THE DETECTION OF MALFUNCTION USING A PROCESS COMPUTER

by

BARRY BELLINGHAM

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of the Loughborough University of Technology.

December, 1975.

Supervisor : F. P. Lees, B.Sc.(Eng.), Ph.D.

Department of Chemical Engineering.

ACKNOWLEDGEMENTS

The author would like to offer sincere thanks to Professor F. P. Lees for his exceptional patience, encouragement and advice during this study.

Thanks are also due to Professor D. C. Freshwater, Head of the Department of Chemical Engineering, the technical staff of the department and to the Science Research Council for its financial support.

Finally, the author is grateful to his wife, Sol, for displaying great understanding and coping admirably throughout this work.

ABSTRACT

A process control computer has been used to detect malfunction in the instrumentation of control loops. A flow control loop is examined and a malfunction detection algorithm is developed which is based upon a comparison of the control valve position and flow. The technique assumes a relatively constant flow-pressure drop characteristic. It is postulated that a flow control loop has inherent measurement redundancy and a simple "static" or "tracking" state estimator is used to obtain an estimate of the flow from the valve position (or control demand) and flowmeter measurements. The check is based upon monitoring changes in the residuals generated by the estimator. The check technique does not require additional process instrumentation, uses little computer time or storage and can be performed while the control loop is operating under direct digital control. The method has been tested by extensive laboratory trials and some limited industrial application.

This malfunction detection method based upon state estimation is generalised to encompass all control loops using a Kalman filter state estimator. The control loops are modelled by linear time invariant transfer functions and it is assumed that the load disturbance is relatively constant or measured. The Kalman filter is designed to yield optimal state estimates by using Mehra's innovation correlation method to account for uncertainty in the system model and statistical properties. The malfunction detection methods are based upon examining changes in the estimator innovation sequence and/or directly estimating loop security parameters associated with control loop malfunction. The loop security parameter estimator is decoupled from the primary Kalman filter using Friedland's method and its implementa-

-tion becomes trivial when it is combined with the results generated from Mehra's adaptive estimator.

The checks can be performed on line to monitor conventional analogue and direct digital control loops.

The proposed algorithms have been tested on an experimental laboratory level control apparatus. The results show that the methods detect malfunction and provide some diagnostic information. Experiments also show that Mehra's adaptive estimator fails if the process measurement noise covariance matrix is small.

The merits of malfunction detection and equipment condition monitoring are considered in terms of reliability theory. Reliability is considered in terms of the state of knowledge of the system and is treated from the view point of the plant designer and operator. The reliability is categorised into four regimes depending upon whether or not monitoring is performed and upon the information received from the monitor. Expressions are derived for these four reliability functions for a single equipment. The information needed to calculate these estimates is the conventional reliability function together with the probability density functions for the time to failure and for the monitor signal from the time of initial malfunction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1.

INTRODUCTION

The progress of the chemical process industry over the last decade or so has generally been towards larger, more efficient automated plants employing single streams instead of several parallel ones. This trend has resulted from the commercial pressure for lower unit costs and for better products.

Associated with this phenomenon has been the growth of process computer control systems (1).

As plants have become larger and more complex, some processes have been required to operate at higher temperatures and pressures with much higher concentrations of reactive chemicals than previously. This obviously results in the possibility of greater plant damage under fault conditions. This damage not only causes economic loss, but also human life is endangered. As well as these catastrophic plant failures, it has become evident in recent years that many of the expected economies of large plants can disappear if a plant cannot be operated continuously with uninterrupted production (2).

The realisation of these problems has resulted in an increasing interest in reliability engineering (3). Reliability engineering is essential for safety, and is concerned with predicting, estimating or optimising the probability of survival, mean life, or more generally life probability distributions of components or systems. Other problems considered are those involving the probability of the proper functioning of the system at either a specified or an arbitrary time, or the proportion of the time that the system is functioning properly. Often maintenance such as repair, replacement, or inspection, may be performed so that the solution of the reliability problem may influence decisions concerning maintenance policies to be followed.

The interest in reliability engineering has required probabilistic information on equipment failure rates and modes and has led to the establishment of a central data bank at the U.K. A.E.A. More recently several papers have appeared in the open literature summarising equipment malfunction surveys (4), (5), (6), (7), (8).

The development of process reliability has relied upon the diagnosis of equipment faults. However, there is a growing recognition that future progress cannot rely upon diagnosis alone (1), (9), (10). It is becoming increasingly important to monitor the state of process equipments on either an intermittent or continuous basis.

At the present time a high plant operating efficiency is achieved by introducing maintenance schedules which impose shut-downs at intervals of time. However, despite this preventive maintenance, malfunction can occur in both process equipment and instrumentation. The consequences of such faults depend upon the importance of the malfunctioning equipment to the overall process security and upon the degree of warning. The most basic form of maintenance may be termed "emergency maintenance", where the equipment is allowed to operate until it fails before it is repaired or replaced. This results in maintenance work being done on an emergency basis, which is inefficient in the utilisation of manpower and may lead to excessive process down-time. Thus, it is desirable to be able to detect process anomalies at an early stage, thereby preventing catastrophic failures, improving the operational efficiency of equipment and facilities, reducing maintenance cost and allowing the process to run closer to its intended conditions. The introduction of suitable malfunction detection policies would help to rationalise the approach to plant maintenance

and at the same time provide valuable information on the development of failures in different equipment and instrumentation.

Most of the malfunction detection on chemical plant is currently performed by the process operator. The problems raised and some methods of detection have been reviewed by Edwards and Lees (1), Anyakora (9), Dowson (11) and Trotter (12). However, as process computer control systems have developed it has become apparent that the computer has the capacity to contribute to the improvement of the overall process reliability (13), (14), (15), (16), (17). There are several areas of process reliability to which the process computer may be applied; however, in this thesis the application to malfunction detection is considered.

The success and role of the process operator in detecting process malfunctions has been examined by Edwards and Lees (1). However, even recognising the competence and dedication of process operators and maintenance personnel, human performance is dependent on outside and often indeterminable pressures. Frequently conducted procedures become routine and human frailty of increased indifference to routine procedures is well known. Routine operations and repetitive functions requiring little imagination and offering little satisfaction are best left to a machine, thereby releasing the operator for more rewarding tasks. However, the decision of manual, automatic, or man-machine malfunction detection is not particularly well defined (1). The choice of system should be based upon the comparative strengths and weaknesses of each.

If a process computer is used for automatic malfunction detection, then the experience, judgement, insight and instinct of the human operator are lost. However, balanced against this are the

characteristics of the process computer. These include the ability of the machine to perform more frequent tests without a higher probability of error. Other important features are: the collection of large amounts of process data, the determination of accurate process time transients, the facility of conditioning process data to a more amenable form, the long term memory for immediate comparison and the ability of operational self interrogation (18).

The application of a process computer to malfunction detection, either as an autonomous device, or as an aid to the process operator, is relatively undeveloped (1). The computer's contribution is mainly restricted to monitoring process alarms, to checks made during sequential operations and to some limited instrument tests.

Damon (17), Hoyte (19), Fraade (20) and Thompson (21) have described methods where the computer performs, with the aid of isolation valves, the type of tests normally performed manually by a process operator or maintenance engineer.

More extensive checks on process equipment and instrumentation, which utilise the computation capacity of the computer, have been suggested by Edwards and Lees (1), Lees (4) and Damon (17), and practical applications of these ideas are beginning to appear (22), (23), (24).

A method of detecting instrumentation malfunction by analysing the statistical properties of an instrument signal time series was proposed by Anyakora (9). A similar technique has been developed to monitor the state of an on-line nuclear reactor (25). In this scheme many of the computer characteristics are exploited. For example, the computer calculates a power spectral density of the instrument signal

and displays the resulting characteristic, as well as a standard characteristic recalled from memory, on a colour visual display unit (V.D.U.).

An important part of any modern chemical plant is the control system, which can of course malfunction. The ordinary process control loop is liable to malfunction and faults can develop in any of its constituent parts such as the measurement, controlling or regulating elements.

The seriousness of a control loop failure or malfunction depends upon the application, but types of failure which may have particularly serious consequences include misleading measurement, incorrect control action or valve seizure resulting in wrong control action. The malfunction of instrumentation appears to contribute significantly to serious process incidents. For example, Whitman (14) has reported that failures of instrumentation in ammonia chemical plants account for 10% of major incidents. Some consequences of instrument failure in chemical plant have been given by Lees (4).

However, even if the individual instrument malfunctions do not seriously affect the process, their sum total can result in a degradation of the control system performance as well as a decrease in process efficiency.

The failure rates of instruments in chemical plant are quite high (4), (8) and since modern plants contain large numbers of them the probability of failed instruments at any particular time is high. Skala (8) has suggested that there may be as many as 1% of the process instruments in a failed state at any time.

The detection of instrument and/or control loop malfunction is usually performed by the process operator using an instinctive approach based upon his mental model of the expected process response and measurement time histories (26). This thesis examines the feasibility of using a process computer to detect malfunction in both conventional analogue setpoint and direct digital control loops.

Chapter 2 develops a particular method for the detection of malfunction in a flow control loop based upon a "simple" state estimator. This check exploits many of the computer's capabilities, such as the memory, the display and the calculation capacity. The technique is illustrated by extensive laboratory experiments and some industrial trials.

The application of state estimation to control loop malfunction is generalised in Chapter 3 where a method is presented which is based upon Kalman filtering.

In addition to deriving computer aided methods of malfunction detection, the impact of the knowledge that an instrument or equipment is malfunctioning on the process management's assessment of reliability is considered. Chapter 4 examines this problem and derives a mathematical model whereby a process operator can predict equipment reliability using his current state of knowledge of the equipment condition indicated by a malfunction detection monitor.

Many of the ideas presented in this thesis involve the concepts of linear algebra and probability theory, and the notation and mathematical quantities used are given in Appendix I (27), (28).

CHAPTER 2.

THE DETECTION OF MALFUNCTION IN A FLOW CONTROL LOOP

2.1 List of symbols

A_v	cross-sectional area of control valve.	m^2
a_{ij}	coefficient in constraint equation.	various
C_v	control valve sizing factor	$\frac{\text{U.S.gall}}{\text{min}(\text{p.s.i.})^{\frac{1}{2}}}$
D	pipe diameter.	m
E( )	expectation.	-
e_Q	error in flow	m^3/s
e_s	error in flowmeter signal.	various
e_x	error in control valve stem position.	various
F	Jacobian matrix (mxn); F distribution	-
f_{ij}	i,j^{th} element of F	-
$\underline{f}(\underline{x})$	non-linear vector function of $\underline{x}$	-
G(s)	transfer function	-
g_i	weighting coefficient	-
H	measurement matrix (mxn)	-
H_o	null hypothesis	-
H_{ji}	enthalpy coefficient	$\text{chu}/\text{lb.mole}$
$\underline{h}\ (\underline{x})$	non-linear measurement vector function of $\underline{x}$.	-
$\underline{h}^{(i)}(\underline{x})$	non-linear measurement vector function with i^{th} equation eliminated.	-
$\underline{h}_i$	i^{th} row measurement vector.	-
J	cost function.	-
k	process gain/discrete time counter.	various
k_1, k_2, k_3	constants in flowmeter and control valve equations.	various

k_{m1}, k_{m2}	constants in flowmeter equations.	various
k_{v1}, k_{v2}) k_{v3}, k_{v4})	constants in control valve equations.	various
l	number of constraint equations.	-
$\underline{m}$	vector of process measurements.	-
m	number of measurements.	-
m_i	i^{th} element of vector $\underline{m}$	various
N	normalised test function.	-
n	number of samples; number of system states; iteration number.	-
ΔP	pressure drop.	kN/m^2 (p.s.i.g.)
ΔP_m	pressure drop across flowmeter orifice plate	kN/m^2 (p.s.i.g.)
ΔP_v	pressure drop across control valve	kN/m^2 (p.s.i.g.)
Q	flow	m^3/s
Q_m	flow through flowmeter	m^3/s
Q_{max}	maximum flow	m^3/s
Q_v	flow through control valve	m^3/s
q	number of process disturbances	-
R	measurement noise covariance matrix; weighting matrix.	-
r_{ij}	ij^{th} element of R	various
$\underline{r}$	vector of residuals	-
r	number of control variables	-
r_i	residual of measurement i	various

$\bar{r}_{i,j}$	sample mean of i^{th} residual from j^{th} ensemble.	various
S	weighting matrix	-
s	flowmeter signal; Laplace operator	various
$\bar{s}^2$	pooled sample variance	various
s_j^2	sample variance of j^{th} ensemble	various
$s_{i,j}^2$	sample variance of i^{th} residual from j^{th} ensemble	various
t	student's t distribution; time	-; s
Δt	sampling interval	s
$\underline{u}$	control vector	-
$\underline{v}$	measurement noise vector	-
$\underline{w}$	vector of process disturbances	-
x	valve position; controller signal	cm; V
x_{max}	maximum valve position; maximum controller signal	cm; V
$\underline{x}$	state vector	-
$\Delta \underline{x}$	linearised state vector	-
$\underline{x}_o$	initial iterative value of $\underline{x}$	-
$\underline{x}_n$	state vector at iteration n.	-
$\underline{x}$ (k)	state vector at time k	-
x_i	i^{th} element of $\underline{x}$	m^3/s
$\underline{y}$	vector of measurements	-
$\Delta \underline{y}$	linearised measurement vector	-
y_i	i^{th} element of $\underline{y}$	various
z	statistical test function	-

Greek letters

alpha	α	confidence limit	-
beta	β	probability	-
zeta	ζ	damping factor	-
lambda	$\underline{\lambda}$	vector of Lagrange multipliers	-
mu	μ	population mean	various
rho	ρ_{ij}	correlation coefficient between measurement i and j	-
sigma	σ	population standard deviation; measurement noise standard deviation.	various: various
tau	τ_d	time delay	s
phi	Φ	sum of squares cost function	-
psi	Ψ	vector of constraint equations	-
omega	ω_n	natural frequency	Hz

Subscripts

i	variable i
m	flowmeter
n	iteration n
o	initial value; malfunction free value
s	flowmeter signal
v	control valve
x	valve position
z	statistical test function

Superscripts

i	i^{th} row vector
T	transpose
-1	inverse
-	mean
^	estimate

2.2 Introduction

Modern processes rely extensively upon the correct distribution and control of flowing liquids or gases.

The usual process flow control loop is illustrated schematically in Figure 2.1. The object of this system is to regulate a fluid flow according to some desired requirement. The operation of the control loop depends upon the controller receiving from the measuring instrument a measurement of the controlled variable. The controller compares this with a desired value or setpoint to obtain an error, performs a mathematical operation on the error and sends an output signal to the control valve, which then adjusts the manipulated variable.

The control valve operates as a variable orifice and the rate of flow through the valve depends upon the upstream and downstream fluid pressures and the opening of the valve. The heart of the control valve is the valve trim; its main parts are the seat and the valve plug. The plug usually consists of a seating surface and a characterised portion.

The characterised portion of the valve plug is of special importance, since it is used to vary the flow area between the plug and the valve seat at a controlled rate. This rate of flow change with valve lift or controller signal is called the inherent flow characteristic of the valve.

Control valve manufacturers produce a diverse range of inherent valve flow characteristics. However, the commonest types are the linear and equal percentage characteristics.

A linear inherent characteristic produces a change in flow, under

constant pressure drop, that is linear with valve position or controller signal. For example, at 50% signal, 50% of the flow capacity will pass through the valve.

The equal percentage valve characteristic produces a change of flow, for a given increment of valve position, that is a percentage of the quantity of flow just before the change was made.

The flow control loop can be based upon a conventional analogue controller or a process computer using direct digital control (d.d.c.). The signals around the loop can be pneumatic, electrical or a mixture of the two.

Typical control loops are shown in Figure 2.2. Figure 2.2.a is a conventional analogue loop in which the flowmeter is an orifice plate and differential pressure transmitter, and in which all the instruments are pneumatic. Figure 2.2.b shows a loop under d.d.c. in which the flowmeter is similar to Figure 2.2.a, but the pneumatic output from the differential pressure transmitter is converted into an electrical signal by a pressure/current (P/I) transducer before entering the process computer. The output from the computer is converted from an electrical to a pneumatic signal in a current/pressure (I/P) transducer and then passed to the control valve. Figure 2.2.c is similar to Figure 2.2.b except the flowmeter is an instrument giving an electrical output, such as a magnetic flowmeter.

The reliability of a flow control loop depends upon the state of its components. The precise definition of control loop failure or reliability is not straightforward, but for practical purposes some workers (4) have suggested that failure is defined when an instrument is not operating to the satisfaction of the process operator.

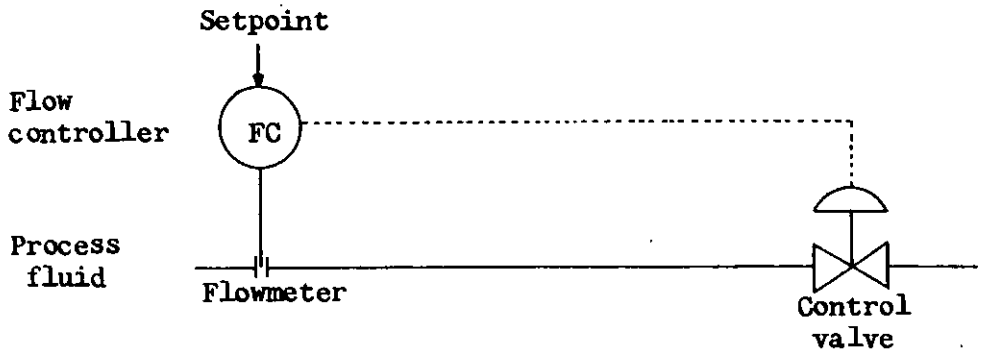


FIGURE 2.1 Flow control by analogue controller.

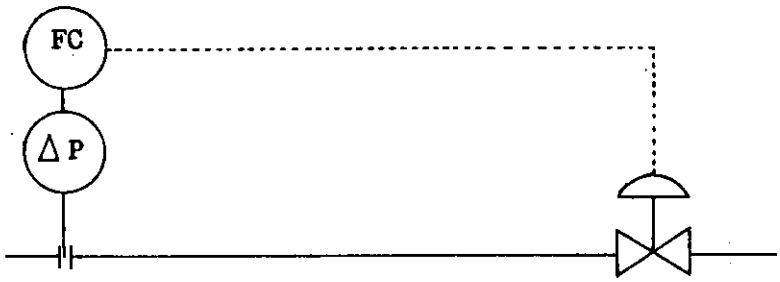


FIGURE 2.2a Analogue flow control loop with orifice/differential pressure transmitter flowmeter.

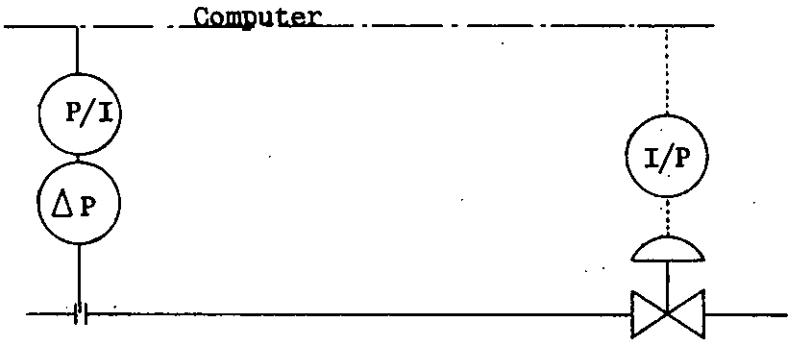


FIGURE 2.2b Direct digital flow control loop with orifice/differential pressure transmitter flowmeter.

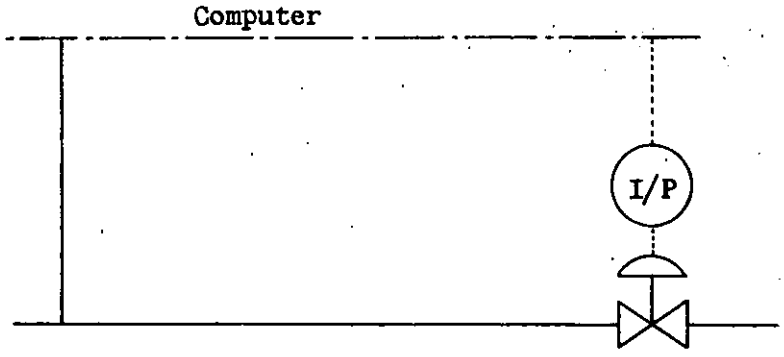


FIGURE 2.2c Direct digital flow control loop with magnetic flowmeter.

The failures detected in this way vary considerably in type and degree. Most failures are determined by inadequate performance of some kind but the acceptability of a given performance depends upon the application. For example, the reliability of a flow control loop may be considered from two points of view:

i) The reproducibility of the system.

ii) The absolute accuracy of the system.

Both of these aspects are interrelated. For normal plant control purposes employing feedback control loops, the primary interest is i) since usually a setpoint is established at which the given system is controlled and the absolute level of this setpoint is only of secondary importance. However, for process performance evaluation such as material and energy balances, optimisation, model building, etc., when a flow has to be compared directly with another, then all the measurements must refer to the same datum level. Absolute accuracy is then required as well as a good standard of reproducibility.

The important point is that what constitutes a failure for one application may be acceptable in another.

Lees (4) and Skala (8) have presented data on process instrumentation reliability. In particular, Table 2.1 details the overall failure rates of the process instrumentation associated with flow control loops.

The overall instrument failure rate gives only limited information and often knowledge of the instrument failure modes are required. Lees has classified failure modes as:

Condition,

Performance,

Safety,

Detection.

Some data on failure modes for various instruments have been analysed by Lees (29), while the results of three independent surveys of data describing malfunctions of valves used in nuclear power plants are given in reference (30). Stiles (31) has investigated the effect of cavitation in control valves and presented several photographs of damaged valve trims caused by operating the system in this way. Environmental effects of temperature and pressure on the failure characteristics of pressure transducers have been examined by Davidson (32).

The problem of detecting malfunction in flow control loops has been examined by several researchers.

A major area of work has been concerned with the problem of data

Instrument	Observed failure rate-faults/year
Control Valve (p)	0.25 - 0.60
Controller (p)	0.29 - 0.38
Flow measurement (fluids)	1.14
Differential pressure transducer (flow) (neglecting impulse lines)	1.73
Magnetic flowmeter	2.18
Current/pressure transducer	0.49
Impulse lines	0.77

TABLE 2.1. Flow control loop instrumentation failure rate data (4).

p denotes pneumatic.

consistency reflecting the definition of flow control loop failure as having occurred when the material and energy balance acceptance criterion is not satisfied.

A set of material and energy balance data from a chemical plant does not usually satisfy the steady state material and energy balance equations. This inconsistency may be accounted for in several ways such as random errors on the measurements, unsteady state process operation, gross measurement errors, and process disturbances.

Kuehn and Davidson (33) and Clementson (34) have proposed methods of obtaining consistent data sets from process measurements containing only small random errors obtained at steady state plant operation.

They introduced the least squares criterion to predict a consistent set of data $\underline{y}$ from a set of measured data $\underline{m}$:

$$\Phi = \sum_{i=1}^{m} \left(\frac{y_i - m_i}{\sigma_i} \right)^2 \qquad (2.1.1)$$

where σ_i^2 is the given error variance on the i^{th} measurement, m_i is the i^{th} observed measurement, y_i is the i^{th} measurement when corrected and m is the number of measurements.

In the analysis the elements of $\underline{y}$ are assumed to be linearly related to one another by equations such as material and heat balances, i.e.

$$\psi_j = \sum_{i=1}^{m} a_{ji} \, y_i = 0 \qquad j = 1,2,\ldots,l \qquad (2.1.2)$$

where a_{ji} is the coefficient of the i^{th} measurement in the j^{th} balance equation, and l is the total number of constraints.

The problem in mathematical form is to minimise Φ, subject to the constraints $\underline{\psi}$.

Notice that with this formulation the m_i's are not the primary measurements such as the ΔP across an orifice plate, but are mass flows computed from the primary measurements.

Ripps (35) considered the problem of the process measurements containing small errors as well as very large errors introduced by complete malfunctioning of an instrument, or very strong instrument bias. He expanded Kuehn and Davidson's procedure by discarding measurements, which were suspected of containing a gross error, in the basic least squares criterion of equation (2.1.1). The method takes advantage of the redundancy present in the material and energy balance equations to estimate the discarded data. By examining the resultant minimum least squares function Φ when measurements are discarded, Ripps demonstrates that this indicates which measurement has the gross error. He states that the minimum Φ often coincides with the most correct data adjustment, but this need not be the case.

Although Ripps' technique is attractive it requires some knowledge about the process in selecting suspect measurements. Also there is no criterion proposed to determine the number of measurements containing gross errors. This makes the result uncertain and alternative results occasionally may be assumed.

Nogita (36) and Nogita and Uchiyama (37) have solved some of these problems by assuming there might be a few systematic errors in the experimental data, and that each of the other measurements is a random sample from a Normal population with unknown correct value (mean) μ_i, known variance σ_i^2 and known correlation coefficient ρ_{ij}.

The least squares formulation of the previous workers was used to determine a set of consistent data $\underline{y}$. Nogita then defined a test

function for the measured data $\underline{m}$ as

$$z = \sum_{i=1}^{m} \left(\frac{y_i - m_i}{\sigma_i} \right)$$

The scalar variable z is shown to be a Normal random variable with a mean μ_z and variance σ_z^2 of:

$$\mu_z = 0$$

$$\sigma_z^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \rho_{ij} \; g_i \, g_j \; \sigma_i \, \sigma_j$$

where the elements g_i are derived as a function of the problem formulation matrices.

z is normalised according to:

$$N = \frac{z}{\sigma_z}$$

so that N has a Normal distribution with zero mean and unit variance. Now by referring to a cumulative Normal distribution table the probability, β, of N being in an interval $\pm \alpha$ can be determined. The test for data consistency is performed by calculating and testing if $|N| > \alpha$. If this criterion is satisfied, then it is possible to say that there exists a gross measurement error, or an unsteady state in the system with β probability of being incorrect in making such a statement.

Nogita uses the serial elimination algorithm of Ripps' technique and the test criterion defined above to determine the suspect measurements.

These techniques of malfunction detection are appealing but suffer from two major disadvantages. Primarily the method assumes the measurements under investigation may be related by linear equations as

shown in equation (2.1.2). Secondly, the formulation assumes the coefficients of the constraint equations, a_{ji}, are deterministic. While this is true in the case of material balances, for energy balances this will not be valid. For example, a process heat balance yields constraint equations of the form:

$$\sum_{i=1}^{m} H_{ji}\, y_i = 0 \qquad j = 1,2 \ldots l$$

where H_{ji} ($= a_{ji}$) is the enthalpy associated with the flowrate y_i.

Now the enthalpy H_{ji} is calculated from temperature and pressure which are themselves process measurements subject to both small random errors and gross errors. However, to use the techniques described it is necessary to assume H_{ji} is known perfectly, which is clearly untrue.

It is therefore concluded that the implementation of these data adjustment techniques requires a deterministic knowledge of the constraint equation coefficients a_{ij}. If there is an element of uncertainty in these coefficients it is suggested that a sensitivity analysis should be performed to determine the feasibility of using these techniques as a malfunction detection algorithm in the particular application considered.

To demonstrate the potential shortcomings of these techniques the method was investigated on an industrial distillation column situated at Works A. Works A was a large tonnage process producing an intermediate organic chemical. A process control computer was used to monitor process performance and perform d.d.c. on some control loops. The problem formulation and sensitivity analysis is given in Appendix II.

The above techniques are concerned with detecting malfunction in particular flow control loops by examining the consistency of groups of measurements. Fraade (20), Flum and Fraade (39) and Garden (40) considered the problem of detecting and correcting malfunctions in individual flow control loops. Fraade has suggested that a process computer may be used to perform an on-line calibration of the flow control loop. He represents the control loop by a characteristic which relates the process flowrate to the control voltage from a process computer. This characteristic changes as the instrumentation in the control loop malfunctions, and so Fraade determines the characteristic on-line by using a calibration tank. At the time of calibration the process computer isolates the main process flow, opens the outlet of the calibration tank and measures the true flow in the line by determining the time for the tank height to change by a specified amount. Now knowing this true flow, which is independent of the system, and the computer control voltage, a new loop characteristic is obtained. This may then be compared with the old characteristic and inferences on loop security drawn. This method was used by Barton et al. (41) to check the feed flow to a cement kiln. However, the applicability of this idea is limited by the need for additional process equipment.

Garden developed similar ideas for updating a flow control loop characteristic, but his method does not use additional process equipment. Instead, Garden uses a learning technique based upon a feedback signal of the control valve position. The technique is applicable to control valve actuators that are operated by discontinuous signals such as an electric drive unit or a solenoid operated pneumatic actuator. Garden suggests that the method provides continuous on-

line adaptation to the changing characteristics of the valve actuators which results in better control than conventional analogue control or its digital equivalent. He also states that his technique may detect deteriorating parts of the loop, although he does not present experimental or theoretical confirmation of this.

A method of detecting malfunction from the "noisiness" of an instrument signal has been developed by Anyakora and Lees (9), (42). They applied their technique to incipient failures of several instruments such as thermocouples and differential pressure transmitters as well as to the incipient stickiness in control valves.

In this thesis a method is developed of detecting malfunction in a complete control loop. It is postulated that a flow control loop has inherent measurement redundancy and the process flowrate can be calculated from the flowmeter or the control valve position measurement (or the control valve demand signal). It will be demonstrated that a state estimator may be used to estimate the flowrate using the flowmeter measurement and the "pseudo" measurement of control valve stem position, and that this provides a data base from which control loop security may be determined. The check is based upon monitoring changes in the residuals generated by the estimator.

The proposed technique does not require additional process instrumentation, uses little computer time and storage and may be performed while the control loop is operating under d.d.c.

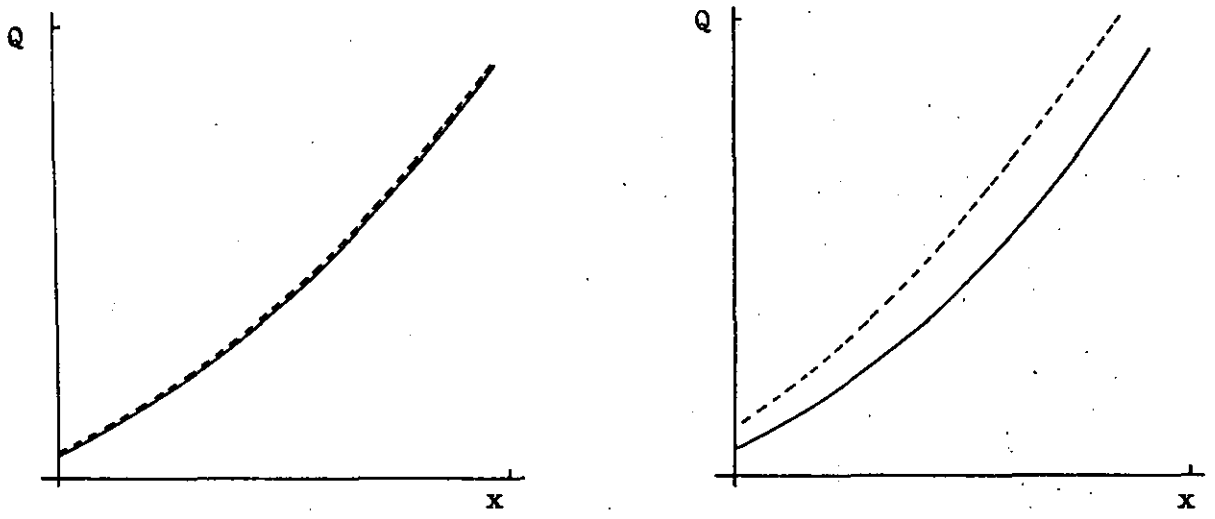
The malfunction detection technique is tested by extensive laboratory experimentation and some industrial tests performed on operating d.d.c. loops at Works A.

2.3 A valve position-flow check

In flow control loops in which the pressure drop across the valve is either constant or varies reproducibly with flow, there is a constant relation between the valve position and the flow, as given by the valve characteristic. Thus the valve position or controller signal can be used to provide an indication of flow which is additional to the measurement given by the flowmeter. As a result of this it is possible to consider a flow control loop as a redundant system since there exist two measurements of the same variable flow.

During normal process operation the control valve stem position (or controller demand signal) may be calibrated against the flowmeter measurement, under the assumption of no malfunction. A divergence between the measured flow and that calculated from this characteristic curve may then be taken to indicate that a loop malfunction has occurred.

This principle is shown schematically in Figure 2.3.



a. Healthy loop. b. Unhealthy loop.

- - - - - - - - flow measured by flowmeter

——————— flow measured by valve position.

Figure 2.3. Relation between valve position and flow.

The flow Q is plotted against the valve position x. Figure 2.3a shows the graph for a healthy loop. Figure 2.3b shows an unhealthy loop. There is now a divergence between the curves of flow as measured by the flowmeter and as estimated from the valve position.

This idea is presented as the first technique for malfunction detection in flow control loops. It is envisaged that a check on loop security could be developed based on this divergence, either by calculating an error or by displaying the plot on a process computer V.D.U.

2.4 Valve position - flow characteristics (43), (44)

In the majority of flow control loops the flowmeter may be classified by two main types; the linear flowmeter and the square root flowmeter.

The equation of a linear flowmeter, such as a magnetic flowmeter or turbine meter may be written as:

$$Q_m = k_{m1} s \qquad (2.4.1)$$

Alternatively, a square root flowmeter, such as an orifice plate and differential pressure transmitter may be used. The equation of flow of an incompressible fluid through a constant cross sectional area orifice is:

$$Q_m = k_1 \sqrt{\Delta P_m} \qquad (2.4.2)$$

The relation for the transmitter is:

$$\Delta P_m \propto s$$

Then the equation describing a square root flowmeter is:

$$Q_m = k_{m_2} \sqrt{s} \qquad (2.4.3)$$

The equation governing incompressible fluid flow through a restriction such as a control valve may be derived from the laws of fluid mechanics and is of the form:

$$Q_v = k_2 \quad A_v \sqrt{\Delta P_v} \qquad (2.4.4)$$

The cross sectional area for flow A_v depends upon the valve stem position x

$$A_v = f(x) \qquad (2.4.5)$$

When this relation is linear equation (2.4.5) may be written as:

$$A_v = k_3 x \qquad (2.4.6)$$

Combining equations (2.4.4) and (2.4.6) results in:

$$Q_v = (k_2 \sqrt{\Delta P_v}) k_3 x \qquad (2.4.7)$$

This relation shows the flowrate through the control valve is directly proportional to the valve stem position provided the control valve pressure drop ΔP_v is constant, and all the other parameters are constant.

Under this condition equation (2.4.7) may be written as:

$$Q_v = k_{v1} x \qquad (2.4.8)$$

Equation (2.4.8) represents the model for a linear control valve characteristic, although in practice some valves are better modelled as:

$$Q_v = k_{v1} + k_{v2} x$$

Another commonly used control valve characteristic is the equal percentage valve whose equation may be derived in an analogous manner to be (43), (44)

$$Q_v = Q_{v\,max} \exp(- k_{v4} (x_{max} - x)) \qquad (2.4.9)$$

Alternatively, a more convenient form is often:

$$Q_v = k_{v3} \exp(k_{v4} x) \qquad (2.4.10)$$

If a control loop is characterised by these models and there are no errors in the system or in the models themselves, then a material balance must constrain the measured flow Q_m and the flow obtained from the valve position Q_v to be equal to the true flow Q

$$Q_m = Q_v = Q \qquad (2.4.11)$$

However, if such an error or malfunction is present in the system then the above equations are no longer valid and in particular

$$Q_m \neq Q_v \neq Q$$

Let e_s and e_x be the errors in the flowmeter signal s and valve position x, then if a control loop is characterised by a linear flowmeter and a linear control valve, the models of equations (2.4.1) and (2.4.8) may be modified as:

$$Q_m = k_{m1} (s + e_s) \qquad (2.4.13)$$

$$Q_v = k_{v1} (x + e_x) \qquad (2.4.14)$$

The effect of these errors on the relation between Q and x may be investigated by simulation.

In this section two types of flowmeter and two control valve characteristics have been described and so there are four types of loop:

i) Type 1: Linear flowmeter, linear control valve.

ii) Type 2: Square root flowmeter, linear control valve.

iii) Type 3: Linear flowmeter, equal percentage control valve.

iv) Type 4: Square root flowmeter, equal percentage control valve.

The errors investigated are:

i) Flowmeter zero error,

ii) Flowmeter range error,

iii) Valve position zero error,

iv) Valve position range error.

The control valve errors are best understood by regarding the valve as an alternative measuring device in which the signal x is equivalent to the signal s in the flowmeter.

The equations, parameters and errors used in the simulation are given in Table 2.2. The results are shown in Figures 2.4 - 2.7.

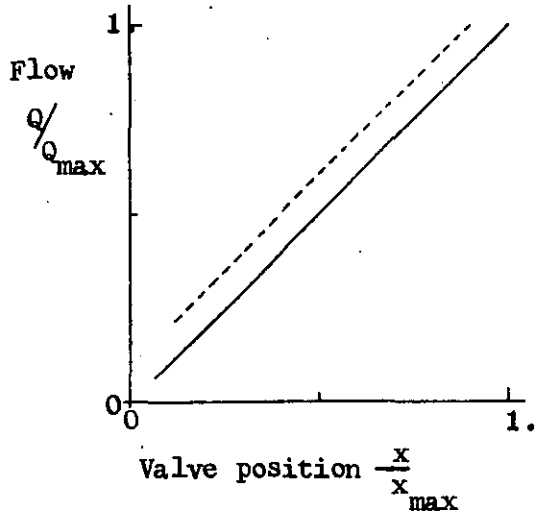
Figures 2.4 - 2.7 show the type of valve characteristic deviation under conditions of control loop malfunction. It can be seen from these Figures for each particular type of loop that there are some errors which are indistinguishable; this is particularly true of flowmeter and valve position range errors. However, some errors give rather distinctive deviations of valve characteristic. For example, the curve for a flowmeter zero error in a Type 4 flow control loop, as shown in Figure 2.7a, is particularly recognisable.

Instrument	Characteristic	Equation	Parameters			Errors	
						Zero	Range
			k_{m_1}	Q_{max}	s_{max}	$\frac{e_s}{s_{max}}$	$\frac{e_{s\ max}}{s_{max}}$
Flowmeter	Linear	$Q_m = k_{m_1}(s + e_s)$	$0.281 \times 10^{-3}\ \frac{m^3}{s.V}$	$2.81 \times 10^{-3}\ \frac{m^3}{s}$	10 V	0.1	0.1
	Square root	$Q_m = k_{m2}\sqrt{s + e_s}$	$0.8886 \times 10^{-3}\ \frac{m^3}{s.V^{\frac{1}{2}}}$	$2.81 \times 10^{-3}\ \frac{m^3}{s}$	10 V	0.1	0.1
			k_{vi}	Q_{max}	x_{max}	$\frac{e_x}{x_{max}}$	$\frac{e_{x_{max}}}{x_{max}}$
Control valve	Linear	$Q_v = k_{v_1}(x + e_x)$	$73.86 \times 10^{-3}\ \frac{m^3}{s.m}$	$2.81 \times 10^{-3}\ \frac{m^3}{s}$	0.0381 m	0.1	0.1
	Equal Percentage	$Q_v = Q_{v\ max}\ \exp(-k_{v4}(x_{max} - (x+e_x)))$	52.2	$2.81 \times 10^{-3}\ \frac{m^3}{s}$	0.0381 m	0.1	0.1

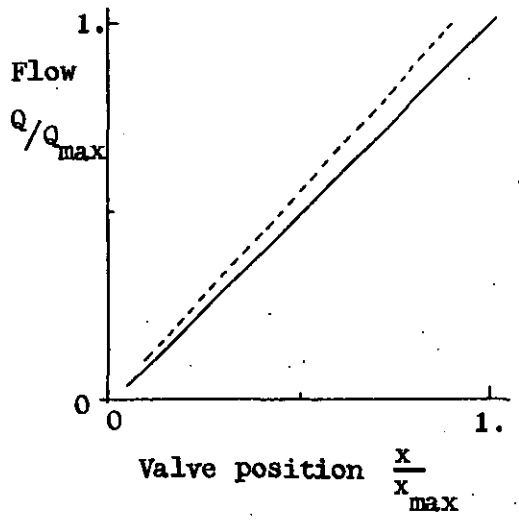
TABLE 2.2. Simple models of flow control loops.

_____ Flow estimated from valve position

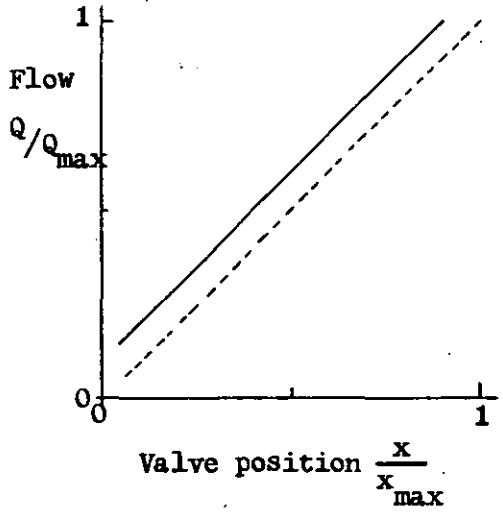
- - - - Flow measured by flowmeter



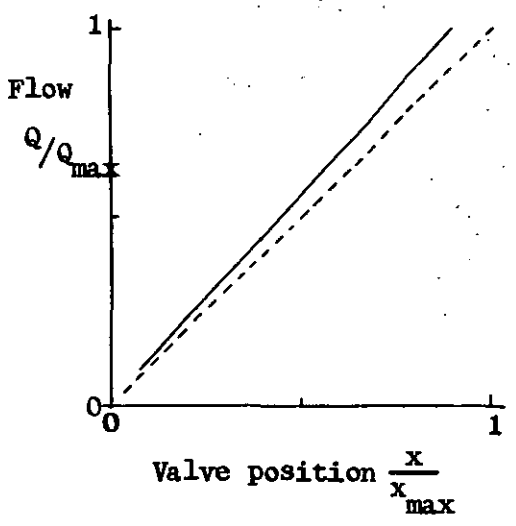
a. Flowmeter zero error



b. Flowmeter range error

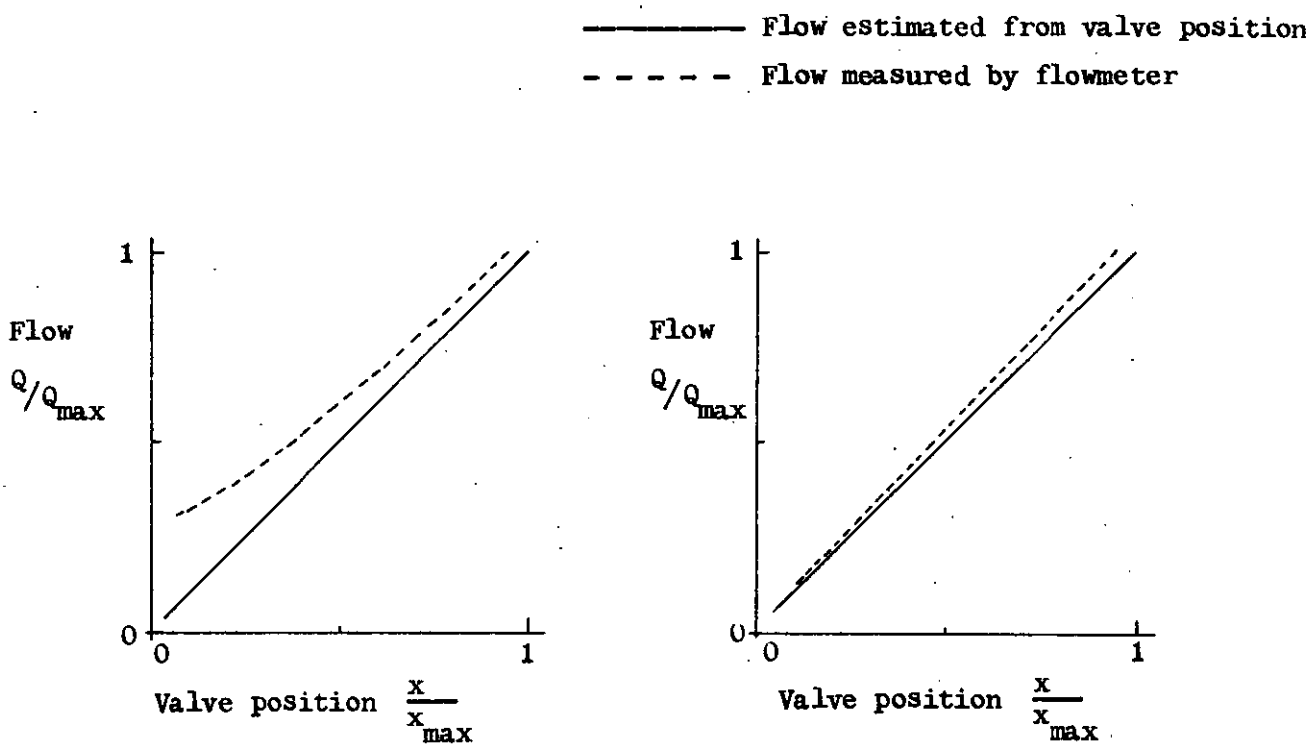


c. Valve position zero error



d. Valve position range error

FIGURE 2.4 Type 1 control loop (linear flowmeter, linear control valve).



a. Flowmeter zero error

b. Flowmeter range error

c. Valve position zero error

d. Valve position range error

FIGURE 2.5 Type 2 control loop (square root flowmeter, linear control valve).

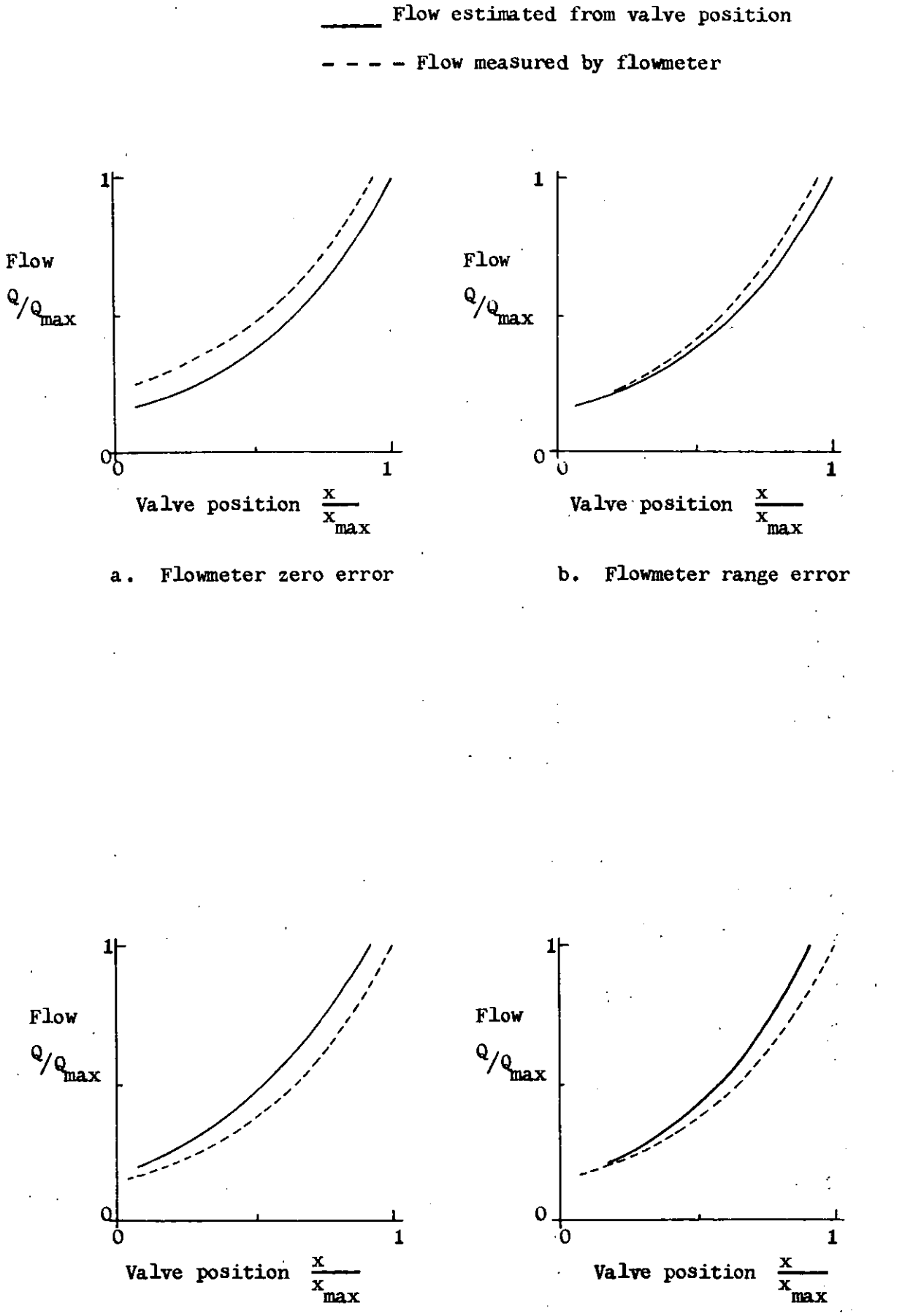
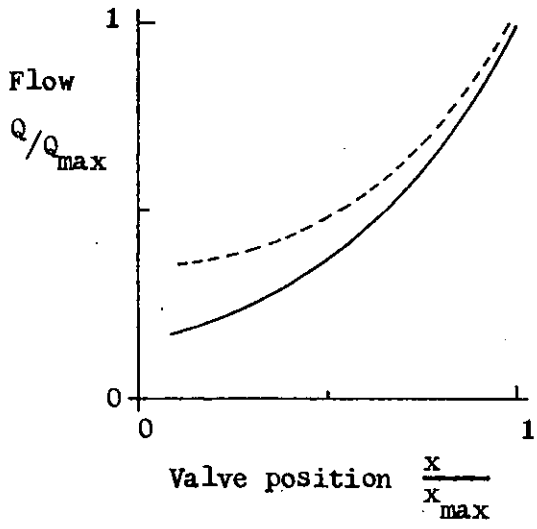


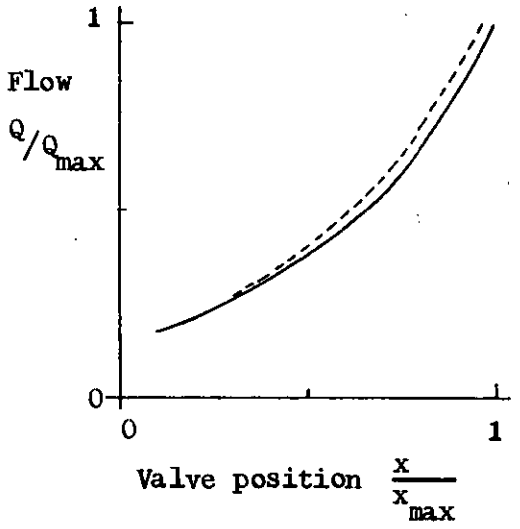
FIGURE 2.6 Type 3 control loop (linear flowmeter, equal percentage control valve).

——— Flow estimated from valve position

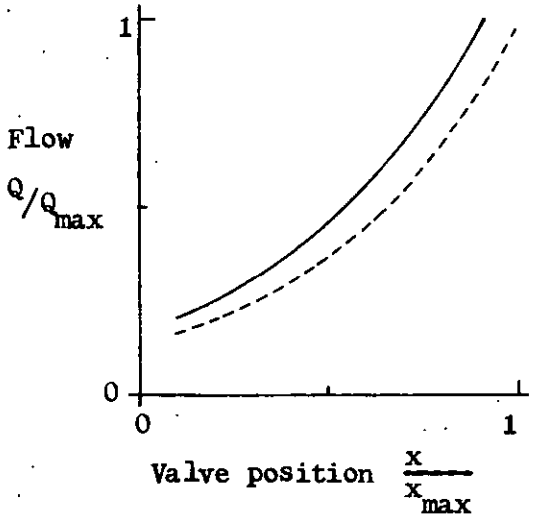
- - - - Flow measured by flowmeter



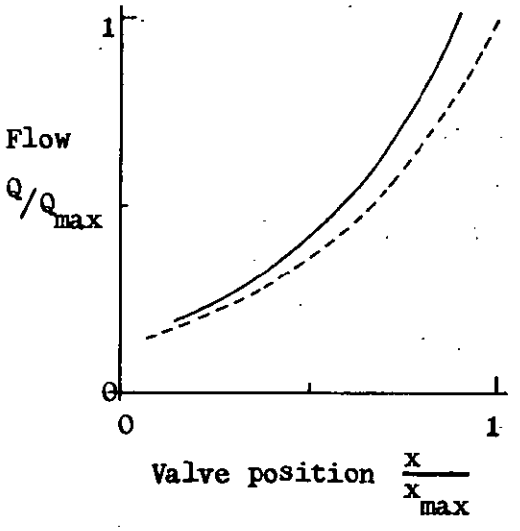
a. Flowmeter zero error



b. Flowmeter range error



c. Valve position zero error



d. Valve position range error

FIGURE 2.7 Type 4 control loop (square root flowmeter, equal percentage control valve).

2.5 Inadequacies of the valve position - flow check

The initial postulation for determining flow control loop security has been based upon the generation of the types of curves shown in Figures 2.4 - 2.7. It is envisaged that these curves could be obtained during normal control loop operation or by "stroking" the valve between the fully open and fully closed position.

However, in practical applications there are several difficulties:

i) Limited valve travel.

Although the stroking of the valve is acceptable in many cases, in others it is not. In these problems the amount of valve travel may also be constrained.

ii) Lags between valve position and flow measurement.

iii) The measurements of valve position and flowmeter signal are not perfect, but depend upon the accuracy of the instrumentation provided. This creates a "noisy" signal. Other sources of noise in the loop may be due to fluctuations in pressure.

These comments are illustrated in Figure 2.8 which shows the valve position and flowmeter measurements logged on a d.d.c. process computer at Works A. The actual control loop was the inner loop of a cascade control system as shown in Figure 2.9.

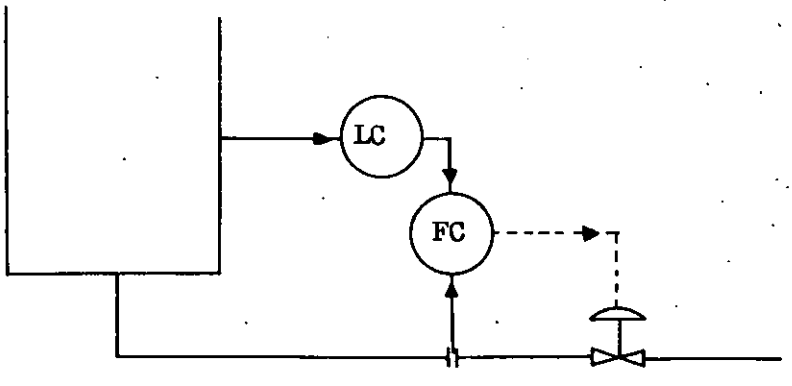


Figure 2.9. An industrial cascade level control system.

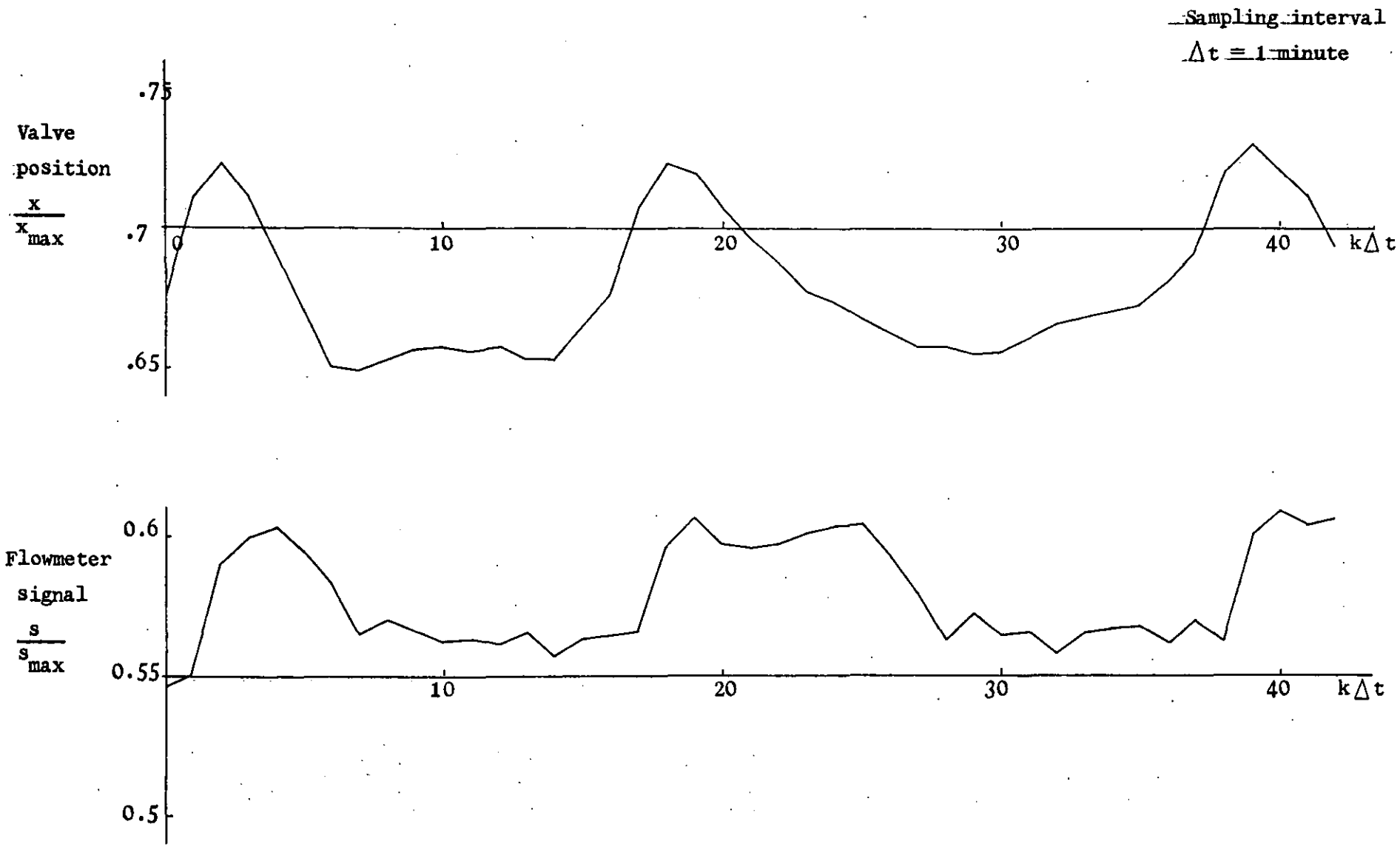


FIGURE 2.8 Relation between valve position and flowmeter signal in an industrial control loop.

The master or outer loop controls level and tended to oscillate so that the setpoint of the flow loop was oscillating. The extent of the lag between valve position and flowmeter signal is derived from Figure 2.8 and may be seen to be of the order of 1 minute.

The philosophy of the valve position-flow check for flow control loop security monitoring is based upon the creation of a model for the system, deviations from which indicate possible system errors. The problems encountered above arise, in part, because of the simplicity of the assumed model. These may be overcome by defining a more flexible model; and to this end the concepts of state estimation are introduced, although the basic philosophy of detecting malfunctions as a result of monitoring system deviations from the assumed mathematical model is the same.

2.6 Least squares estimation theory

A state estimator is a data processing algorithm which computes the state of the system using the following information:

i) Measurements of the system variables.

ii) A mathematical model of the system and its associated instrumentation.

iii) Prior knowledge of some system variables - often called pseudomeasurements.

The output of the state estimator is an assessment of the true system state. Discrepancies between the true system state and estimator output can arise from:

i) Noise in instruments and telemetry channels.

ii) Incomplete instrumentation, in the sense that many variables are unmeasured.

iii) Incorrect mathematical and/or incorrect model parameters.

iv) Delayed measurements reflecting some prior state.

There are many approaches to state estimator formulation (45). It is usual to define a scalar "cost" function which increases with estimation error. An estimator which minimises this cost function is optimal in the sense that it generates the best estimate of the state based on the measurements and pseudomeasurements received.

The optimal estimator not only generates a state estimate but also a covariance or error associated with it. The covariance matrix predicts the magnitude of the estimation error, and hence provides a measure of the confidence for the estimated state. The predicted and the actual estimation errors should be of the same order of magnitude.

The basis for modern estimation theory is the method presented by Kalman (46). His solution to the recursive estimation problem has become known as the Kalman filter, which will be discussed in Chapter 3. Estimation theory is heavily dependent upon the concept of state variables and the widely used dynamic state variable model (46)

$$\dot{\underline{x}} = f(\underline{x}, \underline{u}, \underline{w}) \qquad (2.6.1)$$

$$\underline{y} = \underline{h}(\underline{x}) + \underline{v} \qquad (2.6.2)$$

In this Chapter no attempt is made to model the time behaviour of the flow control loop and steady state operation is assumed, i.e.

$$\dot{\underline{x}} = \underline{0} \qquad (2.6.3)$$

$$\underline{y} = \underline{h}(\underline{x}) + \underline{v} \qquad (2.6.4)$$

Equation (2.6.4) describes how the quantities which can be measured are related to the state variables. Each measured variable

is contaminated with noise. This set of measurement noise is denoted $\underline{v}$, the statistics of which are assumed known.

Specifically

$$E(\underline{v}) = \underline{0}$$

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

If the measurements are assumed to be independent, then R is a diagonal matrix with diagonal terms:

$$r_{ii} = \sigma_i^2$$

and σ_i is the standard deviation on the i^{th} measurement. These values characterise the accuracy of the individual measurements and may be calculated from a knowledge of the instrumentation and telemetry systems.

In least squares estimation theory the optimal state estimate $\hat{\underline{x}}$ is defined as the estimate which minimises the performance criterion:

$$J = (\underline{y} - \underline{h}(\underline{x}))^T R^{-1} (\underline{y} - \underline{h}(\underline{x})) \qquad (2.6.5)$$

The cost function J is a quadratic function of the difference between the actual and estimated measurements, which is weighted according to the expected accuracy of the corresponding measurements. Thus the estimate fits the measurement made by an accurate instrument better than the measurement by an inaccurate instrument.

If the state variables are linear functions of the measured variables then equations (2.6.4) and (2.6.5) become

$$\underline{y} = H\underline{x} + \underline{v} \qquad (2.6.6)$$

$$J = (\underline{y} - H\underline{x})^T R^{-1} (\underline{y} - H\underline{x}) \qquad (2.6.7)$$

The minimisation of equation (2.6.7) may be performed analytically by differentiating with respect to $\underline{x}$ and equating to zero. The resulting

minimisation yields the well known Best Linear Unbiased Estimator (BLUE) (47), (48), (49),

$$\hat{\underline{x}} = (H^T R^{-1} H)^{-1} H^T R^{-1} \underline{y} \qquad (2.6.8)$$

Equation (2.6.8) is attractive since it solves for the state estimate directly.

If the measured variables are non-linear functions of the state then equation (2.6.5) cannot be minimised analytically. However, a standard approach is to linearise $\underline{h}(\underline{x})$ about some initial value $\underline{x}_o$ using a Taylor expansion which yields:

$$\underline{h}(\underline{x}) = \underline{h}(\underline{x}_o) + F(\underline{x}_o)\,\Delta\underline{x} + \ldots . \qquad (2.6.9)$$

when higher order terms are neglected. $F(\underline{x}_o)$ is a Jacobian matrix whose elements are:

$$f(i, j) = \left.\frac{\partial \underline{h}_i(\underline{x})}{\partial x_j}\right|_{\underline{x} = \underline{x}_o} \qquad (2.6.10)$$

and

$$\Delta\underline{x} = \underline{x} - \underline{x}_o \qquad (2.6.11)$$

Equation (2.6.4) may now be rewritten, using equation (2.6.9) as:

$$\underline{y} = \underline{h}(\underline{x}_o) + F(\underline{x}_o)\,\Delta\underline{x} + \underline{v}$$

or

$$\Delta\underline{y} = \underline{y} - \underline{h}(\underline{x}_o) = F(\underline{x}_o)\,\Delta\underline{x} + \underline{v} \qquad (2.6.11)$$

Equation (2.6.11) represents a linear estimation problem with a cost function:

$$J = (\Delta\underline{y} - F(\underline{x}_o)\,\Delta\underline{x})^T R^{-1} (\Delta\underline{y} - F(\underline{x}_o)\,\Delta\underline{x}) \qquad (2.6.12)$$

Thus in a similar manner to the derivation of equation (2.6.8) the least squares estimate may be written as:

$$\widehat{\Delta\underline{x}} = (F^T(\underline{x}_o) R^{-1} F(\underline{x}_o))^{-1} F^T(\underline{x}_o) R^{-1} \Delta\underline{y}$$

or $\hat{\underline{x}} = \underline{x}_o + S(\underline{x}_o) F^T(\underline{x}_o) R^{-1} (\underline{y} - \underline{h}(\underline{x}_o))$ (2.6.13)

where:

$$S(\underline{x}_o) = (F^T(\underline{x}_o) R^{-1} F(\underline{x}_o))^{-1} \quad (2.6.14)$$

If the initial guess $\underline{x}_o$ is sufficiently close to the true state $\underline{x}$ and if the linearisation performed in equation (2.6.9) is a sufficiently accurate representation of equation (2.6.4) then equation (2.6.13) gives the desired state estimate $\hat{\underline{x}}$. In practice it is necessary to define an iteration sequence $\underline{x}_n$, n = 1, 2 . . . which is given by:

$$\underline{x}_{n+1} = \underline{x}_n + S(\underline{x}_n) F^T(\underline{x}_n) R^{-1} (\underline{y} - \underline{h}(\underline{x}_n)) \quad (2.6.15)$$

This iteration procedure starts with $\underline{x}_n = \underline{x}_o$ and proceeds until $J(\underline{x}_n)$ approaches a minimum. Stopping rules are to iterate until $|J(\underline{x}_{n+1}) - J(\underline{x}_n)|$ or until the magnitude of all components of $|\underline{x}_{n+1} - \underline{x}_n|$ are less than some predetermined value.

