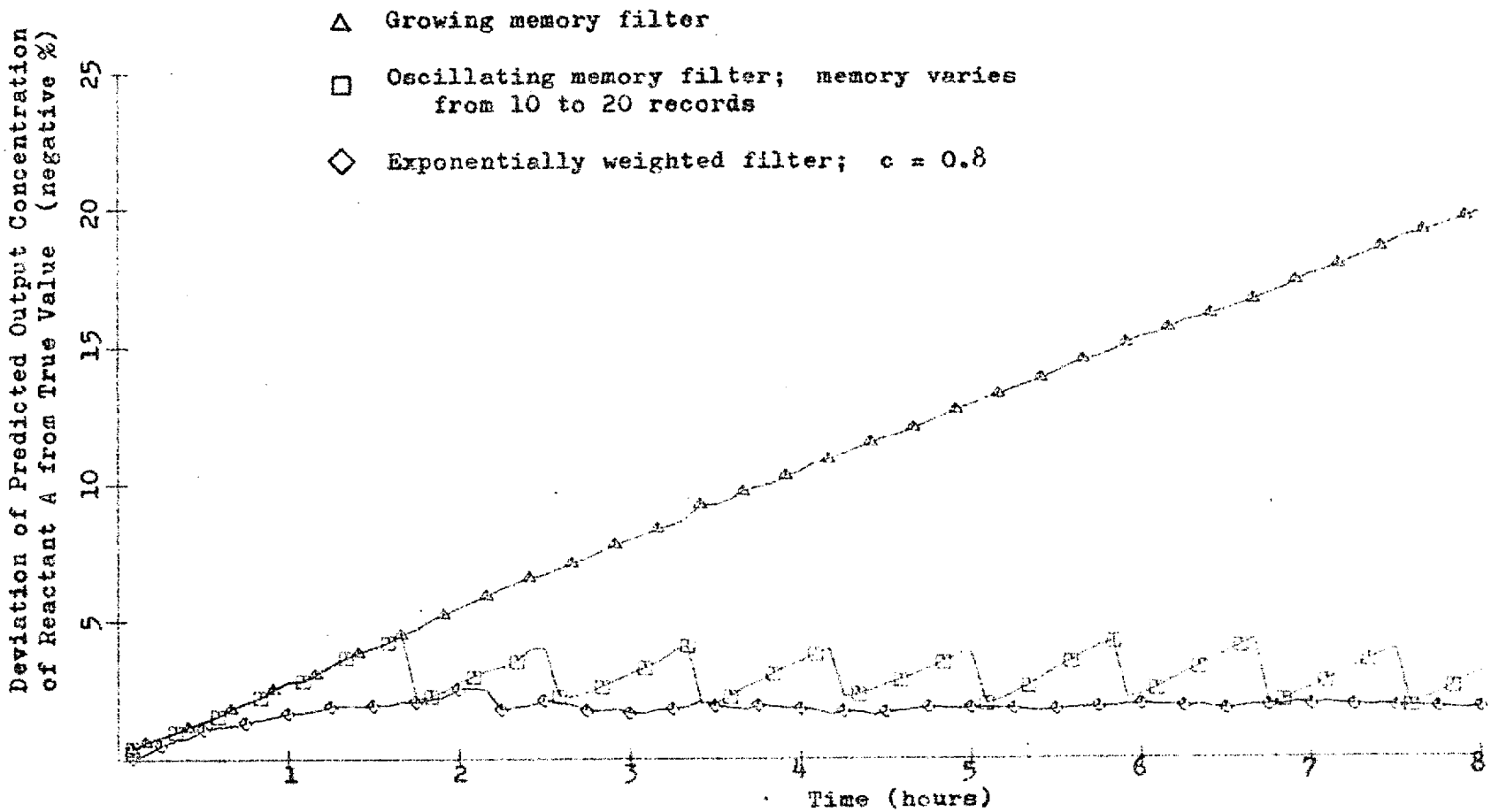


FIGURE 6.2

PREDICTION OF CATALYST ACTIVITY USING FULL MODEL

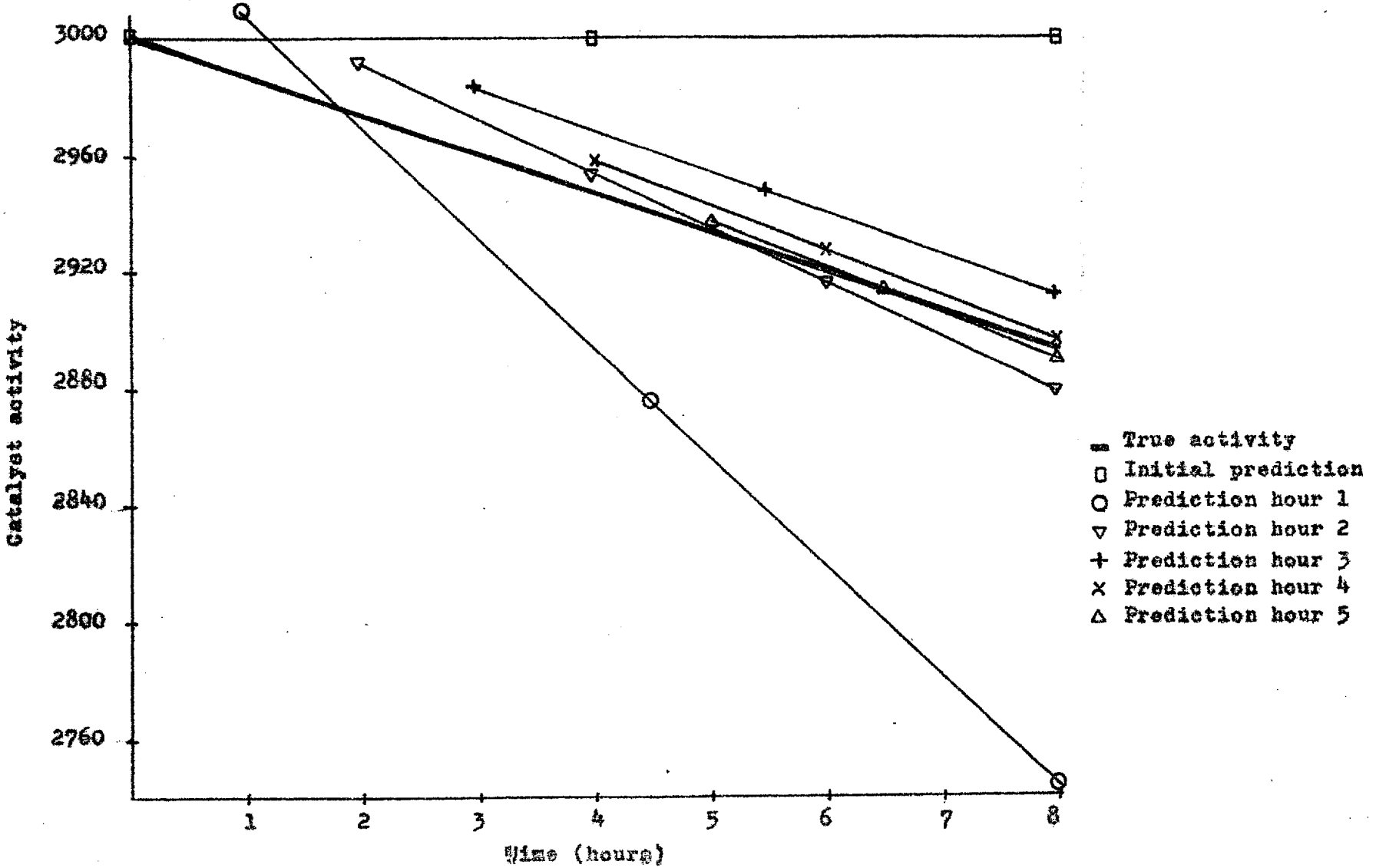


FIGURE 6.3

PREDICTION OF CATALYST ACTIVITY USING SIMPLIFIED MODEL

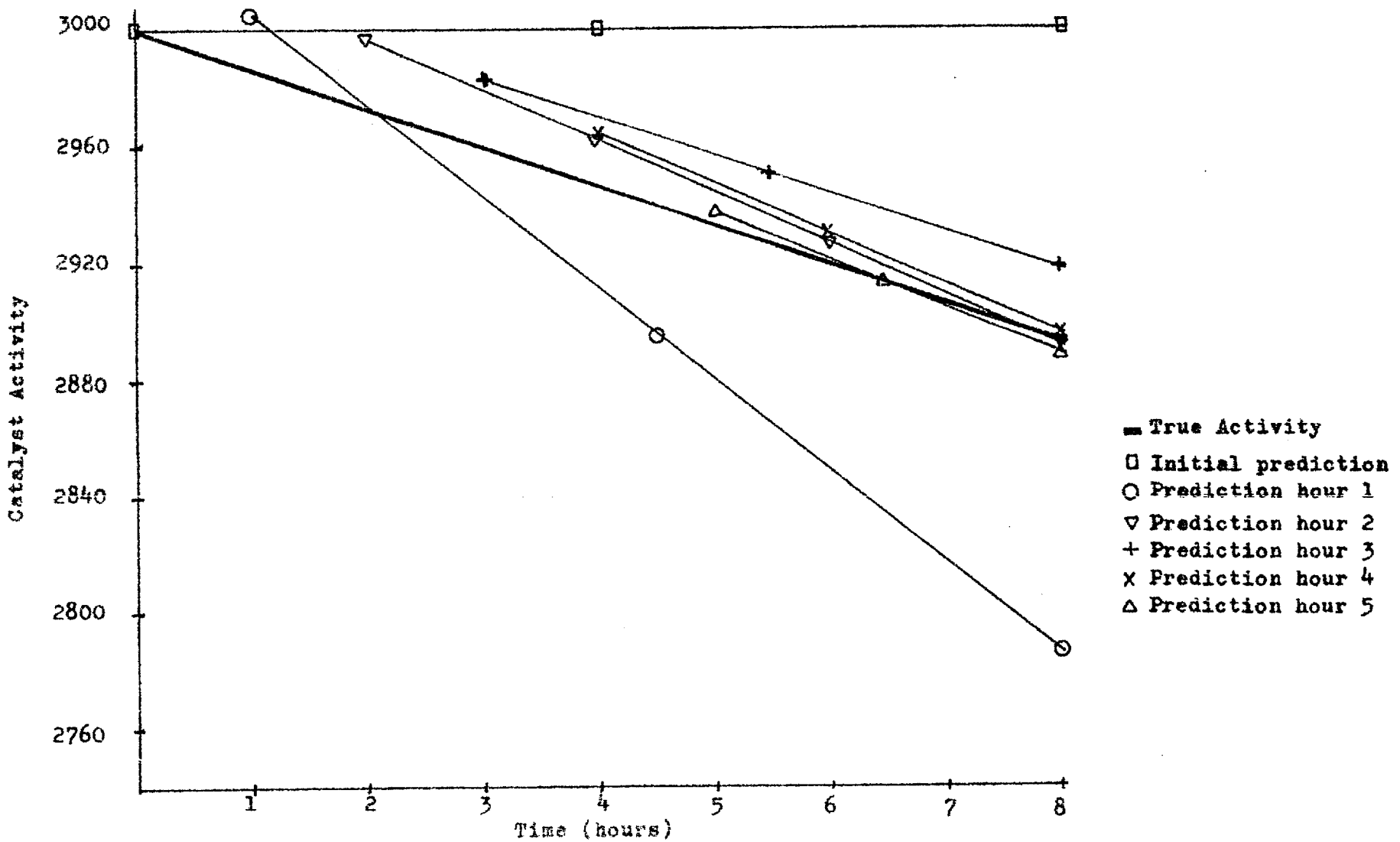


FIGURE 6.4

PREDICTION OF OUTPUT CONCENTRATION
USING FULL MODEL

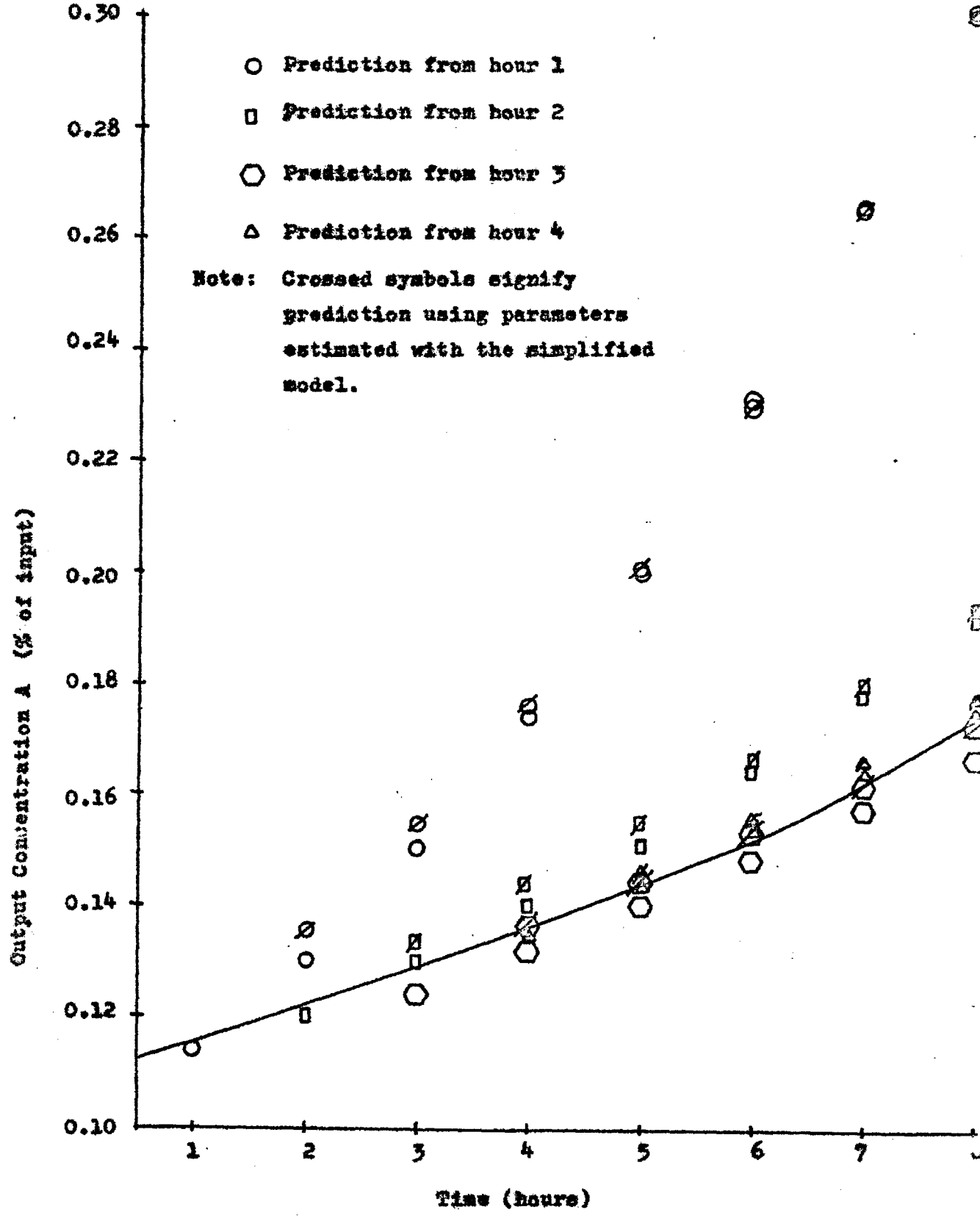


FIGURE 6.5

PREDICTION OF OUTPUT CONCENTRATION USING SIMPLIFIED MODEL

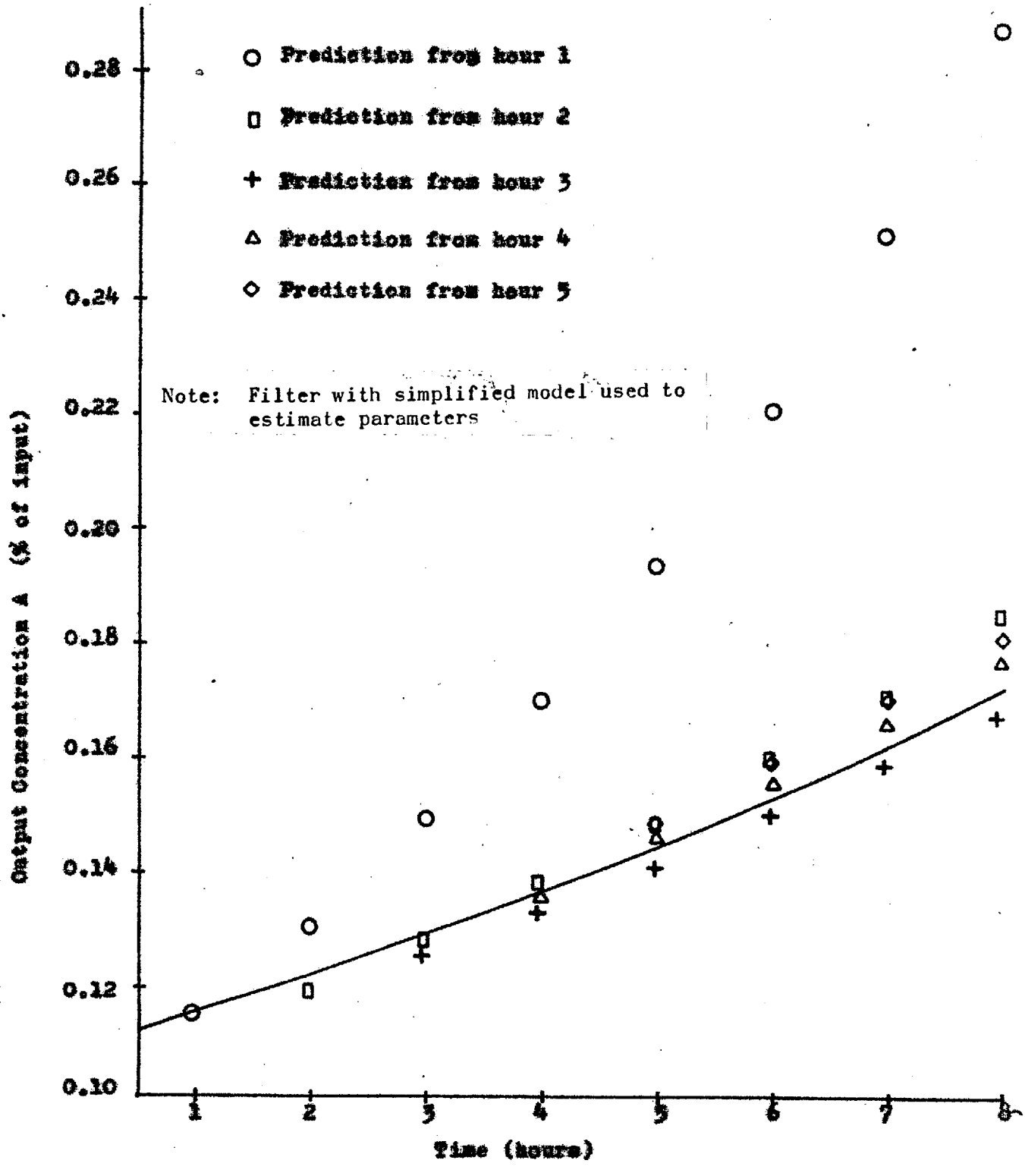


FIGURE 6.6

FILTERING RESULTS USING THE 20-DIMENSIONAL MODEL