It is of course possible for $J(\underline{x})$ to have local minima and thus $\underline{x}_n$ may converge but not to $\underline{x}$. It is also possible for $\underline{x}_n$ never to converge.

This completes this brief introduction to least squares estimation theory. The least squares performance criterion has a long history of providing good state estimates in a wide variety of problems (economics, biology, aerospace trajectories) and in particular is finding current applications in the modelling of electrical networks.

2.6.1 Tracking state estimation

The previous review of least squares estimation theory applies only to systems that can be described by steady state measurement relationships modelled by equation (2.6.4). Such estimators may be denoted "static" state estimators in the sense that the information received at any time is processed without regard to past information. For example, consider two measurement vectors $\underline{y}(k)$ and $\underline{y}(k+1)$ at time k and $k+1$ where $k+1 = k+\Delta t$. $\underline{y}(k)$ contains information about $\underline{x}(k)$ and $\underline{x}(k+1)$, however the static estimator ignores this information when calculating $\hat{\underline{x}}(k+1)$ and only uses measurements $\underline{y}(k+1)$.

This problem is overcome by dynamic state estimation. The model formulation for dynamic state estimation was given in equations (2.6.1) and (2.6.2) and the solution due to Kalman was mentioned. One of the main disadvantages of dynamic state estimation is that modelling of the time behaviour of the system state is necessary. This usually is tedious, time consuming, costly and at best full of uncertainties.

The objective of introducing state estimation to aid malfunction detection was to create a flexible mathematical model to overcome the constraints discussed in section 2.5. The purpose of the estimator is not necessarily to calculate extremely accurate state estimates $\hat{\underline{x}}$ but solely to process the measurements to a more amenable form, from which system security may be determined. With these aims in mind, it is suggested that dynamic state modelling is unnecessary and the philosophy adopted here is that the "best" malfunction detection technique is the simplest one that works.

In a flow control loop operating at a particular setpoint, the

valve will operate over a limited range of positions and there will be some dynamic lag between the valve stem position and flowmeter signal measurements.

Under these conditions it is suggested that a "tracking" state estimator yields an adequate assessment of the state. The tracking state estimator (50) is derived from the static state estimator of equation (2.6.15), i.e.

$$\underline{x}_{n+1} = \underline{x}_n + S(\underline{x}_n) F^T(\underline{x}_n) R^{-1} (\underline{y} - \underline{h}(\underline{x}_n)) \qquad (2.6.1.1)$$

Suppose a series of measurements have been taken at the discrete time k, i.e. $\underline{y}(k)$ yielding $\hat{\underline{x}}(k)$. Equation (2.6.1.1) reveals that the iteration sequence can be started from any arbitrary value $\underline{x}_o$. However, if the iteration is to be done at time $k+1$ the logical initial guess is the estimate $\hat{\underline{x}}(k)$ made from $\underline{y}(k)$.

Now if the state has not moved excessively (i.e. $\underline{x}(k+1)$ is close to $\underline{x}(k)$) and if the observation noise is not "too large", it may be expected that only one iteration will be needed and a reasonable estimator would be the recursion:

$$\hat{\underline{x}}(k+1) = \hat{\underline{x}}(k) + S(\hat{\underline{x}}(k)) F^T(\hat{\underline{x}}(k)) R^{-1} (\underline{y}(k+1) - \underline{h}(\hat{\underline{x}}(k))) \qquad (2.6.1.2)$$

with $$S(\hat{\underline{x}}(k)) = F^T(\hat{\underline{x}}(k)) R^{-1} F(\hat{\underline{x}}(k)) \qquad (2.6.1.3)$$

Equation (2.6.1.2) defines the "tracking" state estimator used in this study.

Notice that in equation (2.6.1.1) n refers to an iteration sequence number while in equation (2.6.1.2) k refers to the discrete time interval.

2.6.2 Use of estimator to detect malfunction

This section discusses some of the techniques available for monitoring system security when a state estimator is used.

The state estimation algorithm formulated is based on the assumption that the model $\underline{h}(\underline{x})$ and R is correct. However, assumed models are often inaccurate or may change due to system malfunction and it is necessary to be able to detect such errors.

The notation $\hat{J}$ denotes the value of the optimal cost function evaluated at $\hat{\underline{x}}$, i.e.

$$\hat{J} = (\underline{y} - \underline{h}(\hat{\underline{x}}))^T R^{-1} (\underline{y} - \underline{h}(\hat{\underline{x}})) \qquad (2.6.2.1)$$

One approach to detection is that if the model is correct, then $\hat{J}$ is a random variable whose probability density function can be calculated. If the measurement errors $\underline{v}$ are assumed Gaussian then reference (49) shows $\hat{J}$ is a Chi-squared random variable with $(m - n)$ degrees of freedom. Thus, if a value of $\hat{J}$ is obtained which lies on the tails of this probability distribution, then it can be assumed that the model is incorrect and an error in $\underline{h}(\underline{x})$ or R has been detected.

Another technique which does not require probability distributions requires more calculation. To illustrate this idea suppose there are m measurement equations, and one in particular is known to be inaccurate. Let $\underline{h}^{(o)}(\underline{x})$ be the model containing all m equations, and let $\underline{h}^{(i)}(\underline{x})$ denote the model with one of the measurement equations removed, $i = 1, \ldots, m$. Thus there are $m + 1$ possible models for the measurements given by:

$$\underline{y} = \underline{h}^{(i)}(\underline{x}) + \underline{v} \qquad i = 0, \ldots, m$$

Corresponding to these models there are m+1 possible state estimates $\hat{\underline{x}}^{(i)}$ and optimal cost functions $\hat{J}^{(i)}$. $\hat{J}^{(i)}$ can be regarded as a measure of how good a fit the model structure $\underline{h}^{(i)}(\hat{\underline{x}})$ provides to the measurement $\underline{y}$. If it happened that for some j that $\hat{J}^{(j)} < \hat{J}^{(i)}$ for all $i \neq j$, it would be reasonable to assume that the measurement $\underline{y}$ was made on a model with the measurement equation j in error.

A different approach to the problem is to examine the residual process (47), (48), (49), (51). The residual is defined as:

$$r_i = y_i - \underline{h}_i(\hat{\underline{x}}) \qquad (2.6.2.2)$$

which represents the observed error of the i^{th} measurement if the model is assumed correct. Under normal conditions (i.e. no modelling errors or system malfunction), the properties of $\underline{r}$ should resemble those of the assumed measurement noise $\underline{v}$. There are several methods of inspecting the residual process.

i) Overall residual plot.

This should confirm the assumed probability distribution function of $\underline{v}$.

ii) Time sequence plot.

Again this should confirm the initial assumptions concerning $\underline{v}$. Draper and Smith (51) suggest that time trends can often be used to locate model deficiences.

iii) Residual against independent variable plot.

iv) Residual against dependent variable plot.

2.6.3 Residual analysis via hypothesis testing

The problem of malfunction detection may be formulated as a problem in hypothesis testing (52), (53). The normal operation of the system and hence the "normal" residual time sequence is regarded as the null hypothesis. The actual residual process from the current system operation is tested against this hypothesis at a certain level of significance. Different types of malfunction can develop in the system. Some of these are:

i) Bias errors in instruments,

ii) Excessively "noisy" instruments,

iii) Change in model parameters.

All of these faults cause the residual r_i to depart from its "normal", malfunction free characteristics. It is therefore proposed to use the following statistical tests to observe the residual process. A comparison is made between an ensemble of residual values taken under the initialisation conditions which are assumed to be free of malfunction (ensemble o) and an ensemble taken during a later check (ensemble 1).

i) Test of mean (52):

The null hypothesis to be tested is whether the means of two different ensembles could have come from the same probability distribution or from distributions with the same means, i.e.

$$H_o : \mu_o = \mu_1$$

A Student's t test enables the difference between two means to be tested as:

$$t = \frac{\left| \bar{r}_{i,o} - \bar{r}_{i,1} \right|}{\bar{s}\,(r_i)\,\sqrt{\frac{1}{n_o} + \frac{1}{n_1}}} \qquad (2.6.3.1)$$

where the sample variances are assumed unknown but equal and:

$$\bar{s}^{\,2}(r_i) = \frac{(n_o - 1)\, s_{i,o}^{\,2} + (n_1 - 1)\, s_{i,1}^{\,2}}{(n_o + n_1 - 2)}$$

$$\bar{r}_{i,j} = \frac{1}{n_j} \sum_{k=1}^{n_j} r_{i,j}(k) \qquad i = 1, \ldots m; \quad j = 0, 1$$

$$s_{i,j}^{\,2} = \frac{1}{n_j - 1} \sum_{k=1}^{n_j} \left(r_{i,j}(k) - \bar{r}_{i,j}\right)^2 \qquad i = 1, \ldots m \quad j = 0, 1 \qquad (2.6.3.2)$$

The value of t is calculated and compared with values tabulated for the number of degrees of freedom $(n_o + n_1 - 2)$. If the calculated value of t is larger than the tabulated t at the preselected significance level and then the hypothesis is rejected and the ensemble mean estimated by $\bar{r}_{i,o}$ is significantly different from that estimated by $\bar{r}_{i,1}$, with α chance of being incorrect in rejecting the hypothesis.

At the 5% significance level the tabulated value of t for infinite degrees of freedom is 1.96.

ii) Test of variance (52)

The null hypothesis, like that for comparing means is:

$$H_o : \sigma_o^{\,2} = \sigma_1^{\,2}$$

The F test defined as:

$$F = \frac{s_{i,o}^{\,2}}{s_{i,1}^{\,2}}$$

measures the difference, in the form of a ratio, of the two estimates of the same variance that can be expected to occur, depending upon the number of degrees of freedom available for each estimate.

Tabulated values of the F probability distribution with $(n_o - 1)$ and $(n_1 - 1)$ degrees of freedom indicate whether one estimated variance is larger than another. At the 5% significance level, with 100 degrees of freedom for each calculation of $s^2_{i,j}$, the tabulated value of F is 1.39. Thus, if the calculated value of F is greater than 1.39, the hypothesis of equal ensemble variance is rejected.

2.7 Experimental apparatus and objectives

Section 2.3 presented a technique for detecting malfunction in a flow control loop. This idea was discussed and criticised and the concepts of state estimation were introduced to solve the inherent difficulties.

In order to test these ideas an experimental test apparatus was designed and constructed.

The objective of the experimental apparatus was to provide a flow control loop using industrial control equipment.

The laboratory flow control rig is shown in Figure 2.10. The rig was 9.4 m high and extended through three levels of the building.

The fluid used was water.

The main components of the test equipment are summarised as follows:

i) Pipework.

The pipework was a mixture of 5.08 cm (2 inch) I.D. mild steel and rigid P.V.C. All the pipework near the pump, control valve and orifice was mild steel.

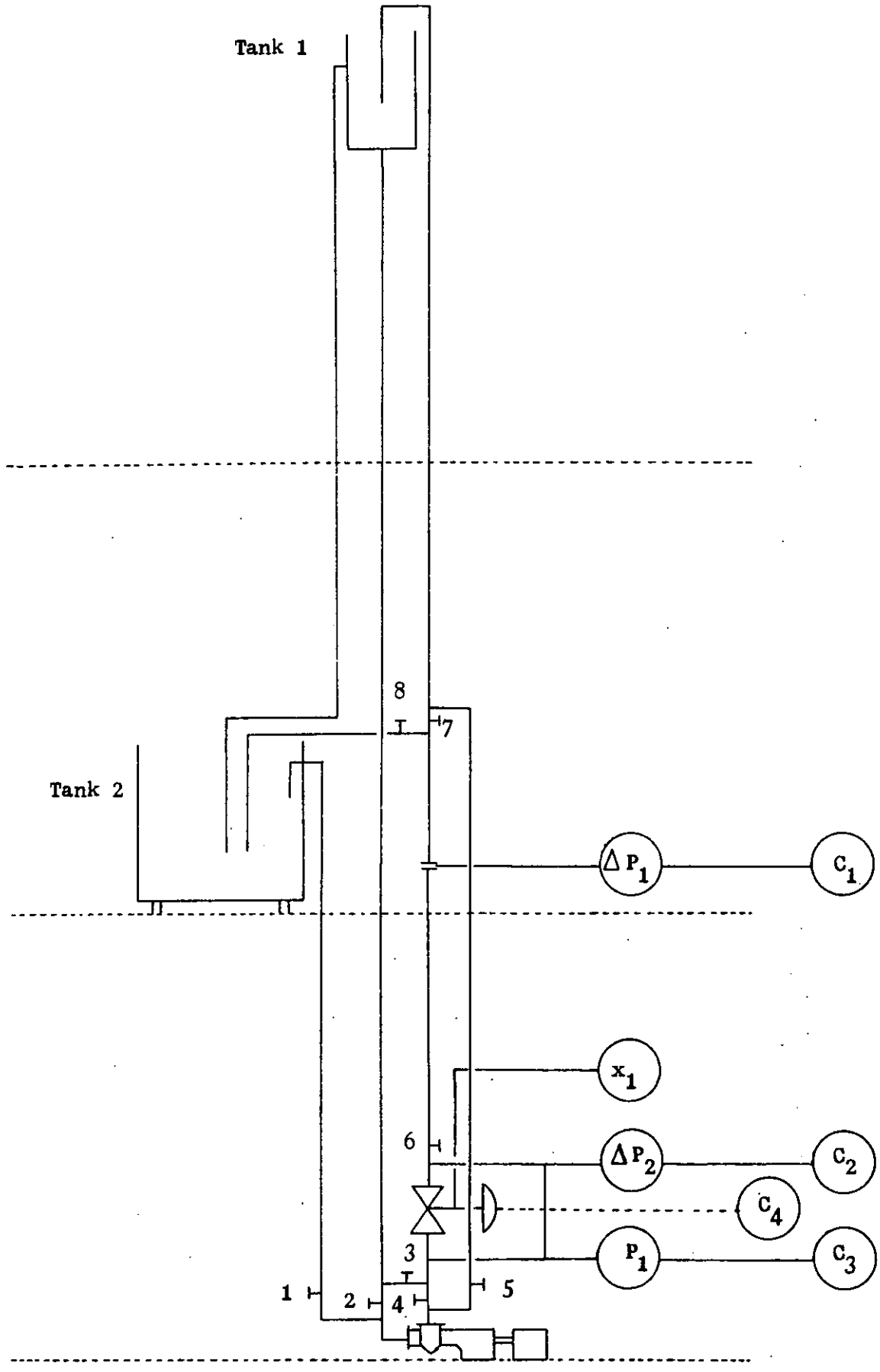


FIGURE 2.10 Experimental flow control rig.

ii) Control valve.

The flow control valve was a characterised pneumatic valve and was manufactured by the Glocon Company Limited. Details of the valve are given in Table 2.3.

Type	Globe
Body size	3.81 cm ($1\frac{1}{2}$ inch)
Trim size	1.905 cm ($\frac{3}{4}$ inch)
C_v (imperial units)	20
Body material	Carbon steel
Trim form	i) Linear ii) Equal percentage
Trim material	316 Stainless steel
Actuator type	Numotor B 40
Positioner	Standard
Signal pressure	20.68 - 103.42 kN/m^2 (3-15 p.s.i.g.)

TABLE 2.3. Details of the control valve used in the experimental flow control rig.

iii) Pump.

A single stage centrifugal pump, supplied by Ingersoll-Rand (Type N - $1\frac{1}{2}$ - 120) was used.

iv) Pump drive.

An electric motor drive was installed which incorporated a starter with overload protection to prevent motor damage if the pump had to deal with impurities or the monometric head was lower than expected.

v) Flow measurement.

The flowmeter was a sharp edged D - D/2 orifice plate designed according to B.S.-1042 (54).

vi) Isolation valves.

The details of the isolation valves are summarised in Table 2.4.

Valve number	Valve type
1	Gate
2	Gate
3	Gate
4	Globe
5	Gate
6	Globe
7	Gate
8	Gate

TABLE 2.4. Experimental flow control rig isolation valves.

vii) Instrumentation.

The location of the measurement instruments are shown in Figure 2.10, and Table 2.5 details their functions.

Instrument code	Instrument type	Maker	Input	Output
ΔP_1	Δ P/P transmitter	Taylor Instruments	0 - 4.98 kN/m^2 (0 - 20 in.W.G.)	20.7-103.4 kN/m^2 (3 - 15 p.s.i.g.)
ΔP_2	Δ P/P transmitter		4.98-62.3 kN/m^2 (20-250 in.W.G.)	20.7-103.4 kN/m^2 (3 - 15 p.s.i.g.)
P_1	P transmitter		69-689.5 kN/m^2 (10-100 p.s.i.g.)	20.7-103.4 kN/m^2 (3 - 15 p.s.i.g.)
C_1	P/I transmitter		3 - 15 p.s.i.g.	5 - 10 V
C_2	P/I transmitter		3 - 15 p.s.i.g.	0 - 10 V
C_3	P/I transmitter		3 - 15 p.s.i.g.	0 - 10 V
C_4	I/P transmitter		0 - 10 V	20.7-103.4 kN/m^2 (3 - 15 p.s.i.g.)
X_1	Linear potentiometer	Penny and Giles	0 - 3.81 cm.	3 - 9 V

TABLE 2.5. Instrumentation of laboratory flow control rig.

The pressure drop across the orifice and that across the control valve were measured by differential pressure transmitters and the pressure between the pump and the control valve by an absolute pressure transmitter. These three transmitters were all pneumatic and pressure/current converters were used to convert their output into electrical signals suitable for process computer inputs. The valve stem position was measured by a linear potentiometer whose output varied linearly with position.

The experimental rig was controlled by a PDP 11-20 process control computer, which was used for logging, signal generation and direct digital control.

2.7.1. System flow-pressure drop characteristics

The experimental rig was designed to investigate two pressure-drop phenomena across the control valve.

A system with pressure drop across the valve approximately constant was obtained with valves 1, 3, 5, 6 and 8 open and valves 2, 4 and 7 closed. Under these conditions the water flowed from tank 2 to tank 1 due to the head difference. A constant head system was achieved by pumping water from tank 1 to tank 2 where an overflow device maintained a constant level. The rig was designed so that under these conditions approximately 70% of the pressure drop occurred at the control valve.

The variation of pressure drop across the control valve with flow for an equal percentage valve trim is shown in Figure 2.11.

With valves 2, 4, 6 and 7 open and valves 1, 3, 5 and 8 closed, a system with variable pressure drop across the valve was obtained. The water flowed from tank 2 through the pump and control valve and back to

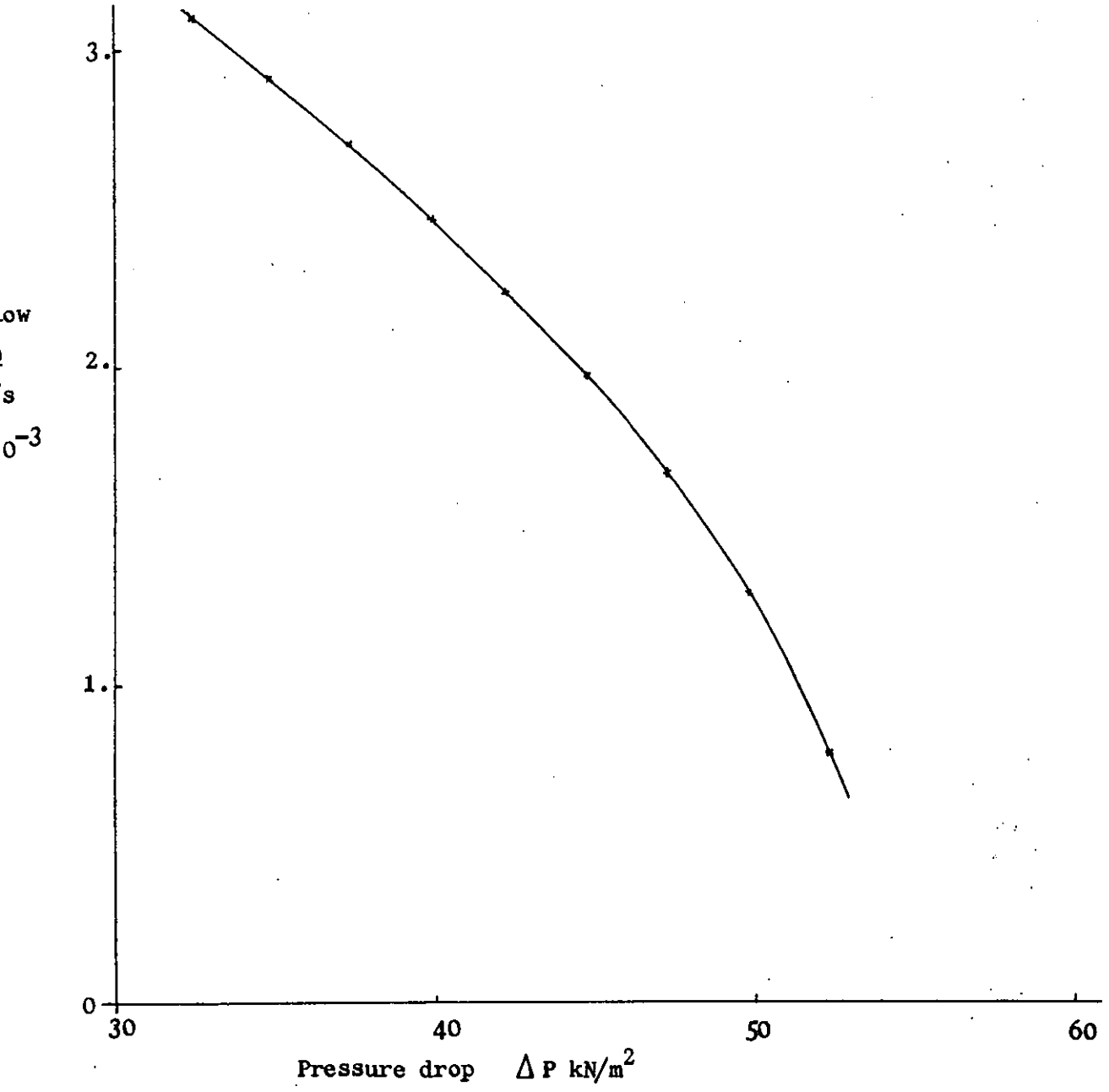


FIGURE 2.11 Variation of pressure drop with flow across equal percentage control valve.

tank 2. Additional changes in system pressure drop could be obtained by throttling valve 4.

2.7.2. System dynamic characteristics

The dynamic characteristics of the control valve and flowmeter were investigated by imposing a step input on to the valve demand signal from the computer. The dynamic responses of the valve and flowmeter are shown in Figures 2.12, 2.13.

Both the control valve and flowmeter can be approximated as second order systems with a transfer function.

$$G(s) = \frac{k\ e^{-\tau_d}}{1 + \frac{2\zeta s}{\omega_n} + \frac{s^2}{\omega_n^2}}$$

The transfer function parameters were calculated to be (55):

i) Control valve
$k = 0.356$ cm/v
$\zeta = 0.8$
$\omega_n = 0.25$ Hz
$\tau_d = 0.2$ s

ii) Flowmeter
$k = 3.22 \times 10^{-3}$ $m^3/s.V$
$\zeta = 0.4$
$\omega_n = 0.13$ Hz
$\tau_d = 2.6$ s

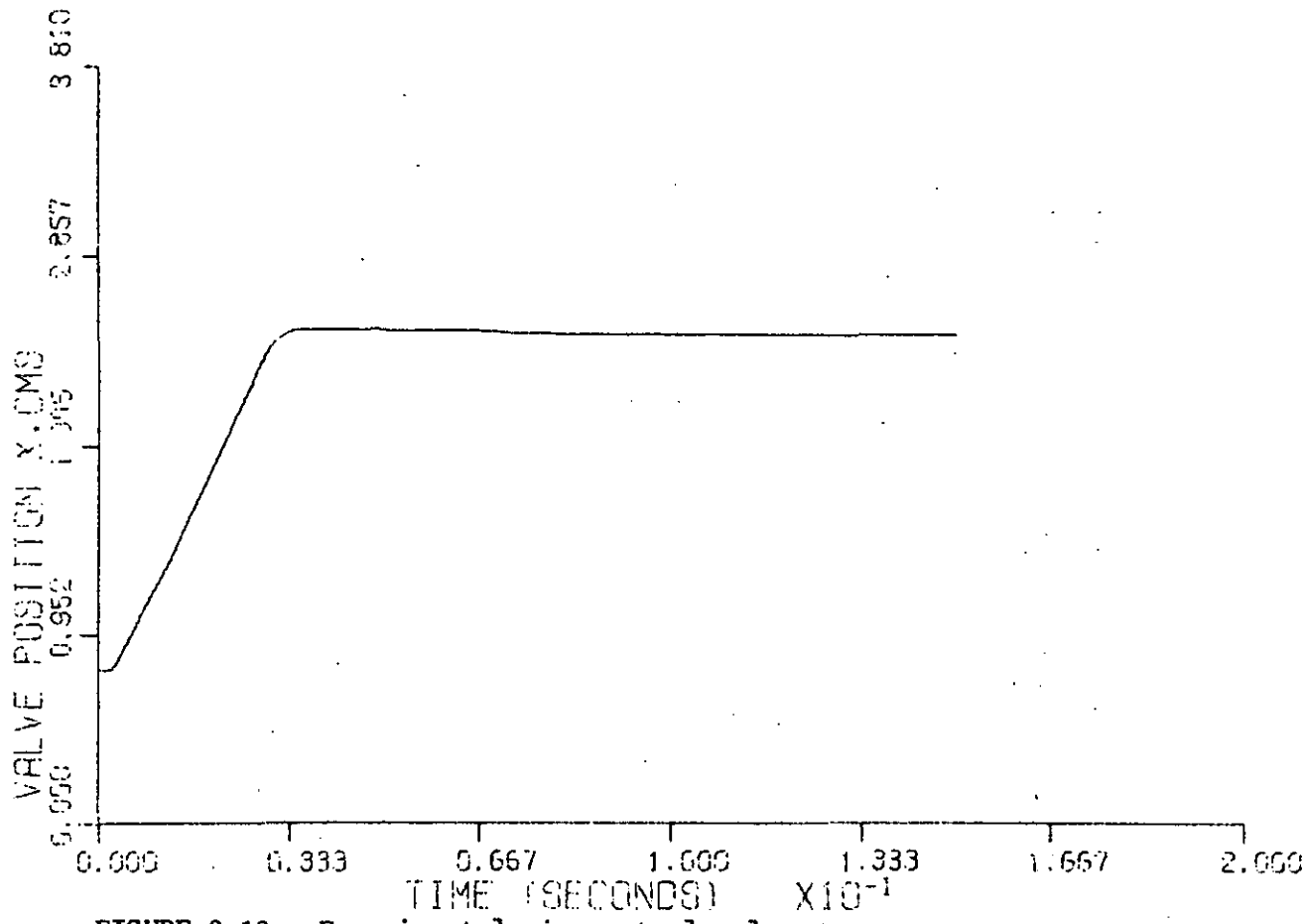


FIGURE 2.12 Experimental rig control valve step response.

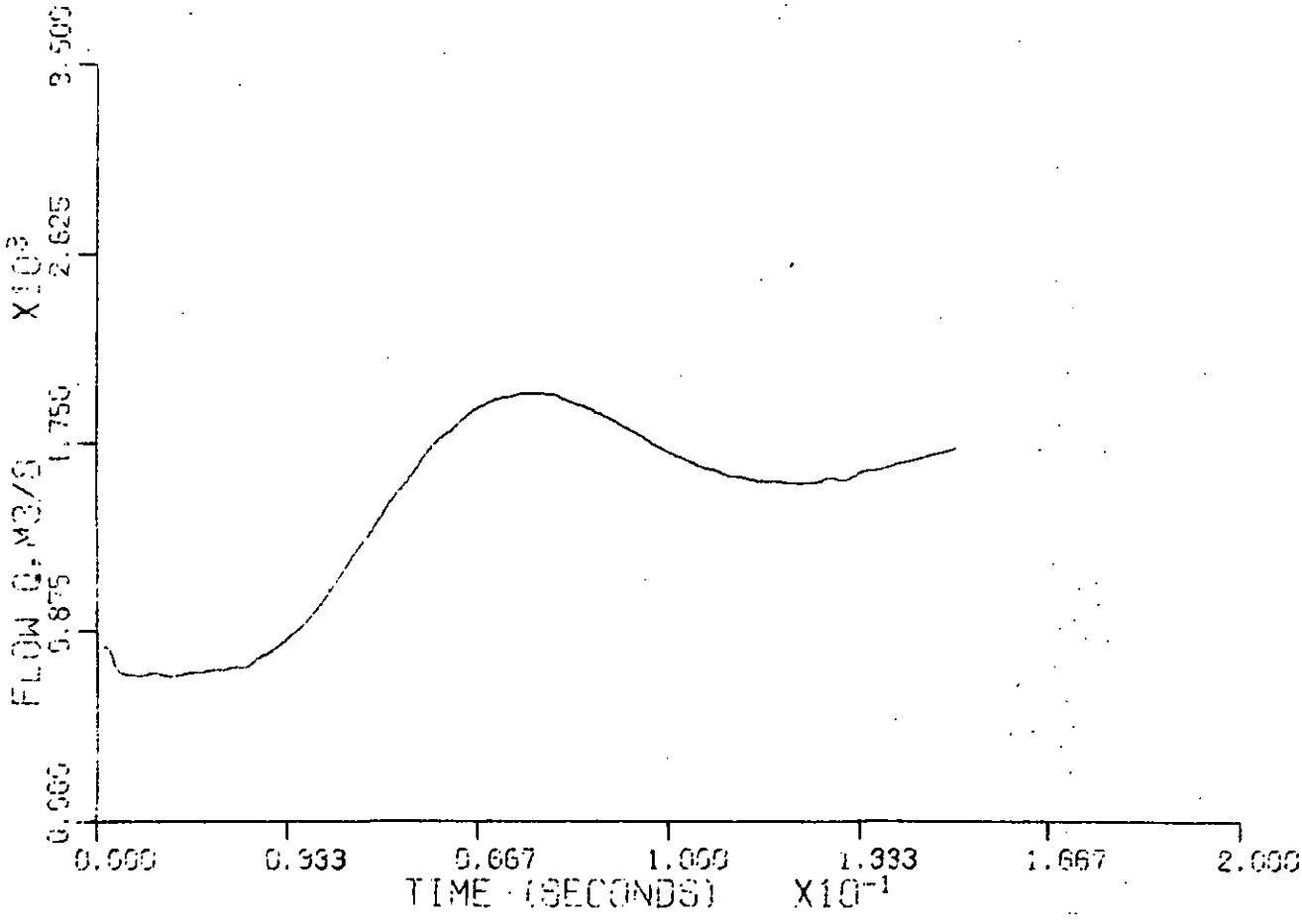


FIGURE 2.13 Experimental rig flowmeter step response.

2.7.3 Estimator design and determination of model parameters

The model describing steady state flow control loop operation was derived in section 2.4. For a Type 1 control loop equations (2.4.1) and (2.4.8) yield:

$$Q_m = k_{m1}\, s \qquad (2.7.3.1)$$

$$Q_v = k_{v1}\, x \qquad (2.7.3.2)$$

The general equation for static state estimation is:

$$\underline{y} = \underline{h}\,(\underline{x}) + \underline{v} \qquad (2.7.3.3)$$

and so equations (2.7.3.1) and (2.7.3.2) may be rewritten in this form by defining the state vector $\underline{x} = Q$, then (2.7.3.3) becomes:

$$\begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} 1/k_{v1} \\ 1/k_{v2} \end{bmatrix} Q + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

This measurement vector applies for the Type 1 and 3 control loop while the Type 2 and 4 loops are characterised by:

$$y = \begin{bmatrix} x \\ \sqrt{s} \end{bmatrix}$$

The Jacobian matrix F is defined by:

$$F = \begin{bmatrix} f_{11} \\ f_{21} \end{bmatrix}$$

with $f_{11} = \dfrac{\partial x}{\partial Q}$

$f_{21} = \dfrac{\partial s}{\partial Q}$ Type 1 and 3 control loop

$= \dfrac{\partial \sqrt{s}}{\partial Q}$ Type 2 and 4 control loop

The final parameter needed to completely define equation (2.7.3.3) is the matrix R. If the system model is perfect R is the measurement

noise covariance matrix of the form:

$$R = \begin{bmatrix} r_{11} & 0 \\ 0 & r_{22} \end{bmatrix}$$

with $r_{11} = \sigma^2_x$

$r_{22} = \sigma^2_s$ Type 1 and 3 control loop

$= \sigma^2_{\sqrt{s}}$ Type 2 and 4 control loop

The equations of the control loop and the parameters defining the measurement equation (2.7.3.3) are summarised in Table 2.6.

Having defined the structure of the system measurement and estimator equations it is necessary to determine the actual parameter values. In an industrial application it is envisaged that these parameters may be determined from plant design manuals or may be obtained experimentally. It is usually less satisfactory to obtain valve parameters from design data. This point will be discussed more fully in section 2.10 when the industrial applications of this technique are presented.

In the laboratory experiments the appropriate parameters were determined by direct calibration on the experimental rig.

Control loop type	Flowmeter equations	Control valve equation	Measurement equations (neglecting noise) $\underline{y} = \underline{h}(\underline{x})$	Measurement noise covariance matrix R		Jacobian matrix F	
				r_{11}	r_{22}	f_{11}	f_{21}
1	$Q_m = k_{m1}\, s$	$Q_v = k_{v1}\, x$	$x = Q/k_{v1}$ $s = Q/k_{m1}$	σ^2_x	σ^2_s	$1/k_{v1}$	$1/k_{m1}$
2	$Q_m = k_{m2}\sqrt{s}$	$Q_v = k_{v1}\, x$	$x = Q/k_{v1}$ $\sqrt{s} = Q/k_{m2}$	σ^2_x	$\sigma^2_{\sqrt{s}}$	$1/k_{v1}$	$1/k_{m2}$
3	$Q_m = k_{m1}\, s$	$Q_v = k_{v3}\exp(k_{v4}x)$	$x = \ln Q/k_{v4} - \ln k_{v3}/k_{v4}$ $s = Q/k_{m1}$	σ^2_x	σ^2_s	$1/k_{v4}Q$	$1/k_{m1}$
4	$Q_m = k_{m2}\sqrt{s}$	$Q_v = k_{v3}\exp(k_{v4}x)$	$x = \ln Q/k_{v4} - \ln k_{v3}/k_{v4}$ $\sqrt{s} = Q/k_{m2}$	σ^2_x	$\sigma^2_{\sqrt{s}}$	$1/k_{v4}Q$	$1/k_{m2}$

TABLE 2.6. Equations and parameters defining the flow control loop estimator.

The control valve characteristic and orifice plate constant were determined by "stroking" the valve; noting the valve position and flowmeter signal and manually measuring the system flowrate. Figure 2.14 shows the resultant characteristic for a control valve with a linear trim. In fact three curves are shown. Curve 1 represents the experimental results obtained from valve stroking. Curve 2 is the model curve obtained by estimating the valve parameters (in this case a model of the form $Q_v = k_{v1} + k_{v2}\, x$ was used) from the experimental curve. The curve is actually fitted by two straight lines to be consistent with the formulation of a linear control valve. There are different parameters for each line. Curve 3 is the result of using the state estimator of equation (2.6.15). The state estimator of equation (2.6.15) used the experimentally determined parameters k_{v1}, k_{v2} and k_{m2} and the matrix R.

In this formulation the matrix R plays the role of a weighting matrix which causes the state estimate $\hat{\underline{x}}$ to fit the "good" measurements closer than the "bad" measurements. If the system model is accurate it has been pointed out that R is the covariance matrix of the measurement noise. However, if the system model is imperfect, then R is more properly regarded as a weighting matrix which takes into account not only measurement noise but inaccuracies in the system model. The larger the noise and model inaccuracies, the larger the elements of R.

In the present work R was determined by optimising directly on the process. It was assumed that R was a diagonal matrix. An initial guess of R was made and using the model parameters described above the static state estimator of equation (2.6.15) was used to estimate $\hat{\underline{x}}$. Now section 2.6.2 has discussed how $\hat{J}$ may be regarded

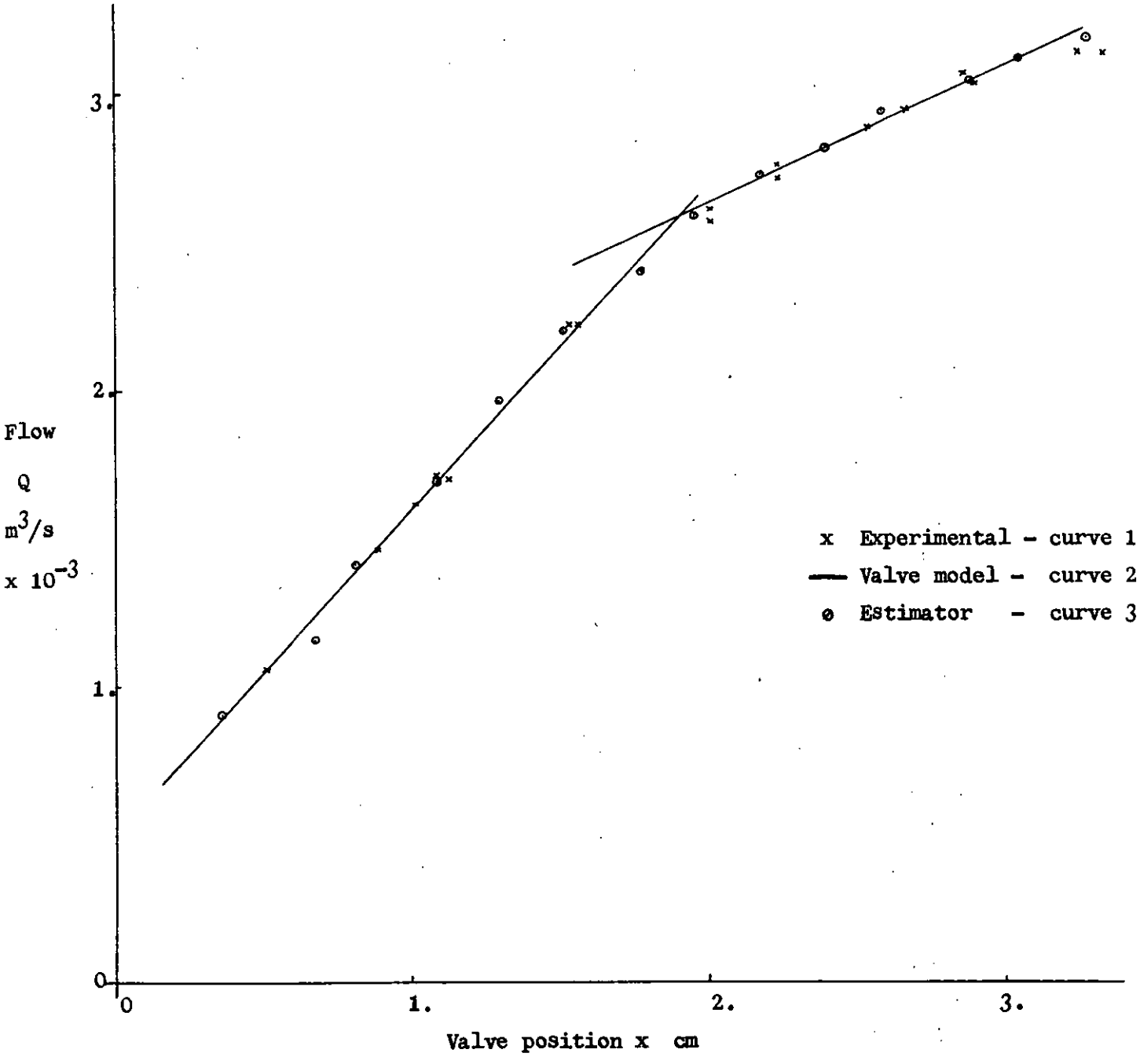


FIGURE 2.14 Linear control valve characteristics: experimental, model and estimator curves.

as a measurement of estimator performance. The matrix R was thus chosen to minimise $\hat{J}$.

The system model ($\underline{h}$ ($\underline{x}$) and R) are now fully defined and the relevant parameters are summarised in Table 2.7 for all four types of control loop.

Figure 2.15 presents the corresponding curves for an equal percentage control valve trim.

2.8 Malfunction detection experiments

The main experiments consisted of three series of runs: open-loop checks with full valve travel and open and closed-loop checks with limited valve travel. All the experiments were performed with the experimental rig in the "constant" system pressure drop characteristic mode, i.e. with isolation valves 1, 3, 5, 6, 8 open and 2, 4, 7 closed. This resulted in an approximately constant control valve pressure drop characteristic as shown in Figure 2.11.

All four types of control loop were investigated. The control valve characteristics were obtained by using linear and equal percentage valve trims. The flowmeter characteristics were obtained by using an orifice plate and differential pressure transmitter throughout the experiments but linearising the signal in the computer to simulate the linear flowmeter case.

The valve position was obtained in two different ways. In some experiments the valve stem position was measured by a linear potentiometer. In other experiments the valve position was taken as the valve demand signal from the computer. In the case where the valve

Control loop type	Flowmeter constant		Control valve constants				R matrix			
	$k_{m1}\ \frac{m^3}{sV}$	$k_{m2}\ \frac{m^3}{sV^{\frac{1}{2}}}$	Valve stem position measured		Valve demand signal measured		Valve stem position measured		Valve demand signal measured	
			$k_{v1}\ \frac{m^3}{s}$	$k_{v2}\ \frac{m^3}{s\ c_m}$	$k_{v1}\ \frac{m^3}{s}$	$k_{v2}\ \frac{m^3}{s\ V}$	r_{11}	r_{22}	r_{11}	r_{22}
1	Simulated	—	0.515×10^{-3} 1.705×10^{-3}	1.092×10^{-3} 0.465×10^{-3}	0.663×10^{-3} 1.8×10^{-3}	0.39×10^{-3} 0.158×10^{-3}	0.0645	0.01	0.25	0.01
2	—	1.393×10^{-3}	"	"	"	"	"	"	"	"
			$k_{v3}\ \frac{m^3}{s}$	$k_{v4}\ cm^{-1}$	$k_{v3}\ \frac{m^3}{s}$	$k_{v4}\ V^{-1}$	r_{11}	r_{22}	r_{11}	r_{22}
3	Simulated	—	0.463×10^{-3}	0.522	0.463×10^{-3}	0.522	0.0645	0.01	0.25	0.01
4	—	1.393×10^{-3}	"	"	"	"	"	"	"	"

TABLE 2.7. Numerical values of parameters in flow control loop estimator.

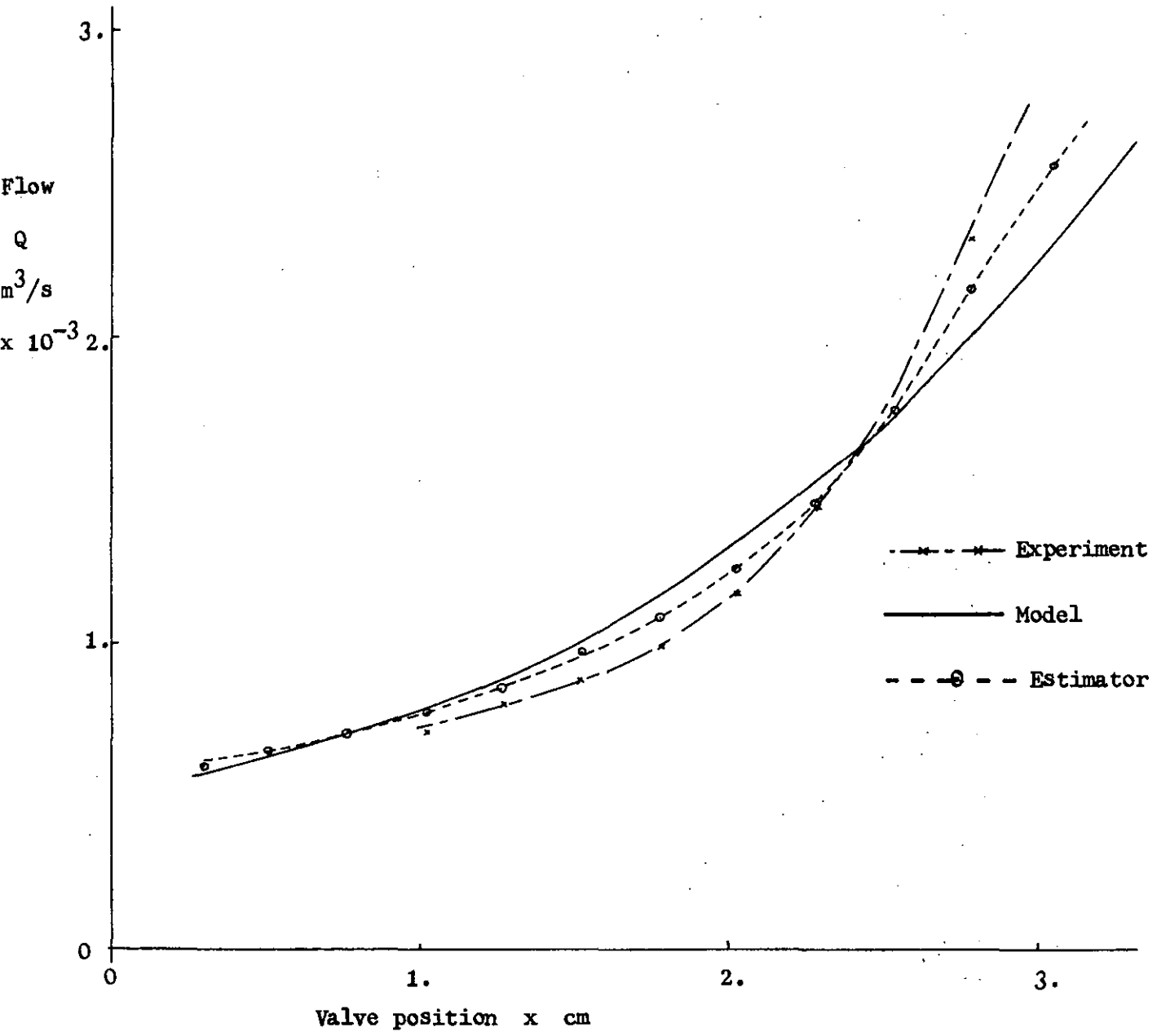


FIGURE 2.15 Equal percentage control valve characteristic: experimental, model and estimator curves.

stem position measurement was taken, errors in this measurement were investigated. The method for the detection of malfunction in the rest of the loop does not depend, however, on a valve stem position measurement.

The form of the estimator, the algebraic parameters and their numerical values have been given in Tables 2.6 and 2.7.

The malfunctions introduced into the system are summarised in Table 2.8 where they are designated by a code letter A-G. The malfunctions were mostly obtained by making appropriate adjustments of the instrument although some experiments involved the introduction of simulated errors by the computer, marked with an asterisk.

One of the malfunctions investigated was that of a damaged valve trim (malfunction G). Figure 2.16 shows the standard equal percentage valve trim while Figure 2.17 shows the damaged trim used in the experiments.

The actual experimental procedure for each particular series of runs is now presented.

2.8.1 Open loop experiments with full valve travel

The first series of runs consisted of open-loop experiments with full valve travel and are designated by the code letter FT. In these experiments the valve was moved by a computer based random number generator to 100 openings over its whole range of travel. At each particular opening the valve was held steady for 25 seconds, the measurement vector $\underline{y}$ recorded and the state estimate $\hat{\underline{x}}$ obtained from the static state estimator of equation (2.6.15).

Experiments were performed on all four types of control loop.

Loop malfunction	Source	Code
Flowmeter zero error	Flowmeter differential pressure transducer	A
Flowmeter range error	Flowmeter pressure/current converter	B
Control valve zero error	Current/pressure converter	C
Control valve range error	Current/pressure converter	D
Control valve zero error	Valve stem linear potentiometer	E
Control valve range error	Valve stem linear potentiometer	F*
Damaged control valve	Valve trim plug	G

TABLE 2.8. Malfunctions introduced into laboratory flow control rig.

* denotes computer simulated error.

FIGURE 2.16 Equal percentage valve trim.

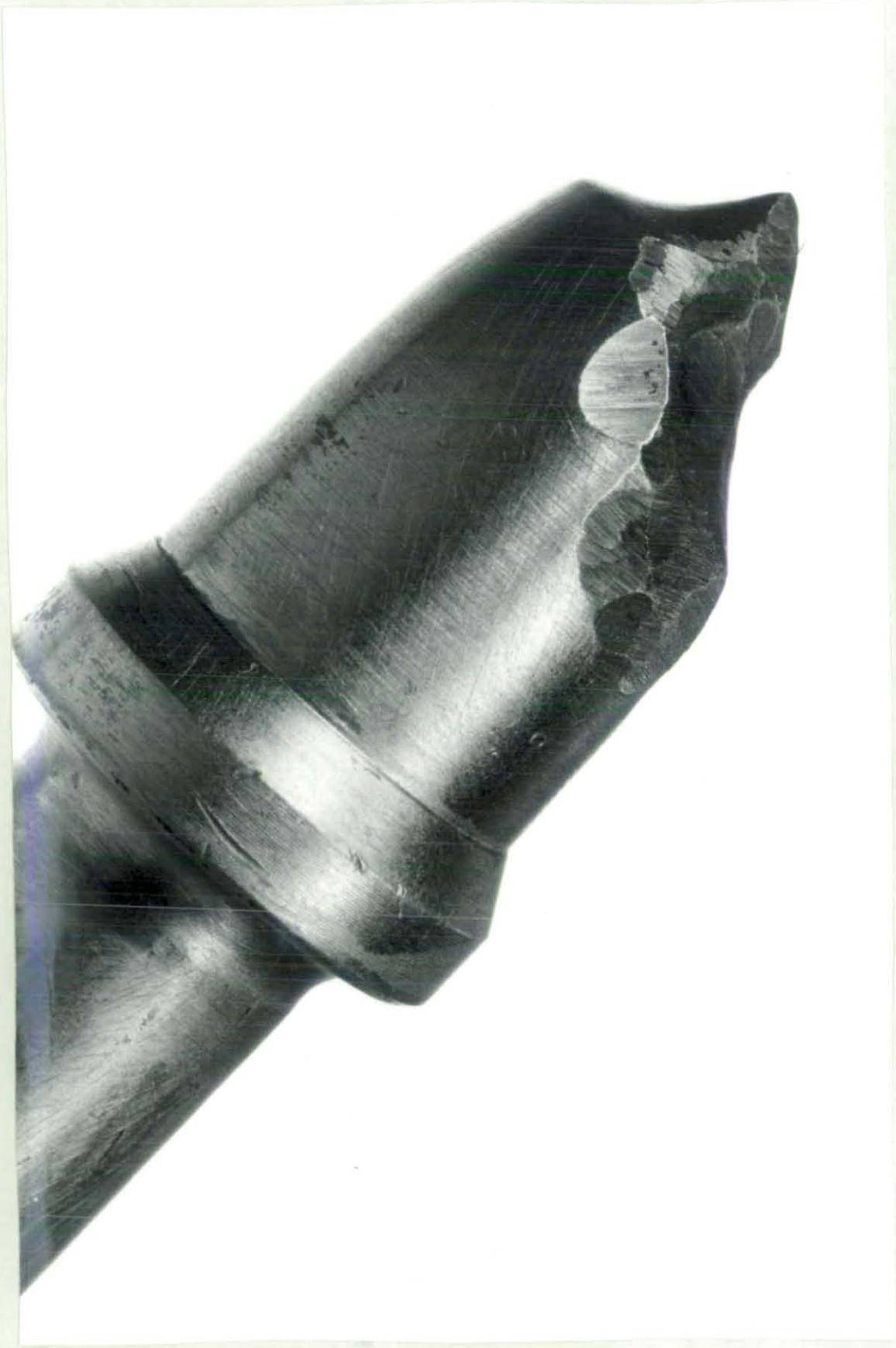


FIGURE 2.17 Damaged equal percentage valve trim.

2.8.2 Open loop experiments with limited valve travel

In this series of experiments the stroking of the valve was not allowed.

The experimental runs were performed on Type 2 and 4 control loops and are coded OL. During the experiments the valve was held at a particular opening. In order to make the system rather more noisy and more realistic, a small pseudo-random binary sequence (PRBS) as shown in Figure 2.18a was imposed on the valve demand in the computer. Figures 2.18 b and c show some typical recorded time series of the valve stem position and flowmeter signal.

The state estimate $\hat{\underline{x}}$ was obtained from the tracking state estimator of equation (2.6.1.2).

2.8.3 Closed loop experiments with direct digital control

The third series of experiments consisted of closed-loop checks with limited valve travel on a Type 4 control loop, denoted CL. In these experiments the control loop was closed under direct digital control from the computer, as shown in Figure 2.2b. The flow was controlled at a particular setpoint and the valve moved over a limited range. The control loop signals obtained are shown in Figures 2.19a and b, which in fact shows a low amplitude oscillation. This arises because of the slight oscillation in the orifice differential pressure transducer output.

The d.d.c. flow control loop employed a proportional plus integral controller with a 3 second sample time. At selected setpoints the state estimate $\hat{\underline{x}}$ was obtained from the tracking state estimator of equation (2.6.1.2).

The complete experimental programme is summarised in Table 2.9.

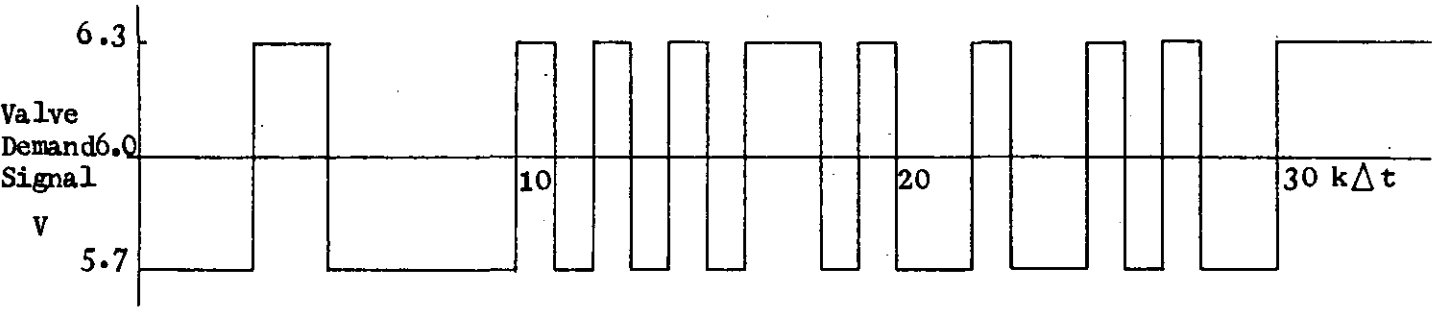


FIGURE 2.18a Pseudo random binary sequence used in open loop limited travel experiments.

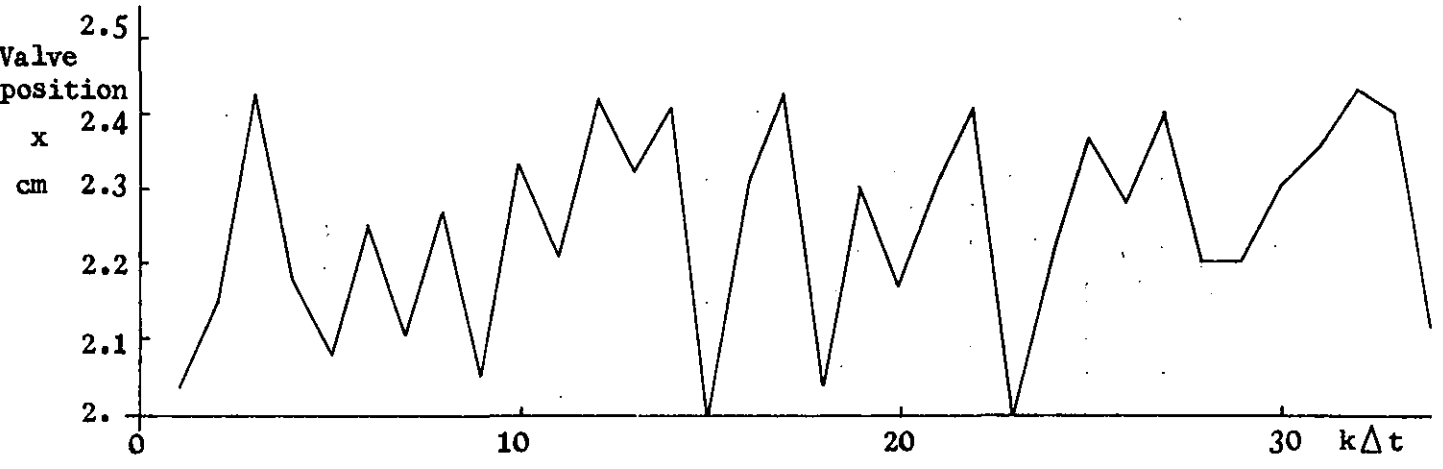
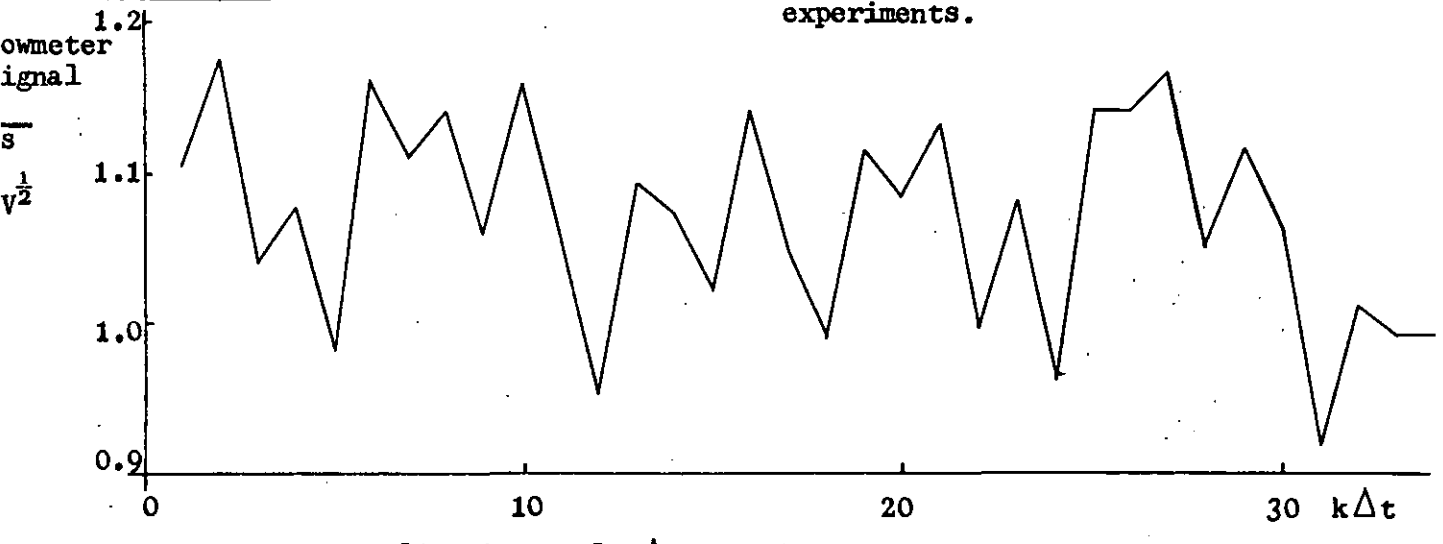


FIGURE 2.18b Typical valve stem position measurement in open loop limited travel experiments.



Sampling interval $\Delta t = 0.5$ s

FIGURE 2.18c Typical flowmeter signal measurement in open loop limited travel experiments.

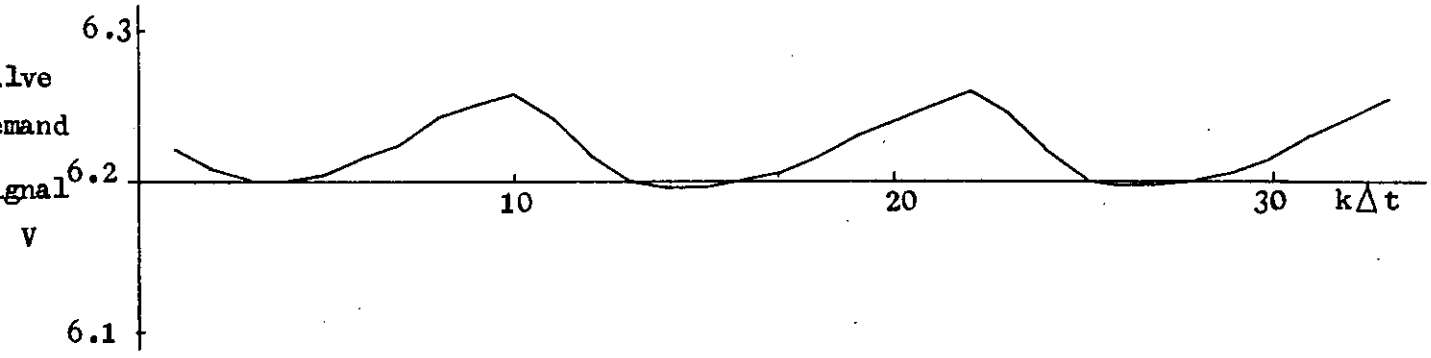


FIGURE 2.19a Typical valve demand signal in closed loop direct digital control experiments.

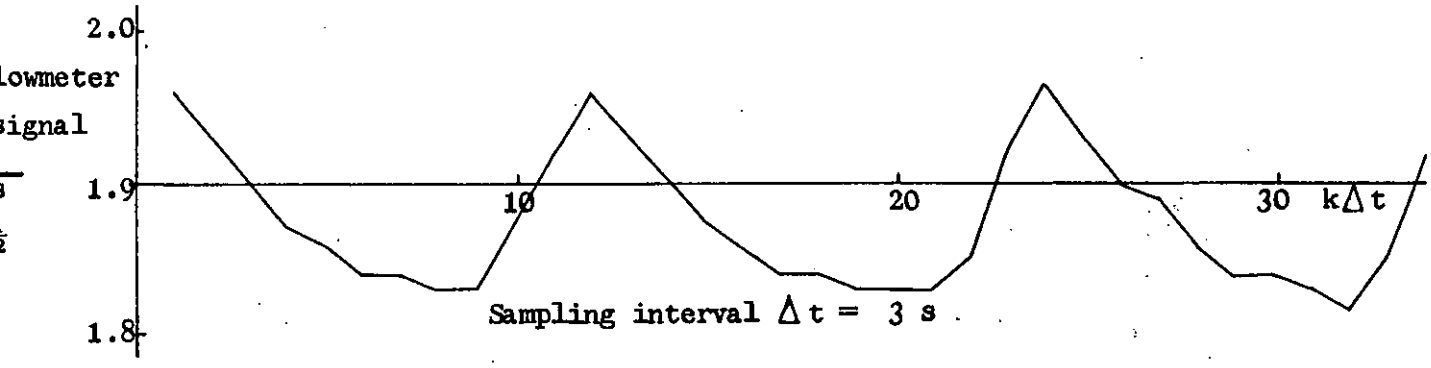


FIGURE 2.19b Typical flowmeter signal measurement in closed loop direct digital control experiments.

Control loop type	Valve stem position measured	Experiment type	Run number	Malfunction
1	Yes	FT	FT/1A *	A + 10%
	"	"	FT/1B *	B - 10%
2	Yes	FT	FT/2A a	A + 10%
	"	"	FT/2A b	A - 10%
	No	"	FT/2B	B - 20%
	"	"	FT/2C a	C + 10%
	"	"	FT/2C b	C - 10%
	"	"	FT/2D	D + 20%
	Yes	"	FT/2E	E + 5%
	"	"	FT/2F *	F + 15%
	"	OL	OL/2A	A + 10%
	No	"	OL/2B	B - 15%
	"	"	OL/2C	C + 10%
3	Yes	FT	FT/3A a *	A + 10%
	"	"	FT/3A b	A - 10%
	"	"	FT/3B a	B + 10%
	"	"	FT/3B b	B - 10%
4	Yes	FT	FT/4A a	A + 5%
	"	"	FT/4A b	A - 5%
	No	"	FT/4B	B - 10%
	"	"	FT/4C a	C + 15%
	"	"	FT/4C b	C - 15%
	"	"	FT/4D a *	D + 10%
	"	"	FT/4D b *	D - 10%
	Yes	"	FT/4E *	E + 8%
	"	"	FT/4F	F + 10%
	"	"	FT/4G	-
	"	OL	OL/4A a	A + 10%
	"	"	OL/4A b	A - 10%
	"	"	OL/4B	B - 10%
	No	"	OL/4C a	C + 10%
	"	"	OL/4C b	C - 6%
	"	"	OL/4D	D + 10%
	Yes	"	OL/4E	E + 10%
	"	"	OL/4F *	F - 5%
	"	"	OL/4G	-
	No	CL	CL/4A	A + 17%
	"	"	CL/4C a	C + 20%
	"	"	CL/4C b	C + 13%
	"	"	CL/4G	-

* Experiments involving computer simulation.