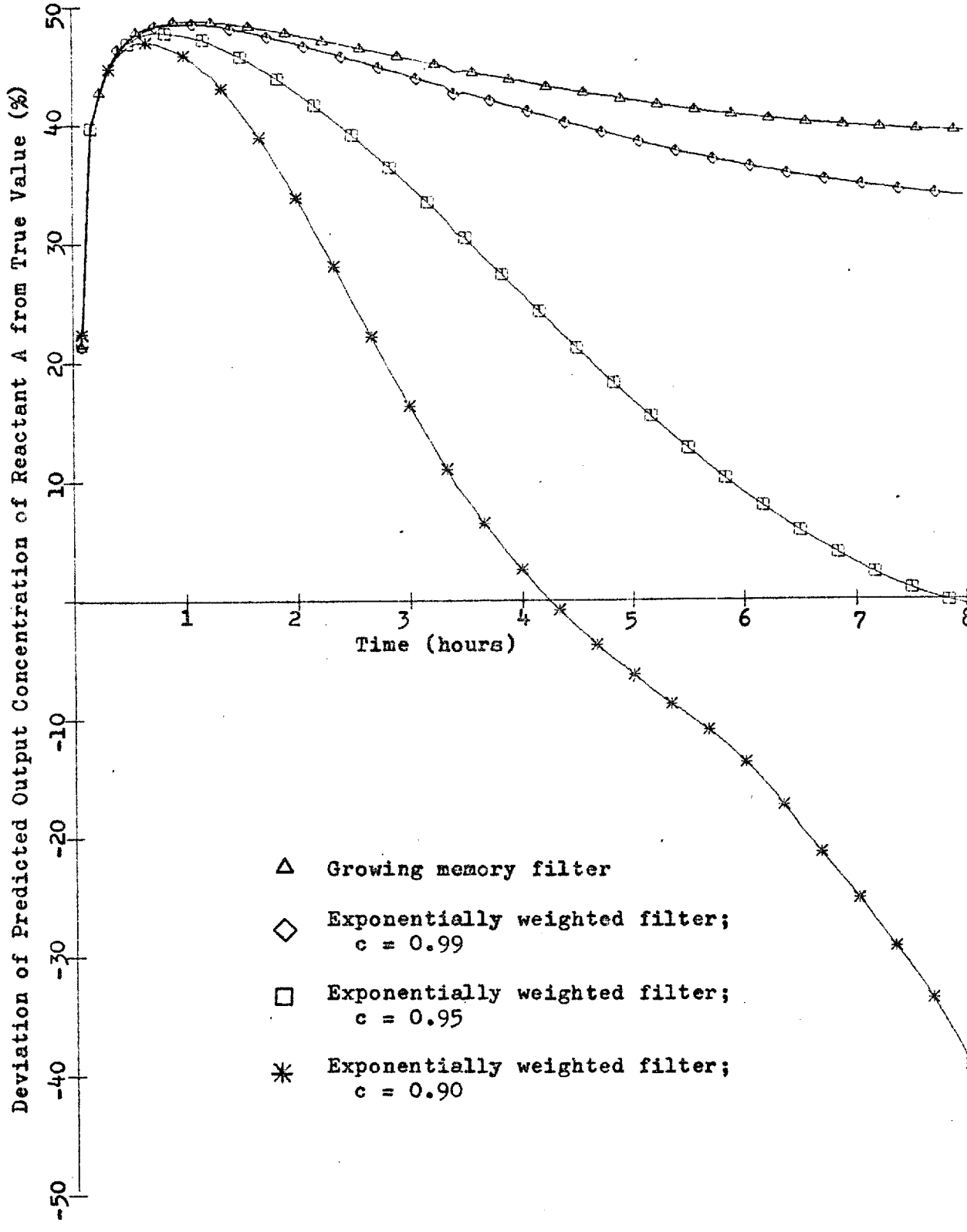
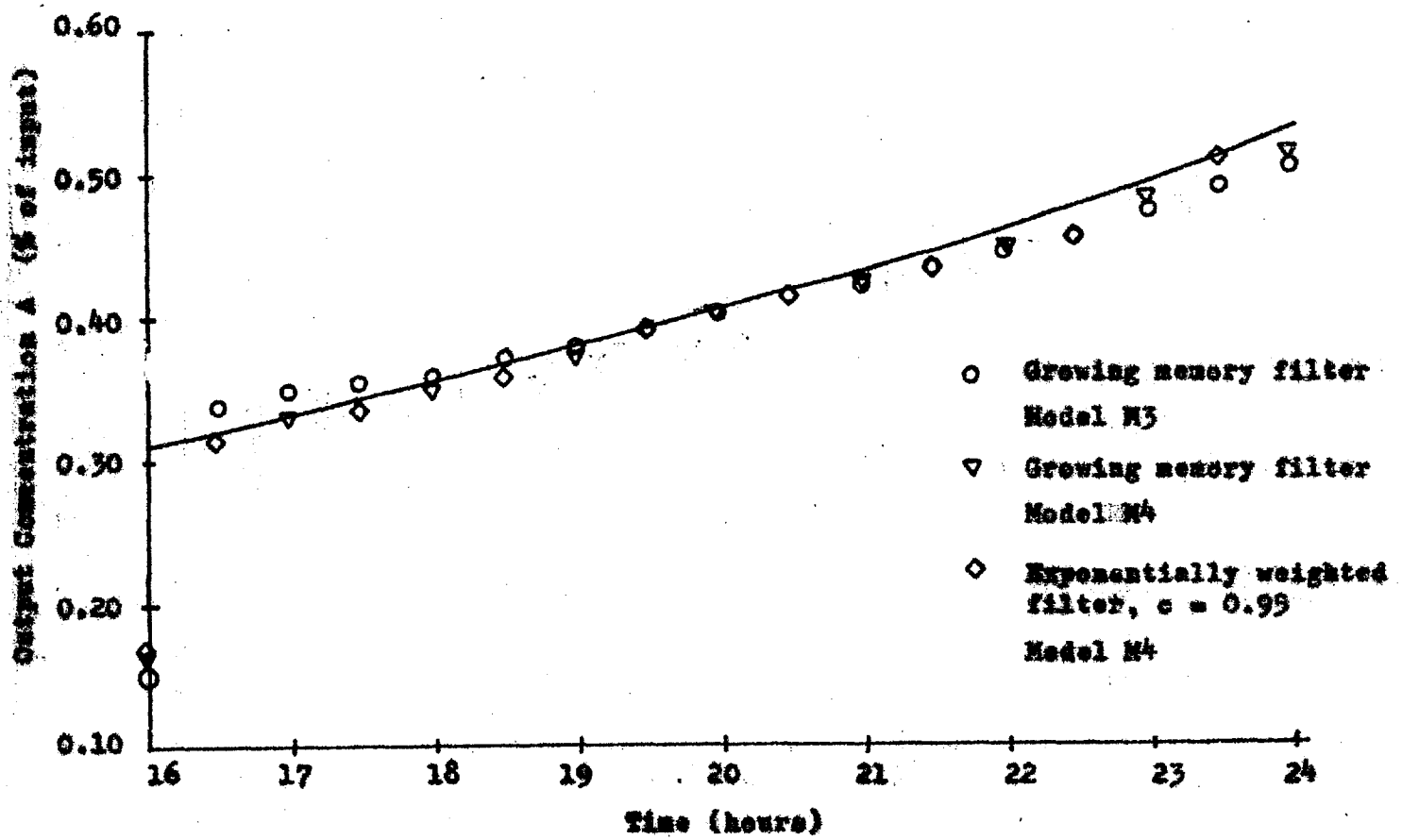


FIGURE 6.7

FILTERING RESULTS FOR REACTOR WITH LOCALLY DECAYING CATALYST



output at some later time. In this example the outlet conditions can be predicted if the concentration of reactant A in the outlet stream is accurately estimated. Thus in the sequel, the filter performance in a given situation will be evaluated by comparing the output concentration of reactant A predicted using the estimated activity with the true output concentration generated by the complete ICI model with recycle.

Figure 6.1 illustrates the adverse effects model deficiencies can have on the performance of the filter. The data generated with decaying catalyst in set S2 was analysed using the complete model in the steady-state predictor, but assuming the catalyst activity was actually constant (model M2). The three curves compare the output of the standard growing memory (Kalman) filter with that of the limited memory filters. The figure illustrates the danger of retaining data that is inaccurately described by the process model. Note that a comparison of the output of the growing memory filter with that of a limited memory filter provides a method for detecting the presence of modelling errors. When the model used in the filter is valid, the estimates resulting from the two different types of filters should be essentially the same.

When the form of the catalyst decay law is correctly postulated, a growing memory filter using either the complete or simplified model provides accurate predictions of future activity. Figures 6.2 and 6.3 show the activity predicted up to the eighth hour, based on the analysis of data set S2 using model M3. A certain amount of time is required for the filter to build up information, but the predictions are quite accurate after the second hour. Figures 6.4 and 6.5 show the composition predictions based upon the filter using the simplified model are very close to those based on the

complete model. The large deviations during the initial period are due to errors in the prediction of the catalyst decay law, and are not caused by the use of the simplified model.