TABLE 2.9. Experiments performed on laboratory flow control rig.

2.9 Malfunction detection experimental results

2.9.1 Open loop experiments with full valve travel

In these experiments two types of check which indicate loop malfunction are possible. Since the valve was 'stroked' over its entire range of travel the estimator valve characteristic could be obtained. This may be then compared with the original malfunction free characteristic.

An example of this check for a Type 1 control loop without malfunction and with a + 10% flowmeter zero error (run FT/1A*) is shown in Figure 2.20. With this type of check a visual display of the malfunction is possible as originally postulated in section 2.4.

An alternative method for security monitoring is based upon the information obtained from the estimator.

The estimator residual process was defined in equation (2.6.2.2). As the valve was 'stroked', the residual time series could be generated, which also corresponded to different valve openings. This may then be compared to an original malfunction-free residual time series and the changes in the individual residuals monitored.

Various techniques for examining the residual process were discussed in sections 2.6.2 and 2.6.3. In these experiments it was found that it is necessary only to monitor the residual mean. Since a malfunction causes both the valve position and flowmeter signal to deviate from their nominal malfunction free characteristics it is sufficient to monitor only one of the residual sequences and this was arbitrarily chosen as the valve position residual.

In each experiment the means of the valve position residual without

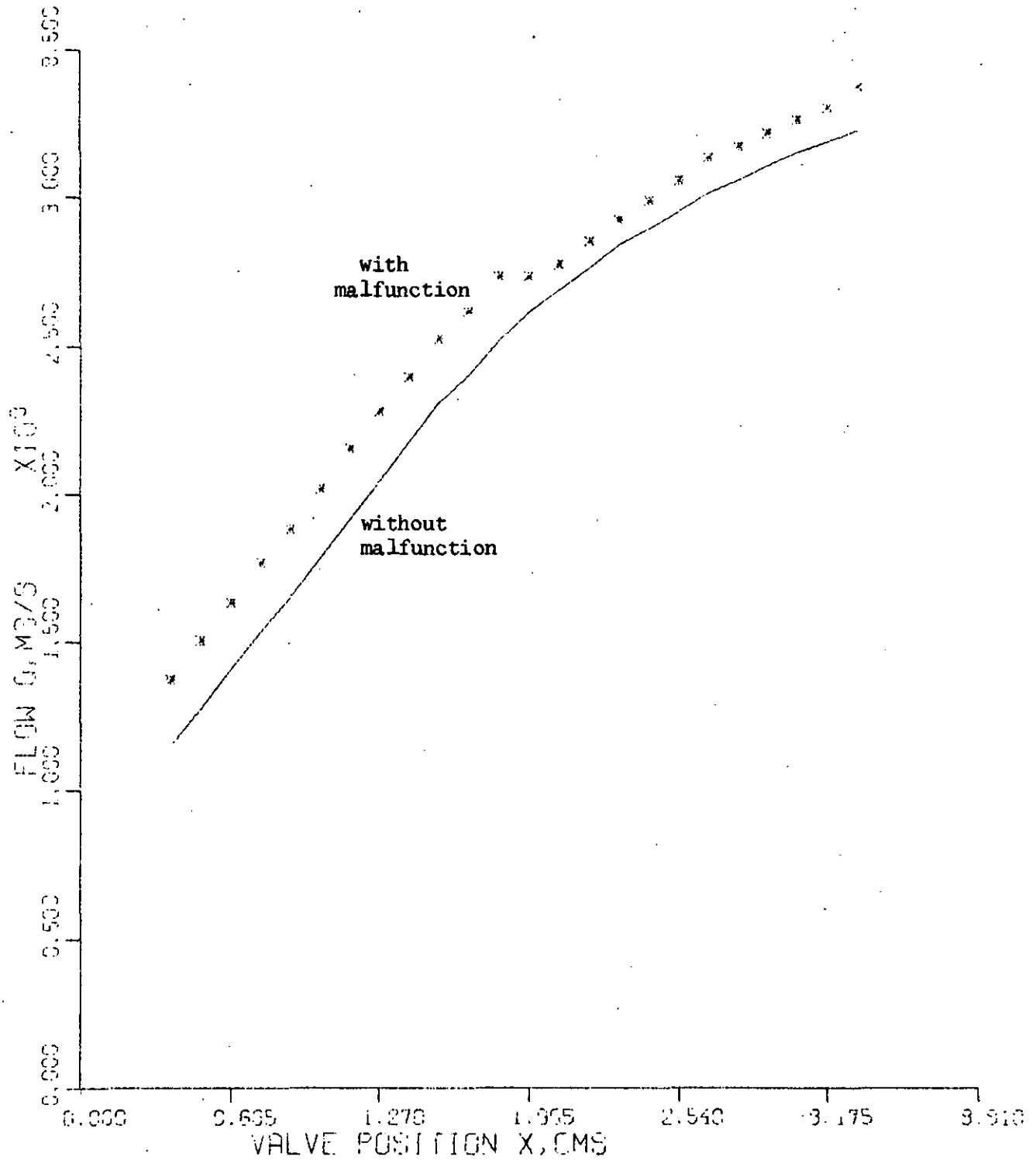


FIGURE 2.20 **Full travel check on Type 1 control loop - run FT/1A***

and with malfunction were determined. The modulus of the change in the means of the residuals were calculated and compared with the significant change at the 95% confidence limit according to Student's t test.

Figures 2.21 a and b illustrate the valve position and flowmeter signal residual time series for a Type 1 control loop without malfunction. In fact, these Figures show normalised residuals defined as:

$$r_i = \frac{y_i - \underline{h}_i(\hat{\underline{x}})}{r_{ii}^{\frac{1}{2}}}$$

A + 10% flowmeter zero error in a Type 1 control loop yields residual sequences as shown in Figures 2.22 a and b.

The statistical characteristics of the normalised valve position residual sequence for these runs are given in Table 2.10.

Run number	Normalised position residual statistics		Modulus of change in mean of normalised valve position residual		Figure
	Mean	Variance	Actual change	Significant change at 95% limit	
-	- 0.46	0.093	-	-	2.21 a
FT/1A*	- 1.37	0.076	0.91	0.18	2.22 a

TABLE 2.10. Behaviour of valve position residual for + 10% flowmeter zero error in Type 1 control loop.

The results of the experimental programme for the open loop tests with full valve travel, detailed in Table 2.8, are given in Figures 2.20 to 2.45. The modulus of the change in valve position residual means are summarised in Table 2.11.

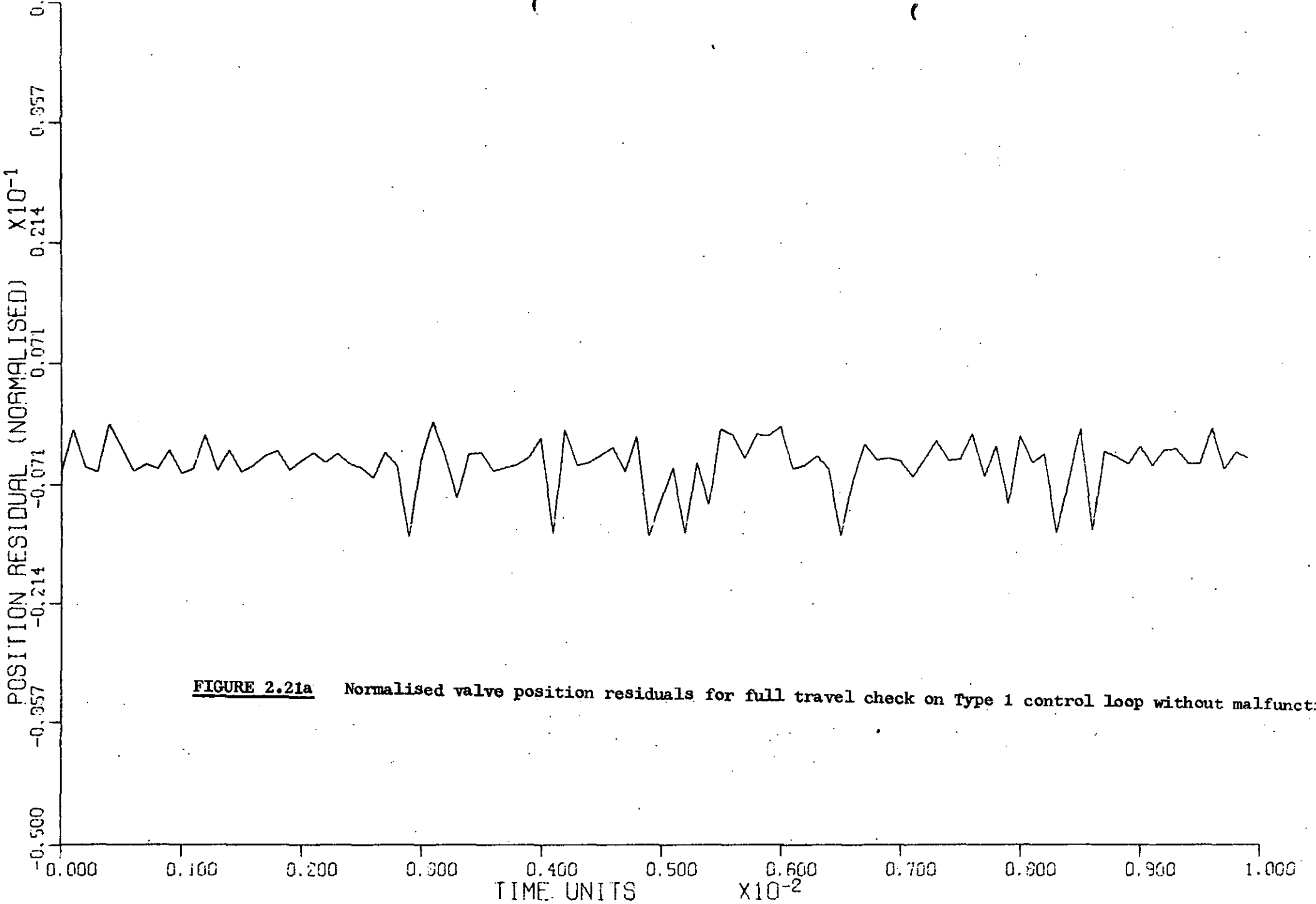


FIGURE 2.21a Normalised valve position residuals for full travel check on Type 1 control loop without malfunction.

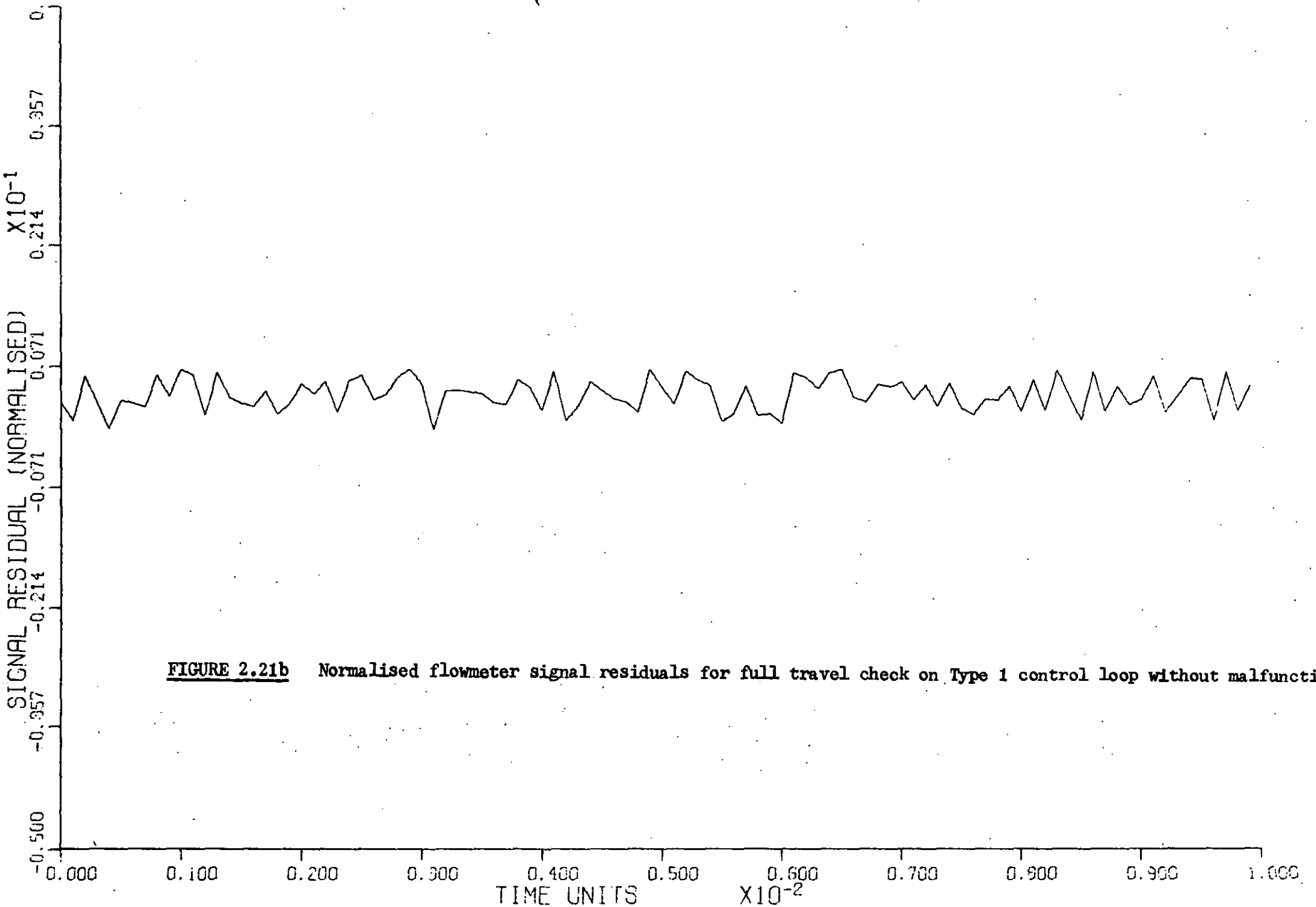


FIGURE 2.21b Normalised flowmeter signal residuals for full travel check on Type 1 control loop without malfunction.

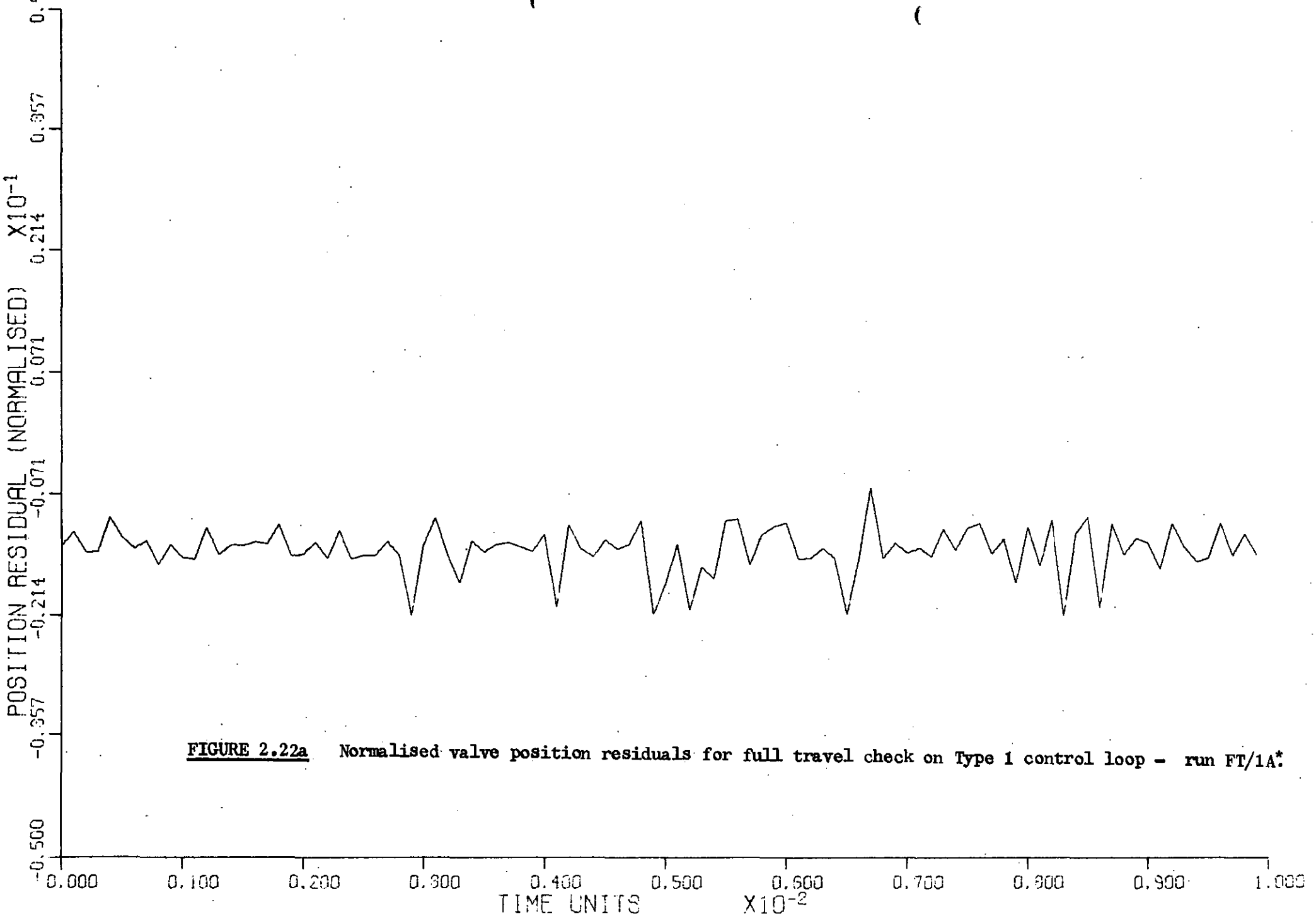


FIGURE 2.22a Normalised valve position residuals for full travel check on Type 1 control loop - run FT/1A*.

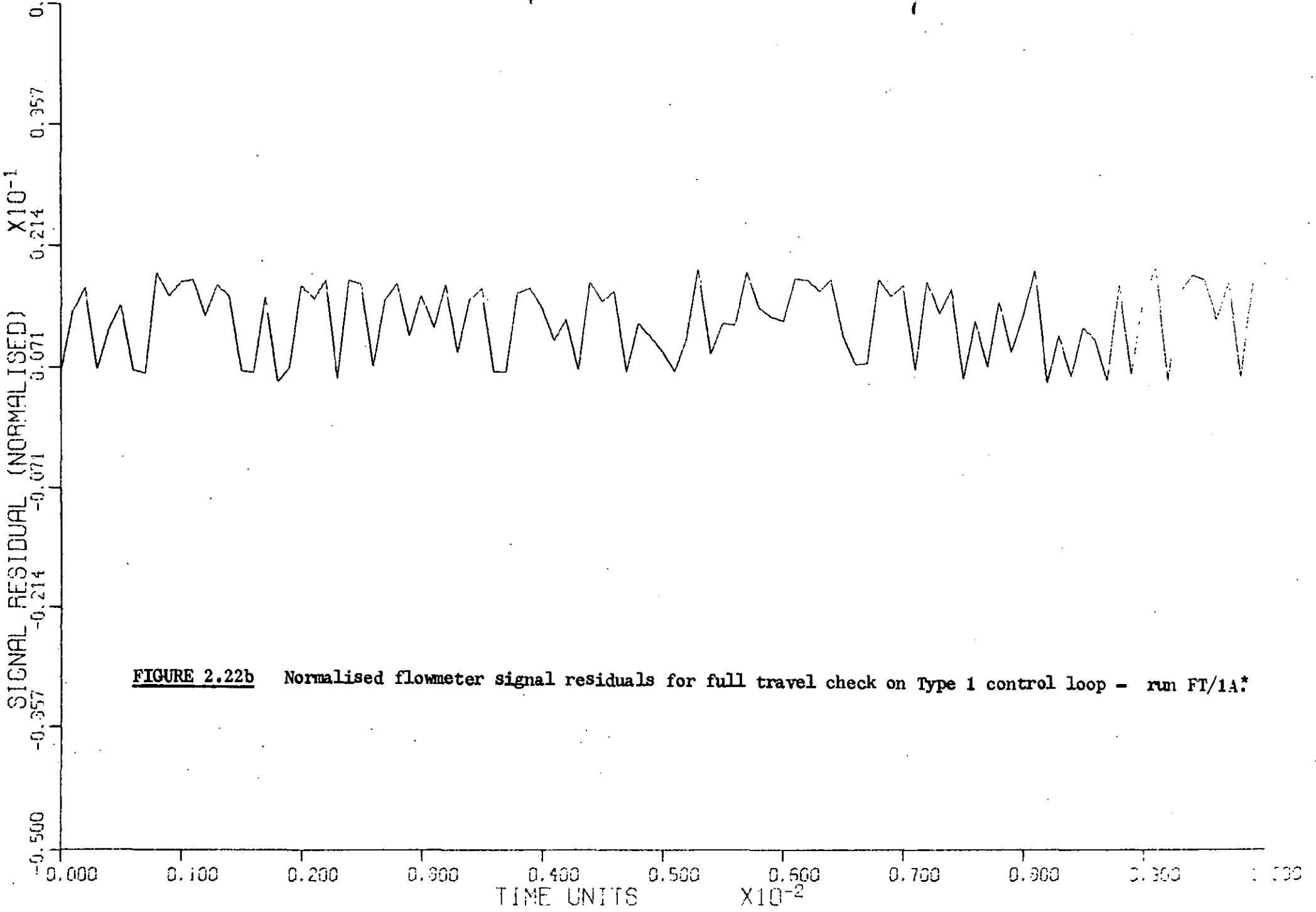


FIGURE 2.22b Normalised flowmeter signal residuals for full travel check on Type 1 control loop - run FT/1A.*

Control loop type	Run number	Modulus of change in mean of normalised valve position residual		Figure
		Actual change	Significant change	
1	FT/1A *	0.91	0.18	2.20
	FT/1B *	0.88	0.14	2.23
2	FT/2A a	1.64	0.11	2.24
	FT/2A b	0.98	0.08	2.25
	FT/2B	1.45	0.15	2.26
	FT/2C a	0.65	0.16	2.27
	FT/2C b	0.83	0.17	2.28
	FT/2D	0.36	0.20	2.29
	FT/2E *	0.43	0.06	2.30
	FT/2F	1.05	0.40	2.31
3	FT/3A a *	0.86	0.14	2.32
	FT/3A b	0.84	0.16	2.33
	FT/3B a	0.42	0.16	2.34
	FT/3B b	0.59	0.15	2.35
4	FT/4A a	1.39	0.17	2.36
	FT/4A b	0.94	0.23	2.37
	FT/4B	0.74	0.19	2.38
	FT/4C a	1.36	0.23	2.39
	FT/4C b	1.28	0.21	2.40
	FT/4D a *	0.72	0.21	2.41
	FT/4D b *	0.58	0.19	2.42
	FT/4E	0.42	0.15	2.43
	FT/4F *	0.79	0.15	2.44
	FT/4G	0.11	0.22	2.45

TABLE 2.11. Behaviour of valve position residual for open loop full travel experiments.

The experimental results of Figures 2.20 - 2.45 confirm the initial method of control loop malfunction detection by the valve position - flow check proposed in section 2.3. The deviations in estimator valve characteristics from the nominal malfunction free characteristics agree with the simulated results of Figures 2.4 - 2.7.

The valve characteristics for a Type 1 and Type 2 control loop shown in Figures 2.20 - 2.31 all show a discontinuity in the deviation due to the system malfunction for $x > 1.9$ cm or $x > 5$ volt (depending upon the process measurement of valve position used in the experiment).

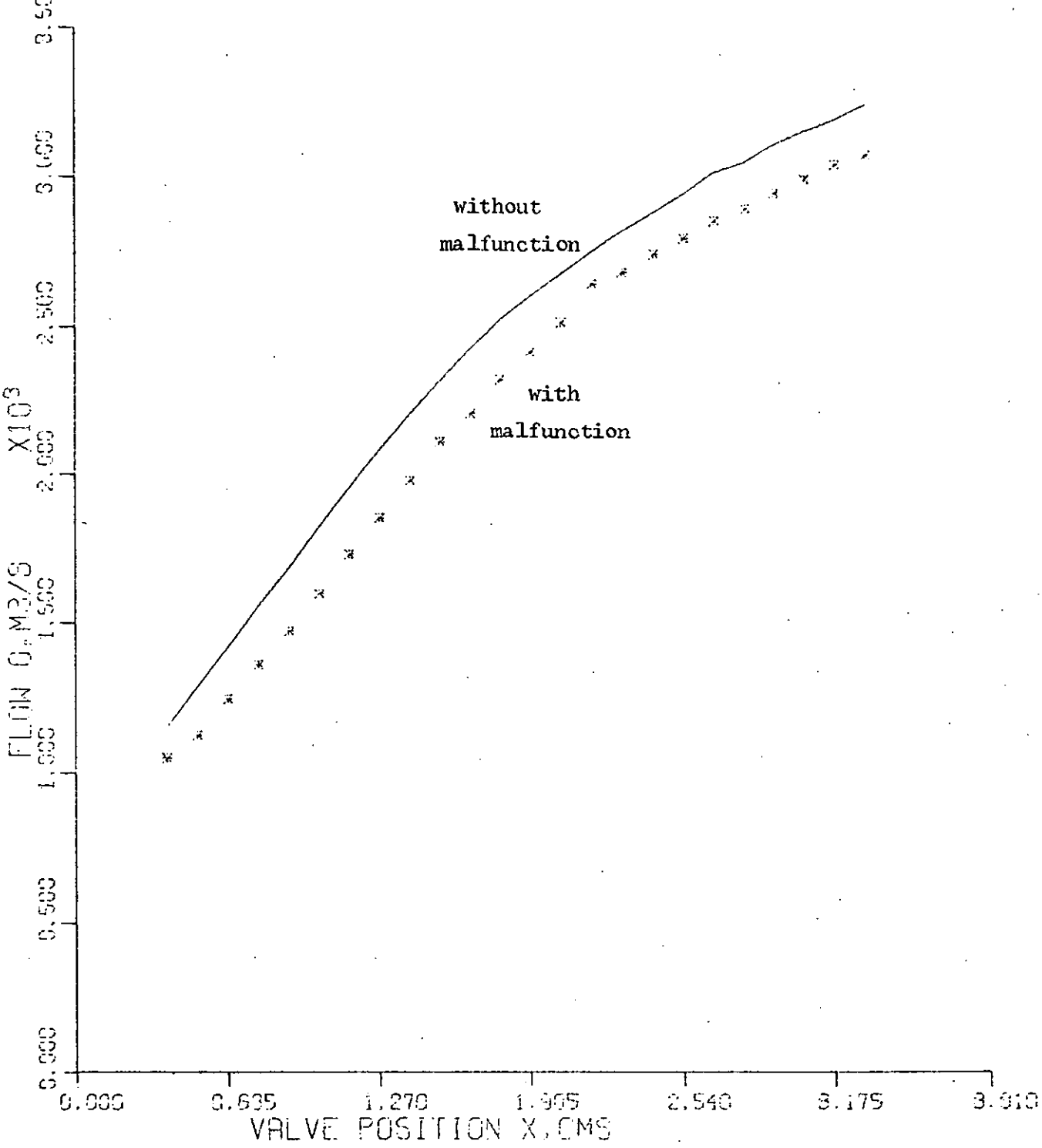


FIGURE 2.23 Full travel check on Type 1 control loop - run FT/1B.*

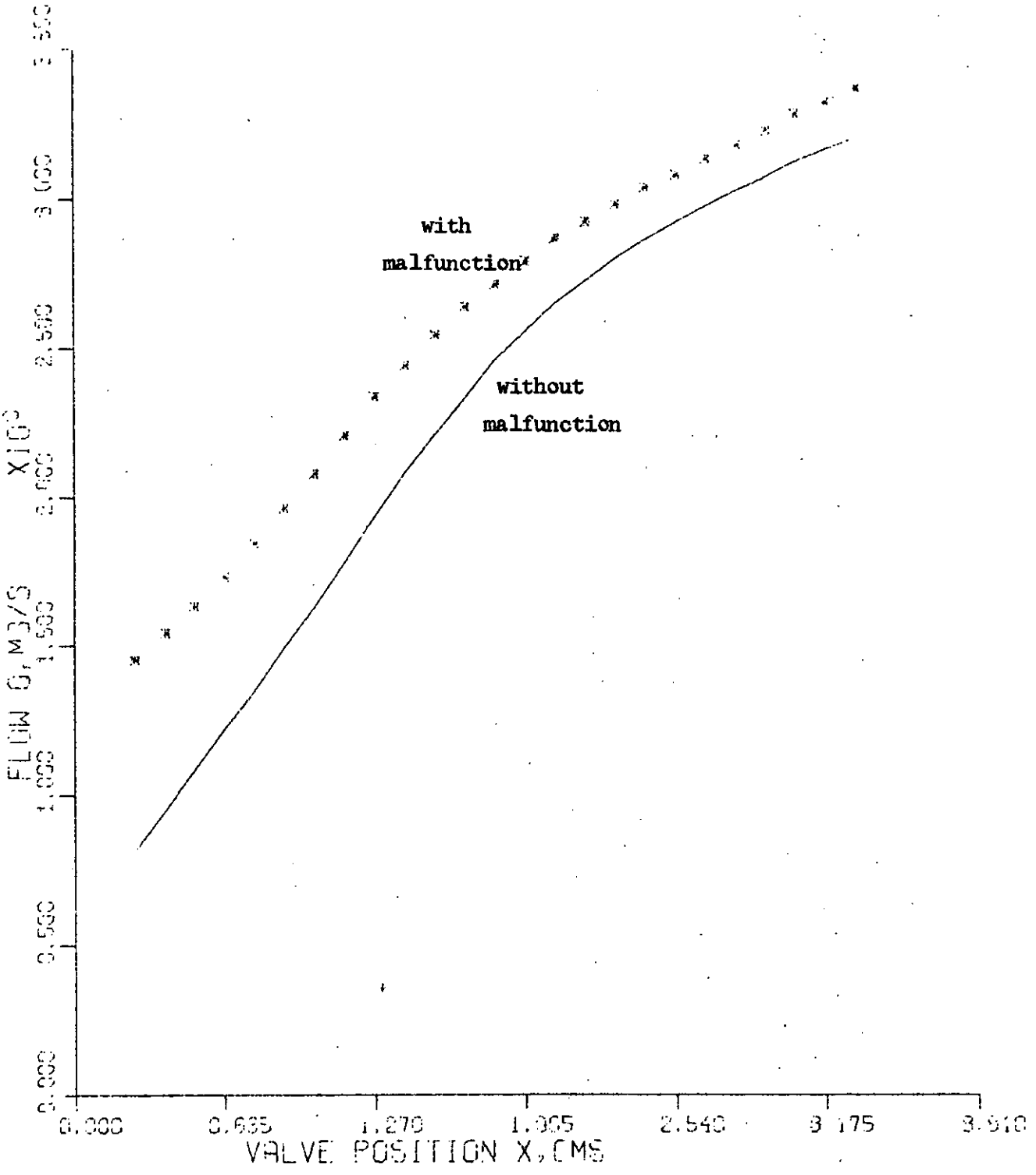


FIGURE 2.24 Full travel check on Type 2 control loop - run FT/2Aa.

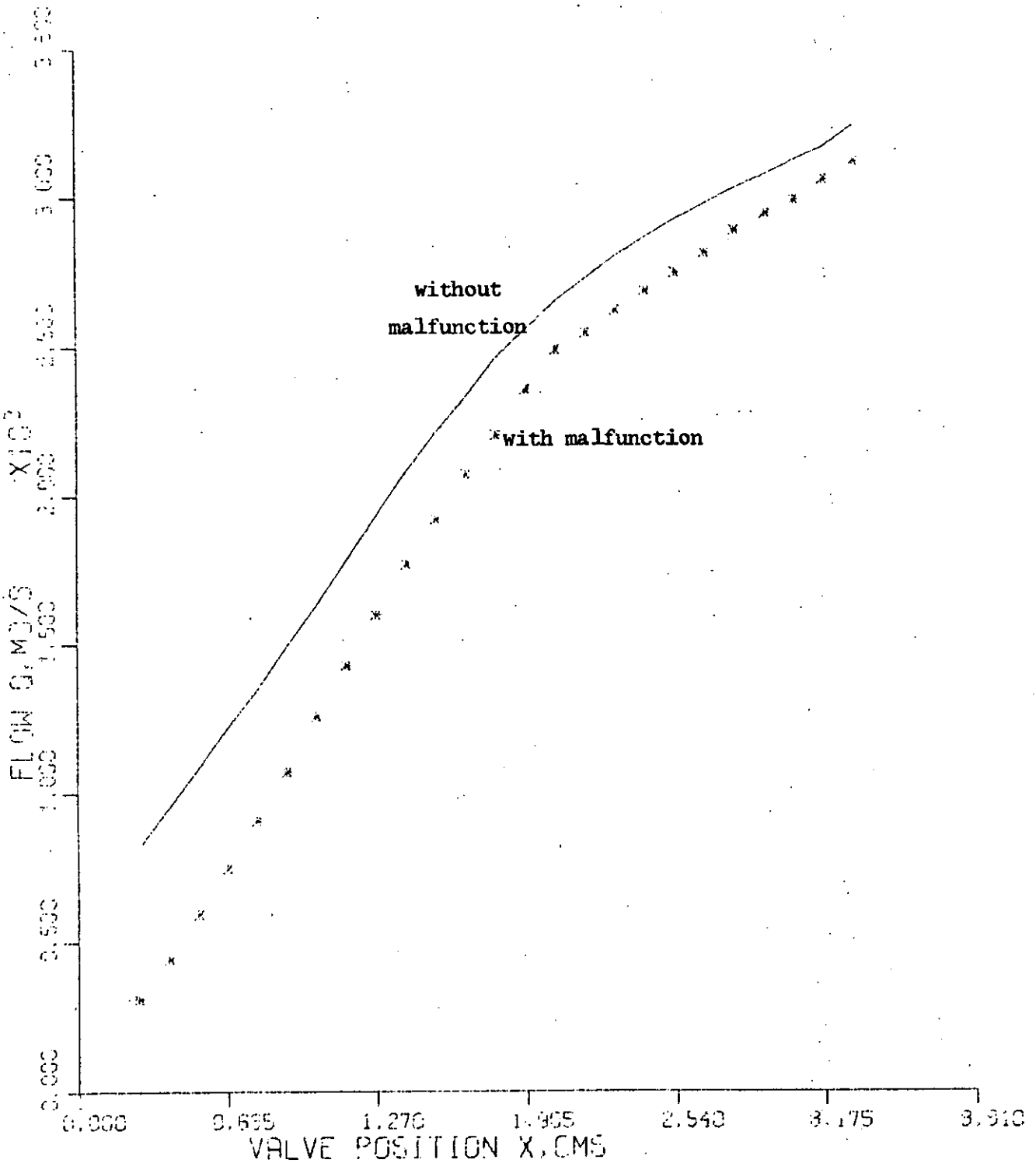


FIGURE 2.25 Full travel check on Type 2 control loop - run FT/2Ab.

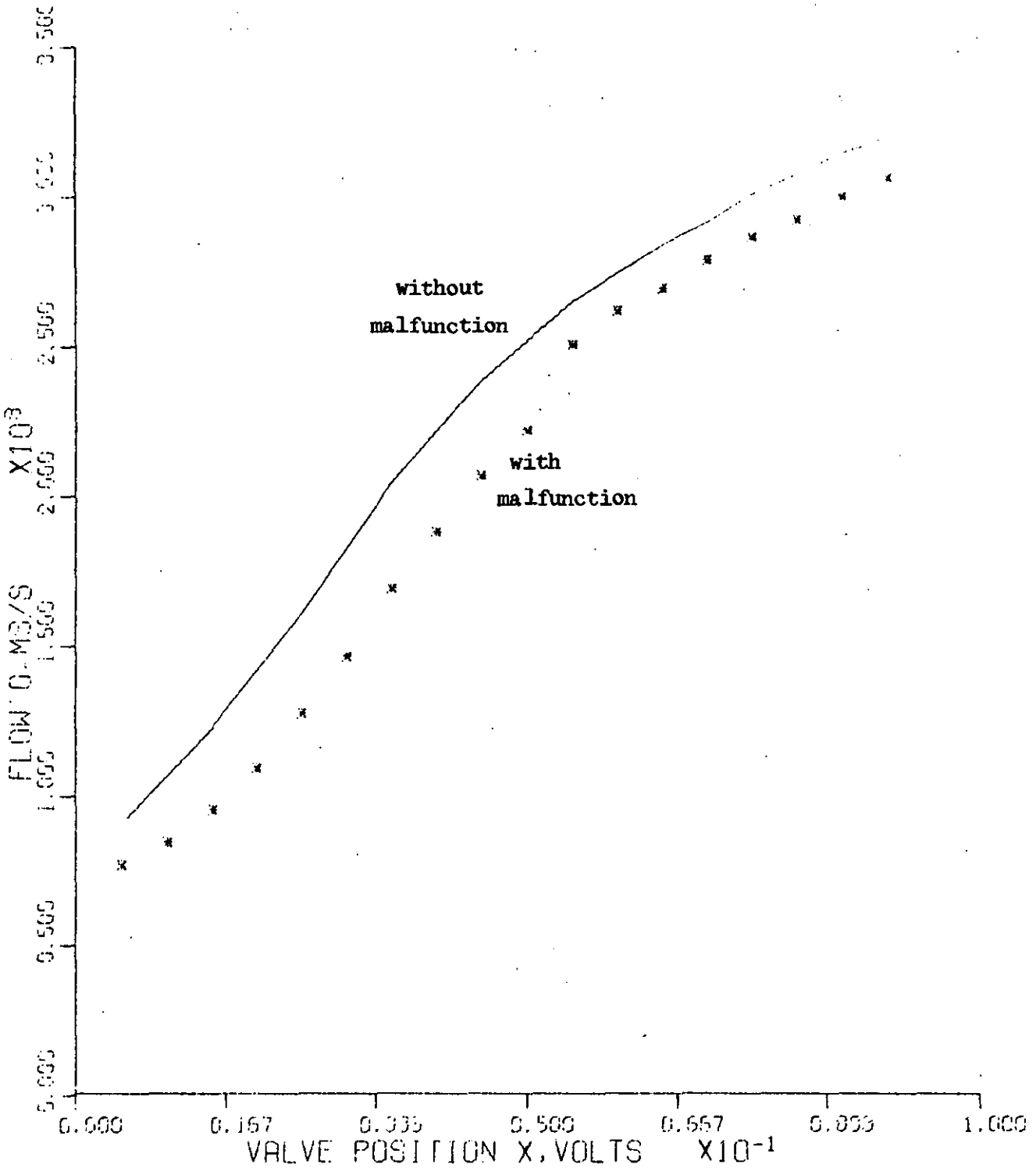


FIGURE 2.26 Full travel check on Type 2 control loop - run FT/2B.

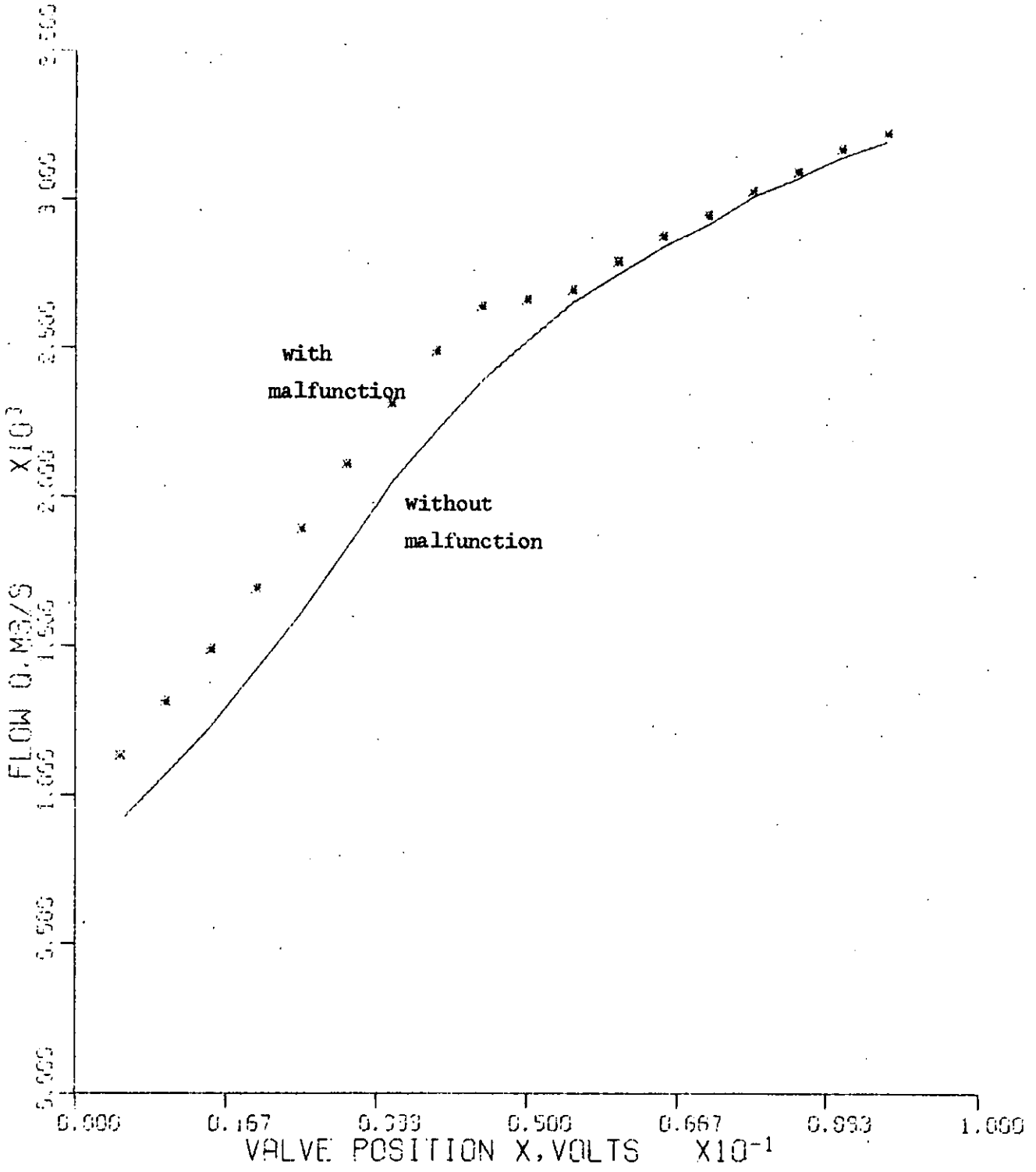


FIGURE 2.27 Full travel check on Type 2 control loop - run FT/2Ca.

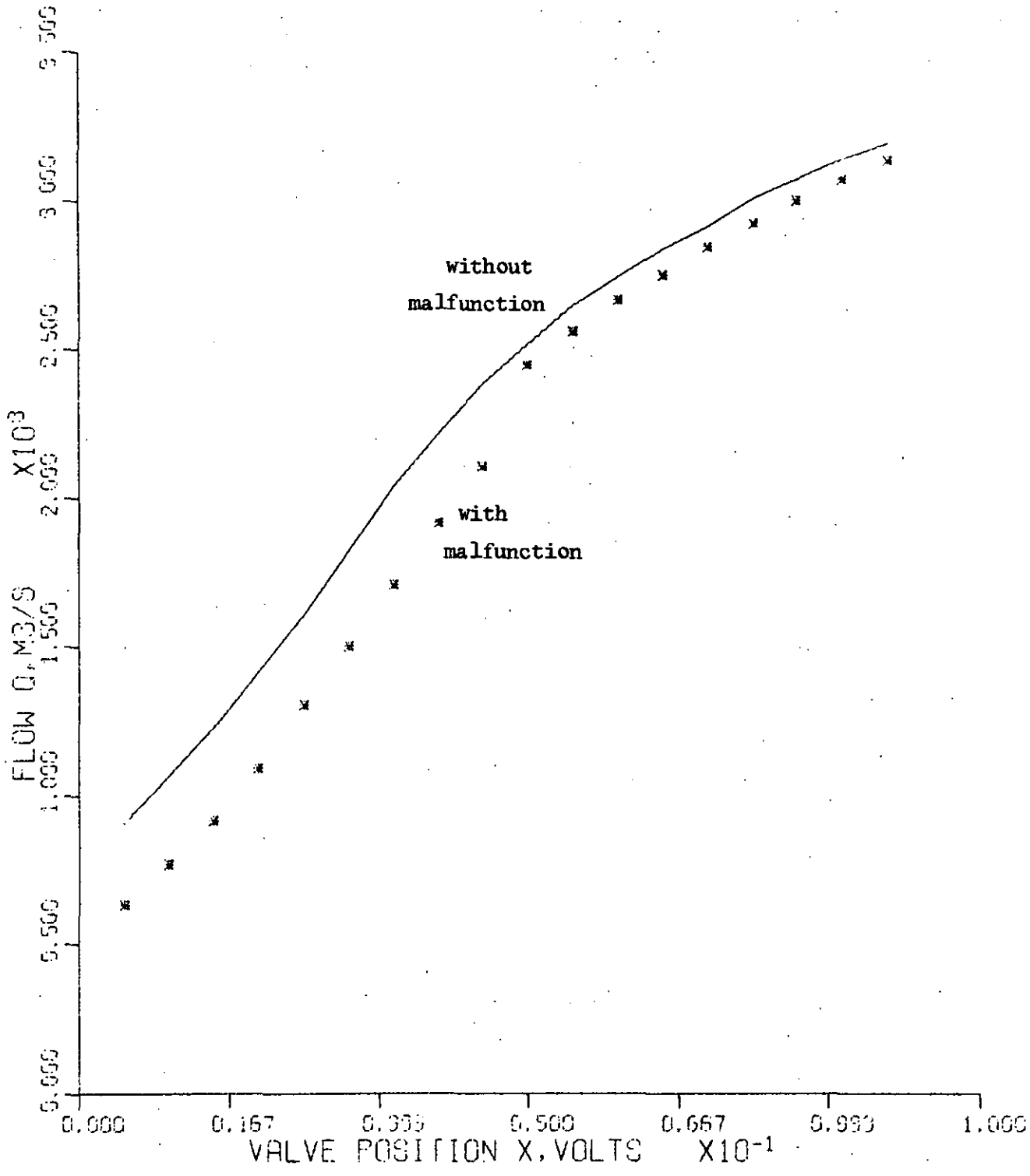


FIGURE 2.28 Full travel check on Type 2 control loop - run FT/2Cb.

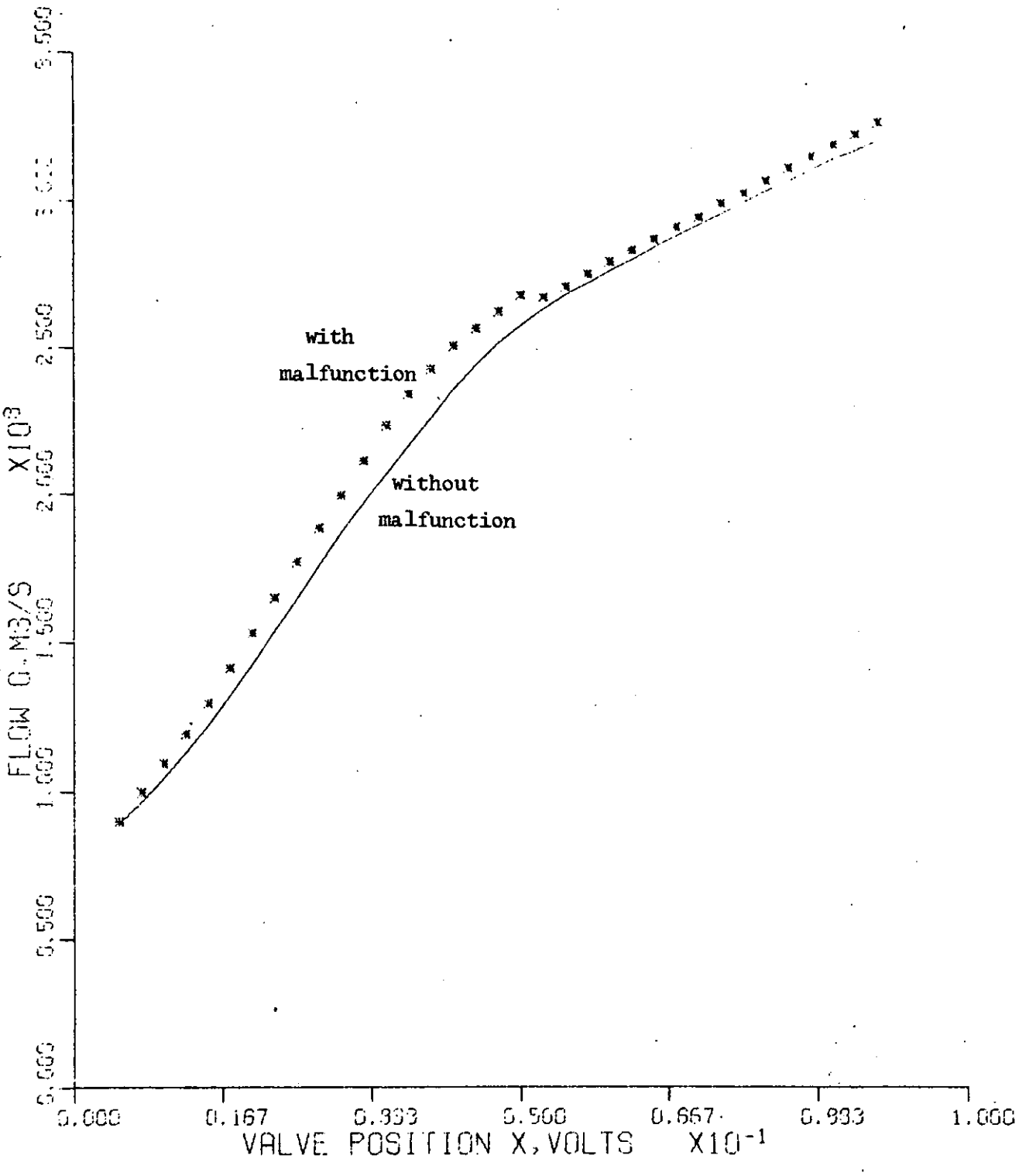


FIGURE 2.29 **Full travel check on Type 2 control loop - run FT/2D.**

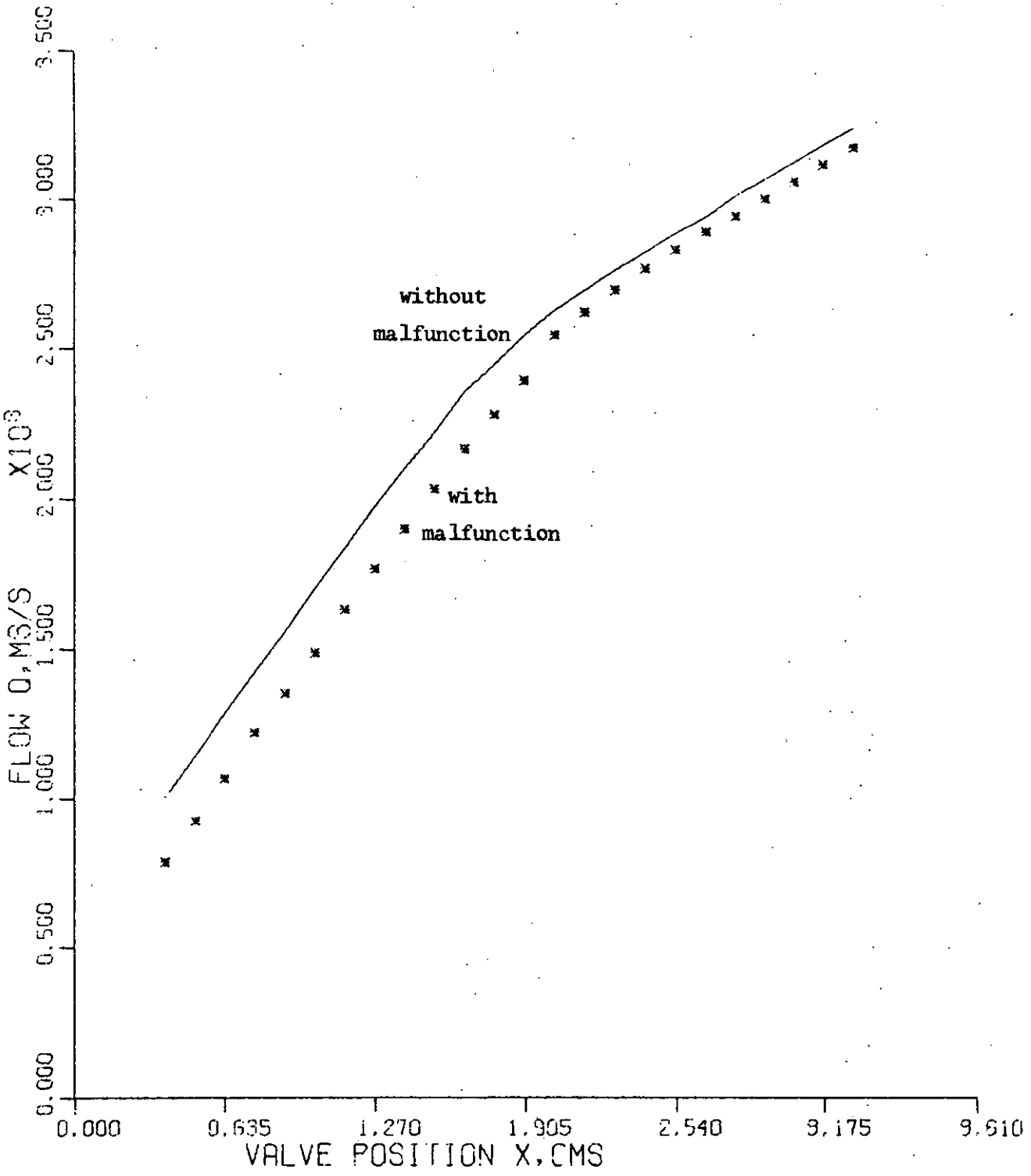


FIGURE 2.30 Full travel check on Type 2 control loop - run FT/2E.

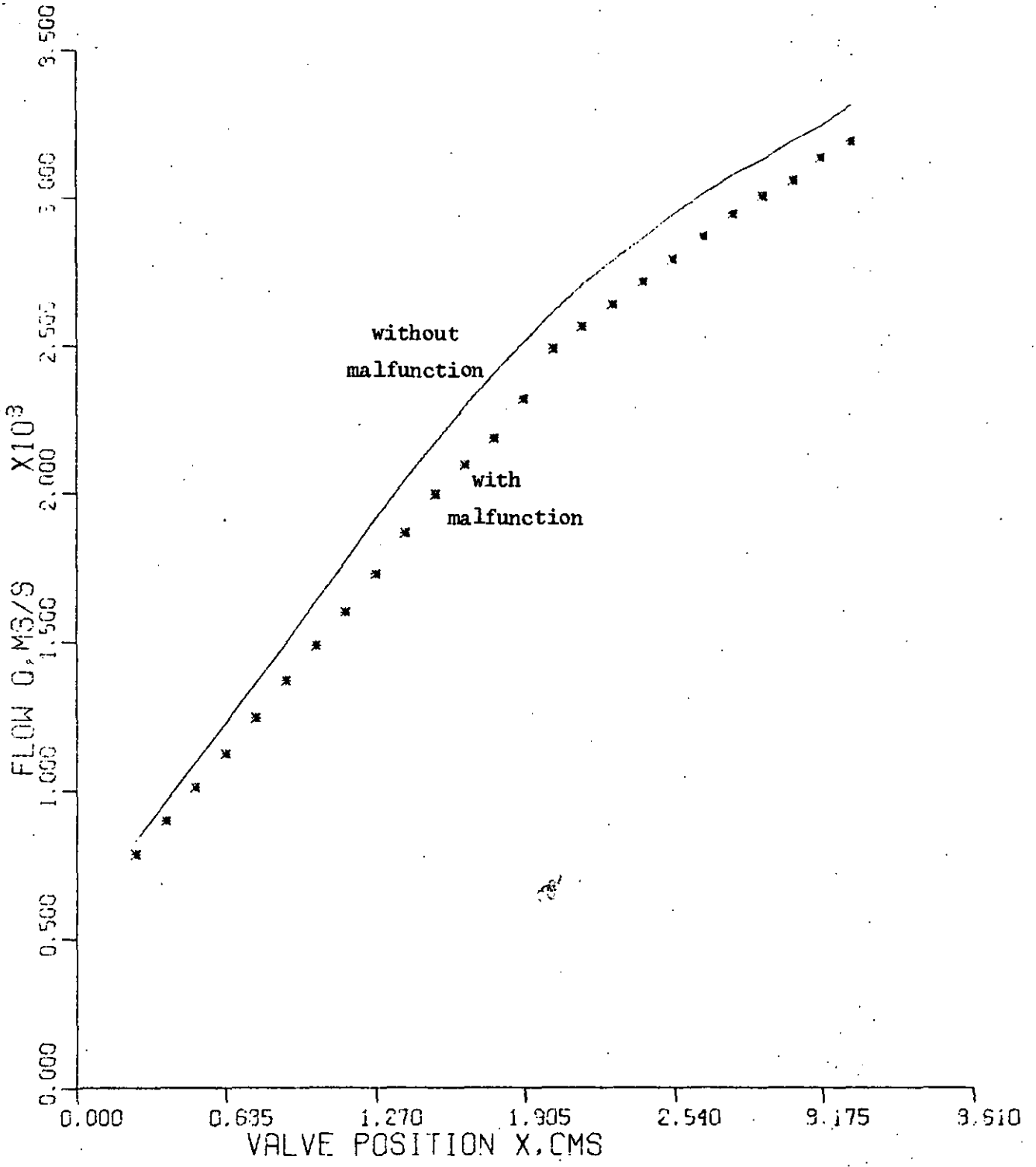


FIGURE 2.31 Full travel check on Type 2 control loop – run FT/2F.*

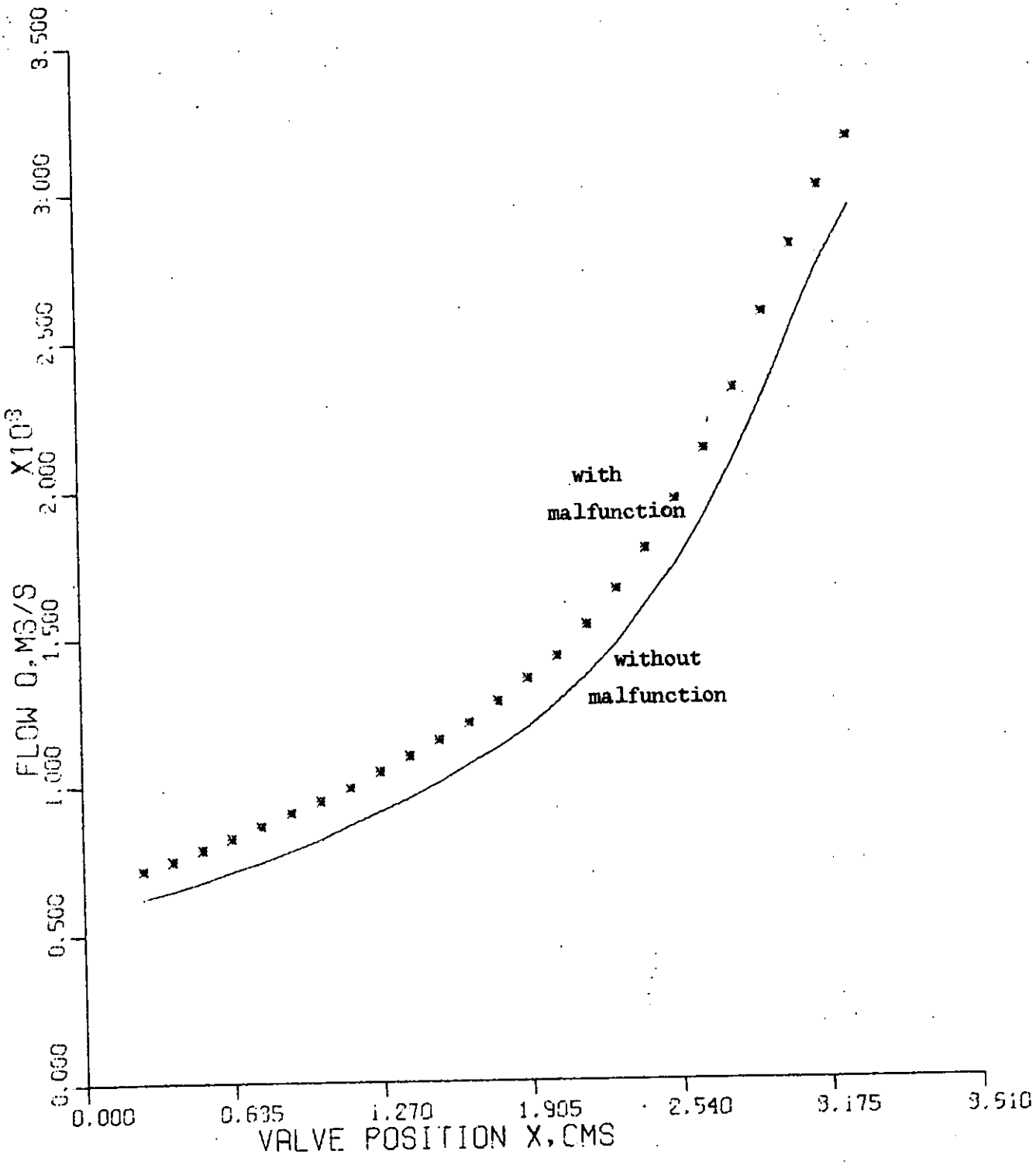


FIGURE 2.32 Full travel check on Type 3 control loop - run FT/3Aa*.

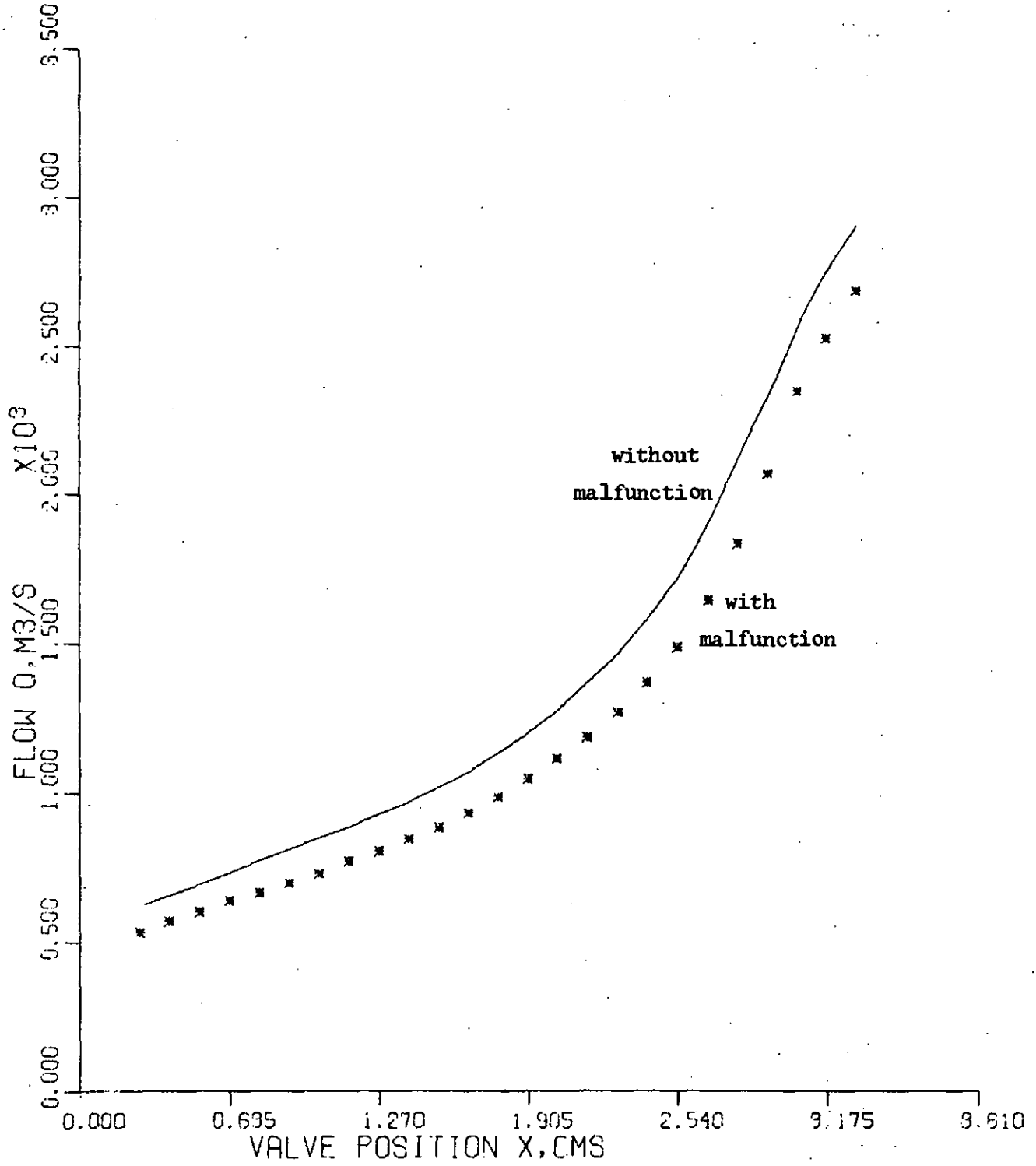


FIGURE 2.33 Full travel check on Type 3 control loop – run FT/3Ab.

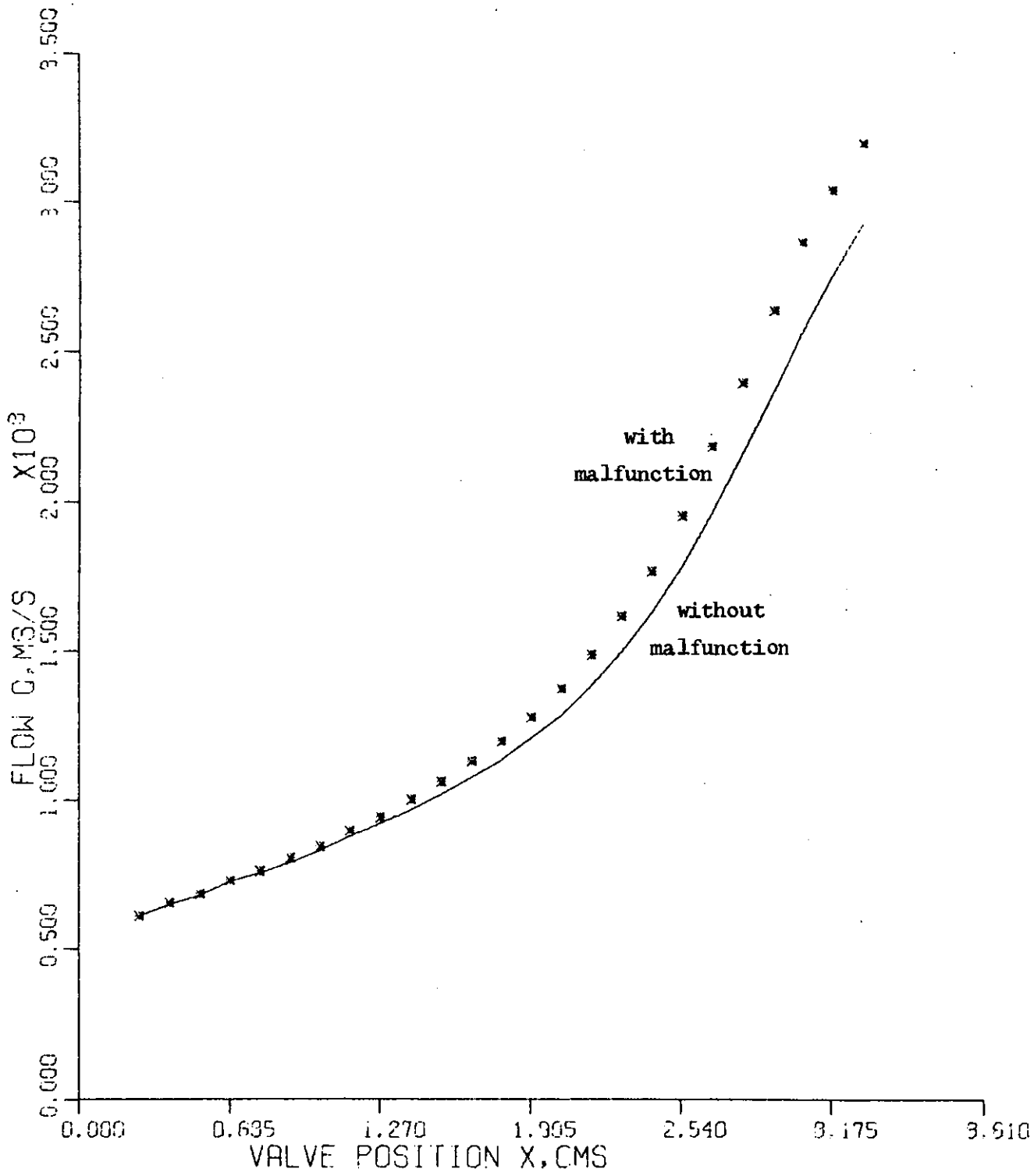


FIGURE 2.34 Full travel check on Type 3 control loop – run FT/3Ba.

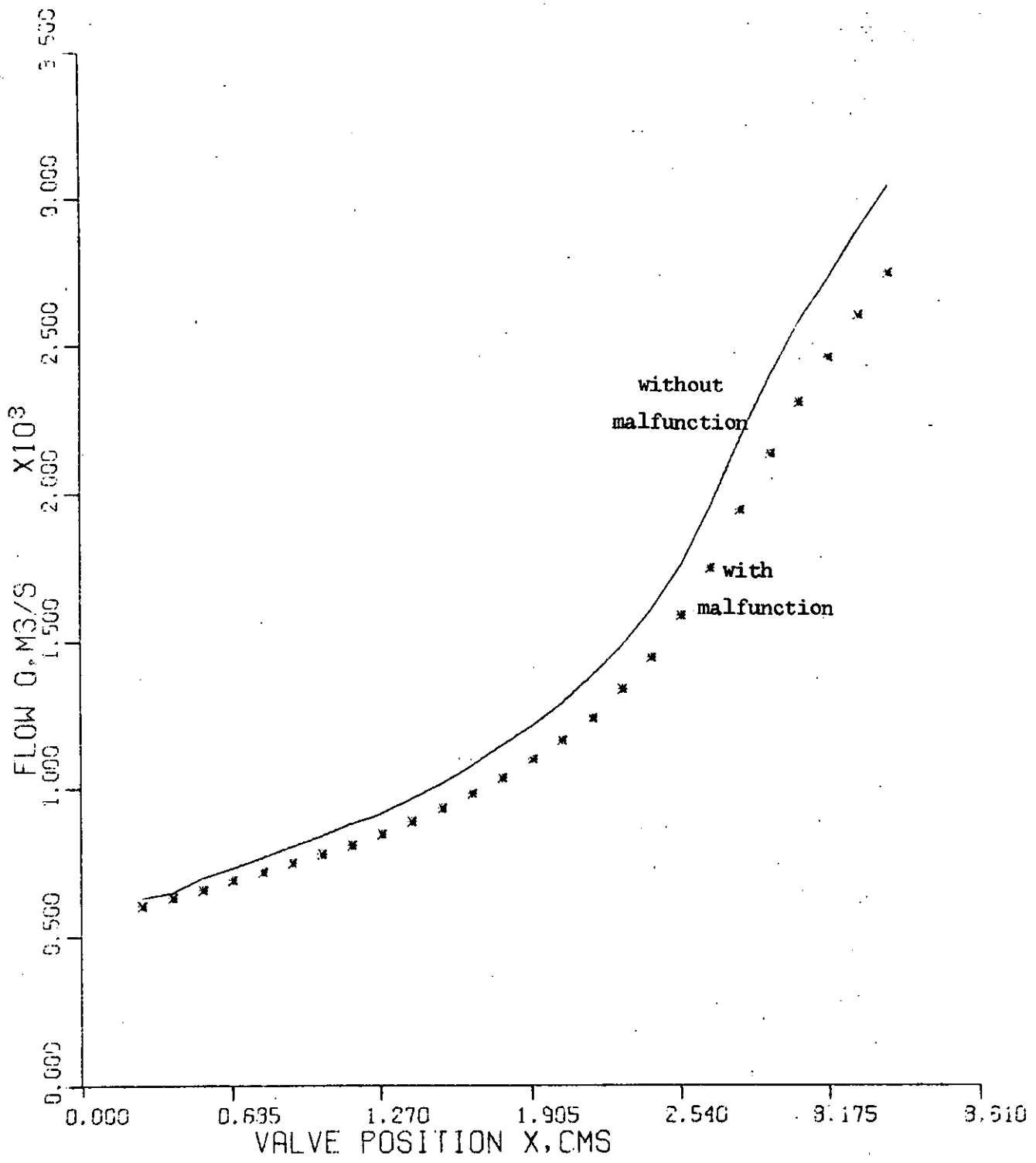


FIGURE 2.35 Full travel check on Type 3 control loop - run FT/3Bb.

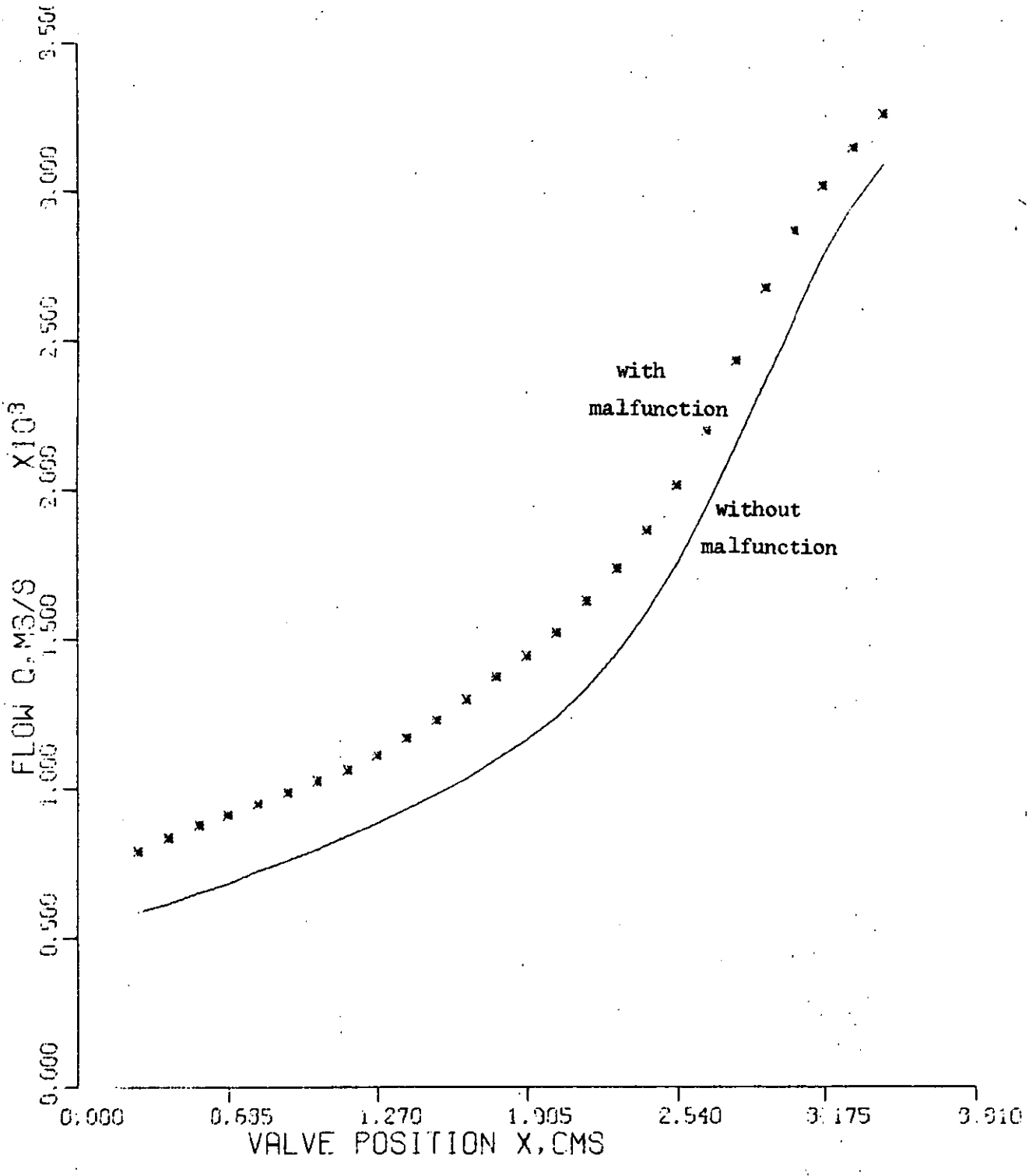


FIGURE 2.36 Full travel check on Type 4 control loop - run FT/4Aa.

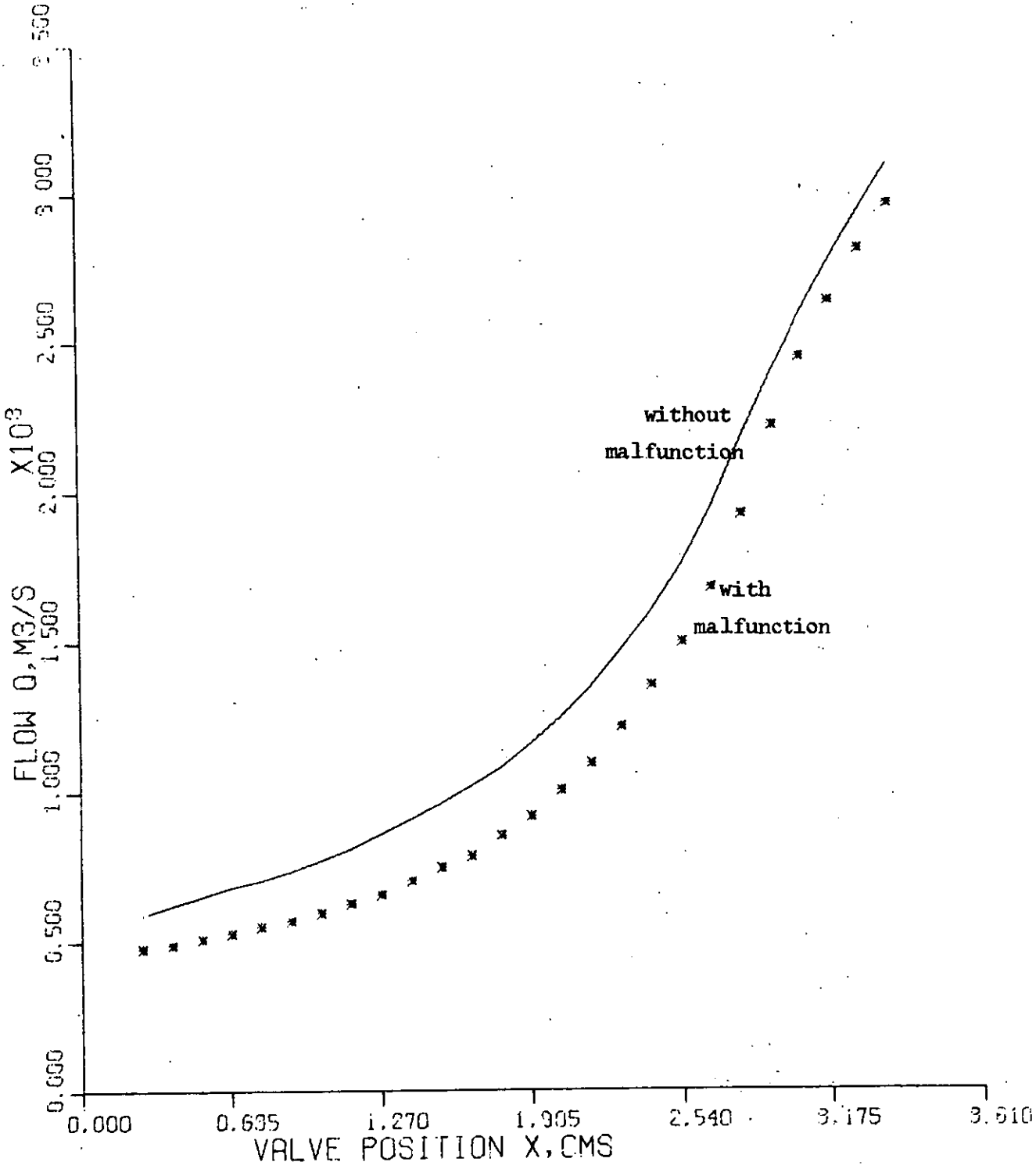


FIGURE 2.37 Full travel check on Type 4 control loop - run FT/4Ab.

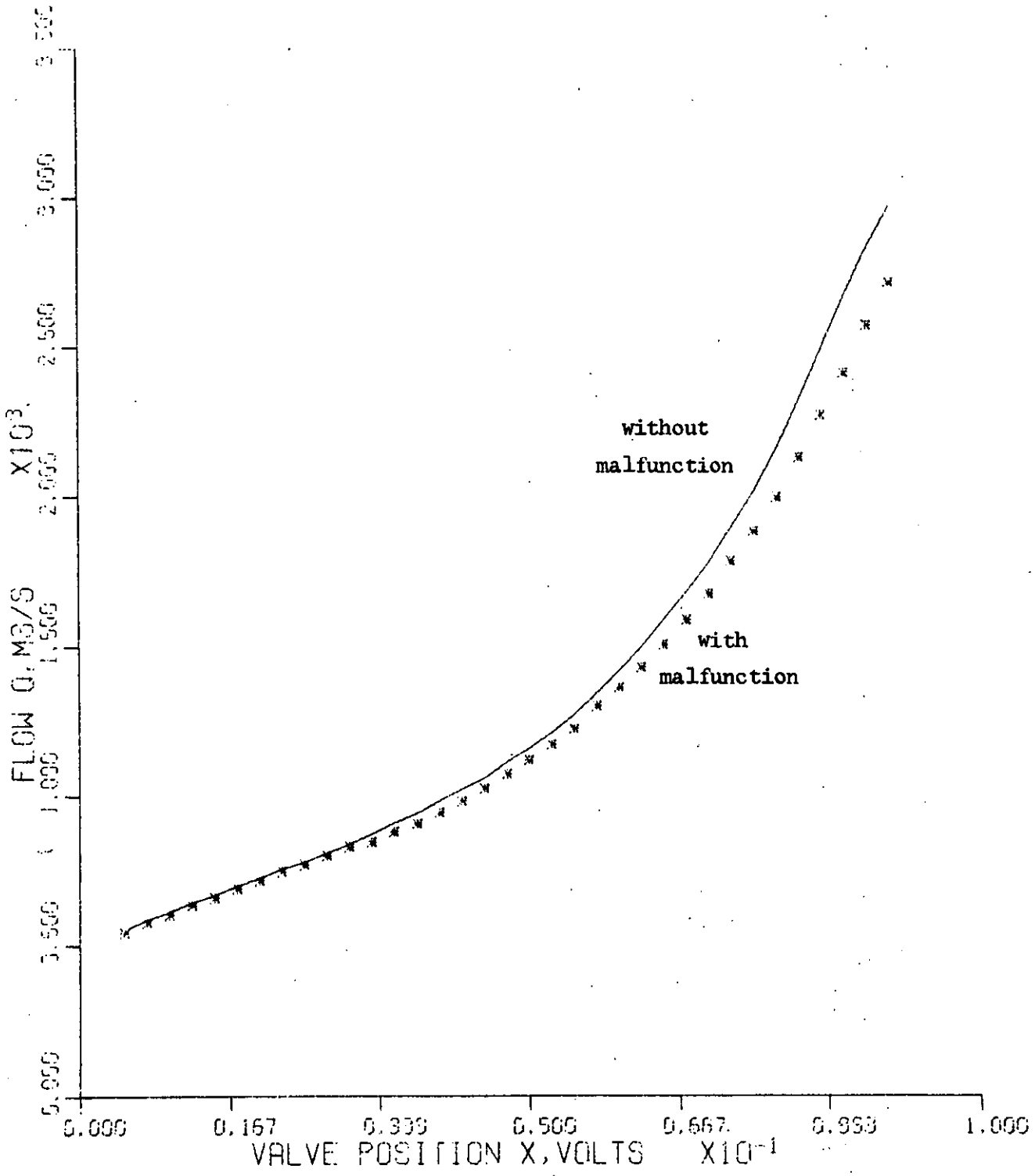


FIGURE 2.38 Full travel check on Type 4 control loop - run FT/4B.

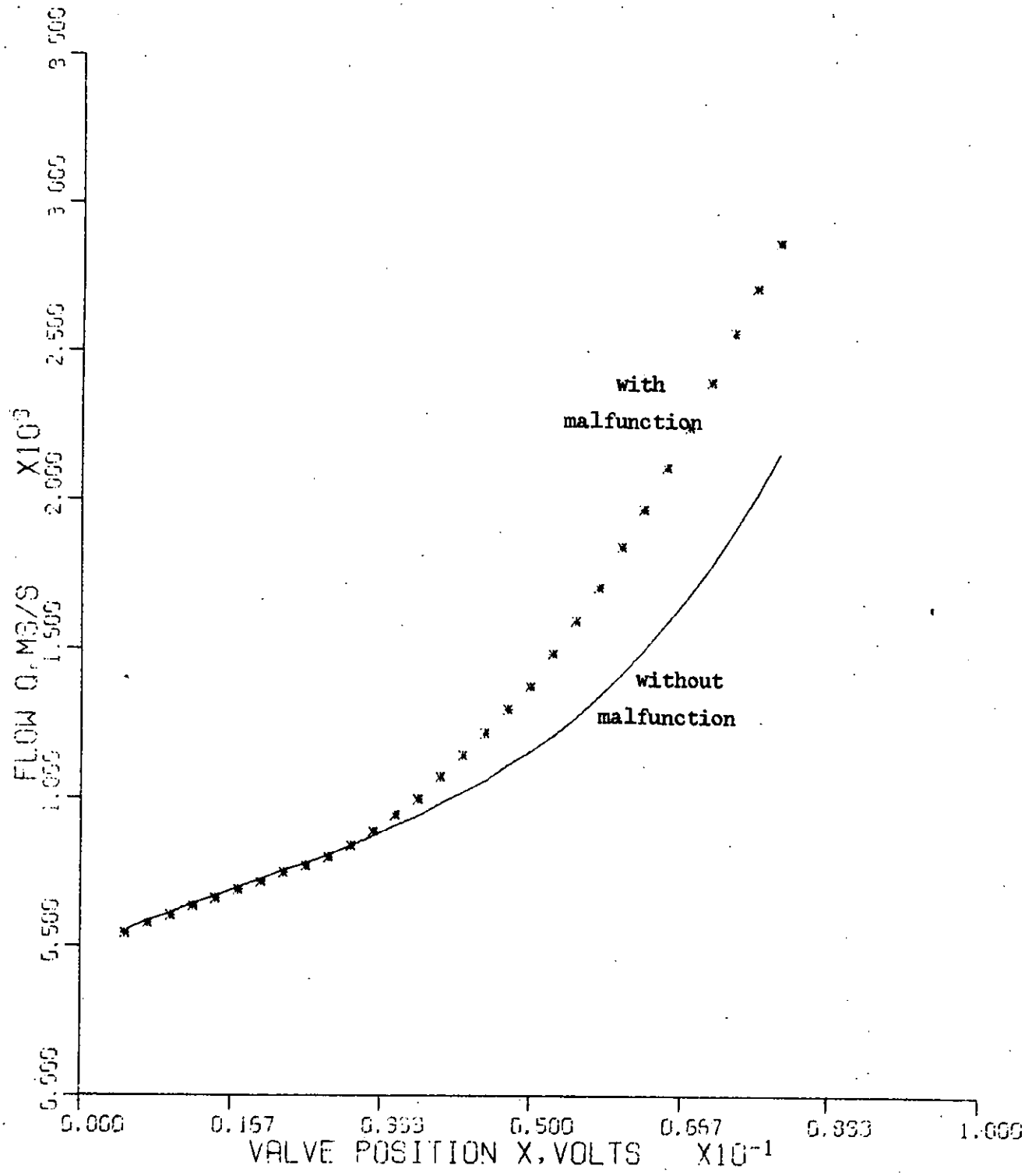


FIGURE 2.39 Full travel check on Type 4 control loop - run FT/4Ca.

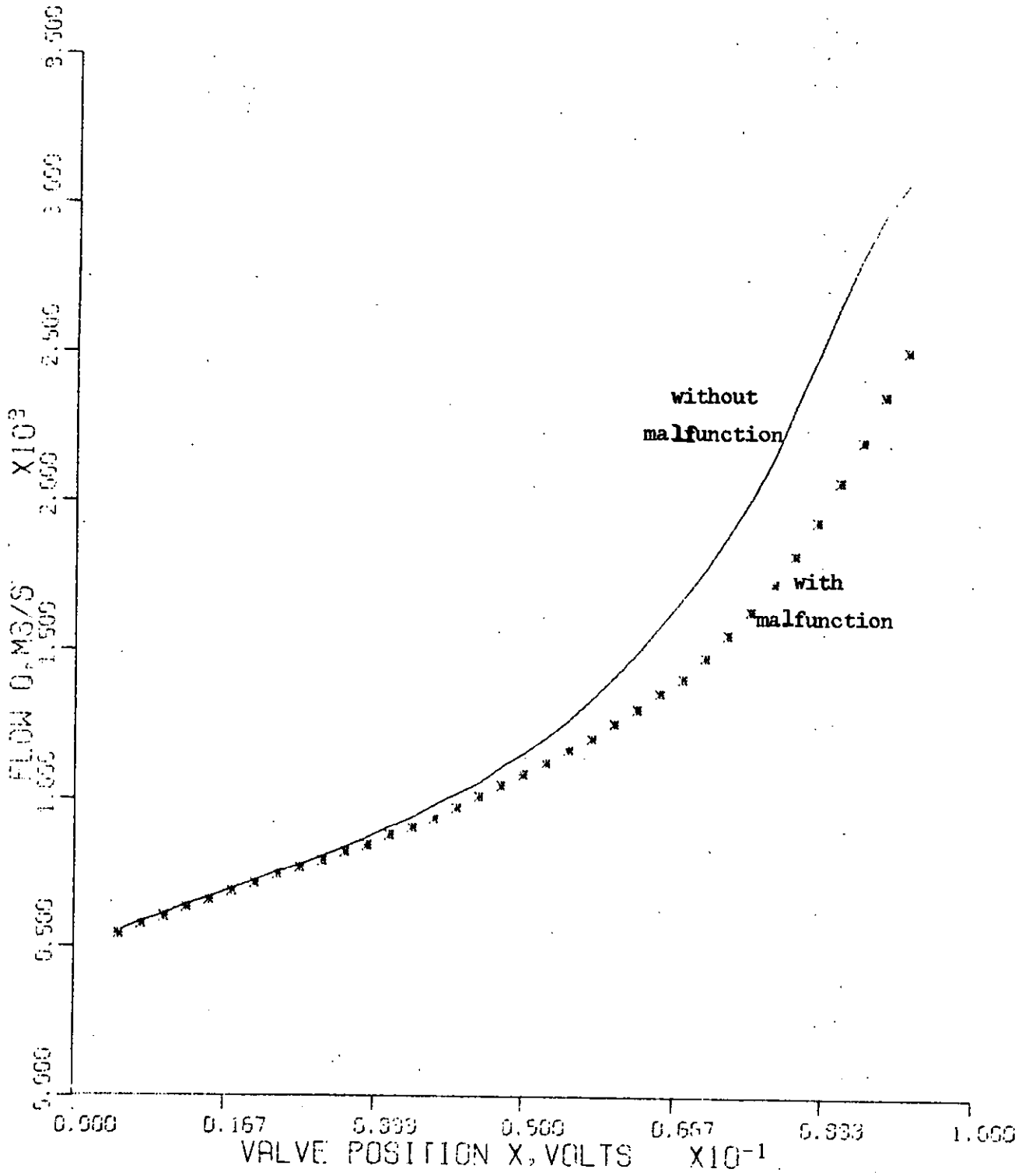


FIGURE 2.40 Full travel check on Type 4 control loop - run FT/4Cb.

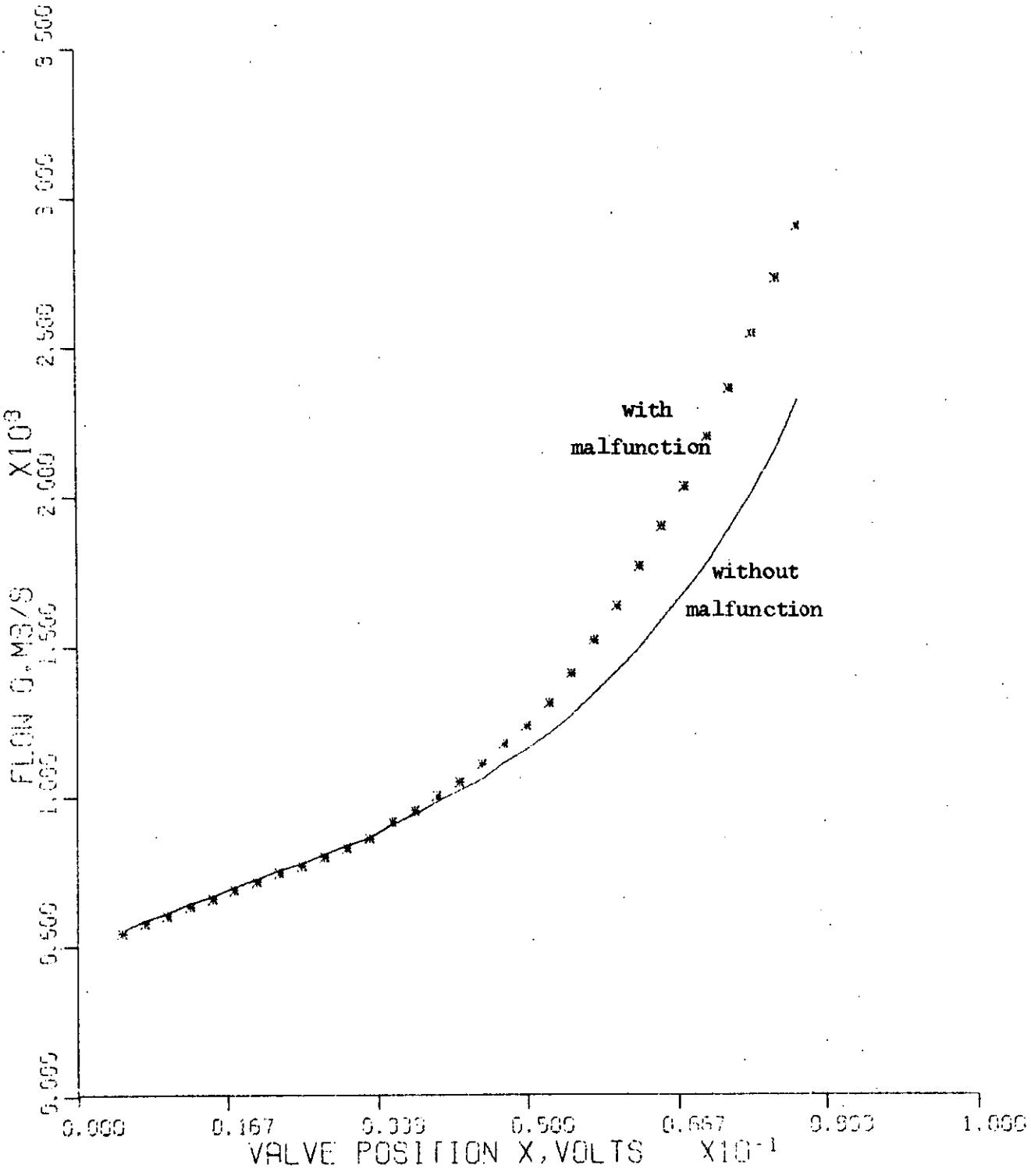


FIGURE 2.41 Full travel check on Type 4 control loop - run FT/4Da.

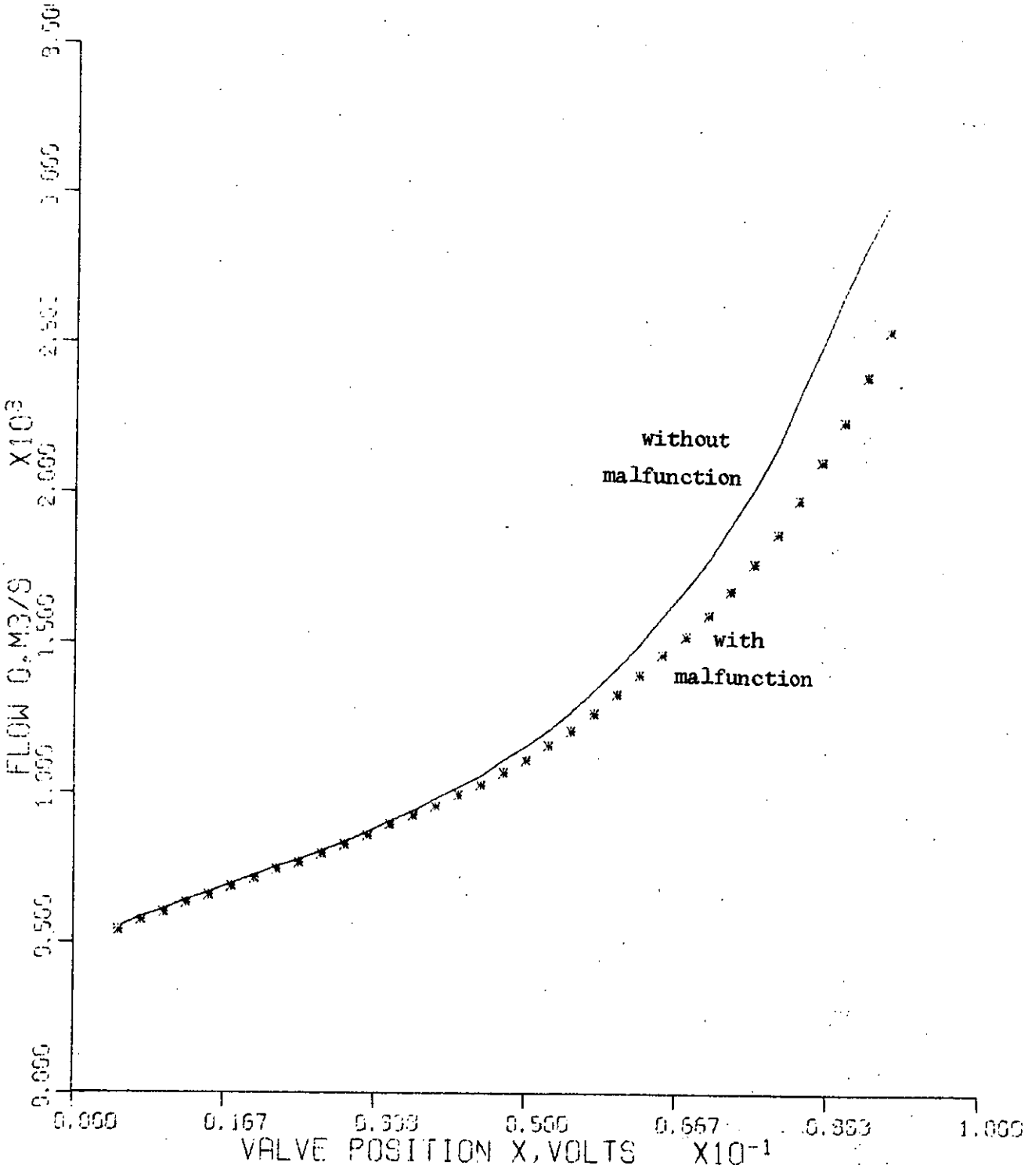


FIGURE 2.42 Full travel check on Type 4 control loop - run FT/4Db*.

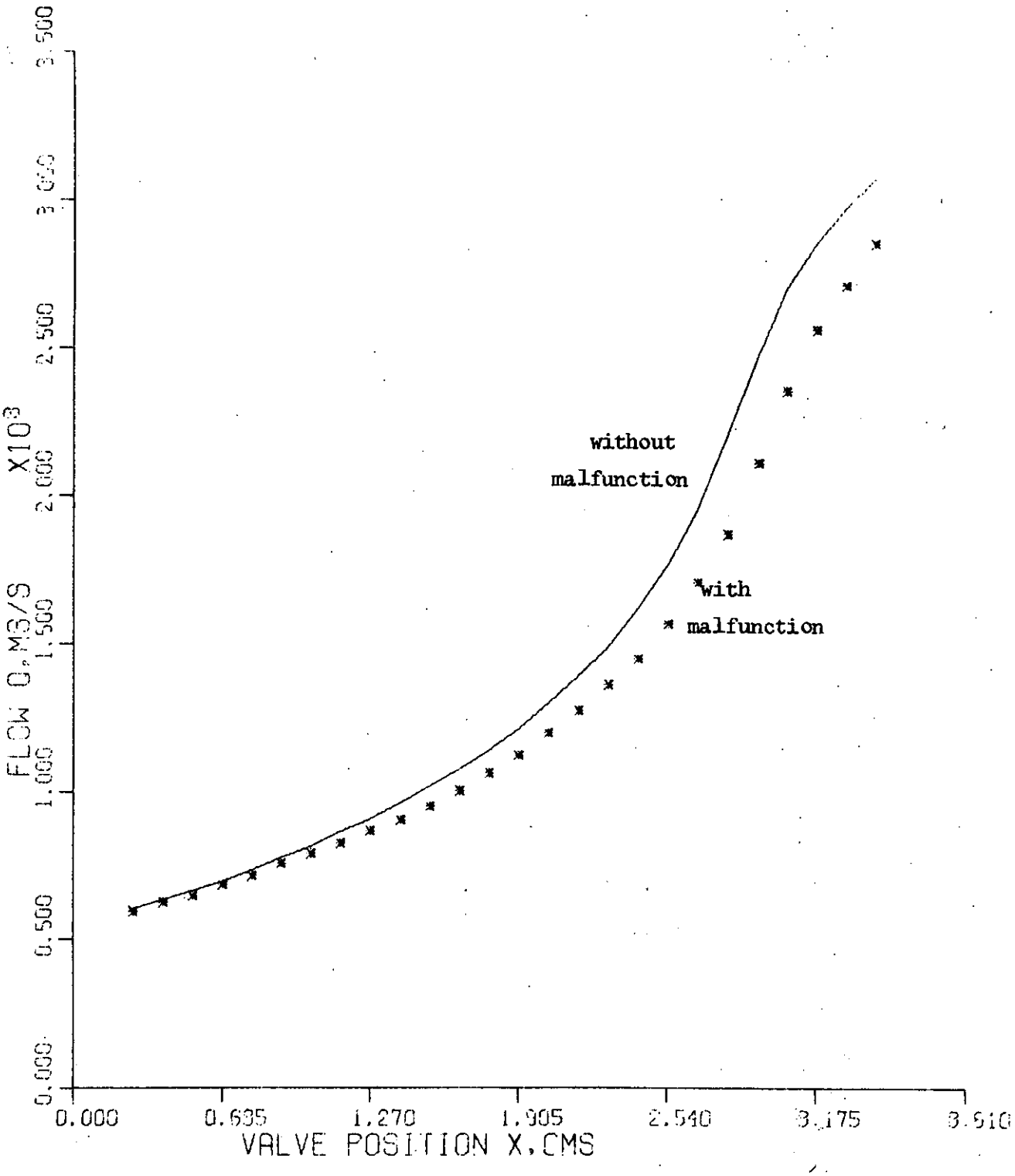


FIGURE 2.43 Full travel check on Type 4 control loop - run FT/4E.

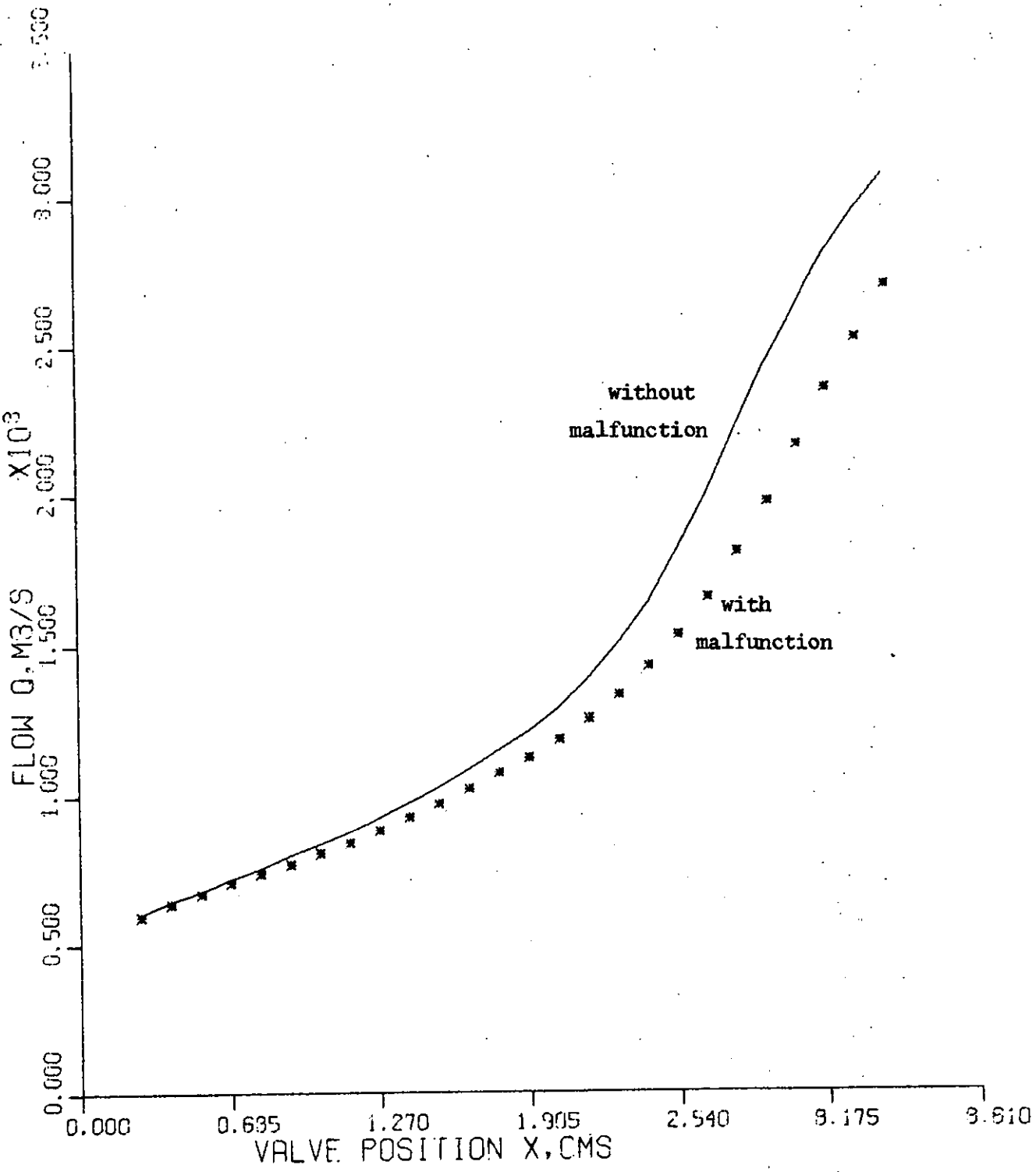


FIGURE 2.44 Full travel check on Type 4 control loop - run FT/4F*

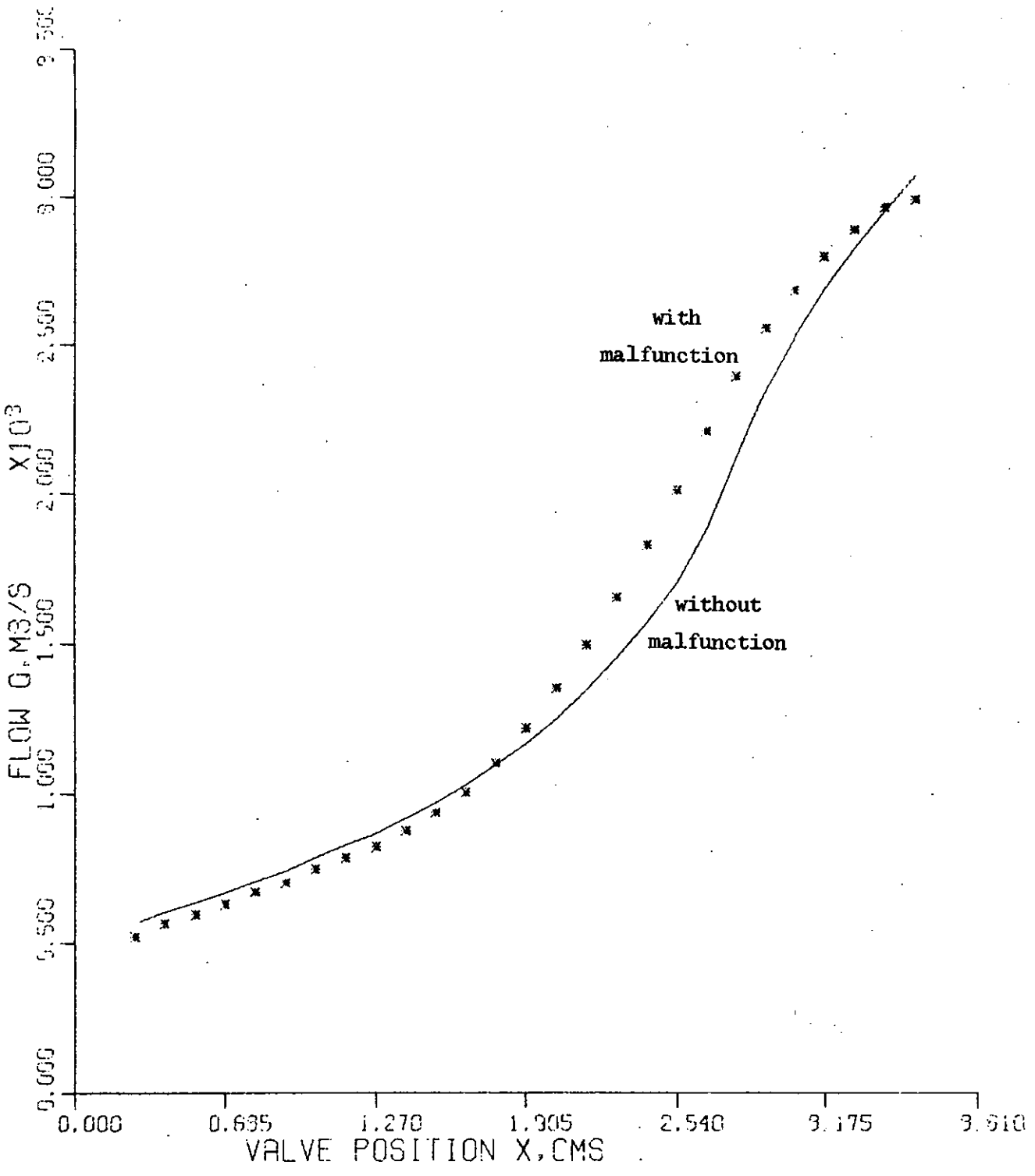


FIGURE 2.45 Full travel check on Type 4 control loop - run FT/4G.

This arises because of the two straight line approximations used to represent the linear control valve characteristic as discussed in section 2.7.3. Equation (2.4.14) may be written in the notation of section 2.7.3, as:

$$Q_v + e_{Q_v} = k_{v1} (x + e_x) + k_{v2}$$

then

$$e_{Q_v} = k_{v1} e_x$$

and so it can be seen that the apparent error in the flowrate calculated from the valve characteristic is a function of the slope of this characteristic. Since in the experimental work k_{v1} had two values, depending upon the degree of valve opening, the apparent error e_{Q_v} changes as k_{v1} changes, thus giving rise to the observed discontinuity in control valve characteristic deviation.

It can also be seen from Table 2.11 that the method of monitoring changes in the residual mean is able to detect the malfunction.

The case of valve trim damage (run FT/4G) is an exception. In contrast to the other malfunctions, which result in a systematic positive or negative deviations in the valve characteristic, this malfunction produces a deviation which is both positive and negative, depending upon the degree of valve opening, as shown in Figure 2.45. In terms of calculating the residual mean this has a cancelling effect.

This type of malfunction can be detected but it requires that the change in the residual be calculated at each valve opening before being averaged; this method was in fact used in the closed loop - limited travel experiments described below.

2.9.2 Open loop experiments with limited valve travel

The malfunction detection method for these experiments cannot be based upon comparison of the control valve characteristic since valve stroking was not possible. The check method is therefore based, as before, on the determination of the valve position residual. The time series of the residual was obtained at the same valve opening. In each experiment the means of the residual sequence without and with malfunction were determined and the modulus of the change in the means was calculated and compared with the significant change at the 95% confidence limit.

Figures 2.46 a and 2.46 b show typical residual time series for a Type 4 control loop without malfunction when the valve demand signal was used as the valve stem measurement. The residuals obtained when a + 10% current/pressure zero error was present (run OL/4C a) are shown in Figures 2.47 a and 2.47 b.

Table 2.12 summarises the results for this series of experiments and it can be seen that the method is able to detect the malfunctions shown in Table 2.9. This includes the valve trim damage malfunction.

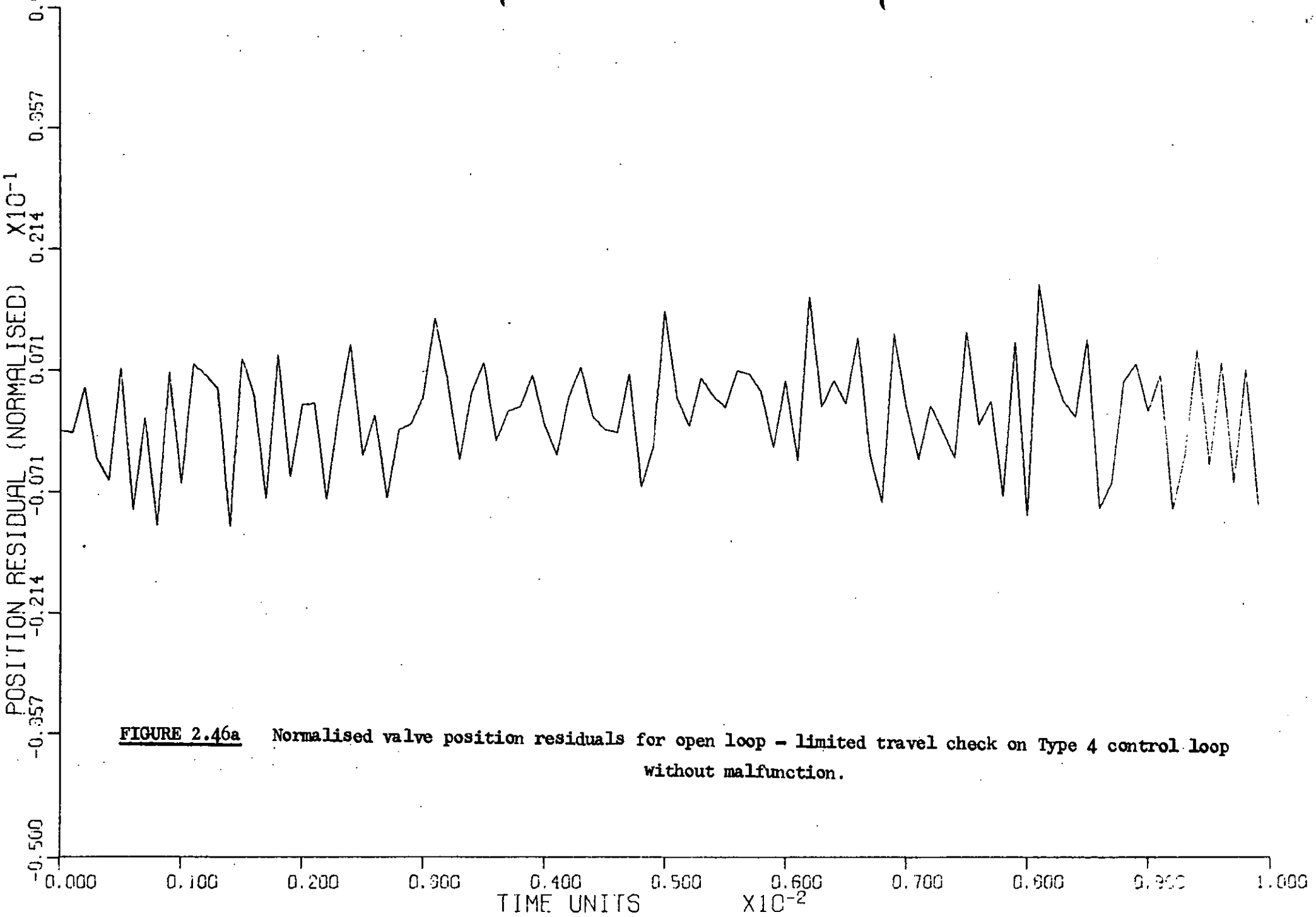


FIGURE 2.46a Normalised valve position residuals for open loop - limited travel check on Type 4 control loop without malfunction.

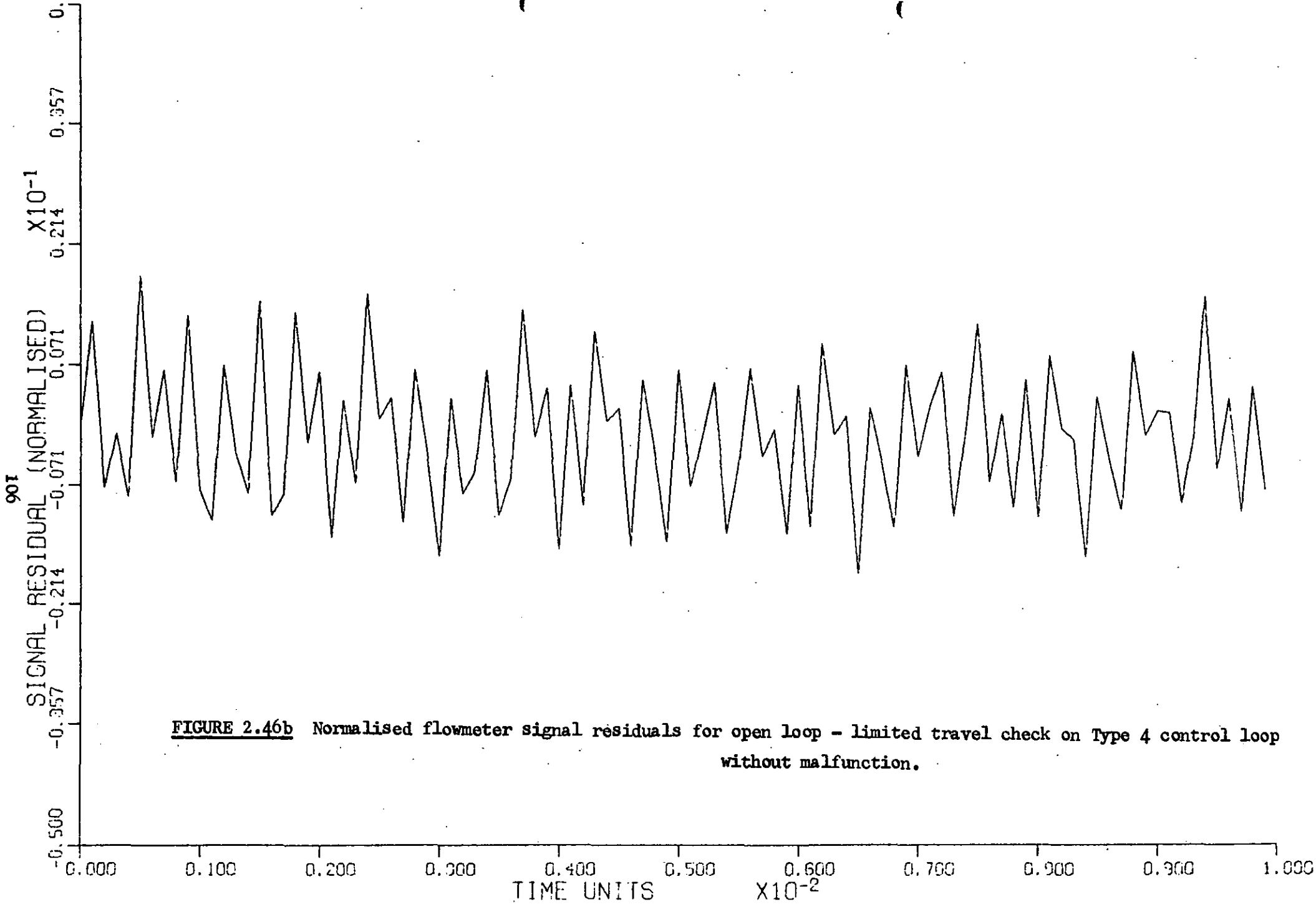


FIGURE 2.46b Normalised flowmeter signal residuals for open loop - limited travel check on Type 4 control loop without malfunction.

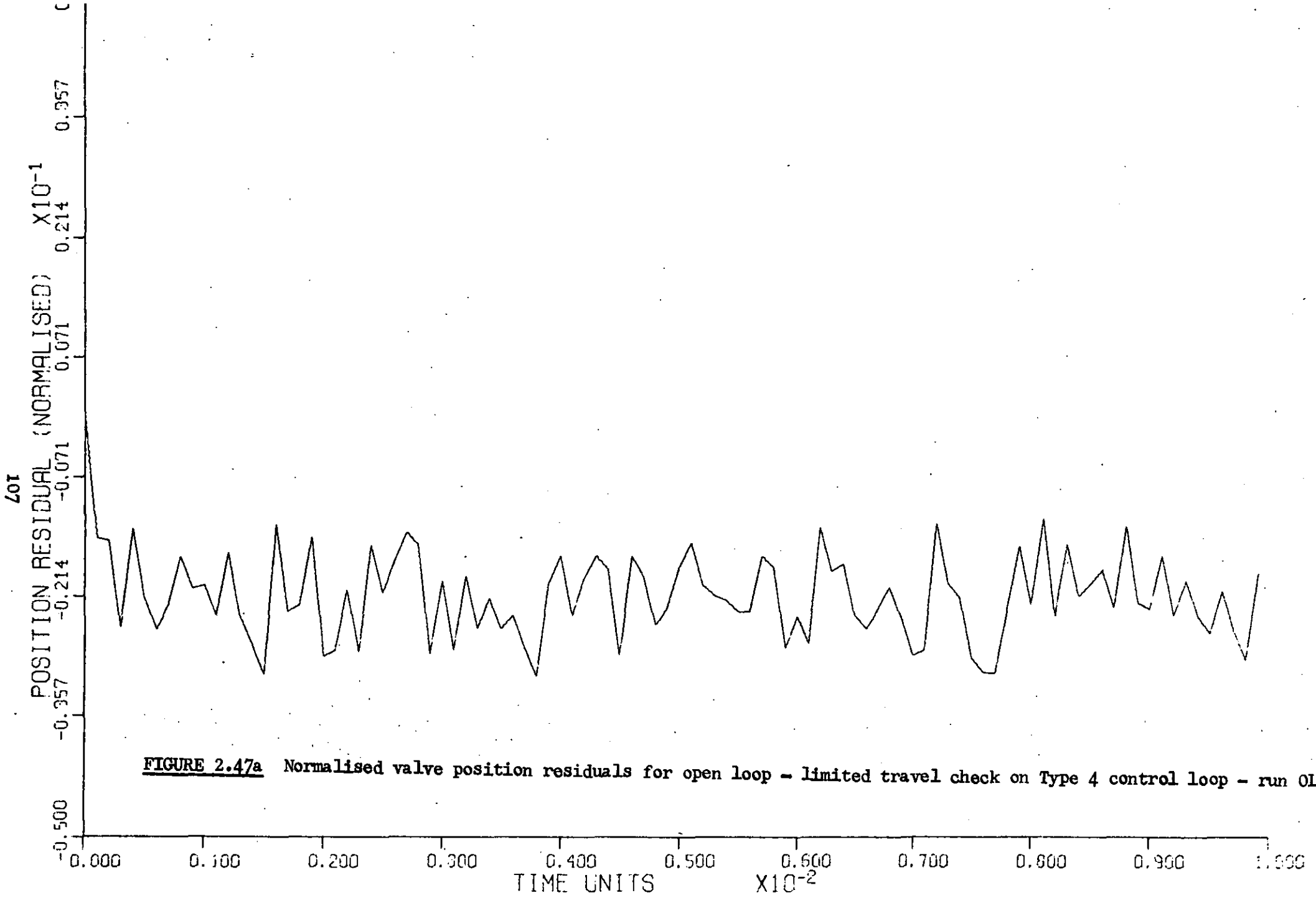


FIGURE 2.47a Normalised valve position residuals for open loop - limited travel check on Type 4 control loop - run OL/4C

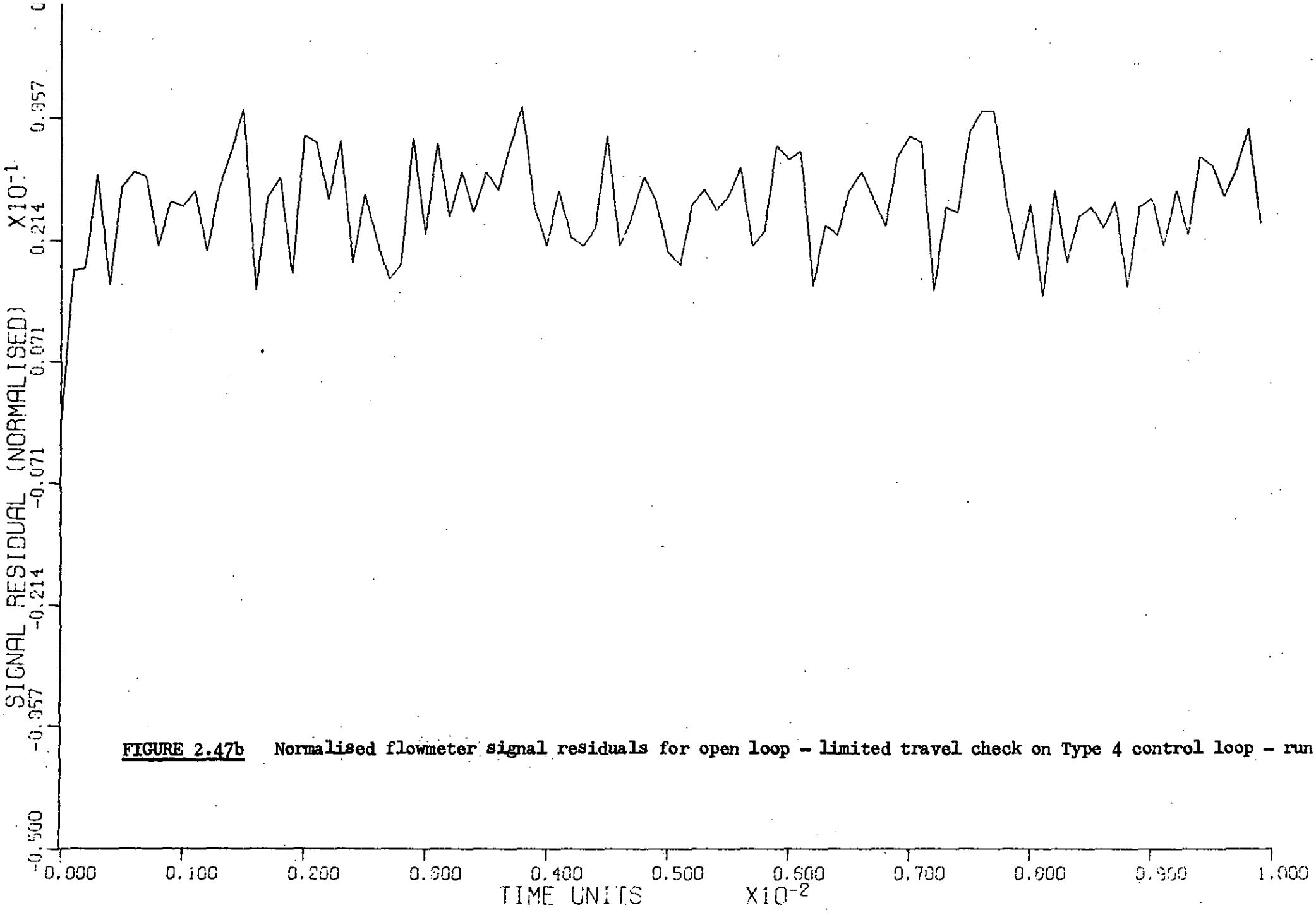


FIGURE 2.47b Normalised flowmeter signal residuals for open loop - limited travel check on Type 4 control loop - run OL/4C

Control loop	Run number	Modulus of change in mean of normalised valve position residual	
		Actual change	Significant change
2	OL/2A	1.28	0.20
	OL/2B	1.16	0.53
	OL/2C	0.49	0.21
4	OL/4Aa	1.60	0.12
	OL/4Ab	2.56	0.21
	OL/4B	2.68	0.28
	OL/4Ca	1.87	0.13
	OL/4Cb	1.48	0.10
	OL/4D	0.48	0.17
	OL/4E	0.46	0.20
	OL/4F*	0.69	0.20
	OL/4G	1.09	0.06

TABLE 2.12. Behaviour of valve position residual for open loop limited travel experiments.

2.9.3 Closed loop experiments with direct digital control

The check method is again based on the determination of the valve position residual. At selected setpoints of the control loop the time series of the residual without malfunction was obtained. In each experiment the means of the residual series without and with malfunction were determined at each particular setpoint, and the modulus of the change in the means of the residuals was calculated and compared with the significant change at each selected setpoint.

The method therefore requires that the mean of the residual without malfunction be stored at the selected setpoints.

The experimental results are summarised in Table 2.13.