In all of the preceding studies, the biases and parameters of the simplified model have been assumed to be constant and equal to the values determined by off-line minimization and listed in Tables 6.1 and 6.2 for $\lambda = 3000$. Table 6.6 shows the error in prediction of the output concentration A that results when the biases in the temperature predictions of the simple model are ignored.

Table 6.6 The Effect of Ignoring Biases in the Simple Model

Time (min)	% Deviation of Predicted A	
	Model M2 with Bias	Model M2 without Bias
5	11.59	26.83
10	8.24	28.32
15	6.59	28.89
20	5.97	29.70
25	4.54	29.07
30	3.53	28.60
40	0.14	25.30
50	0.62	26.59
60	0.70	27.20
75	0.81	25.70
90	0.23	27.33
105	0.32	27.60
120	0.43	27.82

$\lambda_o = 3000$ for both trials.

In practice, the model parameters and biases would vary slowly, and would have to be updated periodically as the local, simplified model becomes invalid. It would be preferable to update them automatically using the on-line filter. To test the feasibility of this technique, a model was constructed to estimate the biases and parameters in the simplified model; nineteen state variables were adjoined to system description M2, consisting of sixteen temperature biases as well as k_3, k_4 and the ratio k_2/k_1 (see equations (6.1) to (6.3)). Figure 6.6 shows the error in the prediction of the output composition of component A made by a standard Kalman filter as well as by some exponentially weighted filters. Though the latter type does, in some cases, reduce the estimation error, it can be seen that the filter becomes unstable, as predicted in Part I. The instability is indicated by the fact that the elements of the P matrix, which is a measure of the uncertainty in the estimate, increase as new observations are made. This implies that further observations actually decrease the amount of information available about the system. One explanation for this is that the filter is being used to estimate too many independent variables about which the steady-state model provides no coupled information. Furthermore, by allowing all of the parameters in the simplified model to vary, the system description becomes too general and in effect forces the estimator to operate on a multi-nodal least-squares surface. Even when only selected parameters were allowed to vary, the model was still too general, and the filter was unstable. In this application, therefore, the biases and parameters must be updated off-line.

The assumption that the catalyst activity depends upon inlet temperature only is intuitively unreasonable, and data set S3 was generated for the evaluation of the estimator when the decay

rate is a function of local temperature. Figure 6.7 compares the results of filtering the data assuming a constant activity along the bed at any time (M3), with those obtained assuming a linear profile (M4). An exponentially weighted filter was also used, since it would be responsive to the slow variation in the slope of the linear profile. It can be seen that for this example, the dependence of the decay on local temperature does not significantly alter the characteristics of the reactor. This, of course, may not be true in other applications, and a more complex profile could easily be accommodated by estimating a further one or two parameters to describe the catalyst profile as a quadratic or cubic function of distance along the reactor.

6.3 Discussion of Results

The feasibility of using a simplified process description for filtering purposes has been demonstrated in a realistic example. The simplified model used was based on engineering judgement rather than detailed analysis, and it is likely that computationally attractive models could be proposed which would allow the solution of a variety of practical estimation problems.

It has been found that the prediction biases due to model deficiencies must be accounted for in using the simplified models. Trials performed when these biases are ignored have been shown to result in errors of about 25% in the prediction of the output concentrations of reactants. In the example presented, it has been necessary to update the model biases off-line, though in theory the updating could be performed by the filter. Whether this is possible depends upon how many independent variables are being estimated, and would have to be investigated for any particular application.

The sensitivity of the filter performance to the initial estimate of the state vector would be expected from the theory. Kalman and Bucy[11] show that the convergence of the estimator to the true state is guaranteed only if the initial estimate, $\underline{x}_{o/o}$, is correct. Otherwise some bias would be expected. However, the amount of bias resulting from an incorrect initial estimate is problem dependent, and it is likely that experience in a given situation will indicate the more critical state variables. It has also been noted that when the initial state estimate is not good, the filter will take a large step after the first observation.

In these cases, the filter should be re-initialized at the new estimate.

This example also demonstrates the utility of limited memory filters for detecting the presence of model deficiencies. When analysis of a set of data by both limited and growing memory filters yields significantly different estimates, a model error is indicated. However, the danger of filter instability when the memory is limited has been demonstrated, and it must be concluded that these filters are most useful as diagnostic tools. Their behaviour is highly unpredictable, since it depends so strongly on the values chosen for the memory parameters.

The major significance of this example is that it demonstrates that realistic dynamic estimation problems can be solved using very simple dynamic models in conjunction with simplified steady-state models, even when the system, in theory, requires a distributed parameter description. The problem of dynamic modelling is greatly simplified, and most of the chemical engineering information available about the system can be utilized in the steady-state model.

CHAPTER 7

CONCLUSIONS

The theory of statistical filtering described in Part I has been developed in many different ways since the presentation of Kalman's classic paper[2]. The unified derivation presented here has the advantage of being inherently simple, and is based on the well known results of least-squares theory. The derivation is unique in that it specifies not only the standard growing memory filter of Kalman, but also the limited memory filters which have been shown to be useful as diagnostic tools for indicating model deficiencies. However, it should be noted that the exponentially weighted filter is not convergent; this fact has been overlooked by several authors.

The theory, however, has been presented merely as a method for solving practical chemical engineering problems. There have been few applications previous to this study, for three major reasons:

1. Stochastic models for chemical process system disturbances are rarely available;
2. Dynamic models for process elements are rarely known;
3. Even when models are known, they typically consist of sets of partial differential equations, too complex for on-line use.