Figures 2.48 a and b show the normalised residual time series for a control loop without malfunction controlling at a setpoint 1.82×10^{-3} m^3/s. Figures 2.49a and b show the corresponding residual sequences obtained with the loop controlling at the same setpoint when the valve trim is damaged (run CL/4G) as shown in Figure 2.17.

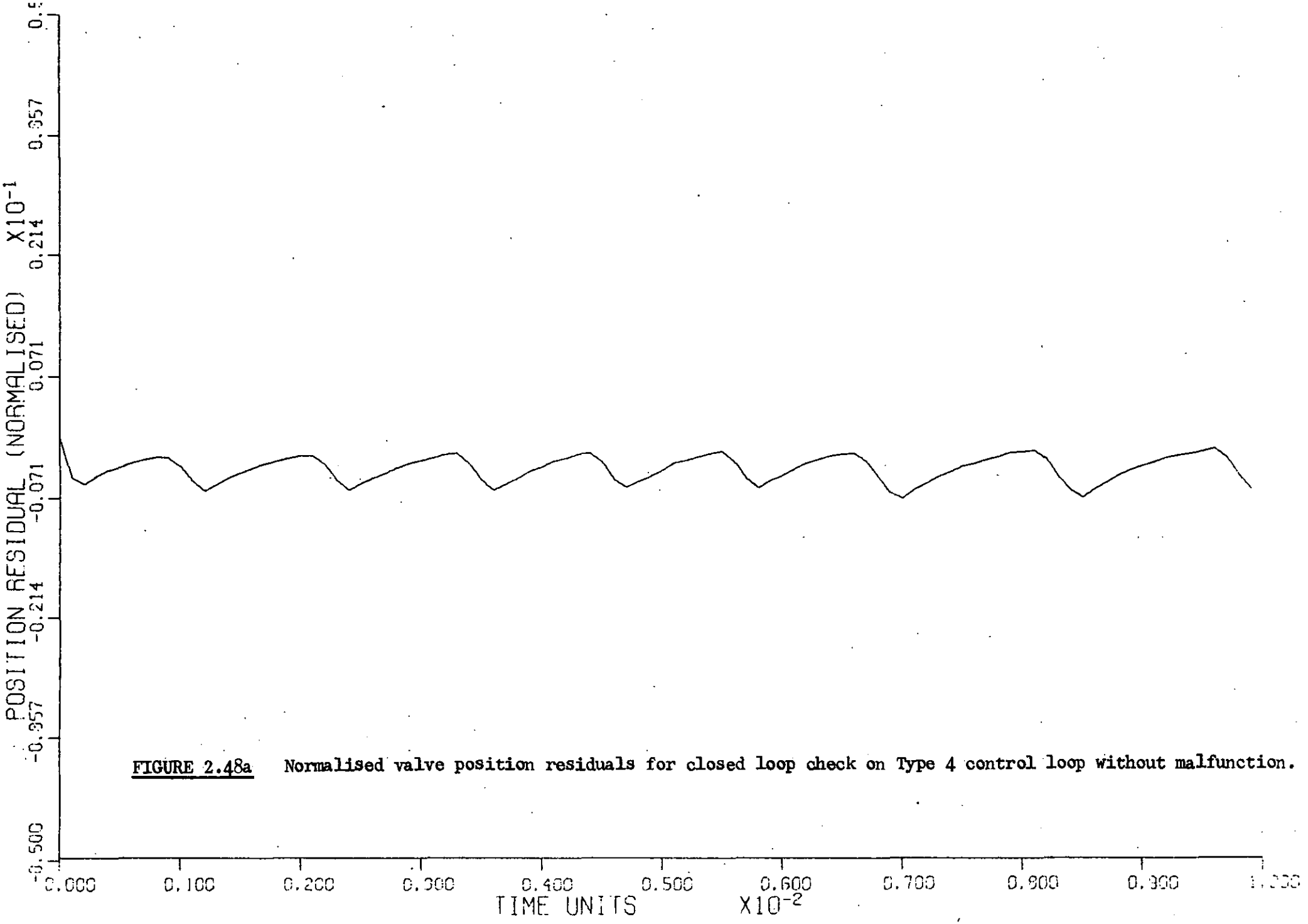


FIGURE 2.48a Normalised valve position residuals for closed loop check on Type 4 control loop without malfunction.

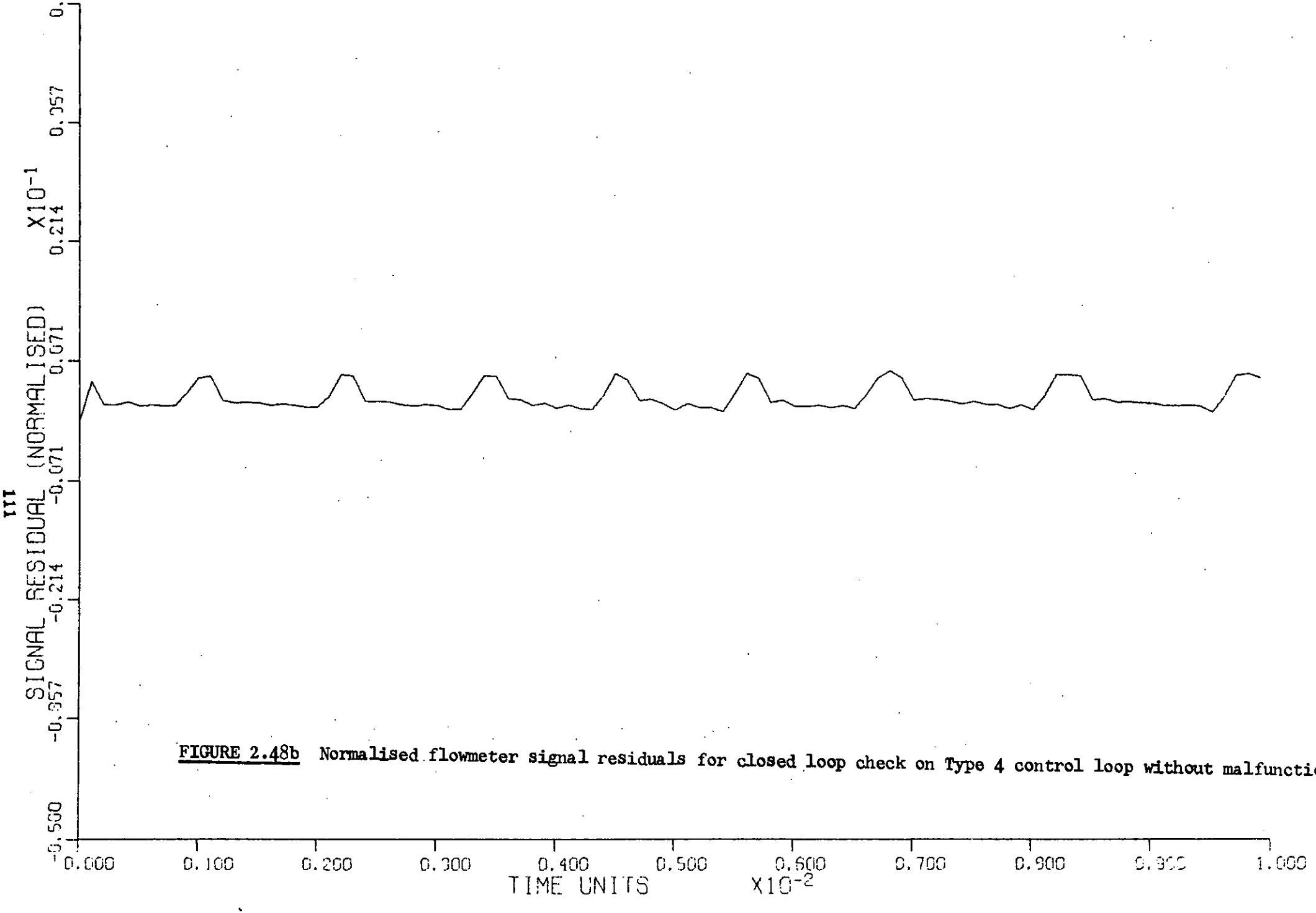


FIGURE 2.48b Normalised flowmeter signal residuals for closed loop check on Type 4 control loop without malfunction.

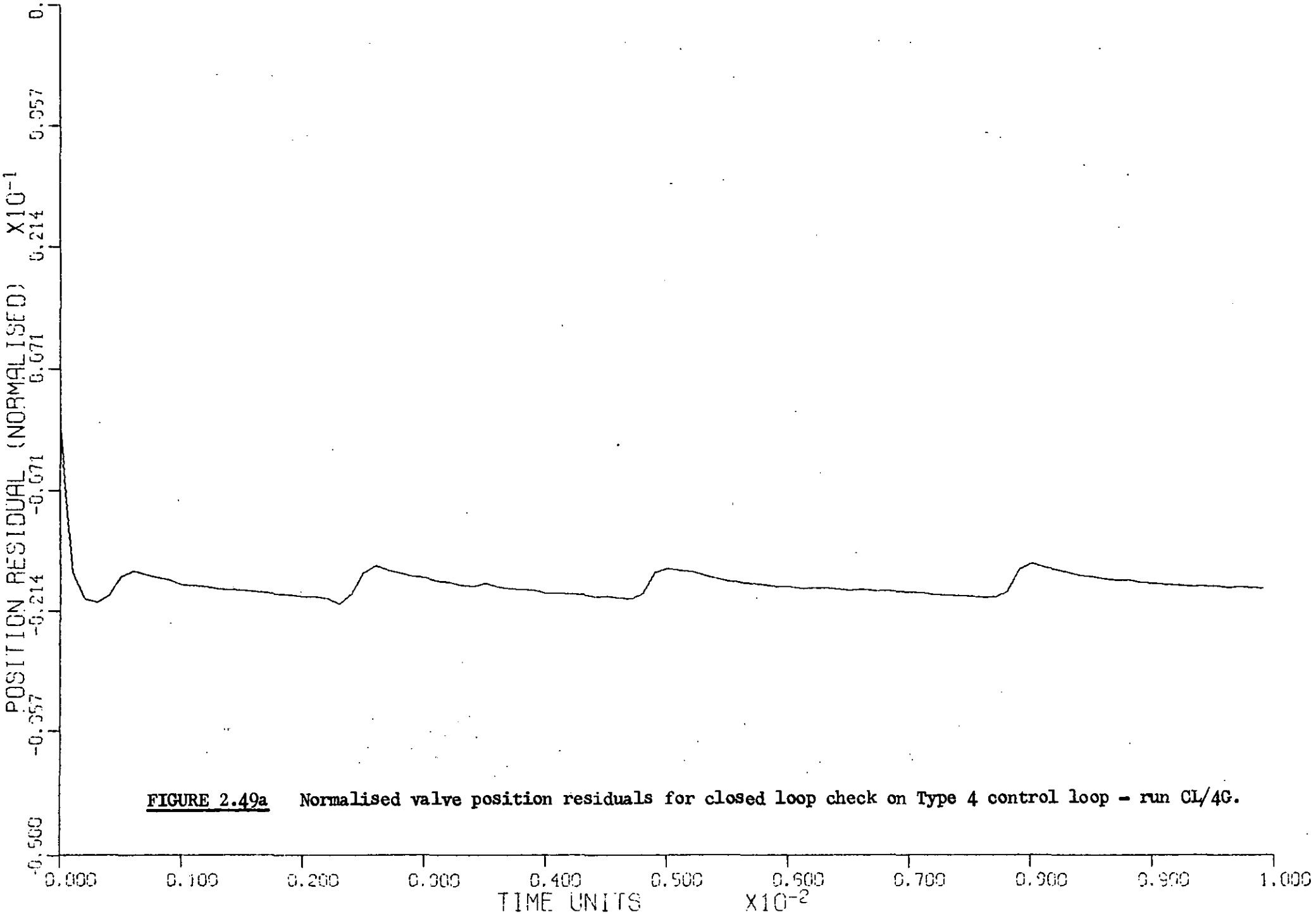


FIGURE 2.49a Normalised valve position residuals for closed loop check on Type 4 control loop - run CL/4G.

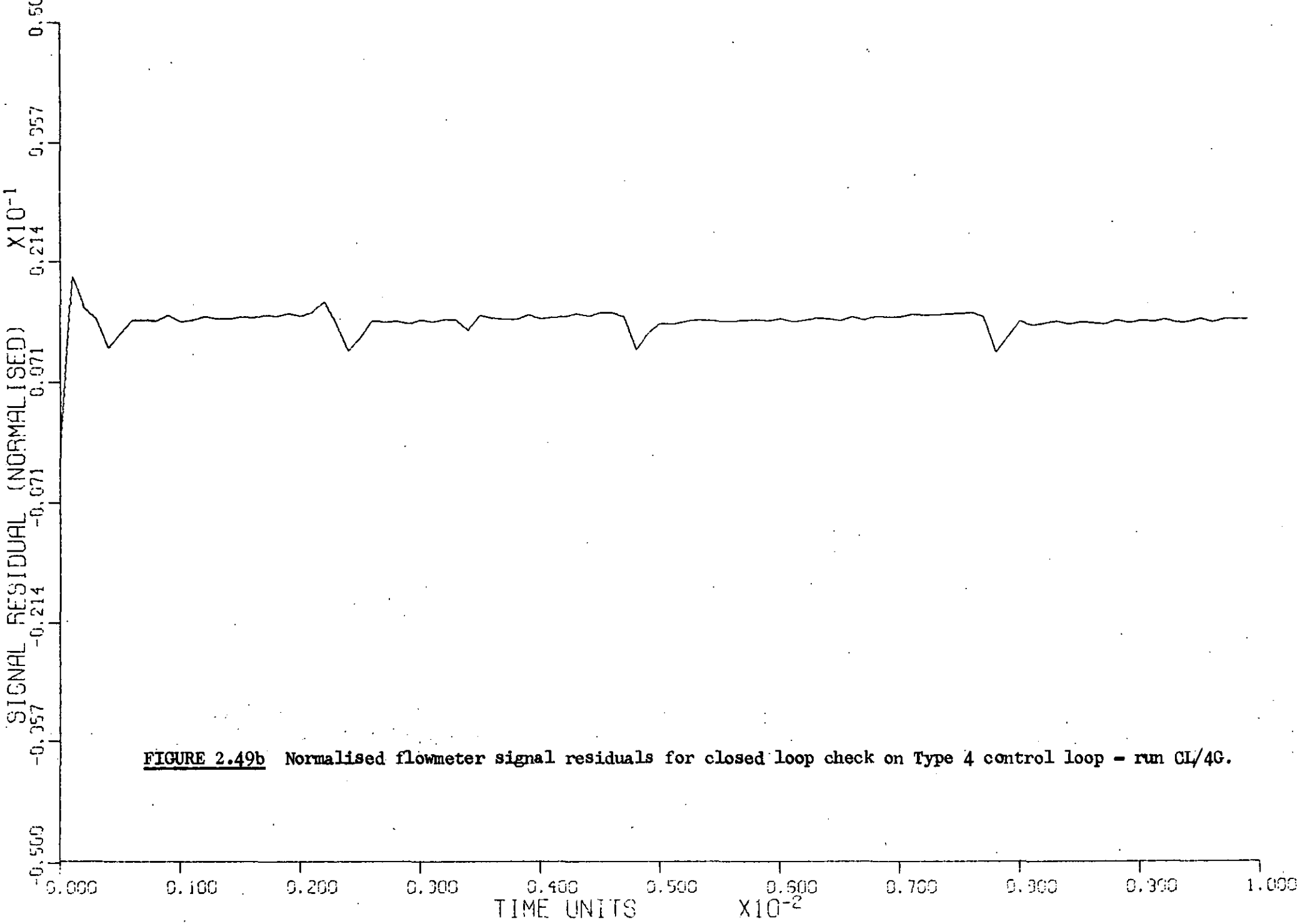


FIGURE 2.49b Normalised flowmeter signal residuals for closed loop check on Type 4 control loop - run CL/4G.

Again, it can be seen from Table 2.13 that the method is able to detect the malfunctions shown in Table 2.9, including the valve trim damage malfunction.

Control loop type	Run number	Setpoint m^3/s	Modulus of change in mean of normalised valve position residual	
			Actual change	Significant change
4	CL/4A	1.82×10^{-3}	0.776	0.05
	CL/4C a	1.82×10^{-3}	2.77	0.07
	CL/4C b	2.65×10^{-3}	2.08	0.08
	CL/4G	1.82×10^{-3}	1.44	0.05

TABLE 2.13. Behaviour of valve position residual for closed loop with direct digital control experiments.

2.10 Industrial experiments

Some limited industrial trials of the proposed method have been conducted at Works A.

The process under investigation operated with the aid of a Ferranti Argus 500 process computer which performed d.d.c. on several control loops.

Two d.d.c. flow loops were selected for analysis, and are denoted FC/1 and FC/2. The loop FC/1 controlled the reflux rate to a stripping column while FC/2 controlled the column feed flow. The physical arrangement of the column with the control loops is shown in Figure 2.50.

In these control loops the computer measured the flowmeter signal (for basic d.d.c.) and also a feedback Electrostep valve positioner signal thus yielding the control valve stem position (although no direct stem position measurement was made).

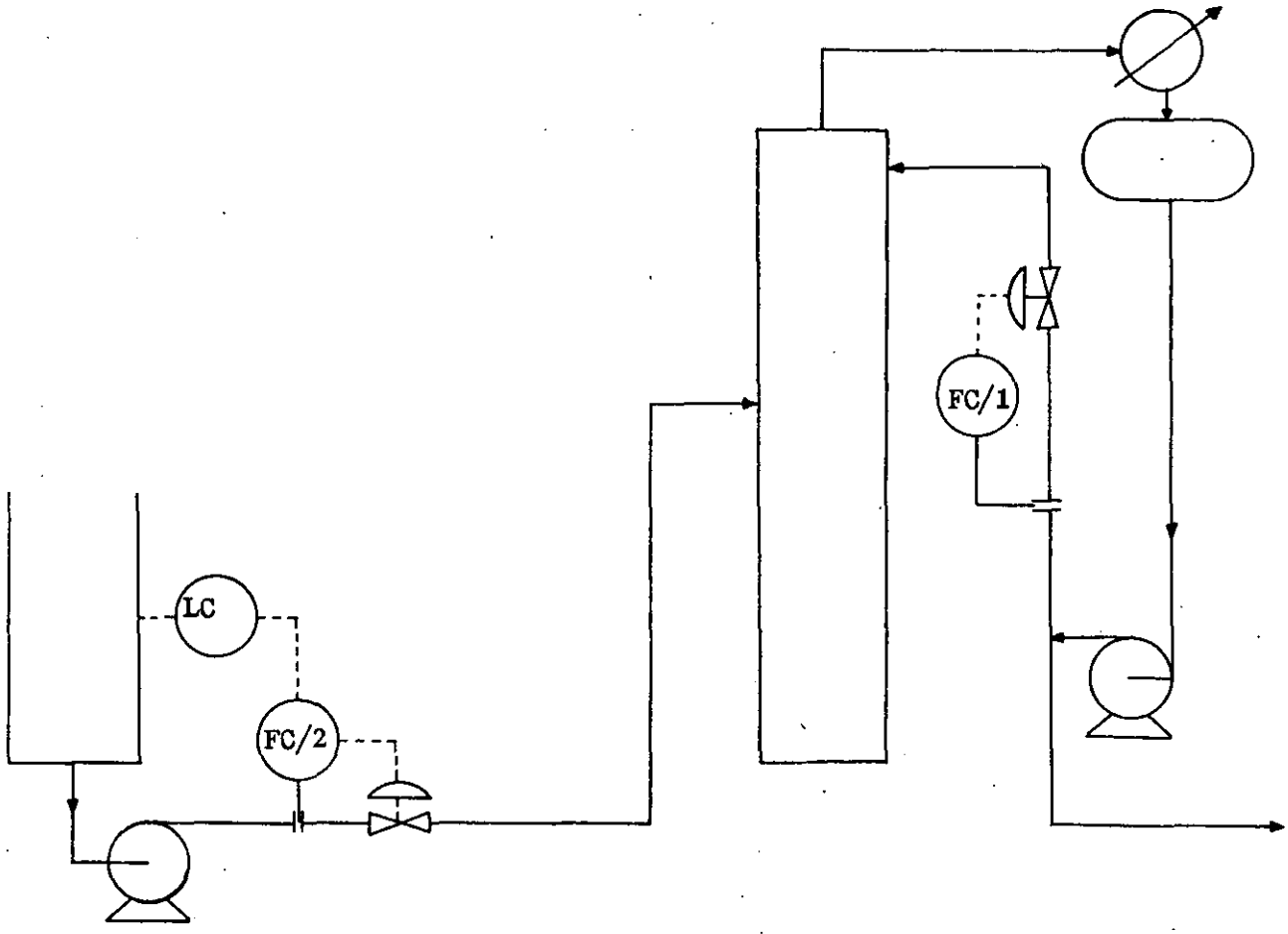


Figure 2.50. Arrangement of industrial control loops FC/1 and FC/2.

For these experiments the process computer at Works A was programmed to store the flowmeter and valve position measurements for each loop at 1 minute intervals on process operator demand. This stored data was then retrieved on to computer punched paper tape and analysed off-line on a PDP 11-20 computer at Loughborough using the malfunction detection algorithm.

2.10.1 Determination of model parameters at Works A.

In order to implement the malfunction detection method via the "tracking" state estimator of equation (2.6.1.2) the measurement equation (2.6.2) must initially be defined, i.e.:

$$\underline{y} = \underline{h}(\underline{x}) + \underline{v}$$

The flowmeter equation was obtained from process design manuals for both control loops FC/1 and FC/2 to be:

$$Q_m = k_{mi}\, s \qquad i = 1,2$$

The quantity s is the measurement used by the computer in its d.d.c. algorithm. The constants k_{m_i} for the loops FC/1 and FC/2 are given in Table 2.14.

Control loop	Flowmeter measurement equation	k_{m_i} $\frac{K\,lb}{h\,V}$
FC/1	$Q_m = k_{m1}\, s$	107.5
FC/2	$Q_m = k_{m_2}\, s$	300.8

TABLE 2.14. Equations and parameters defining the industrial flowmeters.

The information concerning the control valve characteristics was available in the process design manuals and typically is shown in Table 2.15.

Flow control loop	Nominal valve body size (in)	Valve type	C_v Factor	Design Δ P p.s.i.	C_v Normal flow	Normal flow K lb/h	Specific gravity	Valve lift at normal flow - % open
FC/1	3	Equal percentage	140	10	65	79.5	0.59	80
FC/2	4	"	193	17	125	219	0.7	63

TABLE 2.15. Process design specification of the industrial control valves.

To formulate a measurement equation relating the valve position measurement to the fluid flowrate an equal percentage valve characteristic was assumed of the form:

$$Q_v = k_{v1} \exp(k_{v2}\, x) \qquad (2.10.1)$$

Now the information presented in Table 2.15 enables the constants k_{vi} to be determined as follows.

Consider FC/1, then the flowrate when the valve is fully open, i.e. $x = 1.0$ is:

$$Q_{vmax} = \frac{140}{65} \times 79.5 \quad \text{Klb/h}$$

$$= 171.23 \text{ Klb/h}$$

This results in two "points" on the equal percentage valve characteristic corresponding to $x = 1$ and $x = 0.80$. Thus the constants of equation (2.10.1) may be solved simultaneously.

This procedure results in the control valve characteristics defined in Table 2.16.

Flow Control loop	Type of characteristic	k_{v1} Klb/h	k_{v2} 100%	x_{max} %/100
FC/1	$Q_v = k_{v1} \exp(k_{v2} x)$	3.7	3.84	1
FC/2	$Q_v = k_{v1} \exp(k_{v2} x)$	104.5	1.17	1

TABLE 2.16. Equations and parameters defining the industrial control valves.

The final parameter necessary to define the measurement model is the matrix R. Initially nothing was known about the process measurement noise or the accuracy and validity of the assumed control valve equations. Thus an a priori estimate of R was made by assuming $r_{ij} = 0$, $i \neq j$, and setting r_{ii} to be 5% of the maximum values of x and s respectively, giving:

$$R = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}$$

2.10.2 Industrial experiments and objectives

The experiments were designed to investigate several features of the malfunction detection method. These were:

i) To examine the robustness of the estimator;

ii) To examine the validity of the a priori flow control loop models derived in section 2.10.1;

iii) To determine the feasibility of estimating the state via the "tracking" state estimator in industrial d.d.c. flow loops;

iv) To determine the estimator residual characteristics over a period of time;

v) To verify that a control loop malfunction may be detected by examining the estimator residual mean.

To facilitate these experiments the flowmeter and valve position measurements for loops FC/1 and FC/2 were logged at 1 minute intervals over a period of 1 week. Typical measurement time histories are shown in Figures 2.51 a, b and 2.52 a, b for loops FC/1 and FC/2 respectively. There was no indication from the process operators that the loops had malfunctioned during this period.

The logged data was processed off-line on a PDP 11-20 computer. The data for each loop was decomposed into batches of 250 samples and analysed using the "tracking" state estimator of equation (2.6.1.2) to yield the state estimate $\hat{\underline{x}}$, the normalised residual time sequences, $r_{i/r_{ii}^{\frac{1}{2}}}$ $(i = 1, 2)$, and the means of the normalised residual time sequence. The data batches are coded IT/1 a-h and IT/2 a-h for loops FC/1 and FC/2 respectively.

To investigate the change in residual mean when loop malfunction occurs a -10% zero error in the flowmeter signal was simulated and is

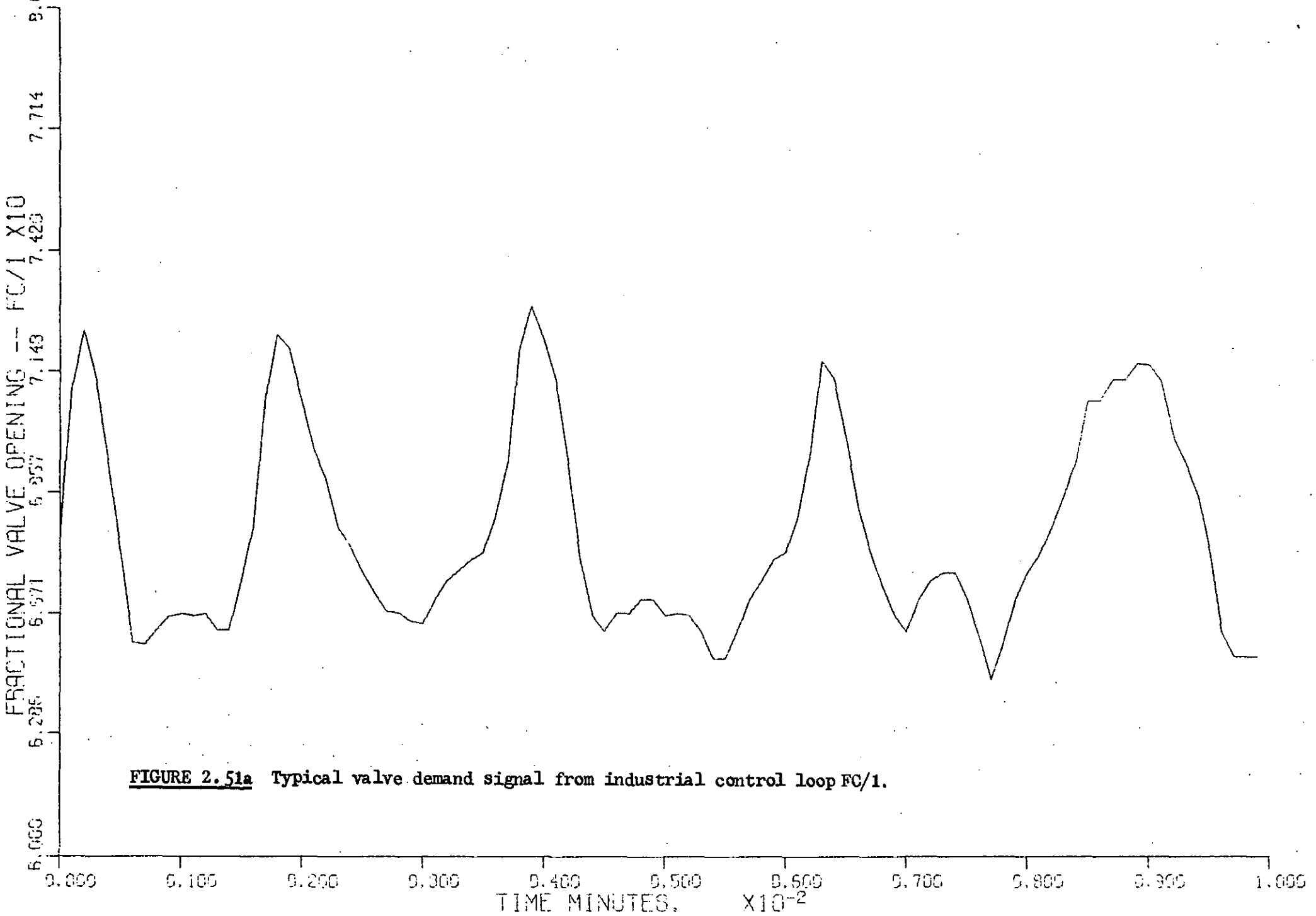


FIGURE 2.51a Typical valve demand signal from industrial control loop FC/1.

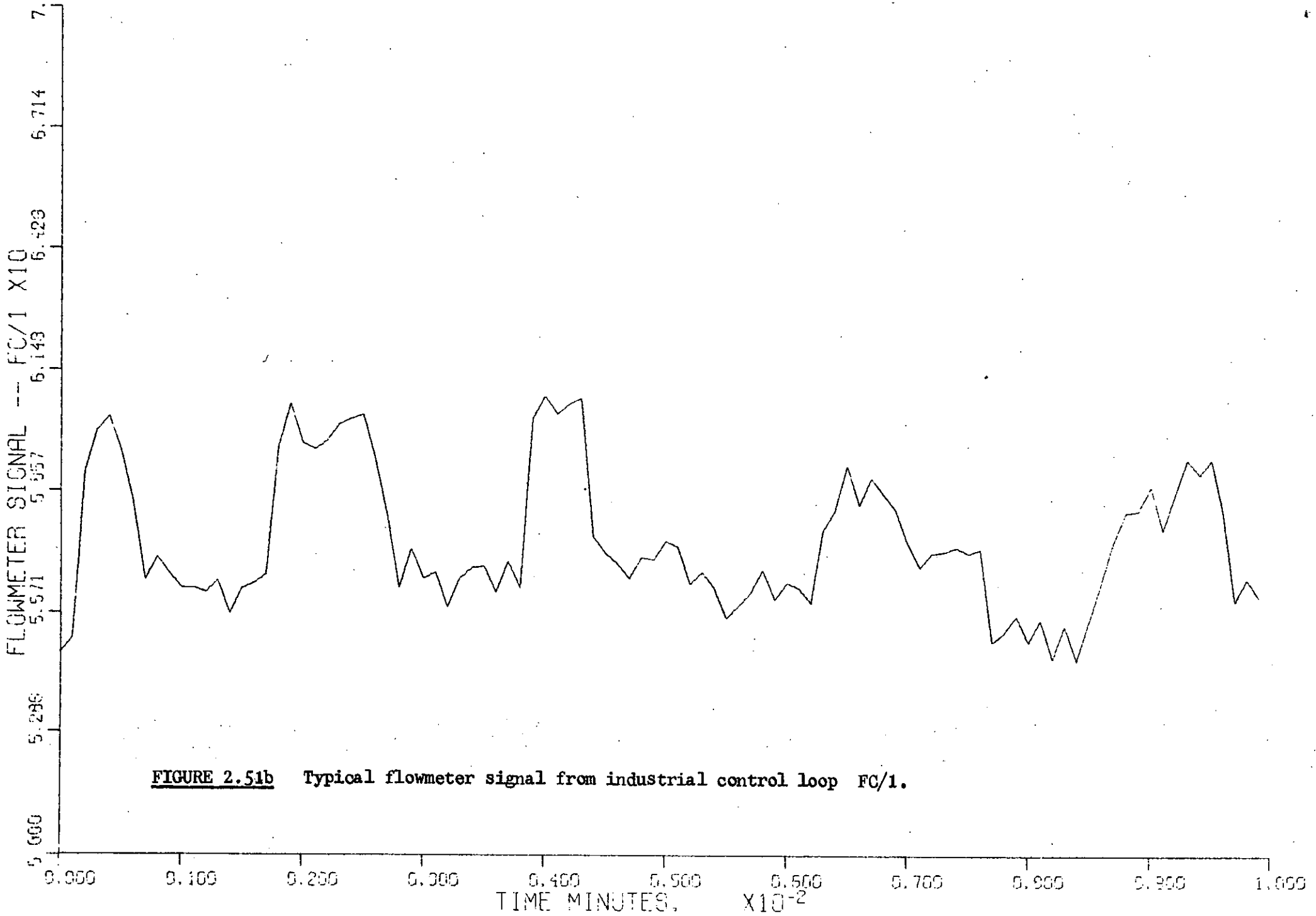


FIGURE 2.51b Typical flowmeter signal from industrial control loop FC/1.

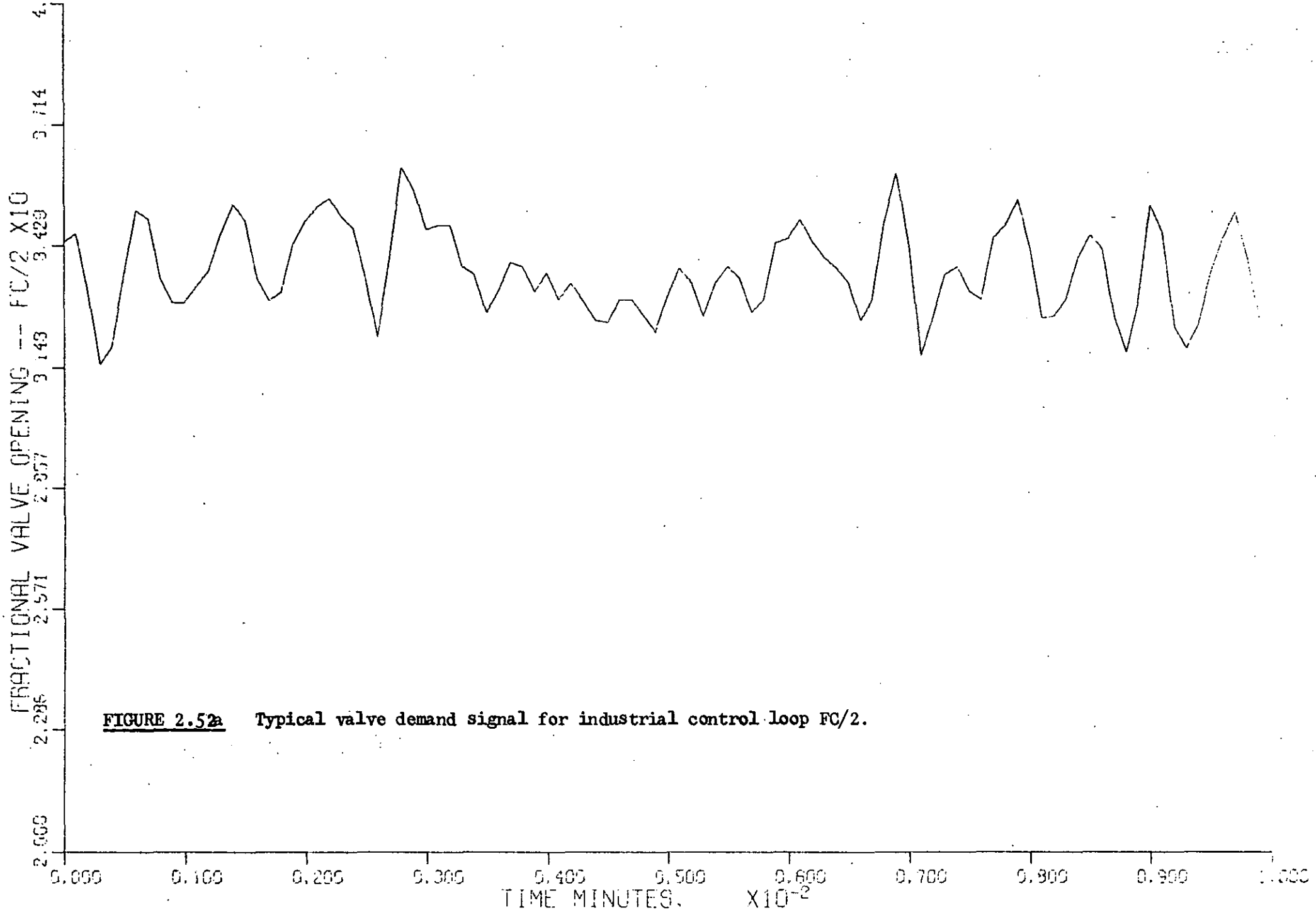
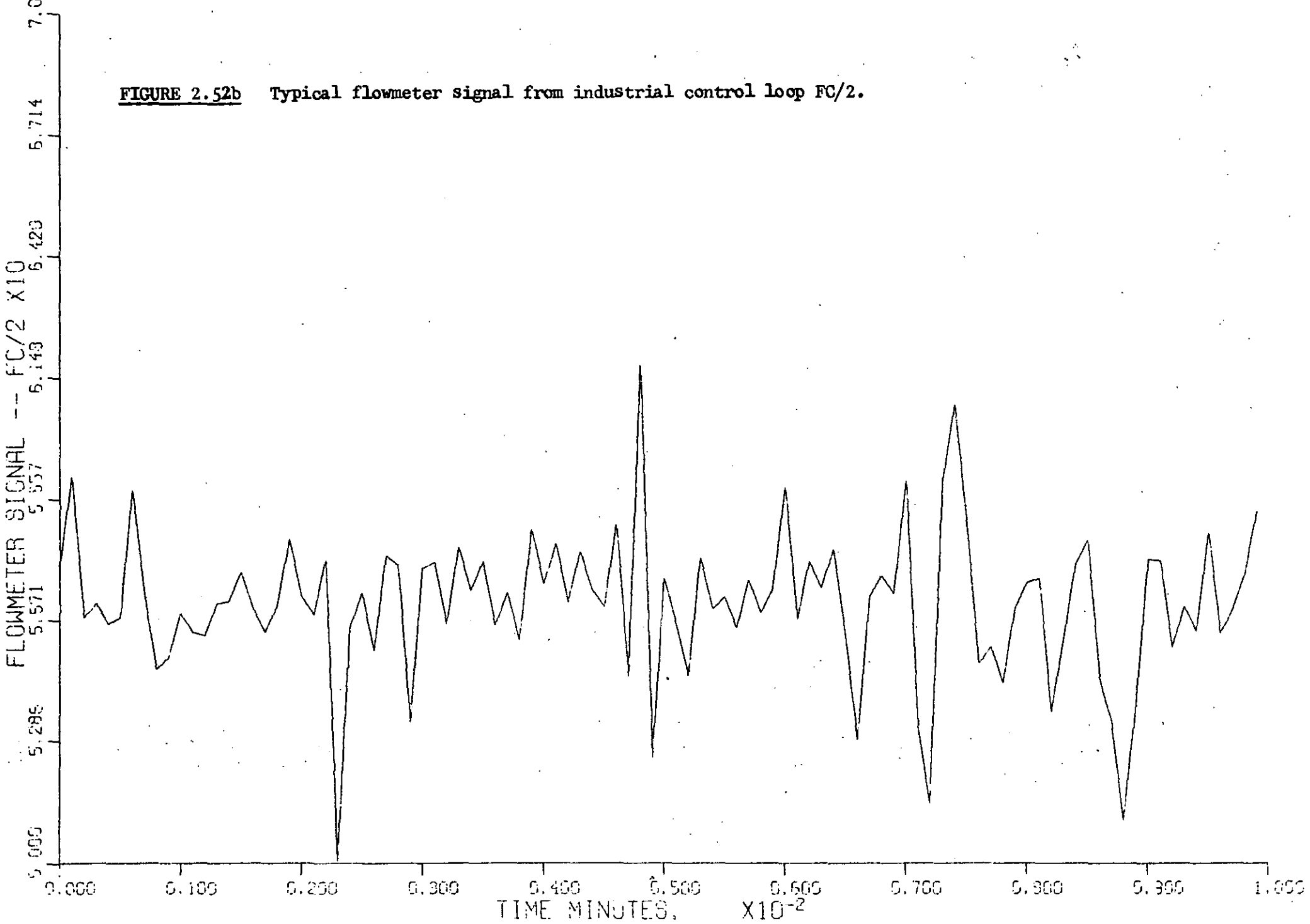


FIGURE 2.52a Typical valve demand signal for industrial control loop FC/2.

FIGURE 2.52b Typical flowmeter signal from industrial control loop FC/2.



denoted by run IT/kA, (k = 1, 2). The modulus of the change in residual mean was calculated and compared with the changes in the loop residual observed over the experimental test period.

2.10.3 Industrial experimental results

Since these flow control loops operated under d.d.c. from the process computer with limited valve travel, and valve stroking was not feasible, then the malfunction detection technique must be based upon examining the characteristics of the residual sequences.

Table 2.17 shows the normalised residual mean for control loops FC/1 and FC/2 over the logging period when the a priori models of section 2.10.2 were used in the estimator.

Table 2.17 shows that the residual means for both flow control loops, assumed without malfunction, are relatively consistent within groups.

Run IT/2A$^{*}_{a}$ which has a -10% zero error in the flowmeter signal s results in a shift of valve position residual mean by approximately 1.0 which clearly demonstrates the presence of a loop malfunction.

There is, however, one disturbing feature of the results in Table 2.17. This is that the loops without malfunction do not yield residual means close to zero. Although this is not a pre-requisite of malfunction detection since the criterion for a loop malfunction is that the modulus of the residual mean shifts from the malfunction free characteristic, it is considered that a nominal zero residual mean for the loop without malfunction is a desirable feature. This is postulated as an aid to result interpretation for the process operator or control engineer, since the information load from monitoring several loops is reduced to the common feature of zero residual mean (1). This desirable

Flow control loop	Sample number	Run number	Normalised residual mean	
			Valve position	Flowmeter signal
FC/1	1 - 250	IT/1 a	-0.87	0.42
	251 - 492	IT/1 b	-0.90	0.46
	743 - 1242	IT/1 c	-0.89	0.44
	2022 - 3211	IT/1 d	-0.84	0.42
	5132 - 5631	IT/1 e	-0.89	0.42
	7136 - 7635	IT/1 f	-0.91	0.46
	8972 - 9221	IT/1 g	-0.83	0.44
	10225 - 10421	IT/1 h	-0.88	0.47
FC/2	1 - 250	IT/2 a	-0.41	0.65
	251 - 492	IT/2 b	-0.53	0.88
	743 - 993	IT/2 c	-0.57	0.92
	1494 - 1743	IT/2 d	-0.58	0.94
	2272 - 2521	IT/2 e	-0.55	0.87
	2772 - 3021	IT/2 f	-0.60	0.94
	5181 - 5680	IT/2 g	-0.38	0.67
	9071 - 9320	IT/2 h	-0.40	0.69
	1 - 250	IT/2A* a	+0.48	-0.81

TABLE 2.17. Behaviour of residuals in industrial d.d.c. flow loops using a priori measurement equations.

feature may be obtained simply by updating the a priori measurement models. This may be done as follows.

Table 2.18 details typical information generated by the state estimator using the a priori models of section 2.10.2. An examination of these results shows that there is a large difference in Q_v and Q_m which indicates that the control valve characteristic is in error thus producing large residuals in the estimator.

Now the philosophy adopted in this Chapter has not been the achievement of accurate state estimates but simply the creation of a mathematical model relating the measured variables in the control loop, from which loop malfunction could be detected.

With these comments in mind and the fact that the control valve characteristics derived from the process design manual are suspect, it was considered justifiable to modify the a priori model in order to achieve approximately zero residual mean when the loop is malfunction free.

The control loop measurement equations were intuitively modified and the a posteriori parameters are given in Table 2.19.

Flow control loop	Flowmeter constant	Control valve constants		R matrix	
	k_{mi} $\frac{Klb}{hV}$	k_{v1} $\frac{Klb}{h\ V}$	k_{v2} $\frac{100}{\%}$	r_{11}	r_{22}
FC/1	90	3.7	3.84	0.05	0.1
FC/2	300.8	104.5	1.45	0.05	0.05

TABLE 2.19. Parameters defining a posteriori measurement models for industrial control loops.

Flow control loop	Valve opening	Flowmeter signal	$Q_v = k_{v1}\exp(k_{v2}\, x)$	$Q_m = k_{mi}\, s$	$\hat{Q} = \hat{x}$	Optimal cost $\hat{J}$	Normalised valve position residual	Normalised flowmeter signal residual
FC/1	0.6532	0.5608	45252.6	60288.2	57055.1	1.847	- 1.14	0.74
	0.6615	0.5559	46717.5	59763.1	57020	1.34001	- 1.04	0.50
	0.6649	0.5584	47334.3	60025.3	57375.6	1.254	- 0.970	0.56
	0.6649	0.5545	47334.3	59605.5	57049.9	1.178	- 1.0	0.42
	0.6688	0.5662	48049.3	60865.9	58220	1.30	- 0.895	0.71
FC/2	0.3381	0.5696	155453	17134	159874	0.81	- 0.56	0.71
	0.34	0.5564	15581	167372	159012	0.44	- 0.44	0.50
	0.3312	0.5628	154210	169283	158348	0.74	- 0.52	0.7
	0.3205	0.57	152276	171487	157499	1.21	- 0.67	0.87
	0.3351	0.5623	154919	169136	158802	0.67	- 0.28	0.77

TABLE 2.18. Tracking state estimator information using a priori measurement models for industrial control loops.

Table 2.20 illustrates the information generated by the estimator using the a posteriori measurement models and Table 2.21 details a residual mean analysis, which confirms that a loop malfunction is characterised by a large shift of residual mean

Flow control loop	Sample number	Run number	Normalised residual mean	
			Valve position	Flowmeter signal
FC/1	0 - 400	IT/1 i	- 0.067	0.071
	493 - 1492	IT/1 j	- 0.074	0.078
	2022 - 3021	IT/1 k	- 0.05	0.056
	7136 - 8135	IT/1 l	- 0.096	0.11
	8972 - 9971	IT/1 m	- 0.092	0.10
	9723 - 9822	IT/1A* b	- 0.63	0.584
FC/2	0 - 400	IT/2 i	- 0.012	0.018
	593 - 1342	IT/2 j	- 0.11	0.145
	1772 - 2521	IT/2 k	- 0.13	0.17
	5181 - 5930	IT/2 l	- 0.067	0.09
	8972 - 9971	IT/2 m	- 0.11	0.16
	5732 - 6731	IT/2A* b	- 1.05	1.32

TABLE 2.21. Behaviour of residuals in industrial d.d.c flow loops using a posteriori measurement equations.

2.11 Concluding remarks

This Chapter has described a control valve position - flow check which is intended to be implemented on a d.d.c. computer. The technique appears promising as a means of detecting the existence of various types of malfunction in flow control loops, though not of identifying the particular malfunction except in certain specific cases.

The method does not require additional process instrumentation, but exploits the capability of the process computer to condition, store, and display information. The computer storage and time requirements

Flow control loop	Valve opening	Flowmeter signal	Q_v	Q_m	$\hat{Q}$	Optimal cost $\hat{J}$	Normalised valve position residual	Normalised flowmeter signal residual
FC/1	0.6414	0.5476	43262.5	49286.8	46230.	0.41	- 0.64	0.04
	0.6092	0.4978	38229.4	44802.1	41081.1	0.48	- 0.68	0.13
	0.6580	0.5501	46108.7	49506.60	47834.8	0.08	- 0.27	0.1
	0.6849	0.5921	51114.6	53287.7	52316.7	0.11	- 0.32	- 0.12
	0.7123	0.5921	56770.2	53287.7	54736.5	0.07	0.13	- 0.23
FC/2	0.3449	0.5852	172356.	176042.	173875.	0.036	- 0.14	0.12
	0.3869	0.5823	183182.	175161.	179664.	0.29	- 0.012	- 0.5
	0.3483	0.5818	173213.	175014.	173956.	0.0099	- 0.089	0.0452
	0.3402	0.5745	168732.	172809.	170374.	0.085	- 0.29	0.032
	0.3376	0.5887	170534.	177071.	173196.	0.123	- 0.13	0.32

TABLE 2.20. Tracking state estimator information using a posteriori measurement models for industrial control loops.

are relatively modest and since most malfunctions appear not to develop very suddenly the check algorithm can be executed at quite infrequent intervals and at low priority.

The technique does not assume a constant pressure drop across the valve but it does assume a constant system flow - pressure drop characteristic. If this condition is not completely met, the level of detectable malfunction will be increased.

It is envisaged that the technique would be used for the detection of relatively gross malfunction which may affect immediate operation of the process rather than the adjustment of fine errors in data sets for subsequent management analysis. The level of error which might be detected in practice is probably about 10%.

Two basic types of check have been described. In the first, a valve characteristic is obtained of flowrate vs valve position. This requires that the valve be moved over its whole range of travel. It is expected that usually this will not be acceptable. However, if this is permitted, it is possible to obtain a full comparison of the original and current characteristic, and if desired to give a visual display. This full valve characteristic comparison enables the source of the malfunction to be determined in some cases.

A second check is based upon the comparison of the original and current residuals generated from a state estimator. This is applicable to all cases including closed loop d.d.c. with limited valve travel and it is anticipated that this check would be more useful.

The method has been tested using a laboratory rig which demonstrated the points discussed above.

Some limited industrial trials have also been conducted. These revealed that models based upon process design manuals were adequate and no serious difficulty was experienced in tracking the operation of control loops and calculating values of residuals.

It is suggested that future work should continue the industrial trials of the method in an attempt to correlate actual flow control loop malfunction to the change in residual, while theoretical development should investigate how the method may be adapted to handle control loops which do not satisfy the assumption of constant system flow-pressure drop characteristic.

CHAPTER 3.

THE DETECTION OF MALFUNCTION IN A GENERAL CONTROL LOOP

3.1 List of symbols

A	cross sectional area	m^2
$A(k+1, k)$	state transition matrix	-
$A_b(k+1, k)$	state transition matrix	-
$\underline{a}$	vector of unknown parameters	-
$\underline{a}_i$	i^{th} vector of unknown parameters	-
$B(k)$	control driving matrix	-
B	measured variable deviation	various
B_T	true measured variable deviation	various
$\underline{b}$	vector of unknown parameters	-
b	unknown scalar parameter	V
$\hat{\underline{b}}(/)$	conditional expectation of $\underline{b}$	-
$C(t)$	measurement matrix	
C_k	innovation autocorrelation matrix	-
$(C_k)_{ij}$	ij^{th} element of C_k	
C	controlled variable deviation/instrument transmitter	-
$\underline{c}$	vector of unknown parameters	-
$E(\ )$	expectation / error	-
$E(/)$	conditional expectation	-
E_1; E_2	error	various
$\bar{E}(s)$	laplace transform of error	-
$\underline{e}$	error vector / Gaussian white noise sequence	-
$F(k+1, k)$	augmented state transition matrix	-
F	continuous time transition matrix	-
G	transfer function	-
$H(k)$	measurement matrix	-
$H_b(k)$	measurement matrix	-
H	tank height deviation	cm

$\bar{H}(s)$	Laplace transform of H	-
$\bar{H}_d(s)$	Laplace transform of tank height setpoint deviation	-
$\bar{H}_m(s)$	Laplace transform of tank height deviation measurement	-
$\underline{h}_i(k)$	i^{th} row of measurement matrix H(k)	
h	tank height	cm
I	unit matrix	-
J	cost function	-
j	smoothing interval	
K(k)	Kalman gain matrix	-
$K_b(k)$	Bias filter gain matrix	-
K_c	controller gain	various
$K_1, K_2, K_3, K_{31}, K_{32}$, K_4, K_f, K_m, K_L, K_v, K_{o1}, $K_{o2,h}$, $K_{o2,x}$	constants in level control loop transfer functions	various
ΔK	incremental change in Kalman gain matrix	-
k	discrete time counter	-
L	covariance matrix/augmented measurement matrix	-
l	discrete time counter	-
M	covariance matrix/control valve demand signal	-
M_r	weighted sum of residuals	various
m	number of measurements	-
N	memory length/counter	-
n	discrete time counter/number of state variables	-

P	pressure	kN/m^2
$\bar{P}(s)$	Laplace transform of controller output pressure	-
$P(\ )$	probability	-
$P(\ /\ )$	conditional error covariance matrix	-
ΔP	pressure drop	kN/m^2
$Q(k)$	process noise covariance matrix	-
Q	flowrate deviation	l/min
$\bar{Q}(s)$	Laplace transform of flowrate deviation	-
q	flowrate/number of process disturbances/scalar process noise covariance	-
$R(k)$	measurement noise covariance matrix	-
R	control loop setpoint deviation	various
R_c	computer generated measurement noise covariance matrix	-
$r_{ij}(k)$	ij^{th} element of $R(k)$	-
$r(k)$	residual	various
$S(k)$	innovation covariance matrix	-
S	scaling parameter/matrix	-
$s_{ij}(k)$	ij^{th} element of $S(k)$	-
s	scaling parameter/Laplace operator	-
t	time/student's t	various
Δt	sampling interval	various
U	matrix/load disturbance deviation	-
$\underline{u}(k)$	control vector	-
V	matrix	-
$\underline{v}$	measurement noise vector	-
$\underline{w}(k)$	vector of process disturbances	-
X	control valve stem position deviation	V

$\bar{X}(s)$	Laplace transform of valve stem position deviation	-
x	control valve stem position	v
$\underline{x}(k)$	state vector at time interval k	-
$\underline{\hat{x}}(/)$	conditional expectation of state vector $\underline{x}(k)$	-
$x_i(k)$	i^{th} element of $\underline{x}(k)$	-
$Y(k)$	sequence of measurement vectors $\underline{y}(1)$ to $\underline{y}(k)$	-
$\underline{y}(k)$	vector of process measurements at time interval k	-
$y_i(k)$	i^{th} element of $\underline{y}(k)$	-
$Z(k)$	estimated innovation covariance matrix at time interval k	-
$z_{ij}(k)$	ij^{th} element of $Z(k)$	-
$\underline{z}(k)$	augmented state vector	-
$\underline{\hat{z}}(/)$	conditional expectation of $\underline{z}(k)$	-
z	z transform operator	-
$\|\cdot\|$	matrix norm	-

Greek letters

α	scaling factor	-
Γ	process noise driving matrix	-
δ	Kronecker delta	-
δP	difference in error covariance matrices	-
$\underline{\theta}$	vector of unknown parameters	
$\underline{\mu}$	mean	various
$\underline{\nu}$	innovation sequence	-
ρ_k	normalised innovation autocorrelation matrix	-

$(\rho_k)_{ij}$	ij^{th} element of ρ_k	-
τ_p	process time constant	min
τ_I	controller integral time	min
$\emptyset$	control loop malfunction gain	
		various

Subscripts

b	parameter vector $\underline{b}$
c	parameter vector $\underline{c}$
i	inlet flowrate
opt	optimal
min	minimum
max	maximum
ss	steady state
x	state vector $\underline{x}$
01	outlet flowrate 1
02	outlet flowrate 2

Superscripts

'	optimum state estimate
-1	inverse
#	pseudo inverse
T	transpose
^	estimate

3.2 Introduction

Chapter 2 has suggested a malfunction detection algorithm which may be implemented on a process computer to determine faults in a flow control loop.

The method was based upon the knowledge of the steady state control valve characteristic relating the process flowrate to the valve demand signal, as well as the primary flowmeter measurement relationship. Thus it was postulated that the flow control loop had inherent measurement redundancy, which was used to detect malfunction. Two techniques were presented which could be applied to open and closed flow loop operation respectively.

The closed loop technique was based upon a simple tracking state estimator derived under the assumptions of steady state operation (or fast process time constants). The technique yielded no diagnostic information and the calculated state estimates were not constrained to be accurate.

In this Chapter it is assumed that a relationship between the process variable and the control valve demand signal exists but is initially unknown. A malfunction detection technique based upon a dynamic closed loop model is proposed which yields "optimal" state estimates when the loop is malfunction free and which also provides some diagnostic information on the location of loop malfunctions when they occur.

The method is based upon Kalman filtering and may be applied to both conventional analogue setpoint and direct digital (d.d.c.)control loops.

To illustrate the proposed algorithm, malfunctions are detected in a laboratory level control loop.

3.3 The control loop malfunction check

In this section two types of control loop are distinguished and the effects of malfunction on their performance are examined.

3.3.1 Conventional analogue setpoint control

Although d.d.c. process computers are finding increasing application on chemical plant, there are still many control loops which are based upon conventional analogue controllers. The role of the computer in such situations is dependent upon the system design. However, it is not uncommon for the computer to access the measured variable in a control loop as a primary input to plant performance, optimisation or management calculation computer programs.

A convenient representation of control loops is the block diagram and Figure 3.1 shows a negative feedback loop whose task is to maintain the process variable C at some desired setpoint R in spite of the load disturbances U. The computer accesses the measured variable B at discrete time intervals via a P/I transmitter and is denoted $y_1(k)$.

Now for a control loop operating at a particular setpoint R, with load disturbances about some nominal value U, there exists a unique relationship between the measured variable B ($\simeq$ R) and the control valve demand signal M (controller output signal or the valve stem position). If this relationship changes, in the absence of setpoint or load changes, then loop malfunction has occurred.

In the control loop of Figure 3.1 there are two types of malfunction which can occur.

The first type of error is a fault in the P/I transmitter. This

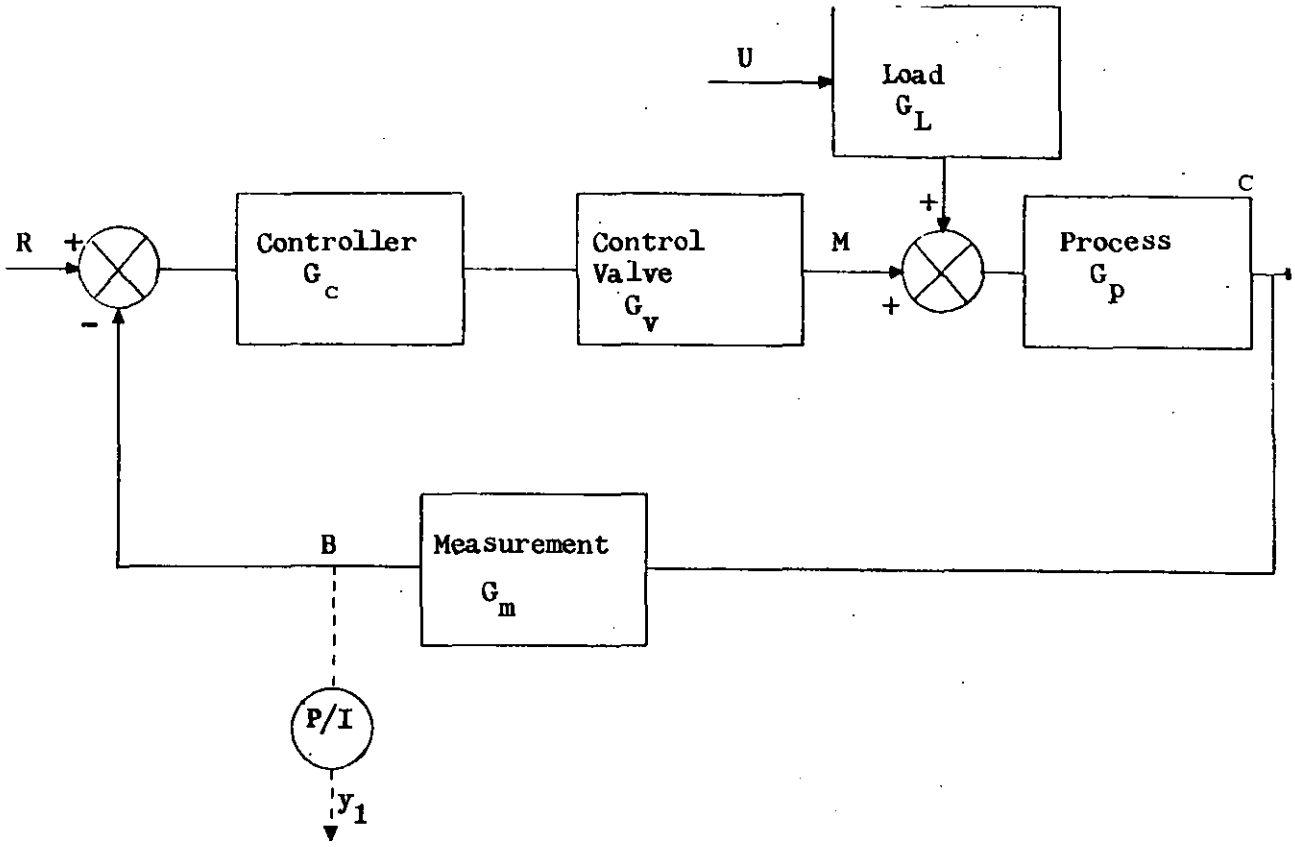


FIGURE 3.1 Block diagram representation of an analogue setpoint control loop.

instrument is external to the control loop and so such a malfunction does not affect control loop operation, i.e. $B \simeq R$ and for the given R and U, M is the expected value. However the computer observation y_1 will be biased and so the relationship between y_1 and M will be altered. Subsequent process calculations involving y_1 will be in error. This form of error will be referred to as a pure measurement error.

The second type of malfunction is an error occurring within the individual blocks of the control loop. This kind of fault does affect loop operation - the end result of which is to cause a change in the relationship between B and M. As an example, consider a malfunction in the measurement block G_m which causes the output B to have a positive error. Denoting the true process measurement as B_T, and the error E(B), then because of the feedback nature of the loop it is

easy to see:

$$R = B = B_T + E(B) \qquad (3.3.1.1)$$

Now an observer, or the process computer, will still register the process measurement as B although the true value is B_T. Since the control valve is essentially being asked to handle different process conditions the valve demand signal will change to some new nominal value $M + E(M)$. This apparent divergence between the expected values of B and M, in the absence of setpoint or load changes, is indicative of loop malfunction.

The precise change in the control valve demand signal E(M) is dependent upon the characteristics of the control loop, and may be derived using conventional block diagram techniques.

For example, the measurement error E(B) may be considered as a load disturbance entering the block diagram after the measurement block G_m.

The transfer function relating the valve demand signal to this load change is (56):

$$\frac{\bar{M}(s)}{\bar{E}(B)(s)} = \frac{-G_c G_v}{1 + G_c G_v G_p G_m}$$

Now by applying the final value theorem to this transfer function (56) the steady state deviation E(M) of the valve demand signal from the nominal value M is:

$$E(M) = \phi \; E(B)$$

where ϕ depends upon the control loop system gains.

The parameter ϕ characterises the magnitude of the malfunction which is detectable in any particular loop.

3.3.2 Direct digital control

The difference between this type of loop and that detailed in Figure 3.1 is that the P/I transmitter becomes part of the feedback control loop and the controller becomes the process computer. Hence there are no pure measurement errors in this loop and consequently all malfunctions affect control loop operation, thereby causing deviations in the expected valve demand signal for a given process measurement in the absence of setpoint or load changes.

The ideas developed in this section have as their basis that, for a given setpoint and load, there is an expected relationship between the process measurement and the valve demand signal. The remainder of this Chapter is devoted to developing an algorithm which will detect malfunction during normal control loop operation. The control loop is represented by a mathematical model which provides a data base from which inferences concerning loop malfunction are made.

3.4 Review of Kalman filtering

The basis for modern estimation theory is the method presented by Kalman (46). His solution to the recursive linear estimation problem has become known as the Kalman filter. Kalman's method has been described extensively in the literature and in addition to Kalman's original paper derivations may also be found in (27), (45) and (57). A statement of the linear estimation problem and the resulting filtering equations which form the solution to the problem are presented in this section.

Consider a linear system whose dynamics are modelled by a linear vector difference equation:

$$\underline{x}(k+1) = A(k+1,k)\,\underline{x}(k) + B(k)\,\underline{u}(k) + \Gamma(k)\underline{w}(k) \qquad (3.4.1)$$

where $\underline{x}(k)$ = n x 1 vector of state variables

$A(k+1,k)$ = n x n state transition matrix

$\underline{u}(k)$ = r x 1 vector of control inputs

$B(k)$ = n x r control driving matrix

$\underline{w}(k)$ = q x 1 vector of dynamic system noise variables

$\Gamma(k)$ = n x q noise driving matrix.

The noise sequence $\{\underline{w}(k)\}$, is assumed to be white noise with statistics given by:

$$E(\underline{w}(k)) = \underline{0}$$

$$E(\underline{w}(k)\,\underline{w}(j)^T) = Q(k)\,\delta(k,j)$$

where $Q(k)$ is the covariance matrix of $\{\underline{w}(k)\}$ and $\delta(k,j)$ is the Kronecker delta, ie.

$$\delta(k,j) = \begin{cases} 0 & k \neq j \\ 1 & k = j \end{cases}$$

The initial state $\underline{x}(0)$, is considered to be a vector of random variables with statistics known to be:

$$E(\underline{x}(0)) = \hat{\underline{x}}(0)$$

$$E(\underline{w}(k)\,\underline{x}(0)^T) = 0 \quad \text{for all } k$$

and $$E\left((\underline{x}(0) - \hat{\underline{x}}(0))(\underline{x}(0) - \hat{\underline{x}}(0))^T\right) = P(0/0)$$

where $P(0/0)$ is the covariance matrix of $\underline{x}(0)$.

At each time instant, k, the available measurements are modelled by

$$\underline{y}(k) = H(k)\,\underline{x}(k) + \underline{v}(k) \qquad (3.4.2)$$

where:

$\underline{y}(k)$ = m x 1 vector of measurement variables

$H(k)$ = m x n measurement weighting matrix

$\underline{v}(k)$ = m x 1 vector of measurement noise variables.

The additive noise sequence, $\{\underline{v}(k)\}$, is assumed to have the following statistics:

$$E(\underline{v}(k)) = \underline{0}$$

$$E(\underline{v}(k)\,\underline{v}(j)^T) = R(k)\ \delta(k,j)$$

where $R(k)$ is the covariance matrix of $\{\underline{v}(k)\}$.

The noise covariance matrices, $Q(k)$ and $R(k)$, are assumed to be positive semidefinite and positive definite respectively. In addition, $\{\underline{v}(k)\}$ and $\{\underline{w}(k)\}$ are assumed to be uncorrelated, i.e.,

$$E(\underline{w}(k)\,\underline{v}(j)^T) = 0 \quad \text{for all } k,j$$

and $$E(\underline{x}(o)\,\underline{v}(k)^T) = 0 \quad \text{for all } k$$

The mathematical model has now been defined and so the estimation problem will be stated.

Recursive Linear Estimation Problem

Given the model described by equations (3.4.1) and (3.4.2), determine an estimate of the state at time $k+1$ which is a linear combination of an estimate at time k and the measurement data $\underline{y}(k+1)$ such that the following criterion is minimised:

$$J = E\left((\hat{\underline{x}}(k+1) - \underline{x}(k+1))^T (\hat{\underline{x}}(k+1) - \underline{x}(k+1)) \right)$$

Kalman (46) showed that the optimal estimate is given by:

$$\hat{\underline{x}}(k/k) = E(\underline{x}(k) / Y(k))$$

where $Y(k) = (\underline{y}(1), \ldots, \underline{y}(k))$

Furthermore, it was shown that the optimal estimate can be generated by the following set of recursive equations, which combine to give the Kalman filter .

$$\hat{\underline{x}}(k/k-1) = A(k,k-1)\,\hat{\underline{x}}(k-1/k-1) + B(k-1)\,\underline{u}(k-1) \qquad (3.4.3)$$

$$P(k/k-1) = A(k,k-1)\,P(k-1/k-1)\,A^T(k,k-1) + \Gamma(k-1)\,Q(k-1)\,\Gamma^T(k-1) \qquad (3.4.4)$$

$$K(k) = P(k/k-1)\, H^T(k)\, (H(k)\, P(k/k-1)\, H^T(k) + R(k)\,)^{-1} \qquad (3.4.5)$$

$$\hat{\underline{x}}(k/k) = \hat{\underline{x}}(k/k-1) + K(k)\, (\underline{y}(k) - H(k)\, \hat{\underline{x}}\,(k/k-1)\,) \qquad (3.4.6)$$

$$P(k/k) = (I - K(k)H(k)\,)\, P(k/k-1)(I - K(k)\, H(k))^T + K(k)\, R(k)\, K^T(k) \qquad (3.4.7)$$

or $P(k/k) = P(k/k-1) - K(k)\, H(k)\, P(k/k-1)$

where

$\hat{\underline{x}}\,(k/k-1) = E(\underline{x}(k)) \,/\, Y(k-1))$

$K(k)$ = Kalman gain matrix

$P(k/k-1)$ = covariance matrix of $(\underline{x}(k) - \hat{\underline{x}}(k/k-1))$

$P(k/k)$ = covariance matrix of $(\underline{x}(k) - \hat{\underline{x}}(k/k))$

The feedback structure of the filter is shown in Figure 3.2, and it is this structure which makes the Kalman filter a very useful tool since it can be realised using a digital computer.

The quantity:

$$\underline{y}(k) - H\, \hat{\underline{x}}\,(k/k-1)$$

which appears in equation (3.4.6) is used extensively in Kalman filtering. Kailath (58) defined this sequence as the innovation sequence and this will be denoted here as $\underline{\vee}(k)$.

Inspection reveals that $\underline{\vee}(k)$ is the difference between the actual process measurements and the predicted measurement, and as such represents the new information available to the filter at each iteration.

Kailath has shown that the innovation sequence for a filter using correct information (i.e. correct system models and noise statistics) with Gaussian white noise inputs is a Gaussian white noise sequence with statistics:

$$E\,(\underline{\vee}(k)) = \underline{0}$$

$$E\,(\underline{\vee}(k)\, \underline{\vee}^T(j)\,) = 0 \qquad j \neq k$$

$$E\,(\underline{\vee}(k)\, \underline{\vee}^T(j)\,) = H\,(k)\, P\,(k/k-1)\, H^T(k) + R(k) \qquad j = k$$

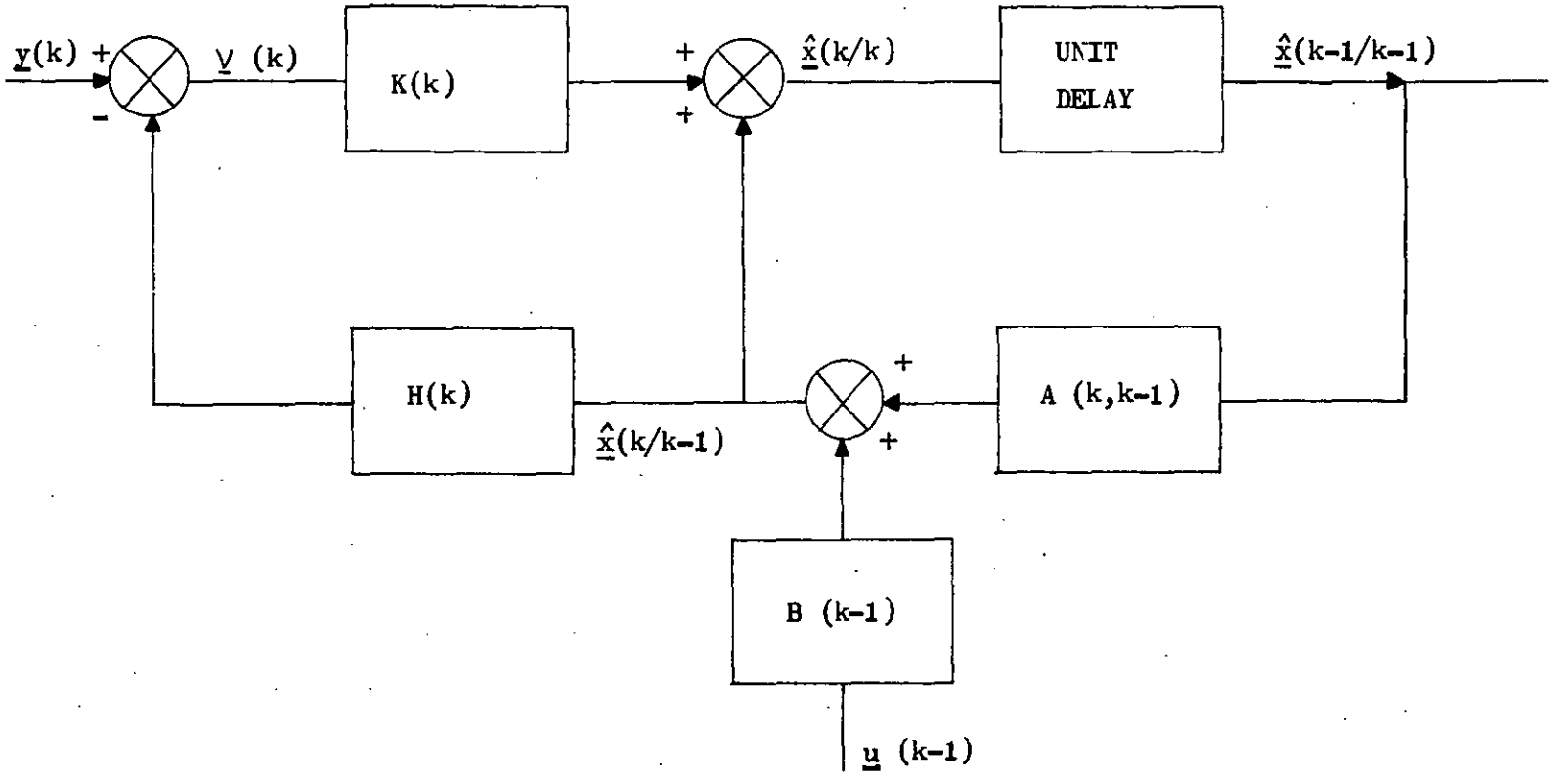


FIGURE 3.2 Feedback implementation of the Kalman filter.

The system model given in equations (3.4.1) and (3.4.2) assumes linearity. Since many problems of practical interest involve non-linear process dynamics, then a linear model must be derived using perturbation techniques. This approach leads to a filter for handling non-linear problems called the "extended Kalman filter". An analysis of this "extended" filter as well as other non-linear approximation techniques may be found in Jazwinski (57).

The results of Kalman are also constrained by the modelling assumption that the noise driving the system and the measurement noise are white. The extension of Kalman's work to problems containing correlated or "coloured" noise was treated by Bryson and Johansen (59) In their paper, the correlated noise was assumed to be generated by a "shaping filter" whose input was white noise. The original state vector was then augmented by the correlated noise sequence and was used to reformulate the problem so that Kalman filtering could be used to obtain a recursive estimate of the new state vector. However the augmentation approach, apart from making the filter more computationally burdensome, leads to matrices which are ill-conditioned.

These problems were eliminated by Bryson and Henrikson (60) for the case of correlated measurement noise. The approach used was to introduce a differencing transformation which yielded a new measurement variable containing uncorrelated noise. This method has the advantages of not increasing the dimension of the state vector and of eliminating the ill-conditioned matrix of the augmented approach.

Theoretically, the Kalman filter gives the unbiased, minimum variance estimate of the system state vector of a linear dynamic system disturbed by additive white noise when measurements of the state vector are linear, but corrupted with white noise. In practice such performance is hardly ever realised. A moment's consideration of equations

(3.4.3) to (3.4.7) reveals that in order to implement the Kalman filter it is necessary to initially specify P(0/0), $\underline{x}(0)$, Q and R (assuming stationary noise statistics), as well as defining the state transition and measurement matrices. If a Kalman filter processes data which are generated by a data generation mechanism, characterised by a structure and/or set of parameter values different from those used in the filter, then suboptimal estimation of the system state vector occurs. In fact several authors have reported that filter divergence can occur due to incorrect a priori information (61), (62), (63). To overcome these problems the Kalman filter should be made adaptive in some sense.

Adaptive Kalman filtering has been the subject of considerable research in recent years and the various adaptive estimation techniques will be discussed in section 3.6.

3.5 Applications of Kalman filtering in chemical engineering

Since Kalman published his solution to the recursive linear estimation problem (46), there has been an extensive research effort devoted to extending his ideas. Although the aerospace industry was not slow to recognise the potential of Kalman filtering, it is only recently that these techniques have been applied to chemical processes.

To use the Kalman filter a state variable representation of the physical process is required. Rarely in the chemical processing industries are accurate process models available, and those that are are usually characterised by complex non-linear differential equations, involving uncertain parameters subject to drift. Some process models may involve pure dead times or be distributed parameter systems.

Another problem arises in describing the statistical nature of the random inputs and measurement errors in industrial processes.

Coggan and Noton's work (64) was concerned with the feasibility of using a non-linear Kalman filter in chemical engineering problems. Their work was based upon computer simulation so that a comparison between the actual and estimated state variables was possible. They reported results for the extended Kalman filter applied to a blending process and a thermal system which were characterised by strong non-linearities, unmeasured disturbances, inaccurate measurements and variable time delay. Their results showed that despite these undesirable features the state estimator was quite stable and converged quickly to within 1 to 2% of the actual state values. This accuracy was considerably better than the accuracy with which any state was actually measured. However, the results obtained from the thermal system indicated that the estimated unknown heat transfer coefficient parameter was considerably biased. The authors attempted to improve this situation by "experimental" runs with:

i) Artificially low or high values of R and P(0/0),

ii) Modifications of the filter Kalman gain,

iii) Various intuitively chosen alterations to the estimation procedure.

None of these proved successful and the results were not reported.

In both of Coggan and Noton's examples, they constrained the values of Q and R to be the same for their simulation and estimation, but they did consider the effect of incorrect a priori information for $\underline{x}(0)$ and P(0/0).

Goldmann and Sargent (65) have performed a detailed study of the factors affecting Kalman filter behaviour for two simulated chemical

processes. In both of their examples they considered only measurement noise.

Their first process consisted of a steady state mass and heat balance around a distillation column, and they showed that the filter was relatively insensitive to incorrect a priori information R, $\underline{x}(0)$ and P(0/0).

Their second example concerned a fixed bed catalytic reactor. The "physical" simulation of the reactor was based upon a detailed mathematical model consisting of mass and energy balances, kinetic equations, mass transfer relations and phase equilibrium relationships applied to a sequence of adiabatic stirred tanks. By assuming particular catalyst decay laws and using the described model the authors could compute the temperature profile along the reactor and its outlet composition, which were subsequently contaminated with Normal random noise to provide process measurements. The filter was used to estimate the catalyst activity and its rate of decay from these measurements. The Kalman filter used a simplified model which ignored all mass transfer effects and assumed a series of first order reactions occurring in an adiabatic plug flow reactor. The catalyst dynamics for the filter were represented by simple functions corresponding to three simulated decay laws, and in particular they approximated a distributed parameter system by a low order polynomial. The noise statistics used for the filter were those used in the simulation.

The authors found that it was possible to predict model parameters and inaccessible variables such as catalyst activity provided that a reasonably accurate steady state model was used. However they did observe that the filter was sensitive, and sometimes divergent, to errors in their assumed models. To overcome this problem the authors

examined more sophisticated filtering schemes in the form of the exponential and limited memory filters (66), (57).

Joffe and Sargent (67) extended the above work on the reactor and developed an optimal control strategy which was a function of the catalyst activity estimated by the Kalman filter. In their formulation they included the effect of input process noise $\underline{w}(k)$ and its covariance matrix $Q(k)$, and by means of simulation demonstrated that the filter was insensitive to the statistical assumptions necessary for its use.

The problem of specifying the a priori statistical information necessary to implement the Kalman filter has been the subject of several studies. This feature was first discussed by Seinfeld et al. (68) and Seinfeld (69). They considered the general problem of the control of a non-linear lumped parameter dynamical system subject to random inputs and measurement errors. A scheme was developed whereby a non-linear Kalman filter was included in the control loop and they illustrated their ideas by a simulation for the proportional control on the temperature of a continuous stirred tank reactor. In implementing the filter the authors remarked that the performance of the estimator depended significantly upon the choices of $\underline{x}(0)$, $P(0/0)$, $Q(k)$ and $R(k)$, and they selected their values by a trial and error approach involving computer simulation. In passing they also commented that if Q was selected too small filter convergence was not obtained.

Wells (70) pursued this point further when he examined the feasibility of extended Kalman filtering applied to an adiabatic continuous stirred tank reactor. He suggested that modelling errors can be accounted for by considering the process noise $\underline{w}(k)$ as a fictitious, zero mean, Gaussian white noise vector whose covariance matrix $Q(k)$

reflects the confidence in the assumed model. The importance of the a priori filter parameters $\underline{x}(0)$, $P(0/0)$, $Q(k)$ and $R(k)$ was discussed and Wells pointed out that $\underline{x}(0)$ and $P(0/0)$ determine the basic speed of response of the filter. He also mentioned the effect of the noise covariances $Q(k)$ and $R(k)$ on the Kalman gain and suggested that a large value of $Q(k)$ should be used if there is uncertainty about the process dynamic model. This causes a large gain and so the filter relies upon current observations to estimate the state vector. This "loosening" effect of the filter causes the steady state estimation error to increase. Increasing the observation noise covariance $R(k)$ has the opposite effect, and consequently the filter tends to disregard measurements containing large errors. Although Wells illustrated the effectiveness of the extended Kalman filter by selecting values of $\underline{x}(0)$, $P(0/0)$, $Q(k)$ and $R(k)$ such that the state estimates were comparable with the simulated system state responses he did not include a sensitivity analysis to illustrate the performance of the filter to poor a priori information.

Hamilton et al. (71) and Seborg et al. (72) examined the linear time invariant form of the Kalman filter by simulation studies and experimental tests on a pilot plant evaporator. They confirmed Wells' work concerning the sensitivity of the Kalman filter to a priori information and advocate that $Q(k)$ and $R(k)$ should be considered as design parameters which are selected to improve filter performance. Although they provided intuitive guides to these matrix selections, they do not present any systematic method for determining them other than by simulation.

Coggan and Wilson (73) realised the problem of information uncertainty in filtering and suggested that one result of it is to cause bias in

the state estimates. They subsequently developed a technique which enables the Kalman filter to detect and inhibit this bias on-line. Their ideas were demonstrated by computer simulations of a concentric tubular countercurrent heat exchanger and an isothermal gas absorption column.

Other references to Kalman filtering which confirm the above remarks may be found in (74) and (75).

The application of Kalman filtering to industrial chemical processes appears to be relatively novel although Gustavsson (76) has reviewed some papers.

Choquette et al. (77) have used extended Kalman filtering to estimate unknown parameters in a reactor system while Wells and Wismer (78) and Thé (79) have reported results from a steelmaking process.

In Sastry and Vetter's (80) work, they modelled the wet-end dynamics of a paper making process and used the Kalman filter to estimate parameters in the resulting non-linear model. King (81) used the Kalman filter for a similar purpose in his mineral flotation pilot plant. A discrete-continuous Kalman filter (57) was used to estimate unknown model parameters, and in particular King imposed two restrictions on the algorithm for practical application. These restrictions were:

i) A lower limit was set on the variance of the state estimate.

ii) In order to enable effective tracking of the parameters after a lengthy period when the parameters were stable, the conditional density $P(\underline{a}/Y_n;\ t_n)$ was never allowed to be less than 10^{-3}.

King gives no guidance of the effect of these restrictions or the filter sensitivity to them.

Applications of Kalman filtering in the nuclear industry may be found in the papers of Godbole (82), Venerus and Bullock (83) and Shinohara and Oguma (84).

3.6 Kalman filtering in uncertain systems

3.6.1 The effect of uncertainty on Kalman filtering

The utilization of Kalman filtering presupposes a known linear dynamic system disturbed by white state and measurement noises of known covariances. However, in actual practice, this knowledge is seldom completely available. The system parameters and noise covariances may only be known approximately and although more accurate modelling is an obvious solution, it is often impractical and time consuming.

The result of using an incorrect system model or incorrect a priori statistics $Q(k)$, $R(k)$ $P(0/0)$ and $\underline{x}(0)$, is to cause large estimation errors, biased state estimates or even divergence of the filter.

Probably the first observed evidence of the sensitivity of the Kalman filter to uncertainties was in the application of the Kalman filter to orbit determination (61). In this work it was observed that modelling errors caused the state estimate to diverge from the true system state, leading to estimation errors much greater than those predicted in theory. The application of the Kalman filter to chemical engineering problems also highlighted this fundamental deficiency of the technique and several suggestions have been presented to overcome this problem (69), (70), (71).

Divergence of the Kalman filter may be explained as follows. For systems which contain no plant noise, i.e. $Q = 0$, the Kalman gain and

computed error covariance matrix both approach zero as time increases (85). This means that after a large number of noise measurements of a deterministic process, the estimator has effectively matched the data with the assumed model and therefore computes each new state estimate using only the preceding state estimate and the assumed state transition matrix, i.e., independent of any new measurements. Since the system model may contain inaccurate parameters or correspond to the true system only over a limited time period, then divergence of the state estimate will occur.

3.6.2 Analysis of suboptimal filter performance

The problem of uncertain a priori information in Kalman filtering has been highlighted. Before analysing methods of overcoming this limitation it is worth considering techniques by which a designer may determine whether his filter is operating optimally. Such tests may be examined before contemplating more complicated estimation schemes, which may be time consuming or even degrade estimation performance.

Berkovec (86), Mehra (87) and Tompretini (88) have proposed tests of the complete system mechanisation using actual system data.

These tests are based upon examining the innovation sequence. The innovation sequence was defined in section 3.4 to be:

$$\underline{\nu}(k) = \underline{y}(k) - H(k)\,\hat{\underline{x}}(k/k-1) \qquad (3.6.2.1)$$

If the filter uses the correct model and noise statistics, then the innovation sequence is a Gaussian white noise sequence with statistics:

$$E(\underline{\nu}(k)) = \underline{0} \qquad (3.6.2.2)$$

$$E(\underline{\nu}(k)\underline{\nu}^{T}(j)) = 0 \qquad j \neq k \qquad (3.6.2.3)$$

$$= H(k)\,P(k/k-1)\,H^{T}(k) + R(k) \qquad j = k \qquad (3.6.2.4)$$

The authors of (86), (87) and (88) process the actual data to determine the statistical properties of the actual innovation sequence. Statistical tests may be performed on the actual innovations and if the results differ significantly from those given in equations (3.6.2.2) to (3.6.2.4), the system model is invalid and the Kalman filter is operating suboptimally.

3.7 The design of Kalman filters for uncertain systems

If the tests of the previous section reveal suboptimal (or even divergent) Kalman filter performance, then the designer may seek methods to overcome the problem of uncertain a priori information.

Mehra (89), Pearson (90) and Weiss (91) have reviewed the literature on Kalman filtering in uncertain systems. Techniques for improving filter performance may be broadly classed as bounding techniques and adaptive estimation.

3.7.1 Bounding techniques

The bounding techniques of improving filter performance tend to be based upon intuition and are usually characterised by a trial and error approach involving considerable computer simulation.

Schlee et al. (61) have discussed several simple methods of eliminating filter divergence. The techniques are based upon trying to limit the decrease in the Kalman gain in order to avoid the filter becoming decoupled from the measurements.

One approach is to artificially increase the plant noise covariance matrix Q. This causes an increase in the error covariance matrices P (k/k-1) and P (k/k), and hence causes the gain matrix to increase.

However in adopting this procedure the amount of noise to be added has to be determined, and this usually becomes a trial and error solution. The filter performance is examined for various values of Q until an acceptable solution is obtained. This technique has been successfully implemented by Fitzgerald (62), Seinfeld (68), Wells (70) and Hamilton et al. (71).

Another commonly used technique is to directly increase the gain matrix K(k) by adding a fixed quantity to it, which again is determined by simulation studies (92).

Tarn and Zaborsky's (66) method of bounding the Kalman gain matrix is based upon increasing the prediction error covariance matrix P(k/k-1) indirectly by exponentially increasing the measurement noise covariance matrix of old observations. If k is the current iteration and n is the iteration at which the measurement noise occurred, then the authors set the noise covariance matrix to be:

$$E\ (\underline{v}(n)\ \underline{v}^T(n)\ /\ t = t_k)\ =\ s^{k-n}\ R(n) \qquad k \geq n\ .$$

where $s\ (\geq 1)$ is an a priori parameter chosen by the designer. This technique has the effect of escalating exponentially with time the covariance matrix of each past observation, thus making past observations have less effect upon current state estimates and the prediction error covariance matrix becomes:

$$P(k/k-1) = s\ A\ (k,k-1)\ P(k-1/k-1)\ A^T(k,k-1) + \Gamma(k-1)\ Q(k-1)\Gamma^T(k-1)$$

This filter has become known as the exponential Kalman filter. Although this idea is computationally simple and attractive, examples have shown that the performance of the modified filter may be quite sensitive to the choice of s, and may even be unstable (65).

Using the same philosophy as Tarn and Zaborsky, Jazwinski (57)

developed limited memory filtering which is based upon the fact that if the assumed system model corresponds to the true system only over a limited period of time, the processing of observations older than the recent past by the Kalman filter will lead to unacceptable state estimation errors. Thus limited memory filtering means discarding the conditioning of the estimate on the distant past. Jazwinski approached the problem from a probabilistic viewpoint and derived a limited memory filter which requires two Kalman filters and a predictor for its implementation. However the resulting estimator is not necessarily stable and requires excessive computer storage, and so Jazwinski proposed that the conditioning of the state estimate on old data should be discarded in batches of N. The resulting limited memory filter produces estimates with memory varying between N and 2N. Simulations (57) have shown that this filter is stable and produces less estimation error than the extended Kalman filter when the latter diverges. However there appears to be no general rule for selecting the memory length N, and there is no reported work on the sensitivity of the limited memory filter to the choice of N.

Crump (93) has also developed a limited memory filter which he terms an augmented memory estimator. However, the same problem of memory length selection is exhibited.

Several researchers have attempted to limit Kalman filter divergence by analysing the innovation sequence. In general the basic idea behind the developed techniques is to make the innovations generated from the actual filter consistent with their theoretical covariances. The innovation sequence was defined in section 3.4 and has a theoretical covariance matrix of:

$$E(\underline{V}(k)\,\underline{V}^T(k)) = S(k) = H(k)\,P(k/k-1)\,H^T(k) + R(k)$$

Using actual data Coggan and Wilson (73) estimated this matrix on-line by using a first order exponential filter:

$$Z(k+1) = \alpha Z(k) + (1-\alpha)\, \underline{\nu}\,(k+1)\, \underline{\nu}^{T}(k+1)$$

They compared this estimated covariance matrix $Z(k+1)$ with the theoretical matrix $S(k+1)$ generated from the Kalman filter and replaced $s_{ii}(k+1)$ by $z_{ii}(k+1)$ if $z_{ii}(k+1) > s_{ii}(k+1)$. Having made this substitution they modified $P(k+1/k)$ to achieve consistency. This has the effect of increasing the Kalman gain $K(k+1)$ and bringing the actual covariance of $\underline{\nu}(k)$ closer to the theoretical. Although Coggan and Wilson illustrated their technique by a series of simulations there are no results regarding the sensitivity of the method to the choice of α.

Quigley (94) suggested that the quality of the Kalman filter may be assessed by examining a scalar figure of merit given by:

$$J(k) = \underline{\nu}^{T}(k)\, (H(k)\, P(k/k-1)\, H^{T}(k))^{-1}\, \underline{\nu}(k)$$

If this performance criterion does not lie within a predetermined interval $J_{min} < J < J_{max}$, the filter is deemed unsatisfactory and the plant noise matrix Q is increased by a fixed amount Q^{*}, thereby increasing the Kalman gain. To use this technique the designer must specify J_{min}, J_{max} and Q^{*}, which are determined by simulation studies, although Quigley suggested that $J_{min} = 1$ and $J_{max} = 5$ are adequate.

Another technique which attempts to prevent filter divergence by examining the innovation sequence has been proposed by Sriyananda (95). He suggested calculating the scalar quantity $\underline{\nu}^{T}(k)\, \underline{\nu}\,(k)$ and determining if this is less than three times the trace of the matrix $(H(k)\, P(k/k-1)\, H^{T}(k) + R\,(k))$. If this test is not satisfied, then filter divergence is suspected and the Kalman gain is frozen at its current value, while the updating of $P(k/k-1)$ is limited to incrementing

it by Q after each new measurement. This procedure causes P(k/k-1) to increase and continues until the performance criterion is satisfied. At this stage P(k/k-1) is large and so the Kalman gain would increase, thus exhibiting some features of limited memory filtering.

In conclusion, it may be said that these bounding techniques for improving Kalman filter performance rely upon extensive computer simulations. Each of the proposed techniques incorporates one or more "tuning" parameters which are chosen by the designer, and thus in practice, the improvement of filter performance becomes a compromise between the time available for filter "optimisation" and the acceptable accuracy of the state estimate $\hat{\underline{x}}(k/k)$. However, to date, it is these bounding techniques which have found application in the chemical engineering state estimation problems reviewed (65), (68), (70), (71), (81).

3.7.2 Adaptive estimation

Adaptive estimation schemes attempt to improve Kalman filter performance (or prevent divergence) by obtaining recursive state estimates in the presence of unknown or inexactly known system information in real time, thereby eliminating the need for an after the fact assessment of filter performance as in the bounding techniques. Such estimation schemes invariably produce a non-linear filter which requires extensive computations, and so in order to implement the schemes in real time suboptimal estimators have been developed.

Magill (96) investigated the optimal estimation problem when certain of the system parameters are unknown. He represented the unknown parameters as a vector $\underline{a}$ and assumed that possible values of $\underline{a}$ form a finite set of possible stochastic processes ($\underline{a}_i$, i=1,...N),

with a known a priori probability of each process occurring $P(\underline{a}_i)$. By using a Bayesian approach to condition the probability of $\underline{a}_i$ occurring on the measurements $Y(k)$, Magill obtained the optimal state estimate $\hat{\underline{x}}(k/k)$ as the weighted sum of N Kalman filters, one for each process $\underline{a}_i$.

Sims et al. (97) have presented computational algorithms for solving Magill's problem, while Smith (98) has obtained a suboptimal estimator for the method. However in practice it seems doubtful that this Bayesian approach would be useful, due to the inordinate computational burden as well as the difficulty of specifying a suitable a priori probability density function for the vector $\underline{a}$.

If the probability density functions for the system uncertainties are unknown, then adaptive estimation can be accomplished using a maximum likelihood technique, which is based upon the philosophy that the most likely values of the unknown parameters are those which make the probability of their occurrence the greatest, given the measurements $Y(k)$.

Abramson (99) obtained an optimal state estimator when the statistics of the measurement and plant noise are diagonal and time invariant. When no a priori information is available for the noise covariance values, a maximum likelihood approach was used, but when an a priori density function is available a maximum a posteriori technique was used.

The resulting likelihood equations are non-linear and there is no general closed form of solution. To overcome this problem a suboptimal estimation procedure was introduced by Abramson. However, numerical simulations have shown that if the a priori values of Q and R are significantly in error, biased state estimates will result.

Abramson also suggested methods by which his work may be extended to cover non-diagonal, time varying covariance matrices.

Shellenbarger (100) also used a maximum likelihood approach to obtain a suboptimal filter for the case of unknown measurement noise statistics. However, for the case of unknown Q his technique is limited by the fact that H(k) is constrained to have more rows than columns.

Another approach using the maximum likelihood method has been given by Sage and Wakefield (101). They simultaneously estimated the system state (including augmented unknown system parameters) and the Kalman gain for a system characterised by a scalar measurement y(k) and random time varying plant noise covariance matrix Q(k). To overcome the problem of specifying a probability density function for the unknown parameter vector $\underline{\theta}$ (k) the authors assumed that $\underline{K}$(k) was the vector $\underline{\theta}$ (k) to be estimated and that $\underline{K}$(k) evolved from a Markov process according to:

$$\underline{K}\ (K+1) = U(k)\ \underline{K}(k) + \underline{e}(k)$$

where U(k) is an nxn transition matrix and $\underline{e}$(k) is a Gaussian white noise sequence with a known nxn covariance matrix $V_{\underline{e}}$. This formulation with the maximum likelihood approach leads to a non-linear two point boundary value problem which is solved to yield an algorithm for estimating the Kalman gain and the system state. The resulting adaptive estimator requires only one Kalman filter and is sequential and therefore may be used in real time applications. However the method assumes that the matrices U(k) and $V_{\underline{e}}$ are known or, more likely, are chosen by simulation. The authors do comment that the performance of the method is not critically dependent upon the precise values of these matrices although no indication of the solution sensitivity is given.

The properties of the innovation sequence have been used to derive an adaptive Kalman filter. Jazwinski (102) has derived an algorithm which may be used for simultaneous state and plant noise covariance matrix estimation of a linear system. His approach is to approximate modelling errors by a Gaussian white noise input, whose covariance Q, a diagonal matrix, is determined to satisfy a requirement that the filter residuals be consistent with their statistics. Short sequences of residuals are used in the estimation of Q, thus the estimator never learns Q.

Assuming a scalar measurement system and $Q = qI$, the predicted residual is defined as:

$$r(k+l) = y(k+l) - E\,(y(k+l)/Y(k)) \qquad l > 0$$

Using the constraint:

$$r^2(k+l) = E\,(r^2(k+l)) \qquad l = 1,2\ldots n$$

which makes the residual value most probable, Jazwinski derived the single residual, i.e. $l = 1$, or innovation estimate of $Q(k)$.

To overcome the problem of the estimate of $Q(k)$ having little statistical significance, Jazwinski proposed replacing the one predicted residual by N predicted residuals

$$M_r = \frac{1}{N} \sum_{l=1}^{N} r(k+l)/R^{\frac{1}{2}}(k+1)$$

The estimate of $Q(k)$ becomes:

$$\hat{Q}_N(k) = \begin{cases} \dfrac{M_r^2 - E(M_r^2/Q(k) = 0)}{S} & : \quad \text{if positive} \\ 0 & ; \quad \text{otherwise} \end{cases}$$

where S is a normalising parameter derived in the algorithm.

To improve the convergence of $\hat{Q}_N(k)$ Jazwinski used a smoothed estimate of the form:

$$\hat{Q}_N^{\,j}(k) = \begin{cases} \frac{1}{j} \sum_{i=k-j+1}^{k} \hat{Q}_N(i) & ; \text{ if positive} \\ 0 & ; \text{ otherwise} \end{cases}$$

The more general problem when Q is a q x q diagonal matrix has also been examined by Jazwinski while Tompretini(88) extended the method to include the situation when the system is characterised by a vector of observations. In his paper Jazwinski presented several simulations to illustrate his technique however the procedure for selecting the averaging interval N and the smoothing interval j has not been developed for the general case.

Another approach to adaptive estimation which relies upon an analysis of the innovation sequence has been formulated by Mehra (87). This author considered a completely controllable and observable system described by linear time invariant models to which steady state Kalman filtering was applied. The time domain properties of a white noise sequence are used to generate a test of optimality of the filter, and this test also serves as a basis for obtaining a solution to the adaptive filtering problem. Although the primary objective of Mehra's work was to identify Q and R, he showed that it was possible to achieve filter adaptation without formally evaluating these matrices by directly estimating the optimal steady state Kalman gain.

Mehra's technique was based upon the correlation properties of the innovation sequence and his method is discussed in detail in Appendix III.

Since Mehra published his work on the innovation correlation technique of filter adaptation several authors have extended his ideas.

Carew and Bélanger (103) and Neethling and Young (104) used Mehra's

basic technique but have proposed alternative algorithms for the direct estimation of the optimal Kalman gain.

Bélanger (105) has also presented an innovation correlation method which extends Mehra's work to the general time varying stochastic process.

Godbole (106), (107) addressed the problem where the noise sequences are correlated and have unknown non-zero mean, i.e.

$$E(\underline{w}(k)) = \underline{\mu}_w$$

$$E(\underline{v}(k)) = \underline{\mu}_v$$

$$E(\underline{w}(k)\,\underline{v}^T(j)) = S\,\delta(k,j)$$

Godbole showed that Mehra's method can be used to handle this situation if the innovation sequence is defined as:

$$\underline{e}(k) = \underline{\nu}(k) - \underline{\mu}_\nu$$

and $$\underline{\mu}_\nu = \frac{1}{N}\sum_{i=1}^{N} \underline{\nu}(i)$$

This idea was demonstrated by Godbole (82) with the application of the Kalman filter for estimating the non-measurable variables of a nuclear pool-type reactor using noisy measurements of a few variables. He assumed that the noise sequences had zero mean but were correlated. Using Mehra's modified technique Godbole estimated Q, R, and S and found that these were consistent for different initial a priori guesses. The innovation sequence white noise test of optimality corresponding to $\hat{Q}$, $\hat{R}$ and $\hat{S}$ revealed 28% violations of the 95% confidence limits indicating the resulting filter was not optimal. Godbole suggested that this non-optimality may be due to process model errors or to $\underline{w}(k)$ and $\underline{v}(k)$ being non-stationary or non-white as assumed. However, other sources of error may be due to the fact that Godbole does not refine his estimates of $\hat{Q}$, $\hat{R}$ and $\hat{S}$ by further data processing as originally recommended by Mehra and also the sample interval for the example is

0.2 second, which is rather large compared to the process time constants of 0.354 and 0.686 second.

3.8 An overview of Kalman filtering in uncertain systems

A number of methods have been presented leading to adaptive correction of Kalman filters, which because of inexact knowledge of the system characteristics, are not operating in an optimal manner.

The bounding techniques of filter adaptation provide intuitively appealing methods of improving filter performance. In general these techniques are easy to understand, are applicable to both linear and non-linear problems, are simple to implement and do not involve the designer in extensive computer programming.

Balanced against these features are the facts that in general each bounding technique incorporates a "tuning" parameter which adapts the filter and which must be selected by the designer. This parameter is usually chosen by a trial and error search procedure which terminates when acceptable Kalman filter performance is achieved. Normally this search technique will involve the designer in extensive time consuming computer simulations.

Despite these apparent disadvantages of the bounding techniques, to date, especially in chemical engineering applications of Kalman filtering, the methods have been used with considerable success in many cases to achieve improved filter performance. However the reported applications of the techniques give no indication of the time taken to determine the "optimal" tuning parameters or the criterion used to assess the quality of the Kalman filter. Most probably such decisions tend to be subjective.

To overcome the disadvantages of the bounding techniques the concept of adaptive estimation was introduced. The objectives of the adaptive estimators are to estimate the system state in real time, rather than rely on an after the fact analysis of filter performance.

In general the adaptive estimators are characterised by non-linear filtering solutions which require extensive computer storage and computation time.

Both the Bayesian and maximum likelihood adaptive estimators suffer because of the inordinate effort needed to implement them on a computer, and it is doubtful whether they are practical propositions for solving the uncertainty problem.

Two adaptive estimators have been reviewed which rely upon information contained in the sequence of residuals. In each method the residuals are first tested to determine if the Kalman filter operation is satisfactory. If it is found that the filter performance is inadequate, the filter is adapted using estimates of the statistics of the residuals. In Jazwinski's approach, N-step predicted residuals ($N \geq 1$) are used whereas one step predicted residuals (i.e. the innovation sequence) are used in Mehra's method.

The modelling assumed in each method dictates the problem class for which each method is applicable. Mehra's technique is restricted to problems where the system is completely controllable and observable and is described by linear time-invariant models. In addition, the testing scheme and subsequent parameter identification are intended for steady state Kalman filtering. This may limit the application of the method to a great many on-line control problems where a time varying filter is required. The formulation considered by Jazwinski allows for time varying dynamics and measurements. The differences in the two methods are

reflected in the residual testing schemes. Only Jazwinski's technique can be used for testing based on one residual. The more residuals that are tested, the longer the time delay in the filter and to a certain extent the application will dictate the number of residuals tested. In Mehra's method, a large sample size is required to estimate the statistics of the residuals. The large sample size is necessitated by the confidence limit nature of the test criterion.

In conclusion, the literature to date provides numerous techniques for handling the Kalman filtering with uncertainty problem. However, at this point there is little reported work of practical application of these techniques and the relative advantages and disadvantages of the various methods are not well known. It would appear that the time is right for a comparative study to be performed so that some light may be shed on this matter.

3.9 The detection of malfunction using a Kalman filter

Mehra and Peschon (108) have presented a general approach to malfunction detection. They represent the system by linear time invariant models with zero mean Gaussian white noise plant and measurement sequences, as given in section 3.4. The matrices A, B, Γ, H, Q and R are assumed to be known or are identified (87) and a Kalman filter is used to process the system measurements to yield the innovation sequence $\underline{\nu}$ (k). Now as discussed in section 3.6.2, a Kalman filter which uses the correct system model and noise statistics generates a zero mean Gaussian white noise innovation sequence with a covariance matrix given by equation (3.6.2.4). Mehra and Peschon suggest that different malfunctions in the system cause the innovation sequence to depart from its zero mean, known theoretical covariance and whiteness properties,

and so the problem of malfunction detection is formulated as a problem in hypothesis testing. The normal operation of the system (i.e. the "optimal" innovation sequence) is regarded as the null hypothesis and the actual innovation generated by the Kalman filter is tested against this hypothesis (i.e. zero mean, known theoretical covariance, whiteness) at a certain level of significance.

If a particular hypothesis is rejected, then malfunction is suspected. However, there appears to be no systematic method of diagnosing the fault although Mehra and Peschon suggest that special system characteristics can often be used to aid diagnostic procedures.

A technique which has been widely used to estimate unknown or uncertain parameters may be adapted as a malfunction detection method (57).

A system containing uncertain parameters may be modelled as:

$$\underline{x}(k+1) = A(k+1,k)\,\underline{x}(k) + A_b\,(k+1,k)\,\underline{b} + B\,(k)\,\underline{u}(k) + \Gamma(k)\,\underline{w}(k) \qquad (3.9.1)$$

$$\underline{y}(k) = H(k)\,\underline{x}\,(k) + H_b(k)\,\underline{c} + v(k) \qquad (3.9.2)$$

The statistics of the noise sequences $\{\underline{w}(k)\}$ and $\{\underline{v}(k)\}$ are assumed to be known and have been defined in section 3.4, while the a priori statistics of the unknown parameters $\underline{b}$ and $\underline{c}$ are specified as:

$$E\,(\underline{b}) = \underline{0} \qquad E\,(\underline{c}) = \underline{0}$$

$$E\,(\underline{b}\,\underline{b}^T) = L \qquad E\,(\underline{c}\,\underline{c}^T) = M$$

It is assumed that $\underline{b}$, $\underline{c}$ $\underline{x}(0)$, $\{\underline{w}(k)\}$ and $\{\underline{v}\,(k)\}$ are uncorrelated.

Now by regarding the constant parameters as the outputs of the dynamic systems:

$$\underline{b}(k+1) = \underline{b}(k)$$

$$\underline{c}(k+1) = \underline{c}(k)$$

then an augmented state variable model may be defined as:

$$\underline{z}(k+1) = \begin{bmatrix} A(k+1,k) & A_b(k+1,k) & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \underline{z}(k) + \begin{bmatrix} B(k) \\ 0 \\ 0 \end{bmatrix} \underline{u}(k) + \begin{bmatrix} \Gamma(k) \\ 0 \\ 0 \end{bmatrix} \underline{w}(k) \qquad (3.9.3)$$

$$\underline{y}(k) = \begin{bmatrix} H(k) & 0 & H_b(k) \end{bmatrix} \underline{z}(k) + \underline{v}(k) \qquad (3.9.4)$$

and $\underline{z}(k) = \begin{bmatrix} \underline{x}(k) \\ \underline{b}(k) \\ \underline{c}(k) \end{bmatrix}$

The linear Kalman filter of section 3.4 may be applied directly to this "new" system and generates state estimates:

$$\hat{\underline{z}}(k/k) = \begin{bmatrix} \hat{\underline{x}}(k/k) \\ \hat{\underline{b}}(k/k) \\ \hat{\underline{c}}(k/k) \end{bmatrix}$$

Although straightforward in theory this method has several disadvantages. Primarily if the system model of equations (3.9.1) and (3.9.2) is assumed time invariant and completely controllable and observable then the Kalman filter is guaranteed to be stable and convergent (46). However, by inspection, the augmented system (time invariant) of equations (3.9.3) and (3.9.4) is no longer controllable and observable, thereby invalidating these filter convergence properties, although the controllability constraint may be satisfied by assuming:

$$\underline{b}(k+1) = \underline{b}(k) + \underline{w}_b(k)$$

$$\underline{c}(k+1) = \underline{c}(k) + \underline{w}_c(k)$$

The vectors $\underline{w}_b(k)$ and $\underline{w}_c(k)$ are assumed to be uncorrelated zero mean Gaussian noise sequences as usual with statistics:

$$E(\underline{w}_b(k)) = \underline{0}$$

$$E\,(\underline{w}_b(k)\,\underline{w}_b^T(j)) = Q_b(k)\,\delta(k,j)$$

$$E\,(\underline{w}_c(k)) = \underline{0}$$

$$E\,(\underline{w}_c(k)\,\underline{w}_c^T(k)) = Q_c(k)\,\delta(k,j)$$

A paramount drawback of this method arises in the increase in a priori information necessary from the designer. In addition to the usual information the user has to additionally specify:

$\underline{b}(0)$

$\underline{c}(0)$

$P_b(0/0)$

$P_c(0/0)$

$Q_b(0)$

$Q_c(0)$

In view of the comments made in the previous sections concerning Kalman filtering in uncertain systems it would seem likely that this method of parameter estimation would involve extensive computer simulation.

Another disadvantage of this state augmentation technique is that the addition of unknown parameters to the state vector increases the computational load and this is not always desirable.

To overcome this latter problem Friedland (109) has furnished a computationally attractive algorithm. This technique involved partitioning the computational effort into two essentially disjoint tasks, one part is the standard Kalman filter which calculates the state estimate $\hat{\underline{x}}(k/k)$ while the other part is an algorithm to generate the estimate of the unknown parameter vector $(\hat{\underline{b}}(k/k)\ \hat{\underline{c}}(k/k))^T$. This idea is presented in some detail in Appendix IV.

However, in spite of these apparent difficulties, Goldmann and Sargent (65) successfully used the state augmentation technique to detect bias and drift in particular process instruments by considering

a steady state mass and heat balance around a binary distillation column.

Their formulation assumed that the plant matrix A(k+1,k) was the identity matrix and there were no process disturbances, i.e. Q = 0. Using a simulation study they demonstrated that the filter performance was relatively insensitive to the a priori information P(0/0) and R. However, the results presented to illustrate the detection of instrument errors are based upon the "optimum" values of P(0/0) and R, i.e. those values chosen to give the best filter performance.

More recently Davis (112) considered the problem of state estimation in the presence of a fault which occurs randomly. The result of the fault is to cause a plant parameter to change from a_0 to a_1. By applying non-linear filtering he derives a filter which optimally estimates the state and the time of fault occurrence, however the resulting equations have no closed form. This problem is overcome by basing a suboptimal estimation scheme on the Kalman filter.

3.10 Statement of malfunction detection technique and objectives

The review of Kalman filtering has revealed that the problem of uncertain a priori system models and/or statistical information is of paramount importance and several studies were examined which attempted to solve this problem in a variety of ways.

The question of malfunction detection has been solved in two ways. The first technique of innovation hypothesis testing is limited by the fact that little or no diagnostic information is obtained, while the state augmentation technique only seems to compound the uncertainty problem due to the additional a priori information needed for its implementation.

However, under the assumption of a linear, time invariant, completely controllable and observable system, many of these problems may be eliminated by combining Mehra's and Friedland's algorithms.

Recall that Friedland decomposed the state augmentation method of malfunction detection into two tasks. The first was the solution of the basic Kalman filter algorithm - the innovations from which were used in a "bias" estimator to determine the system unknown parameters or malfunction. This secondary or "bias" filter needs the following a priori information:

R

PH^T

$P_b(0/0) = M(0)$

Now an examination of Mehra's algorithm, which may be used to handle the uncertainty in implementing the primary Kalman filter, reveals that estimates of $\hat{R}$ and $\widehat{PH^T}$ are calculated as a natural feature of the technique.

Thus the only uncertainty remaining in implementing Friedland's algorithm is $P_b(0/0)$ which has to be chosen by the designer.

The combination of algorithms is shown schematically in Figure 3.3 which outlines the malfunction detection algorithm proposed in this study.

In stage 1 an arbitrary a priori data set is chosen for $\underline{x}(0)$, $P(0/0)$, Q and R and the usual Kalman filter, defined in equations (3.4.3) to (3.4.7), is used to analyse the process measurements $\underline{y}(k)$. When the Kalman gain reaches its steady state value K_{ss}, as guaranteed by the constraints that the system model is linear and controllable and observable, then a large sample of innovations is stored.

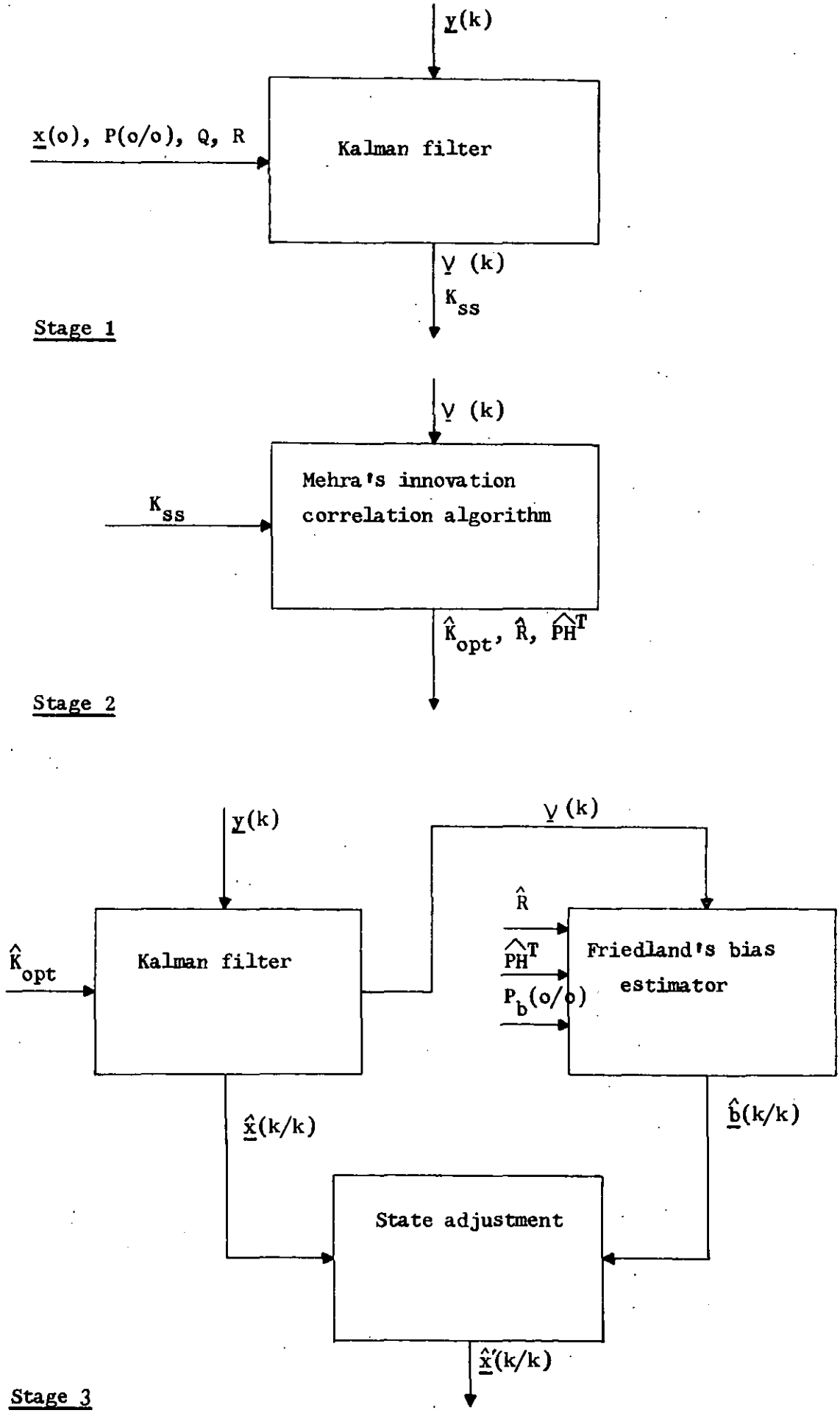


FIGURE 3.3 Implementation of malfunction detection algorithm.

Stage 2 of the algorithm uses the stored innovation sequence and K_{ss} in Mehra's adaptive estimator. Mehra's method tests the innovation sequence for whiteness. If this criterion is not met, the algorithm iterates on K_{ss} to yield an estimate of the optimal Kalman gain K_{opt}, as well as $\hat{R}$ and $\widehat{PH^T}$ ($P = P(k/k-1)_{ss}$). At the end of stage 2 the Kalman filter has been adapted to cope with the initial uncertainty and thus can be used to generate "optimal" state estimates $\hat{\underline{x}}(k/k)$.

The adapted Kalman filter is combined with Friedland's bias estimator in stage 3. The Kalman filter produces state estimates as usual, and the innovation sequence is used as an input to Friedland's filter which estimates the bias caused by malfunction or changing system model parameters. Finally, the Kalman state estimate $\hat{\underline{x}}(k/k)$ and the bias estimate $\hat{\underline{b}}(k/k)$ may be combined to yield a true optimal state estimate $\hat{\underline{x}}'(k/k)$.

This Figure underlines the robustness of the proposed malfunction detection algorithm to the selection of a priori information.

Section 3.3 discussed how control loop security may be monitored by examining the relationship between the measured process variable and the control valve demand signal at a particular setpoint and load, while the above development has shown how biases in a mathematical model may be estimated in real time. It only remains therefore to translate the suggestions in section 3.3 into an appropriate mathematical formulation for use in the algorithm developed above.

The first stage is to represent the control loop by a linear time invariant state space model, which may be derived using standard techniques to be of the form:

$$\begin{aligned} \underline{x}(k+1) &= A\,\underline{x}(k) + \Gamma\,\underline{w}(k) \\ \underline{y}(k) &= H\,\underline{x}(k) + \underline{v}(k) \end{aligned} \qquad (3.10.1)$$

3.10.1 Malfunction detection model formulation for a conventional analogue setpoint control loop

To implement the ideas discussed in section 3.3 it is assumed that the process measurement vector $\underline{y}(k)$ is composed of the control loop measured variable $B = y_1$ and the valve stem position measurement y_2. Thus

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = H\,\underline{x}(k) + \underline{v}(k)$$

Now as mentioned in section 3.3 there are two types of error to consider, pure measurement errors and "internal" loop faults.

The pure measurement errors may be represented by bias terms $b_1(k)$ and $b_2(k)$, since each observation is subject to such errors, and so the observation model becomes

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \left[\begin{array}{c|cc} H & 1 & 0 \\ & 0 & 1 \end{array}\right] \left[\begin{array}{c} \underline{x}(k) \\ \hline b_1(k) \\ b_2(k) \end{array}\right] + \underline{v}(k) \qquad (3.10.1.1)$$

The second type of error is the loop error which caused a deviation, denoted by $b_3(k)$, from the nominal valve stem position measurement $y_2(k)$ for the same apparent process measurement $y_1(k)$ and this may be represented as:

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \left[\begin{array}{c|c} H & 0 \\ & 1 \end{array}\right] \left[\begin{array}{c} \underline{x}(k) \\ \hline b_3(k) \end{array}\right] + \underline{v}(k) \qquad (3.10.1.2)$$

Since the biases $b_2(k)$ and $b_3(k)$ are indistinguishable, the equations (3.10.1.1) and (3.10.1.2) may be combined to yield:

$$\underline{y}(k) = \left[\begin{array}{c|cc} H & 1 & 0 \\ & 0 & 1 \end{array}\right] \left[\begin{array}{c} \underline{x} \\ \hline b_1(k) \\ b_2(k) \end{array}\right] + \underline{v}(k)$$

where $b_1(k)$ represents the deviation of $y_1(k)$ from its nominal value

due to a P/I transmitter error and $b_2(k)$ represents a pure valve stem position measurement error as well as loop errors.

Now $\underline{b}(k)$ is considered to be constant and is treated as an additional state vector to be estimated. The original state space model of equation (3.10.1) is therefore augmented by $\underline{b}(k)$ to give:

$$\begin{bmatrix} \underline{x}(k+1) \\ \hline b_1(k+1) \\ b_2(k+1) \end{bmatrix} = \left[\begin{array}{c|cc} A & 0 & \\ \hline 0 & I & \end{array}\right] \begin{bmatrix} \underline{x}(k) \\ \hline b_1(k) \\ b_2(k) \end{bmatrix} + \begin{bmatrix} \Gamma \\ \hline 0 \end{bmatrix} \underline{w}(k) \qquad (3.10.1.3)$$

$$\underline{y}(k) = \left[\begin{array}{c|cc} H & 1 & 0 \\ & 0 & 1 \end{array}\right] \begin{bmatrix} \underline{x}(k) \\ \hline b_1(k) \\ b_2(k) \end{bmatrix} + \underline{v}(k)$$

This equation (3.10.1.3) is now in the required form for use in the proposed malfunction detection algorithm. The vector $\underline{b}(k)$, in the context of malfunction detection, is termed a loop security vector comprising two loop security parameters (l.s.p's.).

3.10.2 Malfunction detection model formulation for a direct digital control loop

In this control loop there are no pure measurement errors and all of the system malfunctions cause a deviation in the valve demand signal from the expected value at the given setpoint and load.

To implement the malfunction check suggested in section 3.3, the process computer requires measurements of the process variable and the valve demand signal, but these measurements are basic to the d.d.c. algorithm and so no additional process instrumentation is necessary.

The deviation in control valve demand signal is represented by b(k), which is termed a loop security parameter (l.s.p.). Using a

similar approach to that adopted above, the state variable representation of the augmented system is:

$$\begin{bmatrix} \underline{x}(k+1) \\ \hline b(k+1) \end{bmatrix} = \begin{bmatrix} A & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 1 \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \hline b(k) \end{bmatrix} + \begin{bmatrix} \Gamma \\ \cdots \\ 0 \end{bmatrix} \underline{w}(k) \qquad (3.10.2.1)$$

$$\underline{y}(k) = \begin{bmatrix} H & \vdots & 0 \\ & \vdots & 1 \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \hline \hat{b}(k) \end{bmatrix} + \underline{v}(k)$$

The proposed malfunction detection algorithm may now be applied directly to this set of equations to estimate $\hat{b}(k)$ from the process measurement vector $\underline{y}(k)$.

The ideas developed up to this point have not mentioned any particular dynamic system or control loop other than the constraint that the system be linear, time invariant and completely controllable and observable. To demonstrate that the ideas suggested above may be used to detect malfunction in a control loop during normal process operation, a laboratory level control rig was built as a vehicle for experimentation.

The objectives of the experimental work described below were:

i) To investigate Kalman filtering on a practical apparatus when the a priori information is subject to uncertainty;

ii) To investigate the feasibility and robustness of Mehra's innovation correlation method of filter adaptation;

iii) To investigate the feasibility of detecting and diagnosing malfunction in analogue and direct digital control loops using the algorithm presented above.

3.11 Experimental apparatus

The laboratory level control rig is shown schematically in Figure 3.4. The circulating fluid was water and the main components are summarised below.

i) Pipework.

The pipework was fabricated from 1.905 cm ($\frac{3}{4}$ inch) I.D. rigid P.V.C.

ii) Control Valves.

Both control valves were manufactured by the Taylor Instruments Company. The valves were direct acting needle types with 0.635 cm ($\frac{1}{4}$ inch) trims in 1.27 cm ($\frac{1}{2}$ inch) bodies and had a valve stem travel of 1.905 cm ($\frac{3}{4}$ inch).

iii) Level Measurement.

The primary measurement of level in the hold-up tank was made using a pressure difference/pressure transmitter by measuring the difference in pressure between atmospheric and the tank height plus atmospheric.

iv) Instrumentation.

The locations of the measurement instruments are shown in Figure 3.4, and Table 3.1 details their functions.

The rig operated by water flowing from the mains supply via an orifice plate, control valve and rotameter into a hold-up tank. The outlet flow from this tank split into two streams: one flow passed through a fixed restriction to a drain, while the second flow passed through a control valve to drain. Although two control loops are shown in Figure 3.4, the primary task of the experimental rig was level control. The object of this control system was to maintain a constant flow through the fixed restriction to drain despite inlet water flowrate

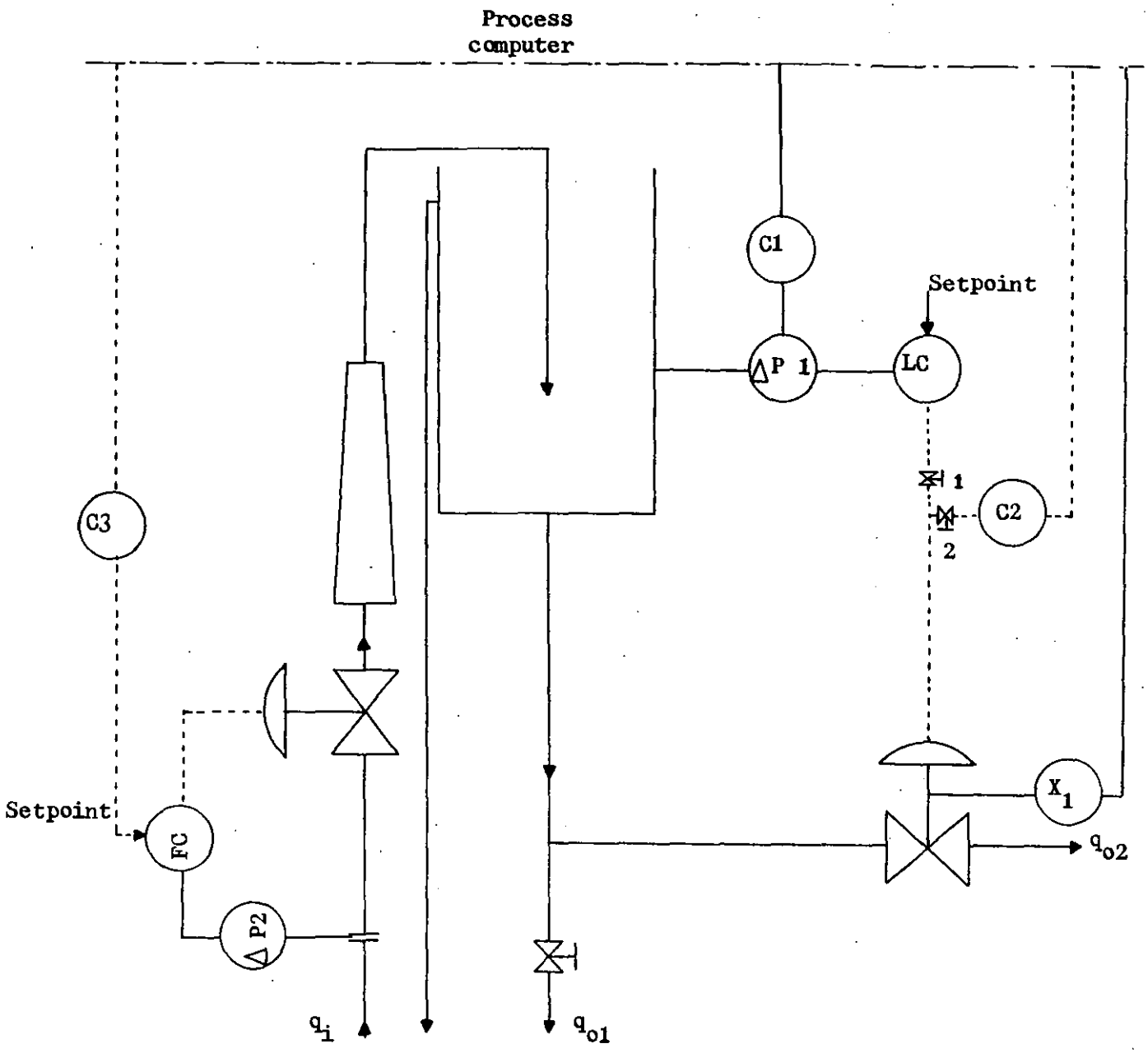


FIGURE 3.4 Experimental level control rig.

Instrument	Instrument	Maker	Input	Output
ΔP_1	Δ P/P transmitter	Taylor Instruments	0–7.47 kN/m^2 (0–30 in.W.G.)	207–103.4 kN/m^2 (3–15 p.s.i.g.)
ΔP_2	Δ P/P transmitter		0–4.98 kN/m^2 (0–20 in.W.G.)	20.7–103.4 kN/m^2 (3–15 p.s.i.g.)
C1	P/I transmitter		20.7–103.4 kN/m^2 (3–15 p.s.i.g.)	5 – 10 V
C2	I/P transmitter		0 – 10 V	20.7–103.4 kN/m^2 (3 – 15 p.s.i.g.)
C3	I/P transmitter		0 – 10 V	20.7–103.4 kN/m^2 (3 – 15 p.s.i.g.)
FC, LC	proportional + integral controller		20.7–103.4 kN/m^2 (3 – 15 p.s.i.g.)	20.7–103.4 kN/m^2 (3–15 p.s.i.g.)
X_1	Linear potentiometer	Penny and Giles	0 – 1.905 cm	4 – 7 V

TABLE 3.1 Instrumentation of laboratory level control rig.

disturbances. This was achieved by adjusting the second flowrate through the control valve so that a constant height was kept in the tank, thereby ensuring a constant flow through the fixed restriction.

The experimental rig was designed to employ two modes of control. With isolation valve 1 open and 2 closed, conventional analogue control using pneumatic hardware was used to control the rig. However, with isolation valve 1 closed and 2 open, a PDP 11-20 process computer could be used to control the system.

In all the experiments performed disturbances were introduced into the inlet water flowrate by using the process computer to alter the set-point of the secondary flow control loop.

3.11.1 Mathematical model of experimental level control rig

The first step in Kalman filtering is the determination of a state variable system model. Initially a dynamic model representing the level control loop will be derived (56). The nomenclature describing the process flowrates is shown in Figure 3.4. An unsteady material balance for the tank yields.

$$\frac{Adh}{dt} = q_i - q_{o1} - q_{o2} \qquad (3.11.1.1)$$

where h = the height of fluid in the tank.

A = cross-sectional area of the tank.

Deviation variables are introduced into the analysis at this point so that a linear transfer function may be derived. Initially, the process is operating at steady state, and so $\frac{dh}{dt} = 0$, thus equation (3.11.1.1) may be written as:

$$A\frac{dh_{ss}}{dt} = 0 = q_{iss} - q_{o1,ss} - q_{o2,ss} \qquad (3.11.1.2)$$

where the subscript ss has been used to indicate the steady state value of the variable.