This study has shown that these problems are not insurmountable. The analysis of the distillation column data has shown that the filter performance is not very sensitive to

either the form of the measurement error distribution or to errors in the initial uncertainty assumed for the elements of the state vector. The column example also illustrates the serious effect that model deficiencies can have on the performance of the filter.

The effects of modelling errors has been examined in detail in the reactor simulation. The filter has been shown to yield accurate predictions of output composition when based on a simplified steady-state model. Furthermore, the measurement scheme was very simple, and the filter did not require observation of the compositions at the outlet, but predicted them and the non-observable catalyst activity on the basis of temperature measurements only.

The reactor example also demonstrates that it is possible to deal with distributed parameter systems without greatly increasing the dimension of the state-vector of the model. Though the particular example is not highly sensitive to the change in decay dependence from inlet to local temperature, the technique used to approximate the distributed parameter system can be used in a variety of situations.

In both examples, the reliance on detailed steady-state models coupled with simplified dynamic models represents a basic change in the approach to estimation problems. It has been shown that complex dynamic models are not required for the analysis of the situations considered, but simple models could be used to describe the dynamic behaviour, while detailed chemical engineering information is utilized in the steady-state model relating the state vector to the observations. This technique could be used to analyze many chemical systems characterized by two major time constants with one

very small in relation to the other. Thus the linear theory developed is applicable to a large class of practical problems.

The studies have been performed under the assumption that there are no random dynamic inputs to the systems considered. It has been noted that this is a moot point. However, if it is felt that random dynamic inputs or random disturbances generated internally do present a serious problem, the theory can be modified to allow for them, except in the case of the oscillating memory filter. The real question is whether these disturbances, consisting in effect of normally distributed perturbations about a mean value, do indeed pose a serious problem in the applications envisaged. It is felt that most random (that is, unpredictable) process disturbances could not be described by a model which postulates additive normal random noise. However, this point cannot be resolved by simulations, but requires the application of the techniques to operating plants.

It must be stressed that the results presented are mere examples of possible applications of statistical filtering theory. Though they are representative of realistic chemical engineering problems, the successful application of the techniques cannot be guaranteed in any particular situation. The theory is complete as presented, and future work should be directed toward practical application of the techniques to operating plants.

REFERENCES

1. WIENER, N., Extrapolation, Interpolation, and Smoothing of Stationary Time Series, New York, John Wiley and Sons, Inc., (1949).

2. KALMAN, R.E., "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, J.Bas.Engr., Series D (March, 1960), 33-45.

3. GAUSS, K.F., in PLACKETT, R.L., "A Historical Note on the Method of Least-Squares", Biometrika, 36 (1949), 458-460.

4. DEUTSCH, R., Estimation Theory, Englewood Cliffs, Prentice Hall (1965).

5. KENDALL, M.G. and STUART, A., The Advanced Theory of Statistics, Vol. 2, London, Charles Griffen and Co., Ltd., (1967).

6. LIEBELT, P.B., An Introduction to Optimal Estimation, Reading, Mass., Addison-Wesley Publishing Co. (1967).

7. BALAKRISHNAN, A.V., "Report on Progress in Information Theory in the U.S.A., 1960-1963", IEEE Trans. on Information Theory, IT-9 (October, 1966), 237-240.

8. ZADEH, L.A., "Report on Progress in Information Theory in the U.S.A., 1957-1960", IEEE Trans. on Information Theory, IT-7 (July, 1961), 139-144.

9. DE RUSSO, P.M., ROY, R.J. and CLOSE, C.M., State Variables for Engineers, New York, John Wiley and Sons, Inc. (1965).

10. MEDITCH, J.S., "Orthogonal Projection and Discrete Optimal Linear Smoothing", Information Processing and Control Systems Laboratory, Tech. Rep. 66-102, Northwestern University (July, 1966).

11. KALMAN, R.E. and BUCY, R.S., "New Results in Linear Filtering and Prediction Theory", Trans. ASME, J.Bas.Engr., Series D (March, 1961) 95-108.

12. HO, Y.C., and LEE, R.C.K., "A Bayesian Approach to Problems in Stochastic Estimation and Control", Proceedings of the JACC, Stanford, California (June, 1964)

13. BELLMAN, R.E., KAGIWADA, H.H., KALABA, R.E., and SHRIDAR, R., "Invariant Imbedding and Nonlinear Filtering Theory", Rand Corp. Report RM-4374-PR (December, 1964).

REFERENCES (continued)

14. PLACKETT, R.L., "Some Theorems on Least-Squares", Biometrika, 37 (1950), 149-157.

15. SWERLING, P., "A Proposed Stagewise Differential Correction Procedure for Satellite Tracking and Prediction", Rand Corp. Report P-1292 (January, 1958).

16. ROSENBROCK, H.H., "On the Connection Between Discrete Linear Filters and Some Formulae of Gauss", Actes du Congres d'Automatique Theorique, Paris (1965), 239-249.

17. SORENSON, H.W., "Kalman Filtering Techniques", Chapter in Advances in Control Systems, 3, New York, Academic Press (1966).

18. MAYNE, D.Q., "Optimal Non-stationary Estimation of the Parameters of a Linear System with Gaussian Inputs", J. Electronics and Control, 14 (January, 1963), 101-112.

19. DENHAM, W.F. and PINES, S., "Sequential Estimation when Measurement Function Nonlinearity is Comparable to Measurement Error", AIAA Journal, 4, No. 6 (June, 1966), 1071-1076.

20. SMITH, G.L., SCHMIDT, S.F., and McGEE, L.A., "Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle", NASA Tech.Rep. R-135 (November, 1961).

21. SORENSON, H.W. and STUBBERUD, A.R., "Recursive Filtering for Systems with Small but Non-negligible Nonlinearities", Int.J. Control, 7, (1968), 271-280.