Subtracting equation (3.11.1.2) from (3.11.1.1) gives

$$A\frac{d(h-h_{ss})}{dt} = (q_i - q_{i,ss}) - (q_{o1} - q_{o1,ss}) - (q_{o2} - q_{o2,ss}) \qquad (3.11.1.3)$$

The deviation variables are defined as:

$$H = h - h_{ss}$$

$$Q_i = q_i - q_{i,ss}$$

$$Q_{o1} = q_{o1} - q_{o1,ss}$$

$$Q_{o2} = q_{o2} - q_{o2,ss}$$

Then equation (3.11.1.3) becomes:

$$A\frac{dH}{dt} = Q_i - Q_{o1} - Q_{o2} \qquad (3.11.1.4)$$

Now the outlet flow variables q_{o1} and q_{o2} may be expressed in terms of the tank height as follows.

Assume the system is at steady state and the tank height is h_{ss}; if there is now a change in level, q_{o1} may be expressed by a Taylor series as

$$q_{o1} = q_{o1,ss} + \left(\frac{d\, q_{o1,ss}}{d\, h_{ss}} \right) (h - h_s) + \ldots$$

In terms of the deviation variables defined above this becomes:

$$Q_{o1} = \left(\frac{d\, q_{o1,ss}}{d\, h_{ss}} \right) H = K_{o1} H \qquad (3.11.1.5)$$

The flowrate q_{o2} is a function of both the tank height h and control valve stem position x, and so by similar reasoning a Taylor series may be written for q_{o2} as:

$$q_{o2} = q_{o2,ss} + \left(\frac{\partial q_{o2,ss}}{\partial h_{ss}} \right)_{x_{ss}} (h - h_{ss}) + \left(\frac{\partial q_{o2,ss}}{\partial x_{ss}} \right)_{h_{ss}} (x - x_{ss}) + \ldots$$

or

$$Q_{o2} = \left(\frac{\partial q_{o2,ss}}{\partial h_{ss}} \right)_{x_{ss}} H + \left(\frac{\partial q_{o2,ss}}{\partial x_{ss}} \right)_{h_{ss}} X$$

$$= Q_{o2} = K_{o2,h} H + K_{o2,x} X \qquad (3.11.1.6)$$

Equations (3.11.1.5) and (3.11.1.6) may be substituted into (3.11.1.4) to yield:

$$A \frac{d\, H}{d\, t} = Q_i - K_{o1} H - K_{o2,h} H - K_{o2,x} X \qquad (3.11.1.7)$$

Now by taking Laplace transforms and noting $H(0) = 0$ equation (3.11.1.7) becomes

$$\left(\frac{A}{K_{o1} + K_{o2,h}} s + 1 \right) \bar{H}(s) = \frac{\bar{Q}_i(s)}{(K_{o1} + K_{o2,h})} - \frac{K_{o2,x}}{(K_{o1} + K_{o2,h})} \bar{X}(s)$$

The following replacements are made:

$$\tau_p = \frac{A}{(K_{o1} + K_{o2,h})} \qquad (3.11.1.8)$$

$$K_L = \frac{1}{(K_{o1} + K_{o2,h})} \qquad (3.11.1.9)$$

$$K_v = \frac{K_{o2,x}}{(K_{o1} + K_{o2,h})} \qquad (3.11.1.10)$$

This results in a process transfer function of the form:

$$\bar{H}(s) = \frac{1}{(1 + \tau_p s)} \left(K_L \bar{Q}_i(s) - K_v \bar{X}(s)\right)$$

3.11.2 Process control block diagram and experimental parameter determination for level control loop

A block diagram for the negative feedback level control loop is shown in Figure 3.5. In formulating this diagram the control valve and measurement transmitter dynamics have been ignored

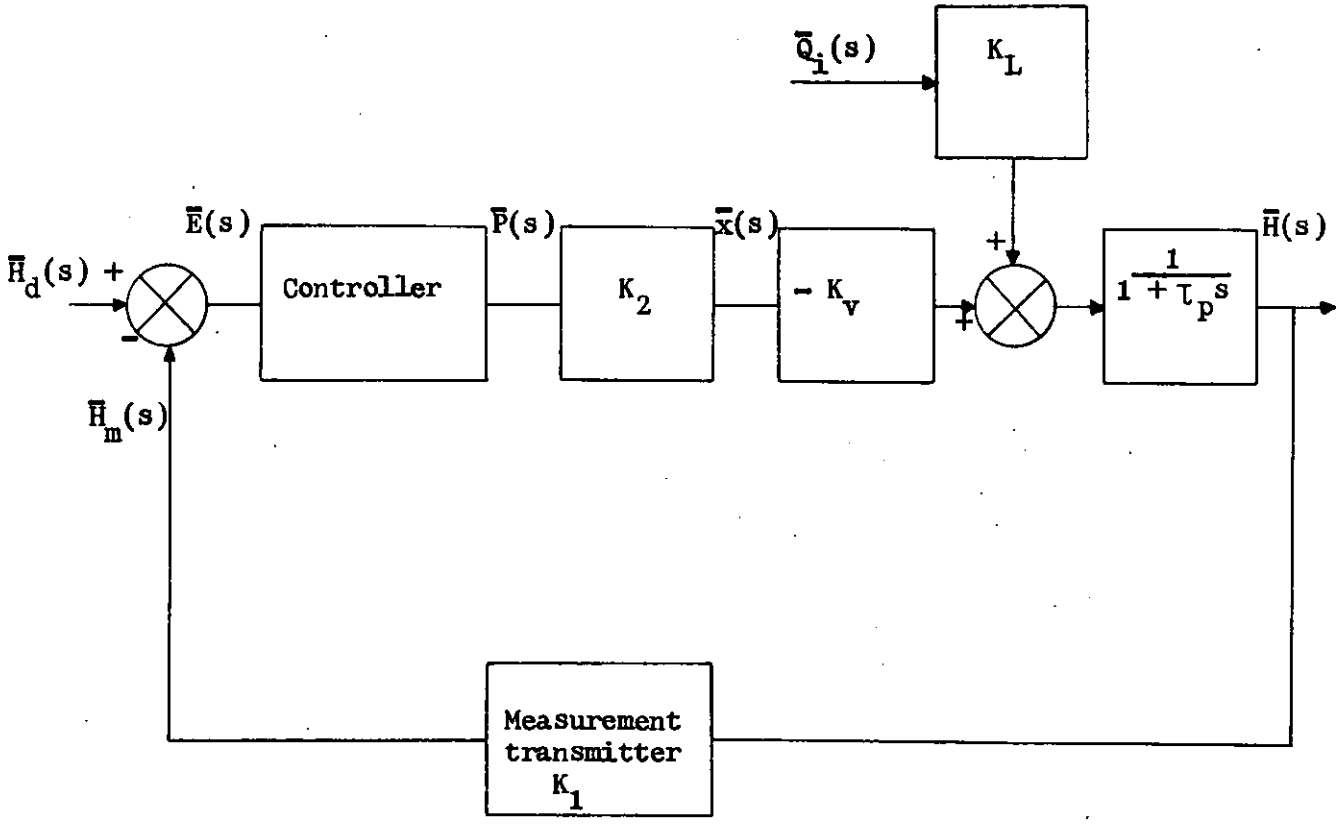


FIGURE 3.5. Block diagram for level control loop.

All of the experiments and system calibrations were performed with the process operating at the steady state process conditions given in Table 3.2

Variable	Steady state value
q_i	4.95 l/min
q_{o1}	1.83 l/min
q_{o2}	3.12 l/min
h	23.25 cm

TABLE 3.2 Steady state process variables for laboratory level control rig.

i) Controller algorithm and constant determination

A proportional plus integral controller was chosen to perform the control task. The continuous time relationship between the controller output and the error input is given by the Laplace transform (56):

$$\frac{\bar{P}(s)}{\bar{E}(s)} = K_c \left(1 + \frac{1}{\tau_I s}\right) \qquad (3.11.2.1)$$

$\bar{P}(s)$ = Laplace transform of the controller output signal

$\bar{E}(s)$ = Laplace transform of the controller input signal

K_c = controller gain

τ_I = integral time

At the steady state process conditions given in Table 3.2 the controller constants were chosen using the method detailed in the Taylor Instruments' pneumatic controller operation manual (113).

ii) K_{o1}; $K_{o2,h}$

The tank height was controlled automatically at the given steady state and it was noted that the controller output signal was 51 kN/m^2 (7.4 p.s.i.g.). The controller was switched to manual control but the same output signal was maintained, thereby ensuring the control valve position was unaltered. Now by altering the tank inlet flowrate q_i, the steady state tank height h_{ss} varied and the corresponding steady state flowrates q_{o1} and q_{o2} were measured.

Graphs of q_{o1} and q_{o2} versus the tank height are shown in Figures 3.6 and 3.7. The constants K_{o1} and $K_{o2,h}$ are determined by measuring the slopes of these graphs respectively.

iii) $K_{o2,x}$

The flow q_{o1} was set to zero by closing the manual hand valve in the outlet pipe. The level was then controlled at the steady state height given in Table 3.2. The inlet flowrate q_i was adjusted, and the process allowed to achieve steady state so that $q_i = q_{o2}$. Now by recording the control valve stem position a calibration of q_{o2} versus stem position was achieved at the specified steady state.

This control valve characteristic is shown in Figure 3.8 and $K_{o2,x}$ is determined from the slope when $q_{o2} = 3.12$ l/min corresponding to the selected steady state process conditions.

iv) A

The hold-up tank cross-sectional area was measured directly.

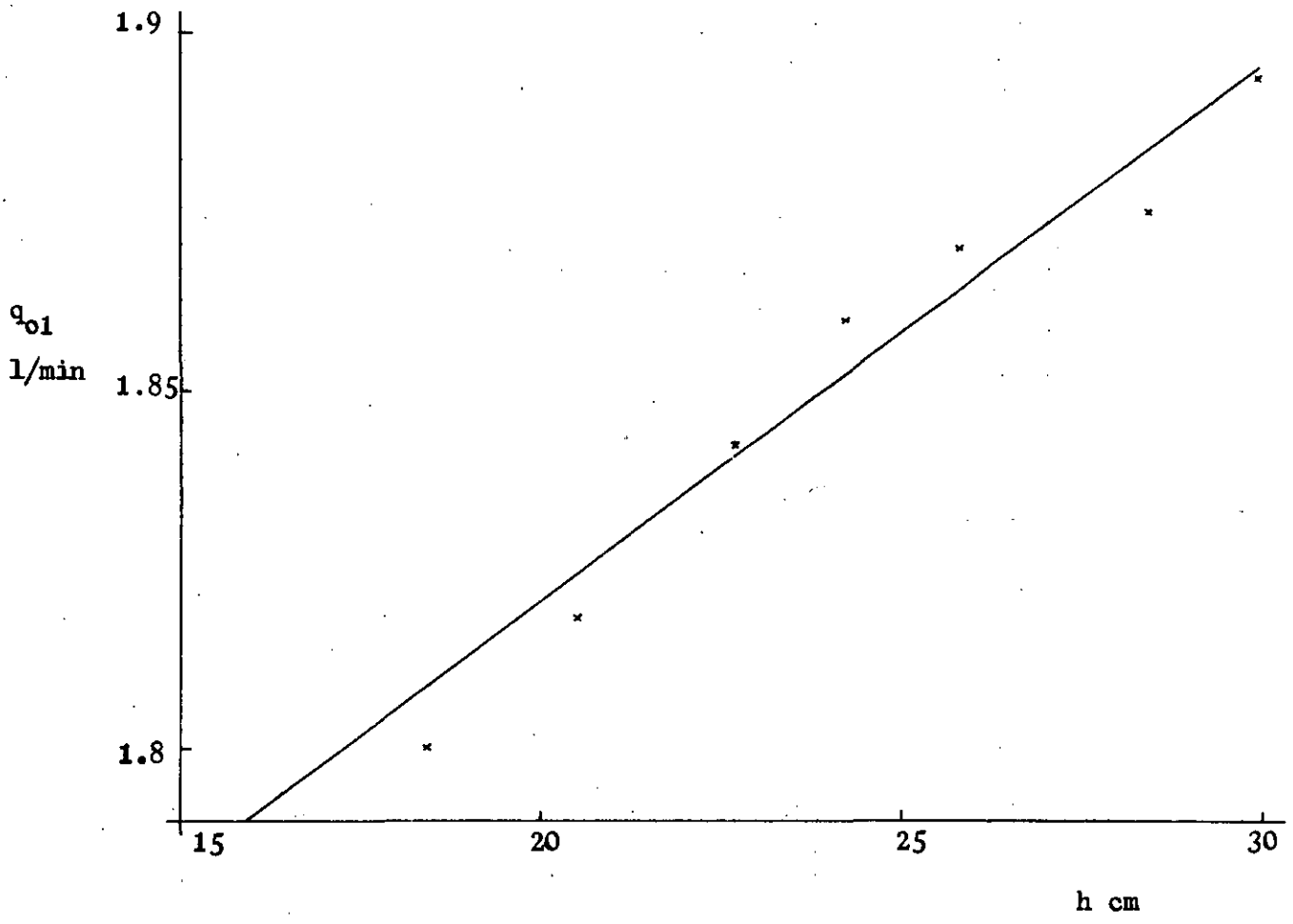


FIGURE 3.6 Experimental determination of K_{o1}

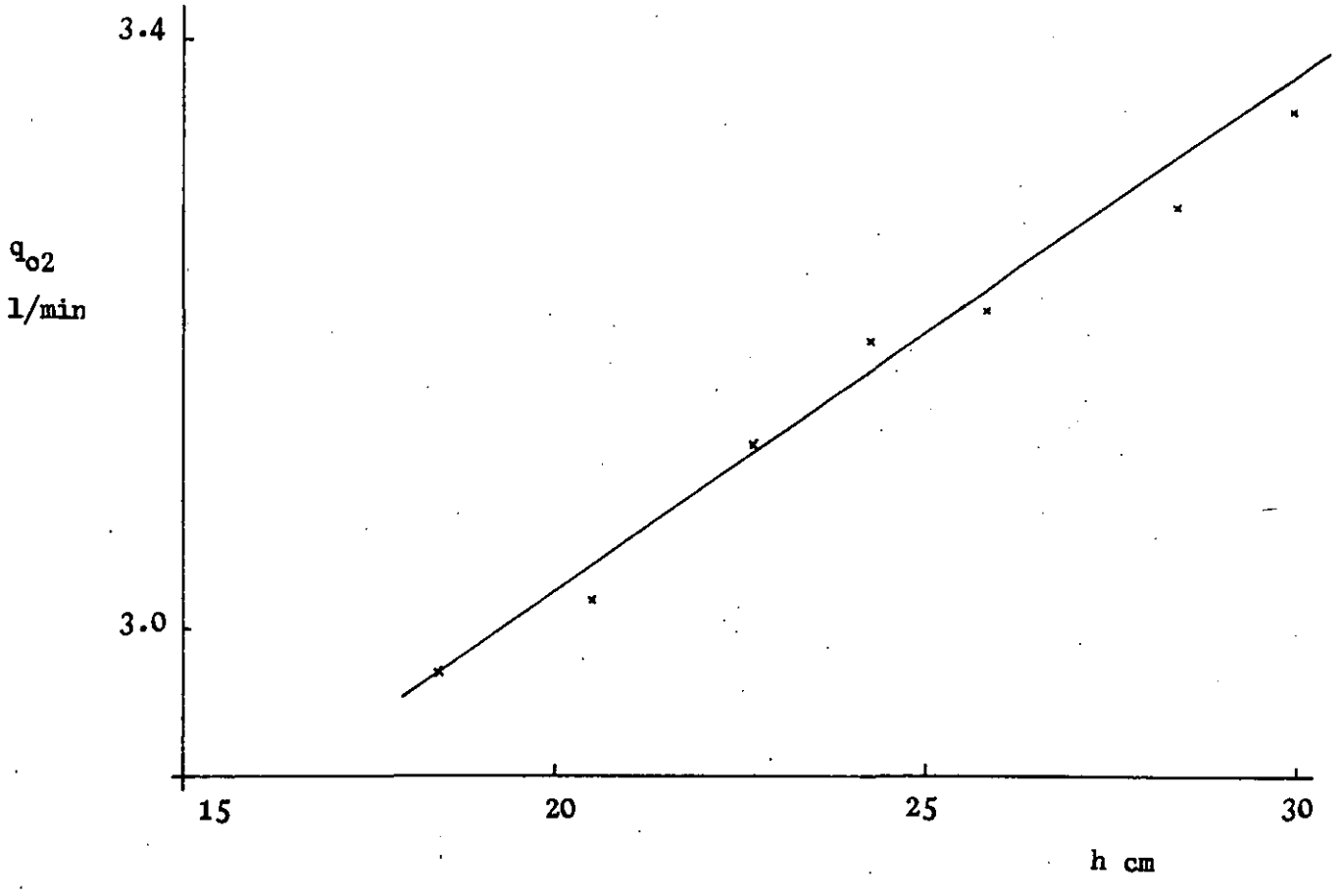


FIGURE 3.7 Experimental determination of $K_{o2,h}$

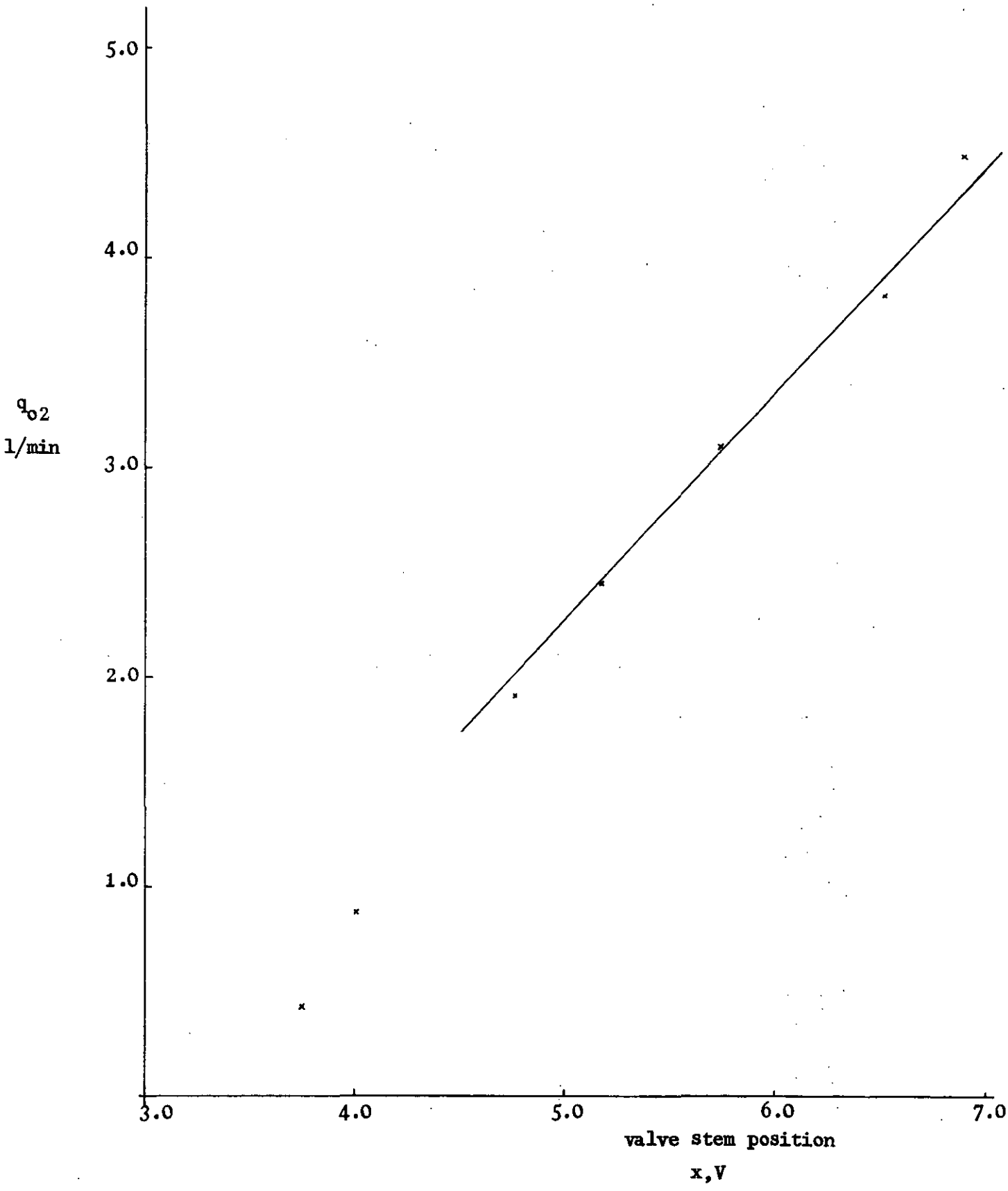


FIGURE 3.8 Level control valve characteristic.

v) K_v; K_L; τ_p

The constants derived above may be substituted into equations (3.11.1.8) - (3.11.1.10) to yield the process transfer coefficients.

vi) K_1; K_2; K_3; K_4

The constant K_1 describes the gain of the tank height measurement transmitter ΔP_1 in Figure 3.4, and is determined by direct calibration on the process.

To ensure consistency of the process variables around the loop in Figure 3.5, a gain term, K_2, relating the controller output in kN/m^2 (p.s.i.g.) to the control valve stem position in volts is required. This characteristic is shown in Figure 3.9.

The coefficient K_3 describes the gain of P/I transmitter, C1, in Figure 3.4. Actually in the experimental rig an amplifier was used to magnify the P/I transmitter output signal and so K_3 is composed of two terms K_{31} and K_{32} ($K_3 = K_{31} \times K_{32}$), which were found by direct calibration on the experimental rig.

K_4 characterises the gain of the I/P transmitter C_2 in Figure 3.4, which was determined by direct calibration.

The numerical values of these parameters are summarised in Table 3.3.

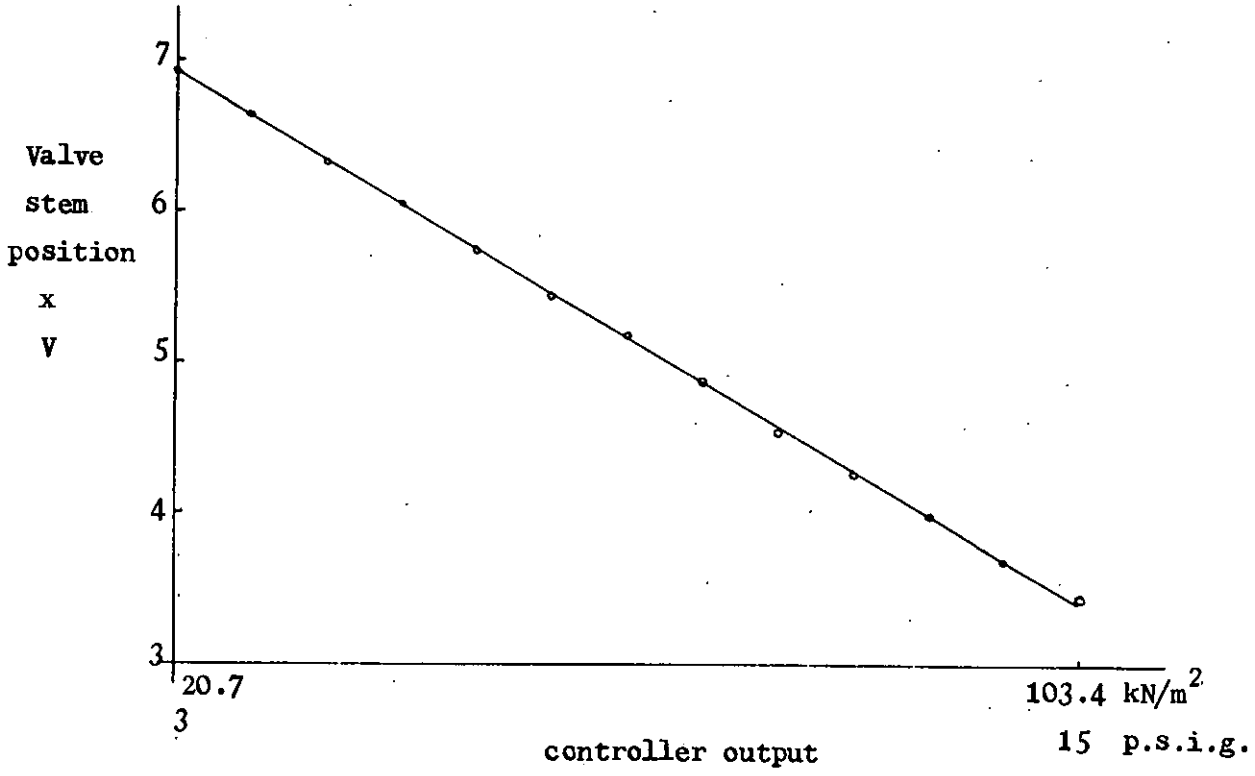


FIGURE 3.9 Experimental determination of K_2.

Variable	Numerical value
K_c	2 p.s.i.g./p.s.i.g. analogue setpoint control
τ_I	0.4 min " " "
K_c	1 V/V d.d.c.
τ_I	0.2917 min "
K_{o1}	7.2 cm^3/min.cm.
$K_{o2,h}$	34.78 cm^3/min.cm.
$K_{o2,x}$	1156.0 cm^3/min. V
A	280 cm^2
K_v	27.46 cm/V
K_L	0.024 min.cm/cm^3
τ_p	6.65 min.
K_1	0.1730 p.s.i.g/cm.
K_2	- 0.3 V/p.s.i.g.
K_{31}	0.388 V/p.s.i.g.
K_{32}	9.4 V/V
K_3	3.65 V/p.s.i.g.
K_4	1.36 p.s.i.g./V

TABLE 3.3. Numerical values of parameters in level control loop.

3.12 State variable model formulation

The fundamental block diagram of the level control system was given in Figure 3.5. The diagram may be redrawn into a more amenable form from which a state variable model may be written directly.

3.12.1 Analogue setpoint control

The pneumatic level controller used in these experiments was of a proportional plus integral type. The transfer function of such a controller, when there is no deviation in loop setpoint is (56):

$$\frac{\bar{P}(s)}{\bar{H}_m(s)} = -K_c\left(1 + \frac{1}{\tau_I s}\right)$$

where $\bar{P}(s)$ is the Laplace transform of the controller output.

This transfer function may be combined with Figure 3.5 to give Figure 3.10a which represents the loop in the time domain. Figure 3.10b is a rearrangement of Figure 3.10a.

Now the state variables are defined as the output of the integrators and so by inspection the state variable continuous time dynamic model is:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{K_1K_2K_cK_v}{\tau_I\tau_p} & \frac{1}{\tau_p}(K_1K_2K_cK_v - 1) \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{K_L}{\tau_p} \end{bmatrix} Q_i(t)$$

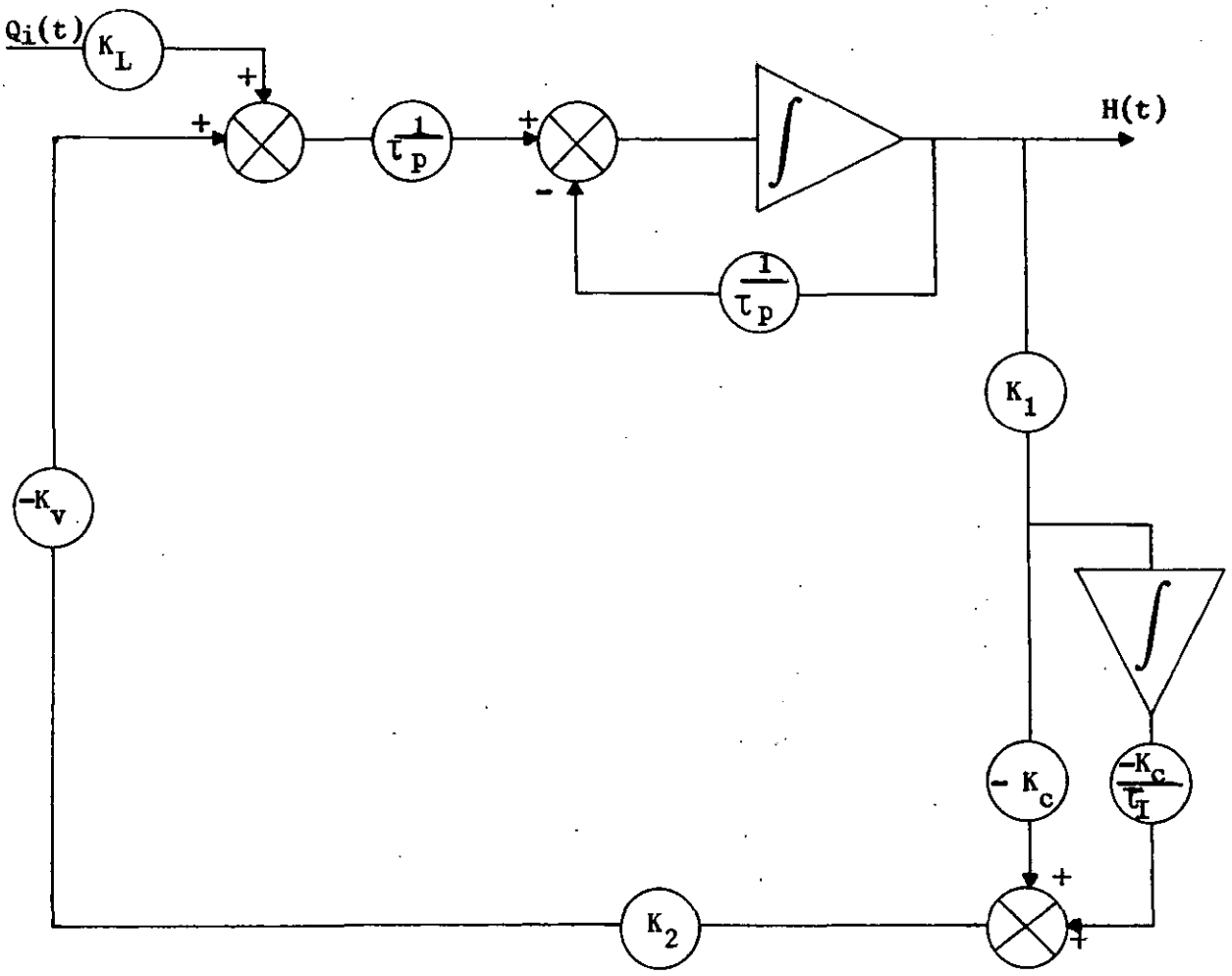
with $x_1(t) = \int H(t)$

$x_2(t) = H(t)$

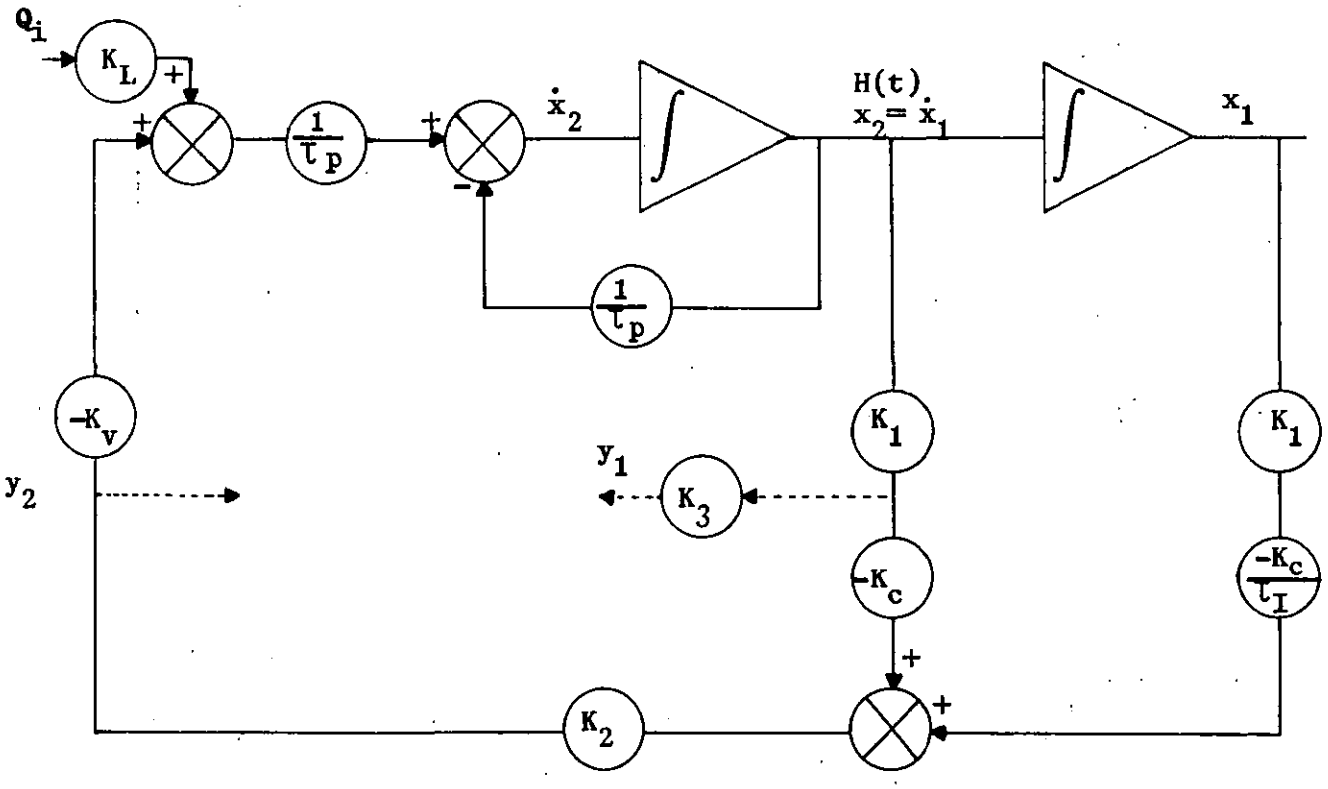
The process measurements are:

y_1 = tank height (V)

y_2 = control valve stem position (V)



a.



b.

FIGURE 3.10 State space representation of level analogue setpoint control loop.

and the observation model is

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 0 & K_1K_3 \\ -\frac{K_1K_2K_c}{\tau_I} & -K_1K_2K_c \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

By substituting the experimental parameters from Table 3.3, the state model becomes:

$$\dot{\underline{x}}(t) = \begin{bmatrix} 0 & 1 \\ -1.071487 & -0.578015 \end{bmatrix} \underline{x}(t) + \begin{bmatrix} 0 \\ 0.003571 \end{bmatrix} Q_i(t) \qquad (3.12.1.1)$$

$$\underline{y}(t) = \begin{bmatrix} 0 & 0.631296 \\ 0.26 & 0.103806 \end{bmatrix} \underline{x}(t) \qquad (3.12.1.2)$$

The above continuous time equations may be written in the general form

$$\dot{\underline{x}}(t) = F\,\underline{x}(t) + B\,\underline{u}(t) \qquad (3.12.1.3)$$

$$\underline{y}(t) = C\,\underline{x}(t) \qquad (3.12.1.4)$$

Now since this work is concerned with a digital computer sampling process measurements at discrete intervals of time Δt, the corresponding discrete time dynamic model is needed. It is well known that the discrete forms of equations (3.12.1.3) and (3.12.1.4) are given by (114)

$$\underline{x}(k+1) = A\,\underline{x}(k) + \Gamma\,\underline{u}(k) \qquad (3.12.1.5)$$

$$\underline{y}(k) = H\,\underline{x}(k) \qquad (3.12.1.6)$$

where $A = e^{F\Delta t}$ (3.12.1.7)

$$= \int_0^{\Delta t} e^{Ft} B\, dt \qquad (3.12.1.8)$$

Equations (3.12.1.1) and (3.12.1.2) were discretised with $\Delta t = 1$ sec

according to equations (3.12.1.5) to (3.12.1.8) to yield

$$\begin{bmatrix} x_1\ (k+1) \\ x_2\ (K+1) \end{bmatrix} = \begin{bmatrix} 0.999852 & 0.0165858 \\ -0.0177715 & 0.990265 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0.59228 \times 10^{-4} \end{bmatrix} Q_i(k) \qquad (3.12.1.9)$$

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} 0 & 0.631295 \\ 0.26 & 0.103806 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \qquad (3.12.1.10)$$

3.12.2 Direct digital control

If a process control computer is substituted for the analogue controller in the block diagram shown in Figure 3.5, then since the computer only executes control at discrete time intervals, the closed loop system may be considered to be composed of discrete and continuous time elements. Such a closed loop system is illustrated in Figure 3.11

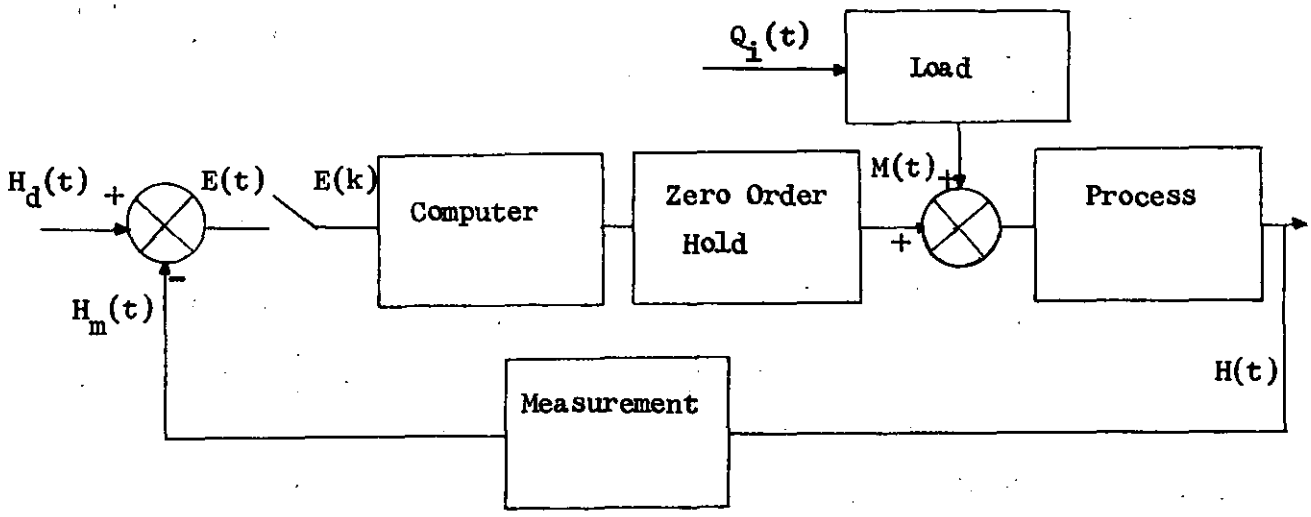


FIGURE 3.11. Block diagram for computer control system.

The sampler or analogue-to-digital converter transforms a continuous time signal E(t) into a sampled signal E(k), and the zero order hold or digital-to-analogue converter is required to maintain control over the system during intervals between data transfers.

Discrete time systems may be analysed using z transforms (114) or by considering the time domain solutions. The latter approach is adopted in this study.

The time domain block diagram for the computer control loop is shown in Figure 3.12.

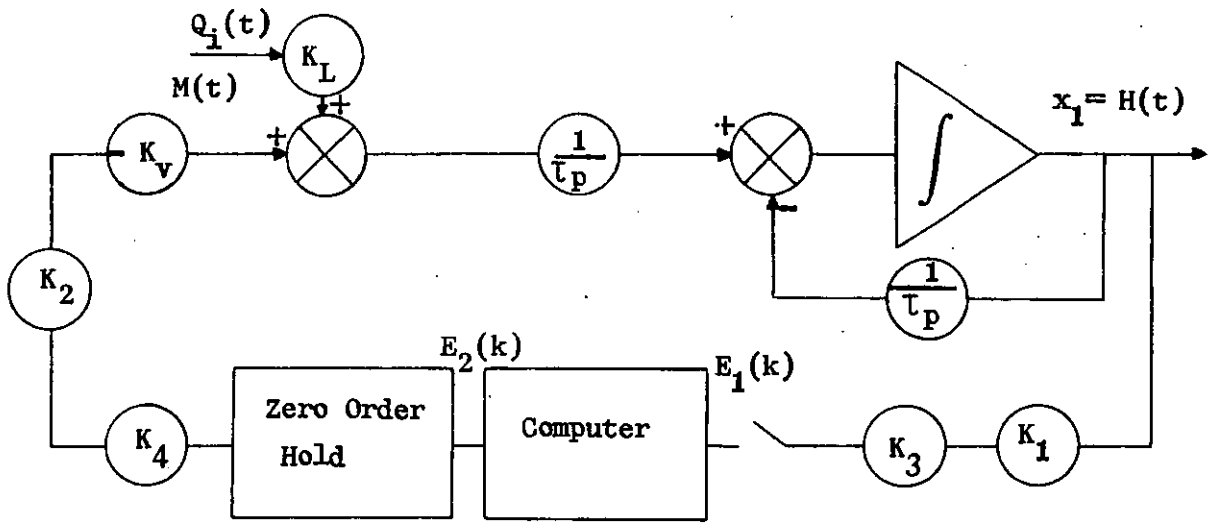


FIGURE 3.12 State space representation of level direct digital control loop.

Now defining x_1 as the output of the integrator, a state model for the open loop process may be written as:

$$\dot{x}_1(t) = \frac{-1}{\tau_p} x_1(t) + \frac{K_L}{\tau_p} Q_i(t) + \frac{M(t)}{\tau_p}$$

$$\text{i.e. } \dot{x}_1(t) = \left[-\frac{1}{\tau_p}\right] x_1(t) + \left[\frac{1}{\tau_p} \quad \frac{K_L}{\tau_p}\right] \begin{bmatrix} M(t) \\ Q_i(t) \end{bmatrix} \qquad (3.12.2.1)$$

Equation (3.12.2.1) is of the same form as equation (3.12.1.3) and so discretising according to equations (3.12.1.7) and (3.12.1.8) with

a discrete sample time of Δt gives:

$$x_1(k+1) = e^{-\Delta t/\tau_p} \; x_1(k) + \left[\left(1 - e^{-\Delta t/\tau_p}\right)\left(K_L - K_L\, e^{-\Delta t/\tau_p}\right)\right] \begin{bmatrix} M(k) \\ Q_i(k) \end{bmatrix} \qquad (3.12.2.2)$$

The computer control algorithm was chosen to be the discrete equivalent of the proportional-plus-integral analogue controller. This may be written as (43):

$$E_2(k) = E_2(k-1) + K_c\,(E_1(k) - E_1(k-1) + \frac{\Delta t}{\tau_I}\, E_1(k))$$

$$= K_c\left(1 + \frac{\Delta t}{\tau_I}\right) E_1(k) - K_c\, E_1(k-1) + E_2(k-1) \qquad (3.12.2.3)$$

Now this is a standard recursion formula of the general form:

$$E_2(k) = b_o\, E_1(k) + b_1\, E_1(k-1) - a_1\, E_2(k-1)$$

Cadzow and Martens (114) have derived a Jordon canonical state space representation of this equation which is:

$$x(k+1) = -a_1\, x(k) + E_1(k)$$

$$E_2(k) = b_3\, x(k) + b_o\, E_1(k)$$

and $$b_3 = b_1 - b_o\, a_1$$

Substituting from equation (3.12.2.3) and defining $x = x_2$ yields:

$$x_2(k+1) = x_2(k) + E_1(k) \qquad (3.12.2.4)$$

$$E_2(k) = \left(\frac{K_c \Delta t}{\tau_I}\right) x_2(k) + K_c\left(1 + \frac{\Delta t}{\tau_I}\right) E_1(k) \qquad (3.12.2.5)$$

The state space flow diagram illustrating the discrete time controller algorithm is shown in Figure 3.13.

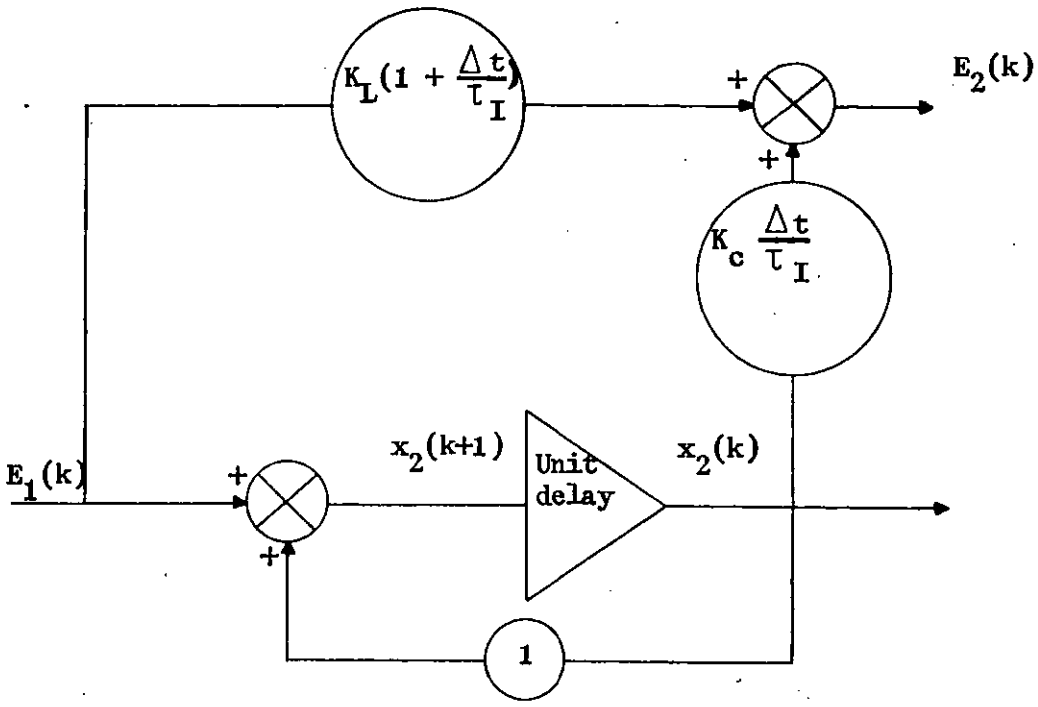


FIGURE 3.13. State space representation of discrete time proportional plus integral controller.

Now remembering

$$E_1(k) = - K_1 K_3 \, x_1(k)$$

$$M(k) = - K_2 K_4 K_v \, E_2(k)$$

and setting $K_f = - K_2 K_4 K_v$; $K_m = K_1 K_3$

equations (3.12.2.2) to (3.12.2.5) may be combined to give Figure 3.14 and the corresponding state space model

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} \left(e^{-\Delta t/\tau_p} - K_m K_c K_f \left(1 + \frac{\Delta t}{\tau_I}\right)(1 - e^{-\Delta t/\tau_p})\right) & \left(\frac{K_f K_c \Delta t}{\tau_I}(1 - e^{-\Delta t/\tau_p})\right) \\ - K_m & 1 \end{bmatrix} \underline{x}(k)$$

$$+ \begin{bmatrix} K_L(1 - e^{-\Delta t/\tau_p}) \\ 0 \end{bmatrix} Q_i(k)$$

The process observations are

y_1 = tank height (V)

y_2 = valve demand signal (V)

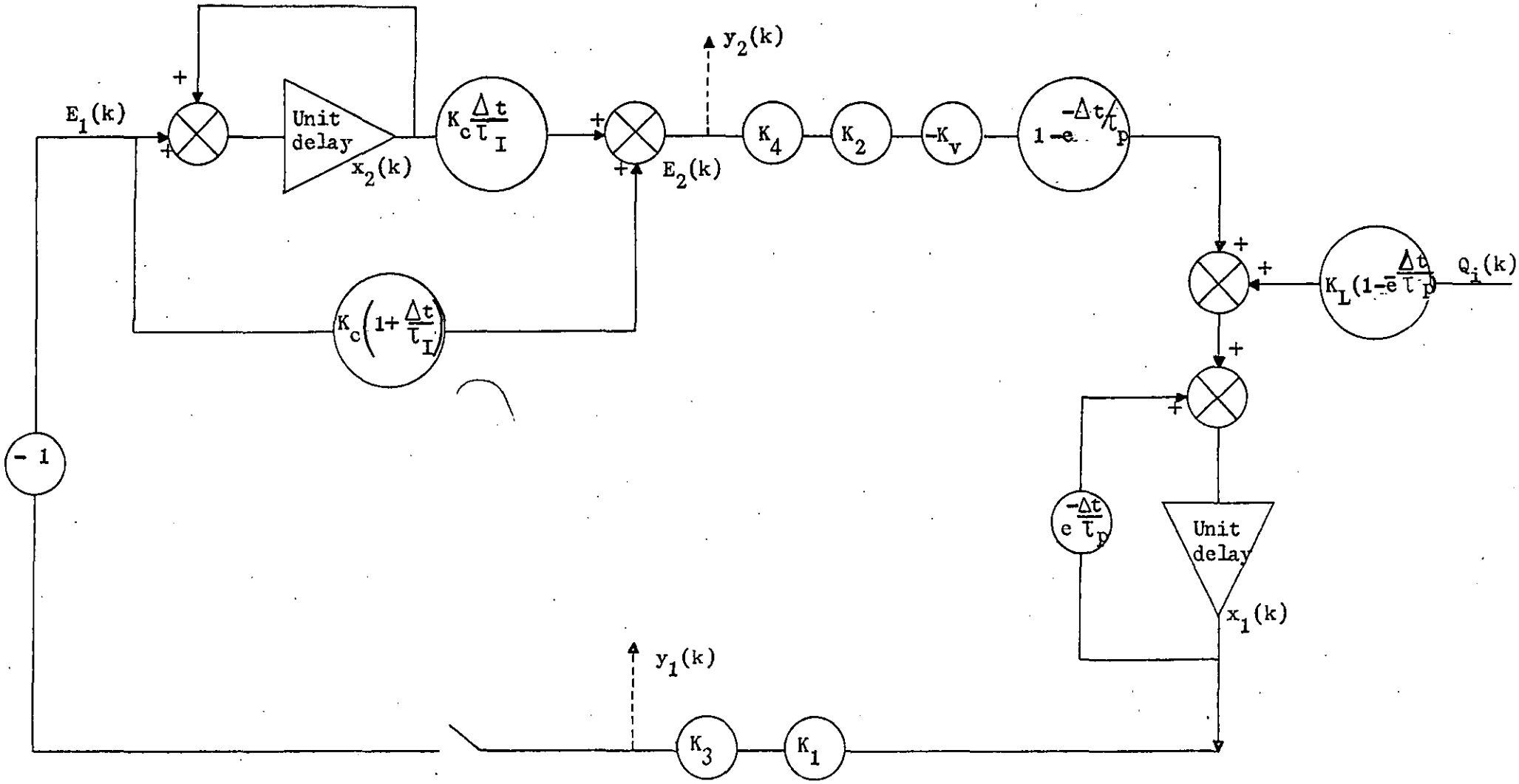


FIGURE 3.14 State space representation of level direct digital control loop.

Notice that the observation y_2 is not a true process measurement, but is in fact the control action determined by the control algorithm. The "pseudo" measurement was chosen because it is calculated in the computer and so avoids the need for a control valve stem position measurement instrument. The observation model is:

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} K_m & 0 \\ -K_c K_m\left(1 + \frac{\Delta t}{\tau_I}\right) & \frac{K_c \Delta t}{\tau_I} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}$$

Using the process parameters of Table 3.3 and a discrete time sampling interval of 2 seconds the state model becomes:

$$\underline{x}(k+1) = \begin{bmatrix} 0.955594 & 0.006402 \\ -0.631295 & 1. \end{bmatrix} \underline{x}(k) + \begin{bmatrix} 0.00011876 \\ 0 \end{bmatrix} Q_i(k)$$

(3.12.2.6)

$$\underline{y}(k) = \begin{bmatrix} 0.631295 & 0 \\ -0.703443 & 0.114288 \end{bmatrix} \underline{x}(k)$$

(3.12.2.7)

The models given by equations (3.12.1.9), (3.12.1.10) and (3.12.2.6), (3.12.2.7) which describe the analogue setpoint and d.d.c. loops respectively are now in the standard form to implement the linear time invariant Kalman filter, and are summarised in Table 3.4.

The process disturbance $Q_i(k)$ is an unmeasured variable and is modelled as a zero mean white noise sequence $\{\underline{w}(k)\}$

For linear time invariant models, tests for controllable and observable systems are given by Wiberg (115).

A system is said to be controllable if the following test is true:

$$\text{rank}\left[\Gamma,\ A\Gamma,\ A^2\Gamma,\ \ldots,\ A^{n-1}\Gamma\right] = n$$

The corresponding test of observability is

Control loop	A	Γ	H	Sample time Δt (sec)
Analogue setpoint	$\begin{bmatrix} 0.999852 & 0.0165858 \\ -0.0177715 & 0.990265 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.59228 \times 10^{-4} \end{bmatrix}$	$\begin{bmatrix} 0 & 0.631295 \\ 0.26 & 0.103806 \end{bmatrix}$	1.
Direct digital setpoint	$\begin{bmatrix} 0.955594 & 0.006402 \\ -0.631295 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.11876 \times 10^{-3} \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.631295 & 0 \\ -0.703443 & 0.114288 \end{bmatrix}$	2.

TABLE 3.4 State variable models for analogue setpoint and direct digital control loops.

$$\text{rank} \begin{bmatrix} H \\ HA \\ HA^2 \\ \vdots \\ HA^{n-1} \end{bmatrix} = n$$

Now applying these tests to the models of Table 3.4 reveals that both the analogue setpoint and d.d.c. loops are completely controllable and observable.

3.13 System dynamic characteristics

The open loop dynamic response of the level control system was determined by setting the system at the steady state specified in Table 3.2, and then imposing a step change of 140 cm^3/min on the tank input flowrate q_i by using the computer to alter the setpoint of the secondary flow control loop.

The experimentally measured tank height response is shown as curve 1 in Figure 3.15 where it is compared with the theoretical response determined from the first order transfer function of equation (3.11.1.11) using the parameters given in Table 3.3.

The first order system time constant is determined when the experimental response is 63.2% complete (56) and is calculated as 7.07 min which compares favourably with the value given in Table 3.3.

The closed loop responses for both the analogue setpoint and d.d.c. loops were experimentally obtained in a similar manner by initially controlling the height at the desired setpoint and then introducing a step load disturbance of 750 cm^3/min.

Figures 3.16a and b show the experimental and theoretical height responses for the analogue and d.d.c. loops respectively. The theoretical responses were calculated from closed loop models defined in Table 3.4.

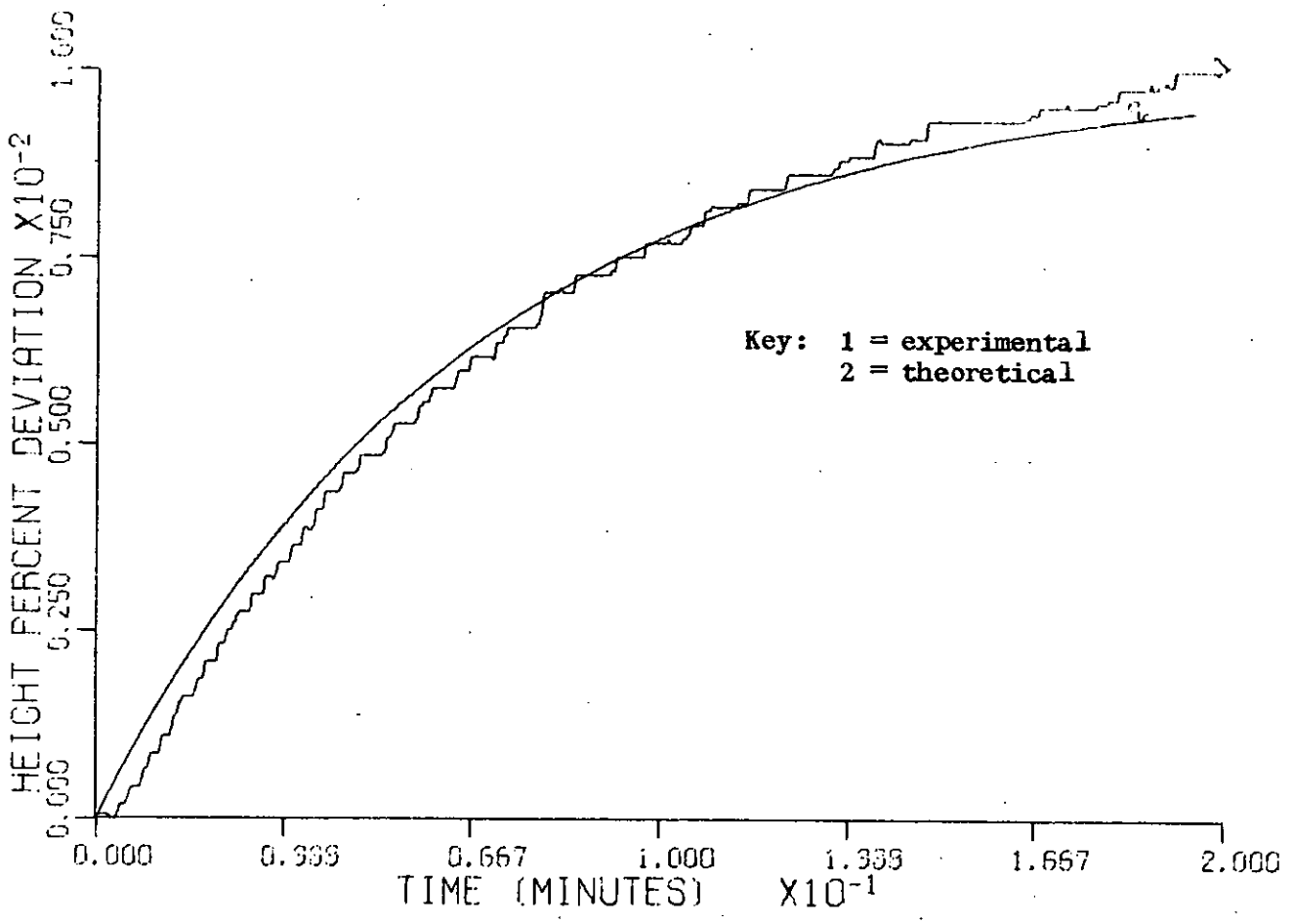


FIGURE 3.15 Level open loop experimental and theoretical responses.

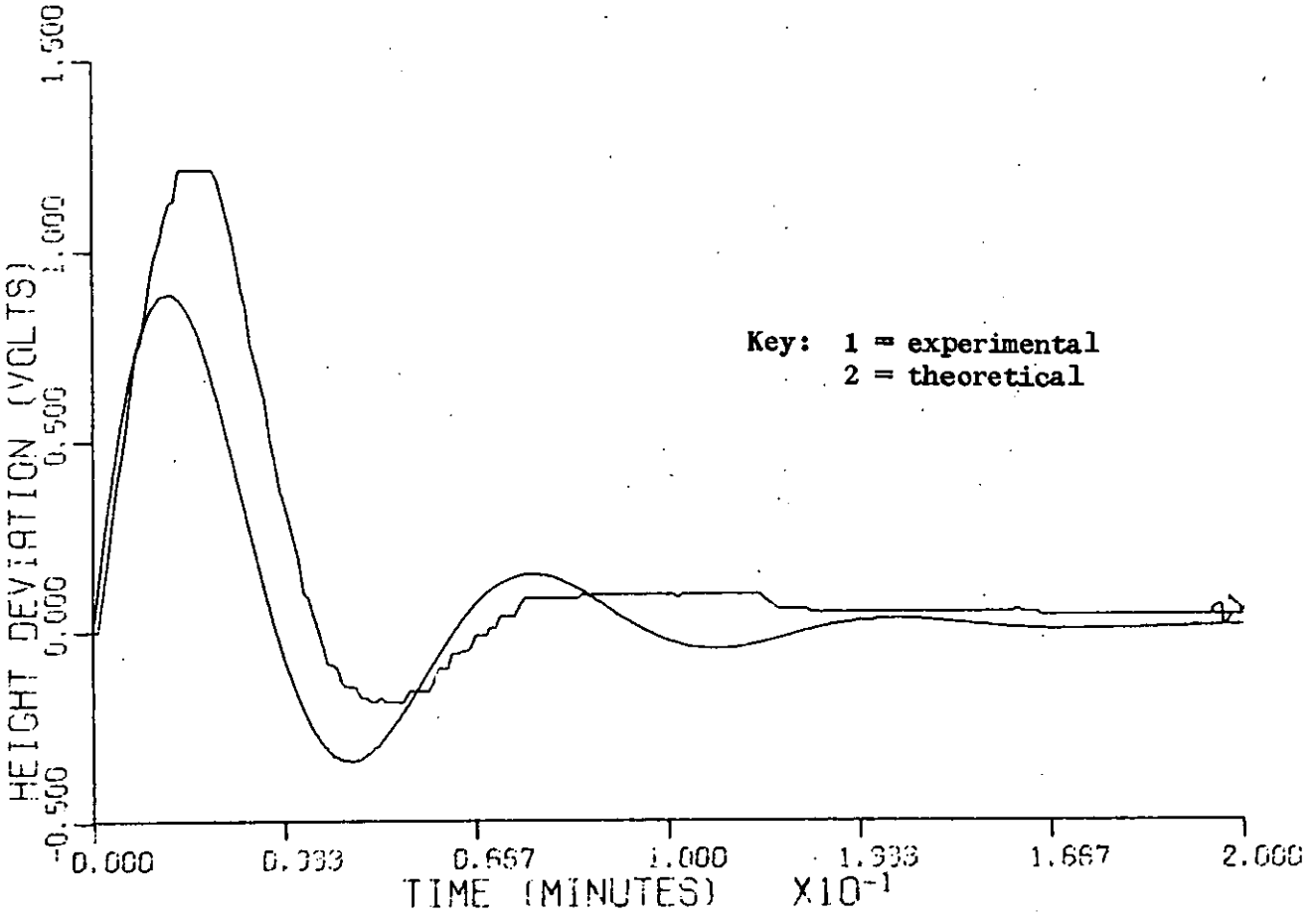
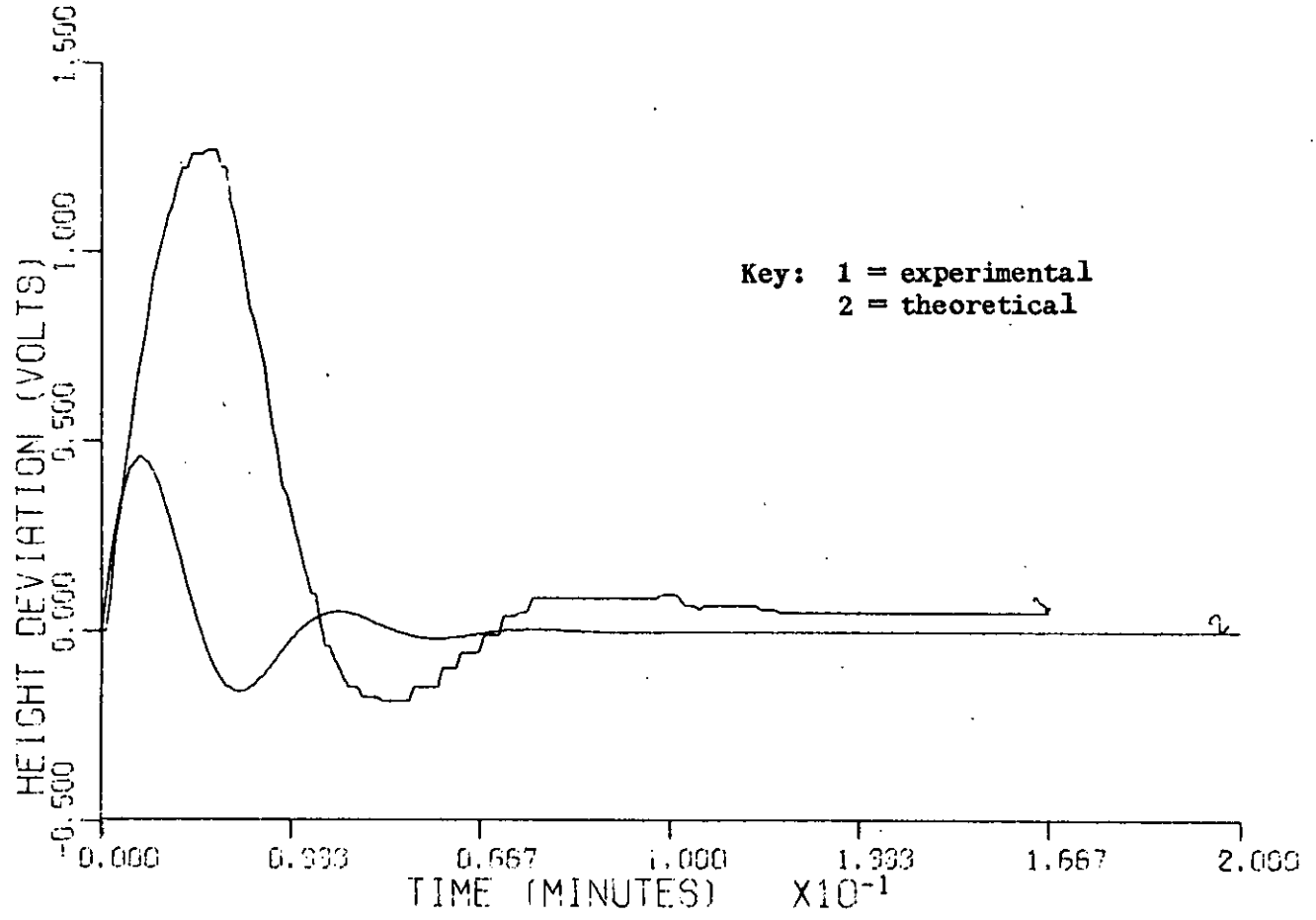
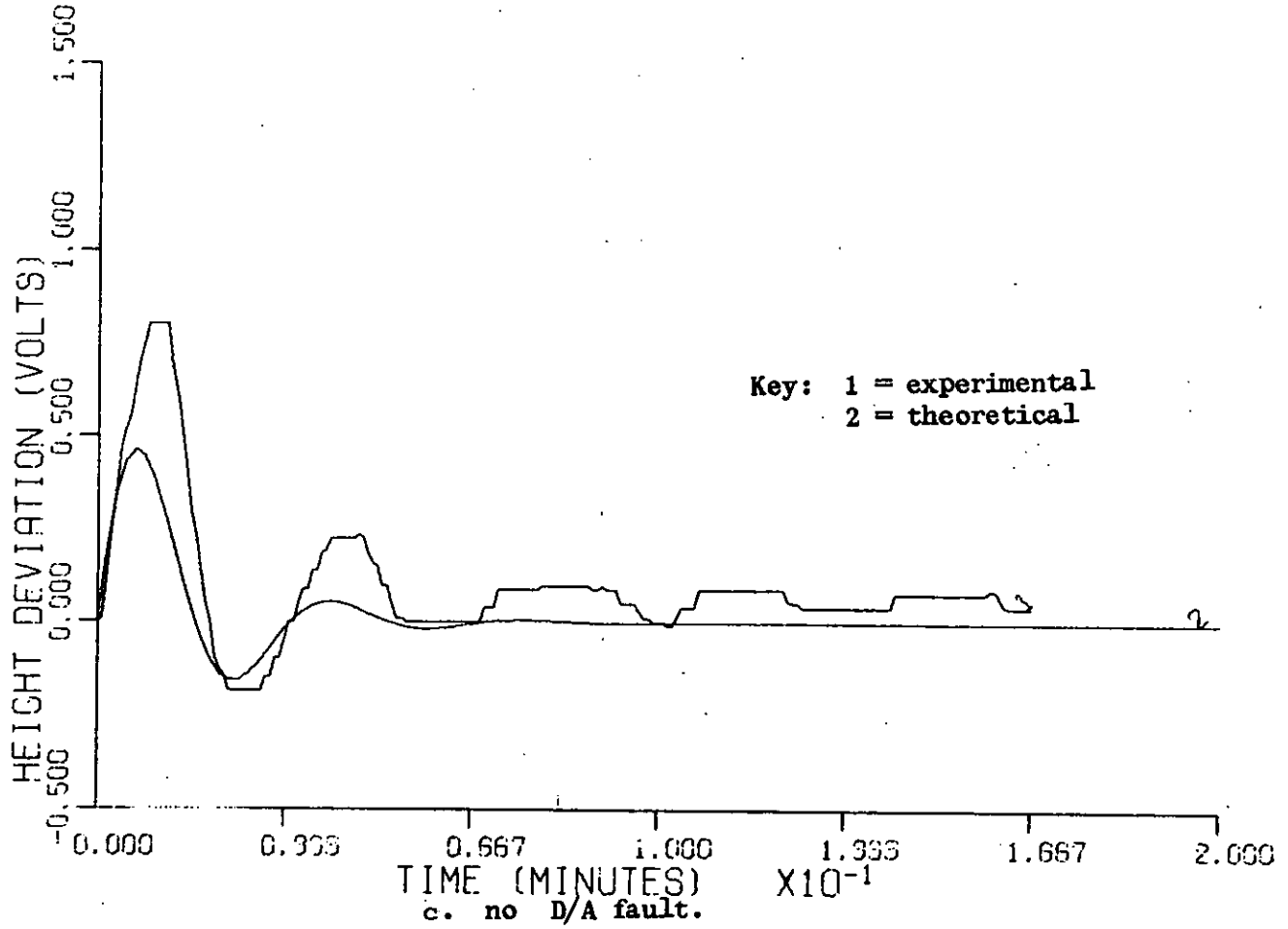


FIGURE 3.16a Analogue closed loop experimental and theoretical responses.



b. D/A fault.



c. no D/A fault.