22. PEARSON, J.B., and SHRIDAR, R., "Digital Estimation of Nonlinear Processes", Proc. IEEE Region 6 Annual Conf., Tuscon, Ariz. (April, 1966).

23. PEARSON, J.B., "On Nonlinear Least-squares Filtering", Automatica, 4 (1967), 97-105.

24. FISHER, J.R., "Optimal Nonlinear Filtering", Chapter in Advances in Control Systems, 5, New York, Academic Press, 1967.

25. SEINFELD, J.H., "Nonlinear Estimation for Partial Differential Equations", Chem. Eng. Science, 24 (1969), 75-83.

REFERENCES (continued)

26. SCHLEE, F.H., STANDISH, C.J., and TODA, N.F., "Divergence in the Kalman Filter", AIAA Journal, 5, No. 6 (June, 1967), 1114-1120.

27. SMITH, G.L., "Secondary Errors and Off-design Conditions in Optimal Estimates of Space Vehicle Trajectories", NASA TN-2129 (January, 1964)

28. BLUM, M., "Fixed Memory Least-squares Using Recursion Methods", I.R.E. Trans, IT-3 (September, 1957), 178-182.

29. JAZWINSKI, A.H., "Limited Memory Optimal Filtering", Proc. of the JACC, University of Michigan (June, 1968).

30. FAGIN, S.L., "Recursive Linear Regression Theory, Optimal Filter Theory, and Error Analyses of Optimal Systems", 1964 IEEE International Convention Record, Part I, New York, 1964.

31. KALMAN, R.E. and BERTRAM, J.E., "Control System Analysis and Design Via the "Second Method" of Lyapunov. I. Continuous Time Systems", Trans. A.S.M.E., J.Bas.Engr., Series D, 82 (1960), 371-393.

32. McCLEAN, J.D., SCHMIDT, S.F., and McGEE, L.A., "Optimal Filtering and Linear Prediction Applied to a Space Navigation System for the Circumlunar Mission", NASA Technical Note D-1208 (March, 1962).

33. EYKHOFF, P., "Process Parameter Estimation", in: Macmillan, et al. (ed.), Progress in Control Engineering, Vol. 2, London, Heywood, 1964.

34. EYKHOFF, P., VAN DER GRINTEN, E.M., KWAKERNAAK, H., and VELTMAN, B.P.Th., "Survey: Systems Modelling and Identification", Third IFAC Congress (London), June, 1966.

35. EYKHOFF, P., "Survey: Process Parameter and State Estimation", Automatica, 4, No. 4 (May, 1968), 205-233.

36. CUNEOD, M. and SAGE, A.P., "Comparison of Some Methods Used for Process Identification, Automatica, 4, No. 4 (May, 1968), 235-269.

37. BALAKRISHNAN, A.V. and PETERKA, V., "Survey: Identification in Automatic Control Systems", IFAC Fourth Congress (Warsaw), 1969.

REFERENCES (continued)

38. GAVALAS, G.R. and SEINFELD, J.H., "Sequential Estimation of States and Kinetic Parameters in Tubular Reactors with Catalyst Decay", Chem.Eng.Science, 24, No. 4 (April, 1969), 625-636.

39. COGGAN, G.C. and NOTON, A.R.M., "Optimal Discrete-Time Recursive State and Parameter Estimation in Chemical Engineering", Symposium: Current Trends with Computers in Chemical Engineering, I.Chem.E, Midlands Branch, Nottingham (21 April, 1969).

40. LEE, E.S., "Quasilinearization and the Estimation of Parameters in Differential Equations", I. and E.C. Fundamentals, 7, No. 1 (February, 1968), 152-158.

41. BRAY, J.W., HIGH, R.J., McCANN, A.D. and JEMMESON, H., "On-Line Model Making for a Chemical Plant", Trans. Soc. Inst. Tech., 17, No. 3, (September, 1965), 65-75.

42. ASTROM, K.J., and BOHLIN, T., "Numerical Identification of Linear Dynamic Systems from Normal Operating Records", IFAC Symposium: The Theory of Self-Adaptive Control Systems, Teddington (September, 1965)

43. ASTROM, K.J., "Control Problems in Paper Making", Proc. IBM Scientific Computing Symposium on Control Theory and Applications, White Plains, New York (1966), 135-165.

44. DAHLIN, E.B., "On-line Identification of Process Dynamics", IBM J. of Research and Development, 11, No. 4 (July, 1967), 406-426.

45. BOX, G.E.P. and JENKINS, G.M., "Models for Prediction and Control", Tech. Reports No. 1-10, University of Lancaster, Department of Systems Engineering.

46. PRICE, R.J., and RIPPIN, D.W.T., "Single Variable On-line Adaptive Optimization of a Water Gas Shift Reactor", Chem. Eng. Science, 23, No. 6 (August, 1968), 593-618.

47. WISE, H.E., Ph.D. Thesis, University of London, 1968.

48. NOTON, A.R.M. and CHOQUETTE, P., "The Application of Modern Control Theory to the Operation of a Industrial Plant", Preprint from IFAC Symposium: Digital Control of Large Industrial Systems, Toronto (1968).

REFERENCES (continued)

49. McCABE, W.L., and J.C. SMITH, Unit Operations of Chemical Engineering, New York, McGraw-Hill (1956), 709.

50. CHU, J.C., GETTY, R.L., BRENNECKE, L.F. and PAUL, R., Distillation Equilibrium Data, New York, Reinhold (1950), 25.

51. GARDNER, R.S., "Rardom Number Subprogram", IBM Share Program 3170 (10 July, 1964).

52. WYLIE, C.R., Jr., Advanced Engineering Mathematics, 2nd. ed., New York, McGraw-Hill (1960), 186.

53. POWELL, M.J.D.,"A Method for Minimizing a Sum of Squares of Nonlinear Functions without Calculating Derivatives", The Computer Journal, 8 (1965), 303-307.