FIGURE 3.16 D.D.C. closed loop experimental and theoretical responses.

The state space model for the analogue setpoint control loop provides a reasonable representation for the experimental response as shown in Figure 3.16a. The discrepancies between the responses may be accounted for in the inaccuracy of actually determining the controller settings from the dials on the pneumatic instrument.

However, Figure 3.16b shows that the d.d.c. model of the control loop is a poor representation of the experimental response. The reason for these differing responses was eventually traced to be in the process computer digital/analogue converter. It was found that the least six bits of the digital/analogue converter were permanently disconnected. This means that the 0-10 V computer output signal was discretised into 16 steps instead of the usual 1024 increments, thereby giving a discontinuous control signal characterised by steps of approximately 0.6 V. The effect of this computer output signal is to produce rather "slack" control.

Unfortunately this process computer fault was not discovered until after the malfunction detection experimental programme was completed. However, as far as the implementation of the Kalman filter is concerned, the effect of this fault is to produce a discrepancy between the system model and the experimental response thereby compounding the uncertainty problem.

When the malfunction detection experimental programme was complete the computer digital/analogue converter was repaired and an experimental step response for the d.d.c. loop was determined as above (using the same controller constants). This experimental response together with the theoretical response derived from the d.d.c. state variable model is shown in Figure 3.16c. As expected, these responses show better agreement than those shown in Figure 3.16b.

3.14 Experimental procedure

3.14.1 Analogue setpoint control

All of the experiments were performed with the process operating at the steady state characteristics detailed in Table 3.2. The tank height was controlled at the desired value using the pneumatic controller by opening the isolation valve 1 and closing 2 as shown in Figure 3.4. The PDP 11-20 process computer introduced an unmeasured disturbance into the flowrate q_i by imposing a pseudo-random binary sequence (PRBS) on to the flow controller setpoint. The PRBS had an amplitude of 1 l/min and a basic switching time of 15 seconds, as well as a programmed facility to start the sequence at randomly selected points in the chain. In addition to creating this disturbance, the computer logged 1000 process measurements of tank height and control valve stem position at 1 second intervals. This logged data was subsequently analysed off-line on an ICL 1904 A computer using the Kalman filter and malfunction detection method discussed earlier.

Some typical time histories of the process variables are shown in Figures 3.17 a-c.

3.14.2 Direct digital control

These experiments followed the same format as those above except that computer control was achieved using a PDP 11-20 computer by closing isolation valve 1 and opening 2, as shown in Figure 3.4. The unmeasured PRBS flow control setpoint disturbance had an amplitude of 0.6 l/min with a switching time of 5 seconds. At 2 second time intervals the process computer performed the control task and logged measurements of tank height and the control valve demand signals.

Typical time histories of the process variables are shown in Figures 3-18 a-c.

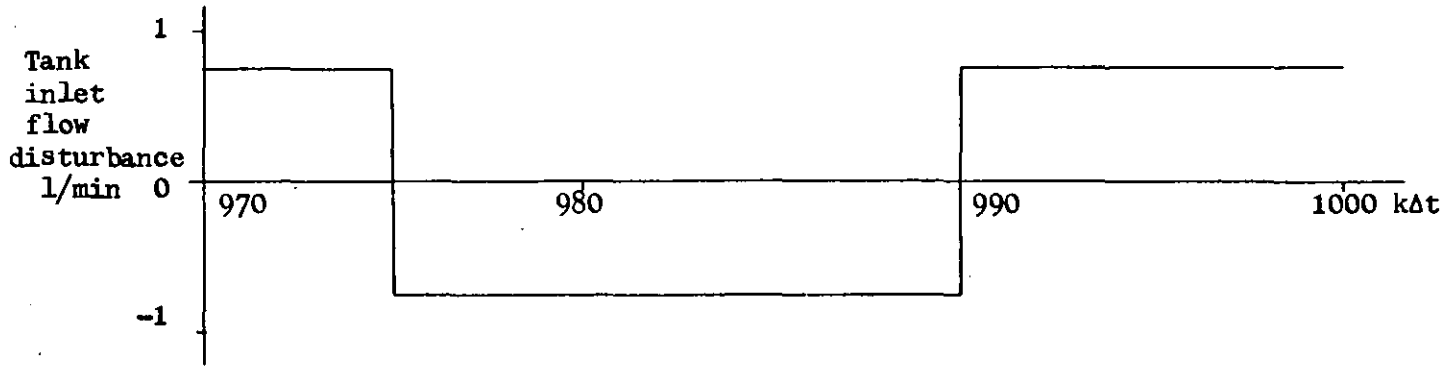


FIGURE 3.17a Pseudo random binary disturbance in analogue setpoint control experiments.

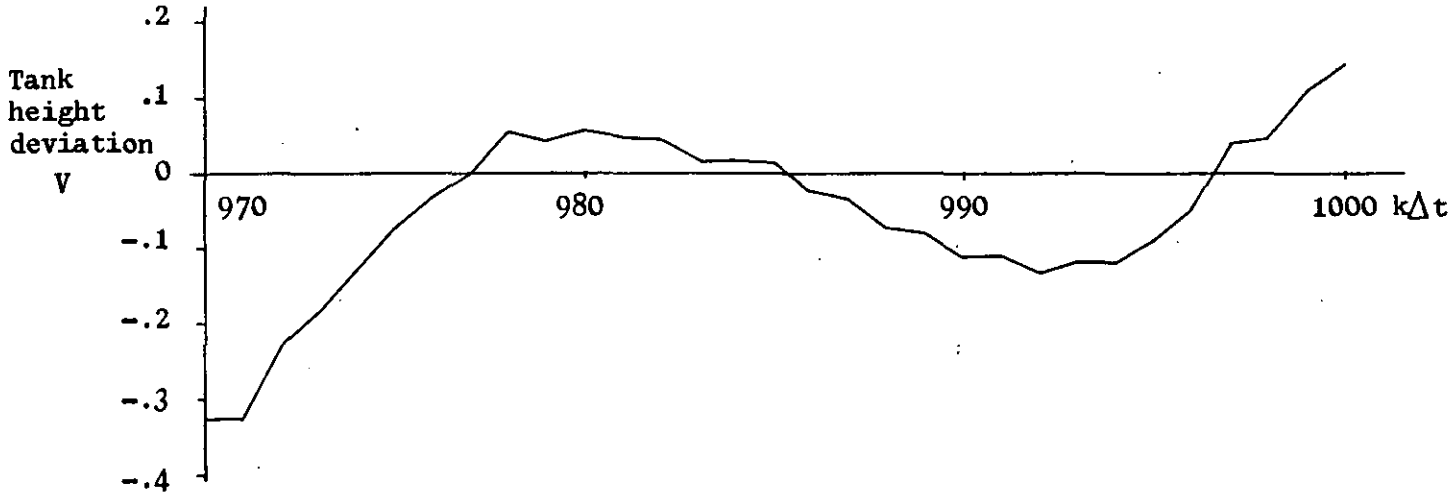


FIGURE 3.17b Tank height deviation measurement in analogue setpoint control experiments.

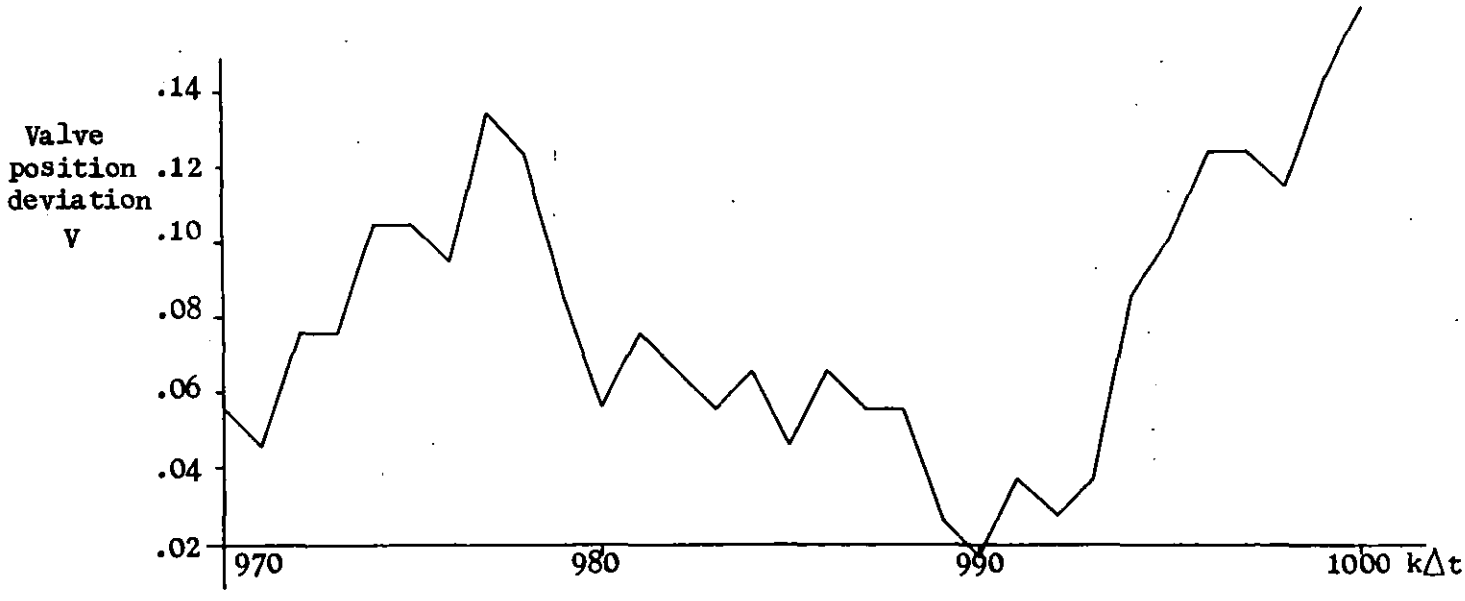


FIGURE 3.17c Valve position deviation measurement in analogue setpoint control experiments.

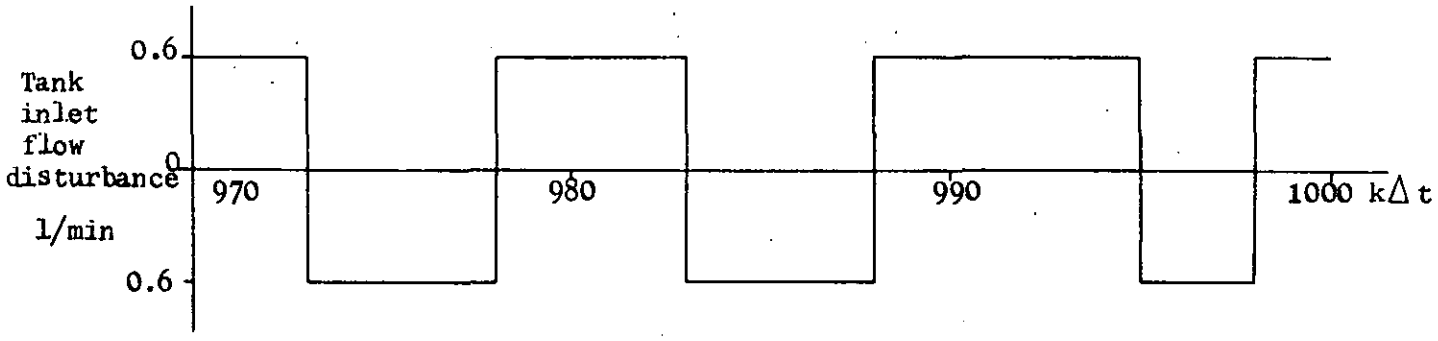


FIGURE 3.18a Pseudo random binary disturbance in direct digital control experiments.

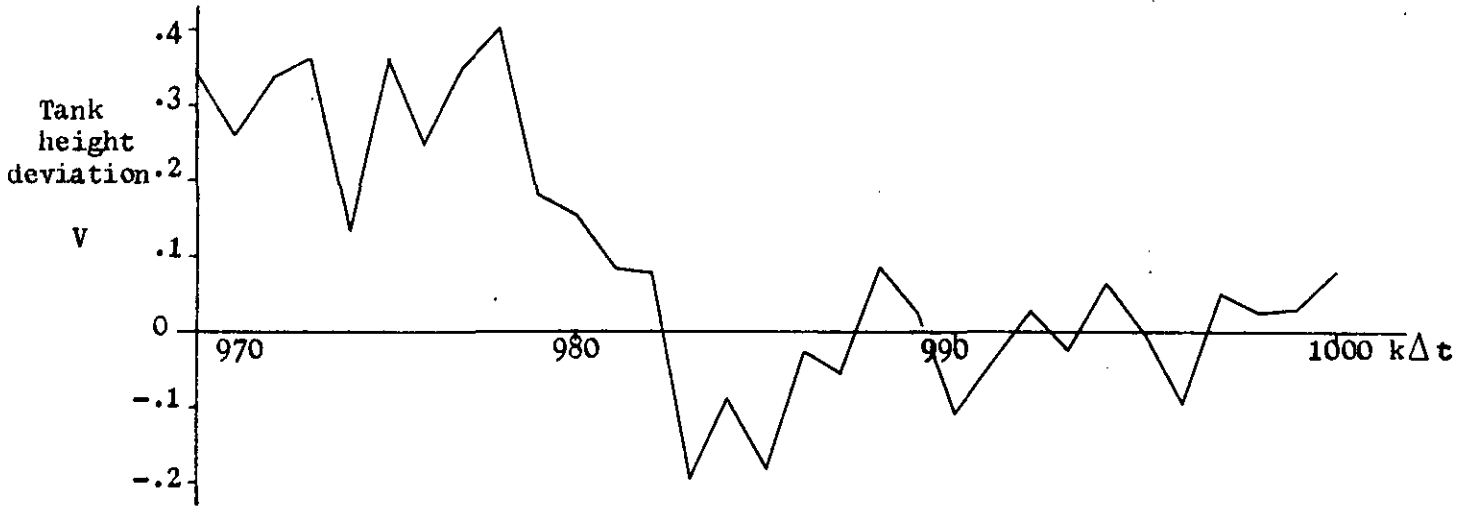


FIGURE 3.18b Tank height deviation measurement in direct digital control experiments.

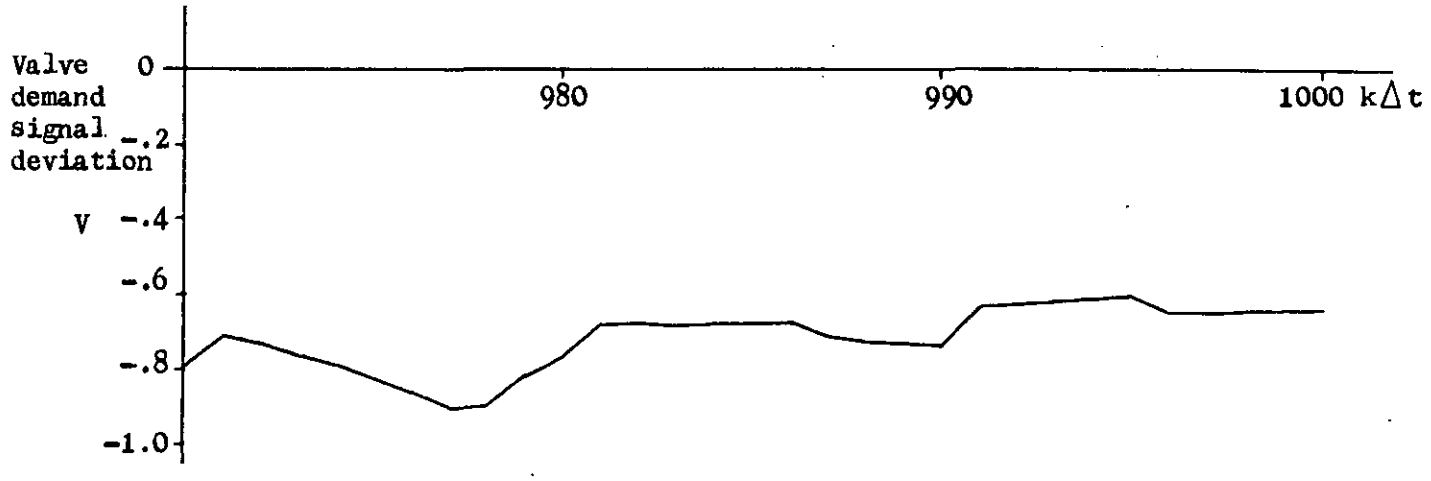


FIGURE 3.18c Valve demand signal deviation measurement in direct digital control experiments.

3.15 Kalman filter design using Mehra's adaptive estimator

Sections 3.11 to 3.12 have been concerned with the creation of a mathematical model to represent the level control system. The function of the Kalman filter is now to relate the process measurements via the mathematical model to yield a data base from which loop security may be assessed.

However section 3.5 has detailed that the mechanisation of the Kalman filter is no trivial matter because of the uncertainty of the mathematical model and difficulty of choosing suitable a priori statistical parameters. To overcome these problems Mehra's adaptive estimator (87), as described in Appendix III, was used in this work.

The level control system was assumed to be operating in a malfunction free condition and a batch of 1000 process measurements was recorded.

Mehra's technique was implemented on an I.C.L. 1904A computer and his direct method of estimating the optimal Kalman gain by measurement refiltering was used. A maximum of 8 iterations was used on the data unless the gain converged according to the criterion:

$$\text{tr} \ (\Delta K \ \Delta K^T) \ \leq \ 0.0001$$

Successive estimates of $\hat{K}_{opt}$, $\hat{R}$, $\hat{PH}^T$ were not obtained by analysing further batches of measurement data as suggested by Mehra.

3.15.1 Analogue setpoint control

The state variable model for this control loop was given in Table 3.4. The Kalman filter was used to process the measurements and two arbitrary data sets of a priori information were chosen to effect this.

The a priori data sets are given in Table 3.5 and are denoted AM1 and AM2. Table 3.5 also shows the steady state Kalman gain and error prediction covariance matrices resulting from the usual Kalman filter algorithm.

Although the K_{ss} and $P(k/k)_{ss}$ matrices in Table 3.5 do not explicitly describe how well the Kalman filter is working, the fact that they are numerically small does indicate that perhaps filter measurement decoupling has occurred.

In each run, for the sample of 1000 points the resultant innovation sequence $\underline{v}(k)$ was generated and the estimates of the autocorrelation matrices $\hat{C}_0, \hat{C}_1 \ldots \hat{C}_k$ $(k = 60)$ were calculated from equation (A.III.7.). The white noise test revealed that the innovation sequence was non-white and so Mehra's algorithm was used to estimate the optimal Kalman gain by iterating on the measurements. The iteration sequence is shown in Table 3.6 from which it can be seen that the algorithm does not converge. This non-convergence results, at the final iteration, in a large incremental Kalman gain correction term, ΔK, and an estimated $\hat{R}$ matrix with a negative diagonal element.

Now one possible source of error in the results presented above lies in the assumptions concerning the noise sequences $\{Q_i(k)\}$ and $\{\underline{v}(k)\}$. Mehra's method assumes that these sequences are uncorrelated with zero mean. The validity of the zero mean assumption in the experiments performed was not necessarily true. However a method due to Godbole (106), (107) generalised Mehra's technique to handle the non-zero mean correlated noise case and was reviewed in section 3.7.2. Godbole's modification was included in the analysis, although it was still assumed that the sequences $\{Q_i(k)\}$ and $\{\underline{v}(k)\}$ were uncorrelated.

Run	Q	R	$\underline{x}(o)$	P(o/o)	K_{ss}	$P(k/k)_{ss}$
AM1	1000	$\begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0.3 \times 10^{-5} & 0.42 \times 10^{-3} \\ 0.1 \times 10^{-2} & 0.18 \times 10^{-3} \end{bmatrix}$	$\begin{bmatrix} 0.16 \times 10^{-3} & 0.5 \times 10^{-6} \\ 0.5 \times 10^{-6} & 0.17 \times 10^{-3} \end{bmatrix}$
AM2	10000	$\begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0.31 \times 10^{-2} & 0.13 \times 10^{-1} \\ 0.43 \times 10^{-1} & 0.84 \times 10^{-2} \end{bmatrix}$	$\begin{bmatrix} 0.46 \times 10^{-3} & 0.5 \times 10^{-4} \\ 0.5 \times 10^{-4} & 0.69 \times 10^{-3} \end{bmatrix}$

TABLE 3.5 A priori information for Kalman filter design in analogue setpoint control loop.

Run AM1 / M			Run AM2 / M	
Iteration number	Percentage of points lying outside the 95% confidence limits.		Percentage of points lying outside the 95% confidence limits.	
	First measurement	Second measurement	First measurement	Second measurement
0	100	100	100	100
1	100	30	100	30
2	63.33	25.	66.67	31.67
3	100.	53.33	100.	51.67
4	93.33	63.33	93.33	63.33
5	100.	58.33	100.	58.33
6	96.67	80.	96.67	83.33
7	100.	63.33	100.	63.33
8	96.67	53.33	96.67	86.67
$\hat{K}_{opt}$	$\begin{bmatrix} -1.10219 & 2.18915 \\ 1.93259 & 1.15811 \end{bmatrix}$		$\begin{bmatrix} -1.09825 & 2.17994 \\ 1.92808 & 1.16669 \end{bmatrix}$	
ΔK	$\begin{bmatrix} 0.355117 & -0.0383561 \\ -0.306376 & 0.204687 \end{bmatrix}$		$\begin{bmatrix} 0.365263 & -0.0262073 \\ -0.311216 & 0.204639 \end{bmatrix}$	
$\hat{R}$	$\begin{bmatrix} -0.346 \times 10^{-3} & -0.508 \times 10^{-3} \\ 0.142 \times 10^{-3} & 0.183 \times 10^{-3} \end{bmatrix}$		$\begin{bmatrix} -0.340 \times 10^{-3} & -0.513 \times 10^{-3} \\ 0.137 \times 10^{-3} & 0.182 \times 10^{-3} \end{bmatrix}$	

TABLE 3.6 Kalman gain estimation using Mehra's direct method on analogue setpoint control measurements - runs AM1/M; AM2/M.

The a priori statistics corresponding to run AM1 and the process measurements were reanalysed using Mehra's modified algorithm and the resultant iteration sequence, denoted as run AM1/G is shown in Table 3.7.

Iteration number	Run AM1/G			
	Innovation sequence mean		Percentage of points lying outside the 95% confidence limits	
	First measurement	Second measurement	First measurement	Second measurement
0	-0.00388702	0.541456	100.	100.
1	0.0236335	0.00427509	50.	23.33
2	0.0145027	0.00406255	33.33	21.67
3	0.0198971	0.00564762	38.33	28.33
4	0.0151437	0.00588248	25.	28.33
5	0.0179015	0.00663521	33.33	31.67
6	0.0151878	0.007284	25.	31.67
7	0.0165251	0.00750063	31.67	25.
8	0.0149736	0.00809775	20.	25.
$\hat{K}_{opt}$	$\begin{bmatrix} -1.117068 & 2.3327 \\ 1.94108 & 0.921385 \end{bmatrix}$			
ΔK	$\begin{bmatrix} 0.0990945 & -0.101863 \\ -0.122053 & 0.0709527 \end{bmatrix}$			
$\hat{R}$	$\begin{bmatrix} -0.343 \times 10^{-3} & -0.337 \times 10^{-3} \\ 0.189 \times 10^{-3} & 0.175 \times 10^{-3} \end{bmatrix}$			

TABLE 3.7 Kalman gain estimation using Mehra's modified method on analogue setpoint control measurements - run AM1/G.

These results show that convergence was still not achieved after 8 iterations and the estimated matrices were similar to those obtained without Godbole's modification in run AM1/M.

Now visual inspection of the process measurement time histories

shown in Figure 3.17 reveals that the process measurement noise is small and possibly unrepresentative of a true control loop. Jazwinski (57) has commented on dynamic systems with perfect measurements (i.e. R = 0) and points out that although the Kalman filter is valid, the matrix P(k/k-1) can become ill-conditioned leading to difficulties in the computation of P(k/k). Although this phenomenon is not directly related to Mehra's algorithm it was surmised that a similar effect was being exhibited.

To investigate this idea the process measurements were artificially made more "noisy" by adding a computer generated zero mean Gaussian random variable to each measurement. The covariance matrix of the computer based measurement noise was :

$$R_c = \begin{bmatrix} 0.0064 & 0. \\ 0. & 0.0004 \end{bmatrix} \qquad (3.15.1.1)$$

The "noisy" measurements corresponding to Figures 3.17 b and c are shown in Figures 3.19 a and b.

Using the initial conditions corresponding to AM1, Mehra's direct method was used to estimate the optimal gain by analysing the "noisy" measurements, and the resulting iteration sequence is given in Table 3.8.

The results of optimising the Kalman gain using these "noisy" measurements in Mehra's algorithm modified by Godbole to account for non-zero noise means are given in Table 3.9 for the a priori information sets corresponding to runs AM1 and AM2.

Tables 3.8 and 3.9 illustrate that Mehra's direct adaptive estimator and the modified method of Godbole show convergence to an optimal Kalman gain $\hat{K}_{opt}$, which is numerically quite similar. The

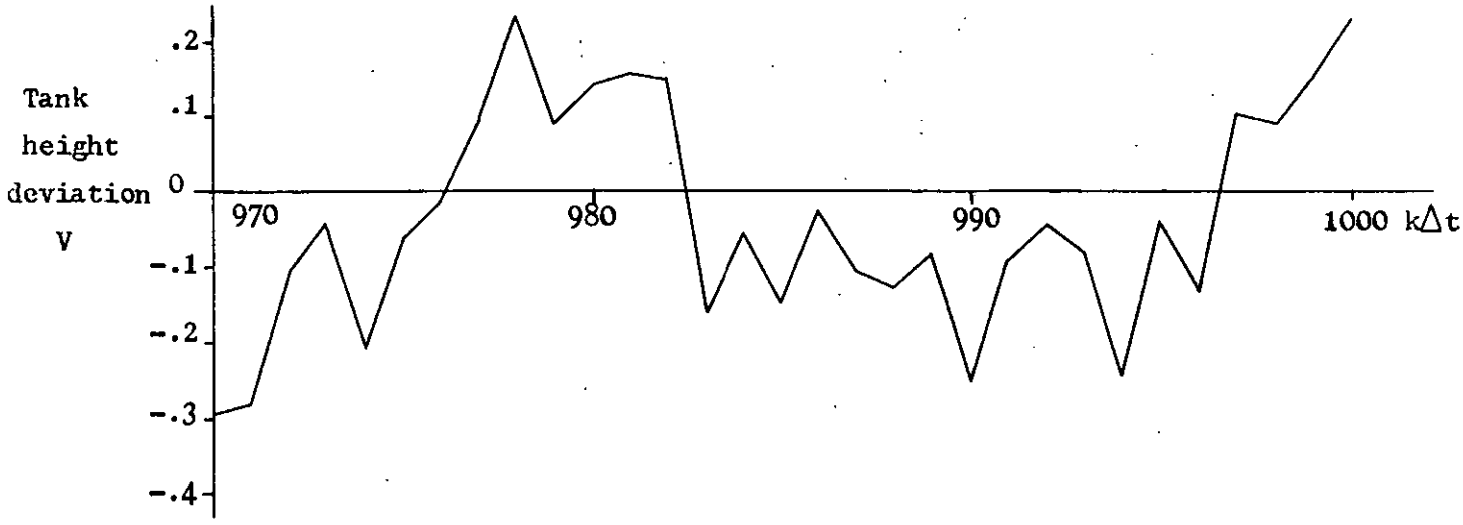


FIGURE 3.19a "Noisy" tank height deviation measurement in analogue setpoint control experiments.

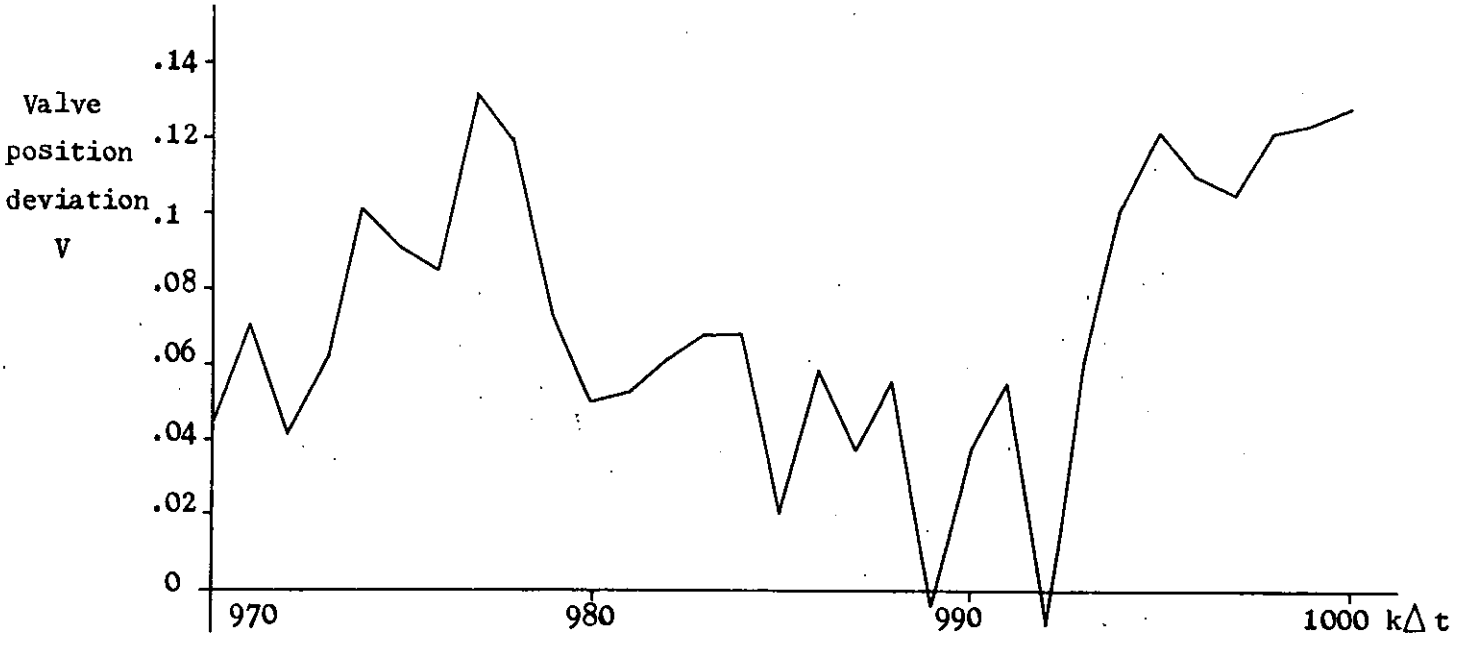


FIGURE 3.19b "Noisy" valve position deviation measurement in analogue setpoint control experiments.

RUN AM1/M/N		
Iteration number	Percentage of points lying outside the 95% confidence limits	
	First measurement	Second measurement
0	100.	100.
1	36.67	15.
2	76.67	21.67
3	41.67	51.67
4	43.33	58.33
5	51.67	56.67
6	50.	56.67
7	51.67	56.67
8	Convergence	
$\hat{K}_{opt}$	$\begin{bmatrix} -0.427082 & 1.17257 \\ 0.803763 & 1.94171 \end{bmatrix}$	
ΔK	$\begin{bmatrix} 0.619 \times 10^{-3} & 0.265 \times 10^{-3} \\ -0.459 \times 10^{-3} & 0.706 \times 10^{-3} \end{bmatrix}$	
$\hat{R}$	$\begin{bmatrix} 0.490 \times 10^{-2} & -0.103 \times 10^{-2} \\ 0.637 \times 10^{-3} & 0.562 \times 10^{-3} \end{bmatrix}$	
$\hat{PH}^T$	$\begin{bmatrix} -0.4164 \times 10^{-2} & 0.1018 \times 10^{-2} \\ 0.1052 \times 10^{-1} & 0.2662 \times 10^{-2} \end{bmatrix}$	

TABLE 3.8 Kalman gain estimation using Mehra's direct method on analogue setpoint control "noisy" measurements - run AM1/M/N.

Iteration number	Run AM1/G/N: Innovation sequence mean		Run AM1/G/N: Percentage of points lying outside the 95% confidence limits		Run AM2/G/N: Innovation sequence mean		Run AM2/G/N: Percentage of points lying outside the 95% confidence limits	
	First measurement	Second measurement	First measurement	Second measurement	First measurement	Second measurement	First measurement	Second measurement
0	-0.0045	0.5418	100.	100.	0.20641	0.367272	100.	100.
1	0.024	0.00426	15.	15.	0.02381	0.004286	18.33	15.
2	0.0350	0.00654	16.67	8.33	0.03487	0.007554	16.67	8.33
3	0.02773	0.0089	13.33	20.	0.02723	0.008927	11.67	20.
4	0.0281	0.01	13.33	20.	0.0284	0.0101	13.33	20.
5	0.0285	0.00972	13.33	20.	0.0285	0.00968	13.33	20.
6	0.0287	0.00878	13.33	20	0.0287	0.00979	13.33	20.
7	0.0286	0.00976	13.33	20	0.0286	0.00975	13.33	20.
8			Convergence		0.0286	0.00977	13.33	20.
$\hat{K}_{opt}$	$\begin{bmatrix} -0.413443 & 1.3154 \\ 0.75243 & 1.64501 \end{bmatrix}$				$\begin{bmatrix} -0.413435 & 1.31525 \\ 0.752738 & 1.64706 \end{bmatrix}$			
ΔK	$\begin{bmatrix} 0.272 \times 10^{-3} & -0.208 \times 10^{-3} \\ -0.700 \times 10^{-3} & -0.287 \times 10^{-2} \end{bmatrix}$				$\begin{bmatrix} -0.102 \times 10^{-4} & -0.174 \times 10^{-3} \\ 0.465 \times 10^{-3} & 0.304 \times 10^{-2} \end{bmatrix}$			
$\hat{R}$	$\begin{bmatrix} 0.522 \times 10^{-2} & -0.788 \times 10^{-3} \\ 0.570 \times 10^{-3} & 0.510 \times 10^{-3} \end{bmatrix}$				$\begin{bmatrix} 0.521 \times 10^{-2} & -0.794 \times 10^{-3} \\ 0.570 \times 10^{-3} & 0.511 \times 10^{-3} \end{bmatrix}$			
$\hat{PH}^T$	$\begin{bmatrix} -0.385 \times 10^{-2} & 0.113 \times 10^{-2} \\ 0.910 \times 10^{-2} & 0.205 \times 10^{-2} \end{bmatrix}$				$\begin{bmatrix} -0.386 \times 10^{-2} & 0.113 \times 10^{-2} \\ 0.910 \times 10^{-2} & 0.207 \times 10^{-2} \end{bmatrix}$			

TABLE 3.9 Kalman gain estimation using Mehra's modified method on analogue setpoint control "noisy" measurements - runs AM1/G/N; AM2/G/N

estimated matrices $\hat{R}$ and $\hat{P}$ (k/k-1) (found by solving $\widehat{PH}^T = P(k/k-1)H^T$) are consistent and in particular $\hat{R}$ is a good estimate of the computer generated measurement noise covariance matrix given in equation (3.15.1.1).

The results of Table 3.9 show the robustness and consistency of the general innovation correlation adaptive estimator since the algorithm converges to the same $\hat{K}_{opt}$ for different a priori statistics.

However, even though the gain optimisation algorithms converge, inspection of Tables 3.8 and 3.9 reveals that the resulting "optimal" innovation sequence fails the white noise test, although the method which includes Godbole's modifications shows less violations of the 95% confidence limits than Mehra's direct method. This means that the resulting Kalman filter is not optimal. There are several reasons which may account for this feature.

Primarily Mehra's innovation correlation adaptive estimator does not explicitly cope with the problem of uncertainty in the system dynamic and measurement models A, Γ and H. The method actually assumes that these models are correct and equal to the true system model. In this work the A, Γ and H matrices used as the process model were certainly subject to parameter uncertainties, and therefore it is reasonable to expect Mehra's adaptive estimator performance to be degraded, as exemplified by the non-whiteness of the resultant innovation sequence.

A second feature is that Mehra's problem formulation is for a linear time invariant dynamic system. However, the level control loop model was in fact non-linear and although a linearised model about a particular steady state was derived, the validity of such a model may have been violated due to the relatively large process disturbances. Another source of non-linearity lies in the control loop pneumatic

hardware. For example, the pneumatic controller was represented by an ideal proportional plus integral transfer function given in equation (3.11.2.1), while in practice such a realisation would be unlikely. Again it is suggested that the consequence of these non-linearities is to cause a degradation in the performance of the adaptive estimator.

A further assumption of Mehra's estimator is that the record length of process observations is long enough to justify the approximation that the estimated innovation correlation matrices $\hat{C}_k$ are equal to the true correlation functions. This is only valid if there are an infinite number of process observations, and so in practice there is an estimation error associated with the calculation of $\hat{C}_k$. Again this error may degrade the innovation correlation algorithm. In this work a sample of 1000 process observations was used to calculate $\hat{C}_k$.

However, in spite of these apparent sources of error, Mehra's unmodified and modified innovation correlation estimator does still converge to an estimate of the optimal Kalman gain $\hat{K}_{opt}$, as illustrated in Tables 3.8 and 3.9. The resultant $\hat{K}_{opt}$ when used in the Kalman filter generates state estimates $\underline{\hat{x}}(k/k)$ which are more accurate than those obtained from ordinary Kalman filtering with the a priori statistics given earlier. This feature is shown in Figures 3.20 a and b for run AM1/G/N. The "noisy" measurements $\underline{y}(k)$ are shown as well as the estimates $\underline{\hat{y}}(k) = H\,\underline{\hat{x}}(k/k)$ resulting from the usual Kalman filter, with the a priori formation AM1, and the optimal filter gain from run AM1/G/N. These Figures illustrate that the estimates resulting from ordinary Kalman filtering using the a priori information AM1 are extremely poor. The adaptive estimator is able to detect this inadequate filter performance, and in spite of all the limitations discussed above manages to estimate a Kalman gain which results in greatly improved estimates $\underline{\hat{y}}(k)$.

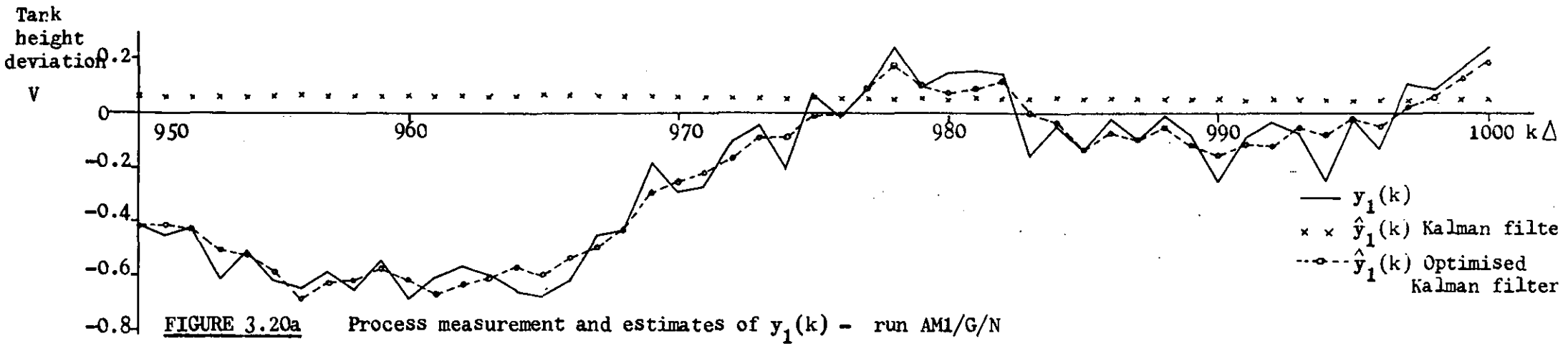


FIGURE 3.20a Process measurement and estimates of $y_1(k)$ - run AM1/G/N

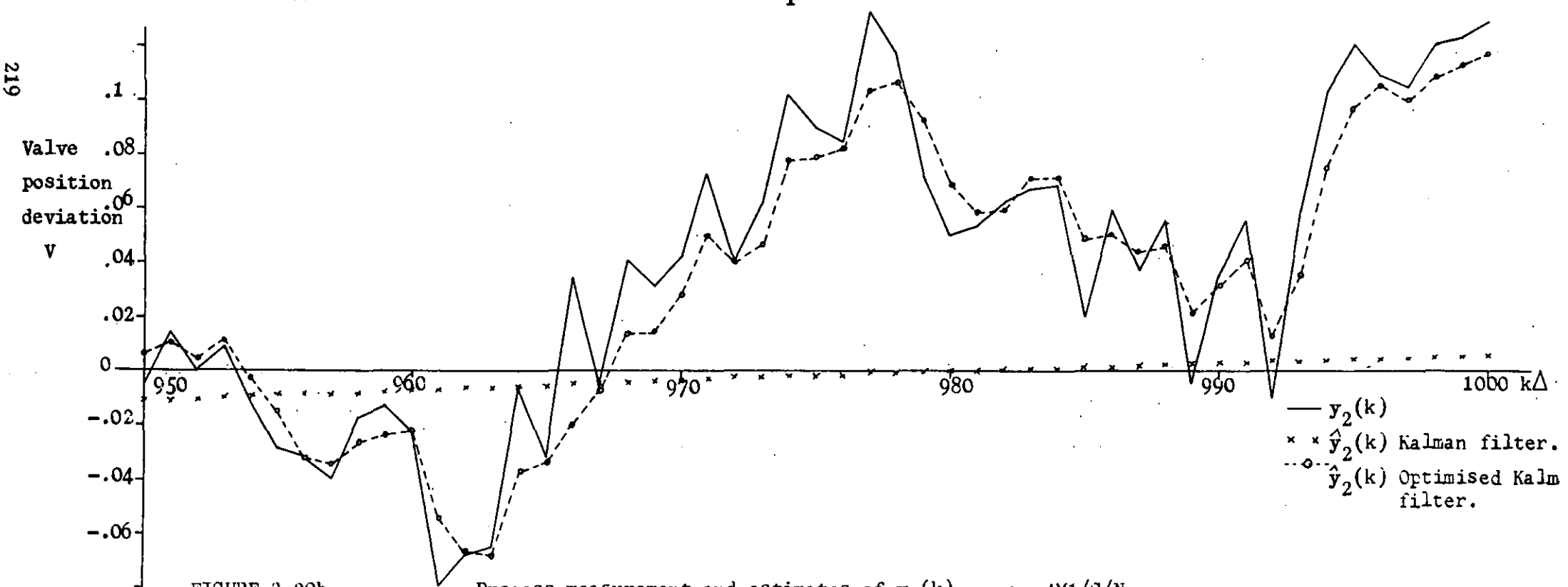


FIGURE 3.20b Process measurement and estimates of $y_2(k)$ - run AM1/G/N.

3.15.2 Direct digital control

The experience of estimating the optimal Kalman gain in the analogue setpoint control loop revealed that Mehra's algorithm was unsatisfactory when the process measurement noise was small. In these experiments the tank height process measurement was made "noisier" by physically altering the damping adjustment on the height measurement Δ P/P transmitter. The second "pseudo measurement" in the d.d.c. loop was the calculated valve demand signal, which of course was deterministic and hence uncontaminated with random measurement noise.

The state variable model describing the d.d.c. loop was given in Table 3.4.

Now 1000 process measurements, sampled at 2 second intervals, were analysed by the ordinary Kalman filter to estimate the system state using the following set of a priori information.

$$Q = 1000$$

$$R = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$$

$$\underline{x}(0) = \underline{0}$$

$$P(0/0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Ordinary Kalman filtering resulted in the steady state matrices:

$$K_{ss} = \begin{bmatrix} 0.108 \times 10^{-2} & -0.117 \times 10^{-2} \\ 0.177 \times 10^{-3} & 0.181 \times 10^{-1} \end{bmatrix}$$

$$P(k/k)_{ss} = \begin{bmatrix} 0.171 \times 10^{-3} & 0.280 \times 10^{-4} \\ 0.280 \times 10^{-4} & 0.160 \times 10^{-1} \end{bmatrix}$$

The 1000 process measurements were analysed using Mehra's direct gain estimator as before and the resulting iteration sequence is shown in Table 3.10.

Run DM1/M		
Iteration number	Percentage of points lying outside the 95% confidence limits	
	First measurement	Second measurement
0	91.67	91.67
1	10.	23.33
2	8.33	10.
3	8.33	16.67
4	5.	16.67
5	5.	13.33
6	5.	16.67
7	5.	13.33
8	5.	16.67
9	5.	13.33
$\hat{K}_{opt}$	$\begin{bmatrix} 0.980574 \times 10^{-1} & -1.86566 \\ 0.904654 & -0.503898 \end{bmatrix}$	
ΔK	$\begin{bmatrix} 0.105 \times 10^{-2} & 0.785 \times 10^{-2} \\ 0.106 \times 10^{-1} & 0.519 \times 10^{-1} \end{bmatrix}$	
$\hat{R}$	$\begin{bmatrix} 0.480 \times 10^{-4} & 0.702 \times 10^{-3} \\ 0.480 \times 10^{-4} & -0.367 \times 10^{-3} \end{bmatrix}$	

TABLE 3.10 Kalman gain estimation using Mehra's direct method on direct digital control measurements - run DM1/M

These results do not show convergence to an optimal gain, although the incremental change ΔK at the last iteration is small compared to $\hat{K}_{opt}$. The resulting innovation sequence of the second measurement fails the white noise test indicating the suboptimality of the "optimised" filter. However, the most disturbing feature of the algorithm

is that $\hat{r}_{22}$ is negative.

This problem was encountered in the previous work on the analogue setpoint control loop where it was suggested that the algorithm failed because of the low process measurement noise. In this d.d.c. loop it is known that the pseudo measurement $y_2(k)$ is noise-free and again it is suggested that this feature causes algorithm breakdown.

To overcome this problem the valve demand signal pseudo-measurement was contaminated with noise generated by the process computer. However, in this experiment rather than generating a zero mean Gaussian random variable, a pseudo random binary sequence was formed. This PRBS represents an approximation to a zero mean white noise sequence, and was chosen because the computer generation of such sequences is trivial. The PRBS had a basic period of 1023 and an amplitude of 0.15. At each time interval the measurements were sampled and the computer generated PRBS was added to $y_2(k)$ to yield a "noisy" signal.

Now using the "noisy" measurement vector and the model with the a priori statistics given earlier, both Mehra's direct method and the algorithm which includes Godbole's modification were used to adapt the Kalman filter to account for uncertainty. These runs are denoted DM1/M/N and DM1/G/N respectively, and the resulting algorithm iteration sequences are given in Table 3.11.

The results show that both algorithms fail to converge to an optimal Kalman according to the criterion given earlier. After 8 iterations the resultant $\hat{K}_{opt}$ yields an innovation sequence which fails the white noise test in each case, although Godbole's modification results in fewer violations of the 95% confidence limit. Some reasons for the failure of the white noise test were discussed in the previous section. A further contribution to the performance

Iteration number	Run DM1/M/N		Run DM1/G/N			
	Percentage of points lying outside the 95% confidence limits.		Innovation sequence mean		Percentage of points lying outside the 95% confidence limits	
	1st measurement	2nd measurement	1st measurement	2nd measurement	1st measurement	2nd measurement
0	91.67	91.67	0.003313	0.18076	91.67	93.33
1	10.	18.33	−0.006757	0.007093	8.33	18.33
2	10.	36.	−0.01044	0.013617	6.67	30.
3	6.67	28.33	−0.009773	0.0135706	6.67	31.67
4	15.	28.33	−0.01018	0.013611	10.	31.67
5	20.	20.	−0.01013	0.01365	10.	31.67
6	20.	20.	−0.009684	0.01482	10.	28.33
7	20.	20.	−0.009576	0.01511	10.	31.67
8	20.	20.	−0.00983	0.01449	10.	28.33
$\hat{K}_{opt}$	$\begin{bmatrix} 0.68647 & -0.225944 \\ -0.11488 & 0.417491 \end{bmatrix}$		$\begin{bmatrix} 0.703524 & -0.206078 \\ -0.391664 \times 10^{-1} & 0.338993 \end{bmatrix}$			
ΔK	$\begin{bmatrix} 0.228 \times 10^{-2} & 0.100 \times 10^{-2} \\ 0.223 \times 10^{-1} & 0.560 \times 10^{-2} \end{bmatrix}$		$\begin{bmatrix} 0.756 \times 10^{-2} & 0.375 \times 10^{-2} \\ 0.850 \times 10^{-1} & 0.298 \times 10^{-1} \end{bmatrix}$			
$\hat{R}$	$\begin{bmatrix} 0.523 \times 10^{-2} & 0.154 \times 10^{-2} \\ 0.984 \times 10^{-3} & 0.229 \times 10^{-1} \end{bmatrix}$		$\begin{bmatrix} 0.531 \times 10^{-2} & 0.153 \times 10^{-2} \\ 0.178 \times 10^{-2} & 0.227 \times 10^{-1} \end{bmatrix}$			
$\hat{P}H^T$	$\begin{bmatrix} 0.851 \times 10^{-2} & -0.110 \times 10^{-1} \\ -0.326 \times 10^{-2} & 0.141 \times 10^{-1} \end{bmatrix}$		$\begin{bmatrix} 0.842 \times 10^{-2} & -0.918 \times 10^{-2} \\ -0.109 \times 10^{-2} & 0.110 \times 10^{-1} \end{bmatrix}$			

TABLE 3.11. Kalman gain estimation on direct digital control "noisy" measurements. Runs DM1/M/N; DM1/G/N.

degradation of the adaptive estimator in this work arises because of the measurement noise on the observation $y_2(k)$. Recall that Mehra's problem formulation defines the measurement noise to be a zero mean Gaussian random variable. In this work the contaminating noise added to the measurement $y_2(k)$ was an approximation to white noise in the form of a PRBS. Since this is not a true zero mean Gaussian random variable, then adaptive estimator performance may be expected to be degraded.

To illustrate this point a simulation, denoted DM2/G/N was performed corresponding to run DM1/G/N by contaminating the measurement $y_2(k)$ with a computer generated zero mean Gaussian random variable of variance equal to the PRBS. The results of using Mehra's algorithm with Godbole's modification are shown in Table 3.12.

Comparing runs DM1/G/N and DM2/G/N in Tables 3.11 and 3.12 does in fact confirm the above postulation concerning the degradation of filter performance due to the PRBS measurement noise. Table 3.12 for run DM2/G/N reveals that the Kalman gain does converge to an optimal value and the resulting innovation sequence white noise test has fewer violations of the 95% confidence limit than run DM1/G/N.

Despite the suboptimality of run DM1/G/N the resultant estimated matrices $\hat{R}$ and $\hat{P}$ (k/k-1) are consistent and the Kalman filter using $\hat{K}_{opt}$ provides improved state estimates $\underline{\hat{x}}(k/k)$ over those obtained from ordinary filtering using the assumed a priori information. This is illustrated in Figure 3.21 a-b, which compares the "noisy" measurement vector $\underline{y}(k)$ with the estimated measurements $\underline{\hat{y}}(k)$ resulting from ordinary and "optimal" Kalman filtering.

The results presented in section 3.15.1 and 3.15.2 have highlighted several features of Mehra's innovation correlation adaptive estimator.

Run DM2/G/N				
Iteration number	Innovation sequence mean		Percentage of points lying outside the 95% confidence limit.	
	First measurement	Second measurement	First measurement	Second measurement
0	0.0039763	0.17037	91.67	93.33
1	-0.006359	0.006803	10.	8.33
2	-0.010546	0.010409	8.33	6.67
3	-0.009686	0.012148	8.33	5.
4	-0.009819	0.011894	6.67	5.
5	-0.009928	0.011622	10.	3.33
6	-0.009951	0.0011568	10.	3.33
7	-0.0099565	0.001155	10.	3.33
8	Convergence			
$\hat{K}_{opt}$	$\begin{bmatrix} 0.737577 & -0.178306 \\ 0.544531 & 1.08129 \end{bmatrix}$			
ΔK	$\begin{bmatrix} -0.129 \times 10^{-3} & -0.961 \times 10^{-4} \\ 0.362 \times 10^{-2} & 0.278 \times 10^{-2} \end{bmatrix}$			
$\hat{R}$	$\begin{bmatrix} 0.523 \times 10^{-2} & 0.592 \times 10^{-3} \\ 0.1401 \times 10^{-2} & 0.185 \times 10^{-1} \end{bmatrix}$			
$\widehat{PH}^T$	$\begin{bmatrix} 0.878 \times 10^{-2} & -0.836 \times 10^{-2} \\ 0.773 \times 10^{-3} & 0.277 \times 10^{-1} \end{bmatrix}$			

TABLE 3.12 Kalman gain estimation on direct digital setpoint control "noisy" measurements - run DM2/G/N.

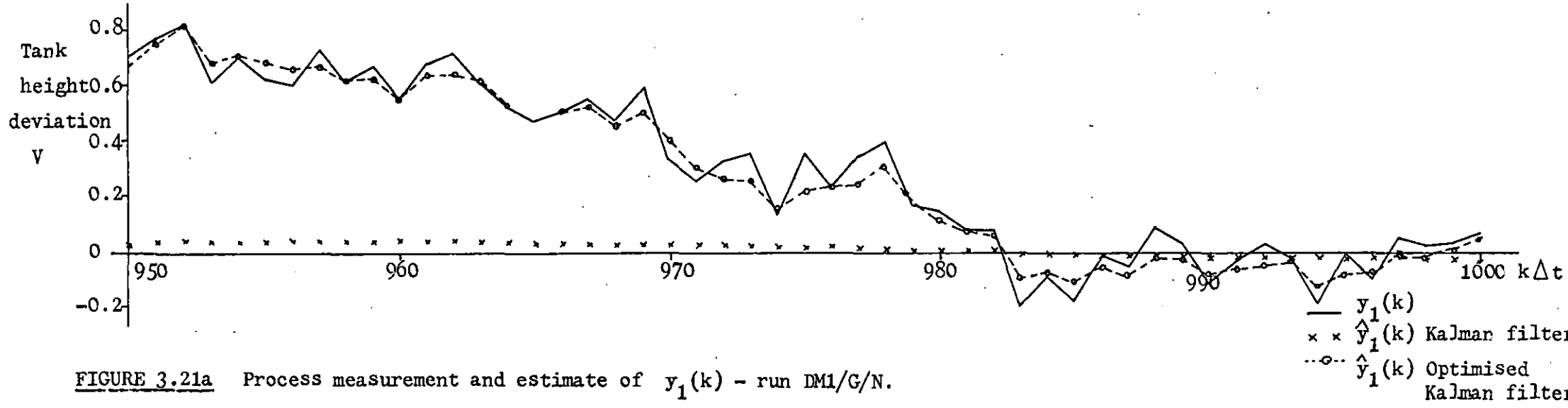


FIGURE 3.21a Process measurement and estimate of $y_1(k)$ - run DM1/G/N.

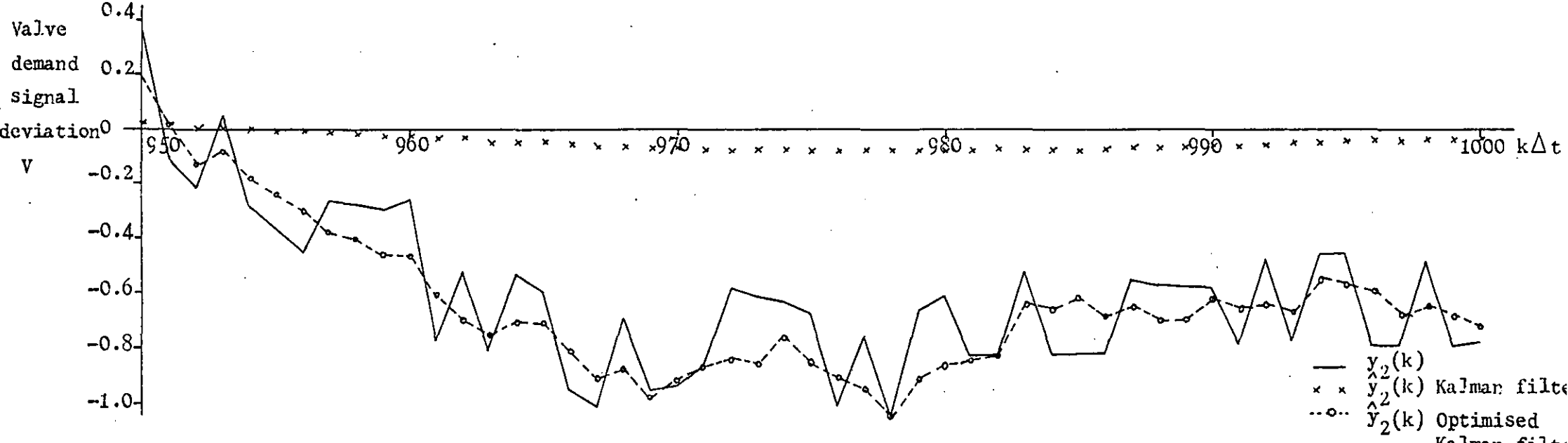


FIGURE 3.21b Process measurement and estimate of $y_{2(k)}$ - run DM1/G/N.

Fundamentally it has been found that for systems in which the measurement noise is small (i.e. the covariance matrix R is small), the adaptive estimator fails to converge and the resultant parameter estimates are not consistent. In this study this problem was solved by artificially increasing the measurement noise using a computer based random variable generator.

In all of the experiments performed the resultant innovation sequence did not satisfy the white noise test of Kalman filter optimality, although in each case Godbole's modification to Mehra's basic method yielded fewer violations of the 95% confidence limit. Some reasons for this unsatisfactory feature were discussed. However, in spite of the uncertainty in the system models, the inherent non-linearity of the process, and the non-whiteness of a measurement noise sequence, the resultant estimated matrices $\hat{K}_{opt}$, $\hat{R}$ and $\widehat{PH}^T$ were consistent and in particular, $\hat{R}$ was close to the true value R. The innovation correlation algorithms were shown to be robust and consistent and converged to the same estimated matrices for different a priori information.

In spite of the approximations used in implementing the adaptive estimator, the resultant "optimised" Kalman filter yielded accurate state estimates $\hat{\underline{x}}(k/k)$ as illustrated by comparing $\underline{y}(k)$ and $\hat{\underline{y}}(k)$

3.16 Malfunction detection experiments and determination of the loop malfunction gain ϕ

Experiments were performed on both the analogue setpoint and direct digital control loops, and are denoted AS and DS respectively.

The experimental procedure was detailed in section 3.14 and the estimator of Figure 3.3 was used to calculate the loop security parameters $\underline{b}(k)$ from the process measurements. The numerical values of the parameters used in the bias estimator were derived from the Kalman filter optimisation procedure of the previous section (runs AM1/G/N and DM1/G/N) and are summarised in Table 3.13.

The malfunctions introduced into the system were similar to those investigated in the flow control loop experiments, and are given in Table 3.14, while the complete experimental programme is summarised in Table 3.15.

Table 3.15 also includes the values of the control loop malfunction gain $\emptyset$ which was defined in section 3.2.

To illustrate the derivation of this function an example for the analogue setpoint control loop is considered.

The control loop block diagram of Figure 3.5 is redrawn as Figure 3.22 for convenience.

Loop	A	Γ	H	$\hat{K}_{opt}$	$\hat{R}$	$\hat{P}H^T$	$P_b(0/0)$
AS	$\begin{bmatrix} 0.999852 & 0.0165858 \\ -0.017771 & 0.990265 \end{bmatrix}$	$\begin{bmatrix} 0. \\ 0.59228\times10^{-4} \end{bmatrix}$	$\begin{bmatrix} 0. & 0.631295 \\ 0.26 & 0.103806 \end{bmatrix}$	$\begin{bmatrix} 0.413443 & 1.3154 \\ 0.75243 & 1.64501 \end{bmatrix}$	$\begin{bmatrix} 0.522\times10^{-2} & 0. \\ 0. & 0.51\times10^{-3} \end{bmatrix}$	$\begin{bmatrix} -0.385\times10^{-2} & 0.113\times10^{-2} \\ 0.910\times10^{-2} & 0.205\times10^{-2} \end{bmatrix}$	$\begin{bmatrix} 10. & 0. \\ 0. & 10. \end{bmatrix}$
DS	$\begin{bmatrix} 0.955594 & 0.006402 \\ -0.631295 & 1. \end{bmatrix}$	$\begin{bmatrix} 0.11876\times10^{-3} \\ 0. \end{bmatrix}$	$\begin{bmatrix} 0.631295 & 0. \\ -0.703443 & 0.114288 \end{bmatrix}$	$\begin{bmatrix} 0.703524 & -0.206078 \\ -0.0391664 & 0.338993 \end{bmatrix}$	$\begin{bmatrix} 0.531\times10^{-2} & 0. \\ 0. & 0.227\times10^{-1} \end{bmatrix}$	$\begin{bmatrix} 0.841\times10^{-2} & -0.918\times10^{-2} \\ -0.108\times10^{-2} & 0.110\times10^{-1} \end{bmatrix}$	[10.]

TABLE 3.13. Numerical values of parameters in loop security parameter estimator.

Loop malfunction	Source	Code
Level measurement zero error	Level differential pressure transducer	A
Level measurement zero error	Level pressure / current converter	B
Control valve zero error	Current / pressure converter	C
Control valve zero error	Valve stem linear potentiometer	D*

TABLE 3.14. Malfunctions introduced into laboratory level control rig.

* denotes computer simulated error.

Loop	Valve stem position measured	Run number	Malfunction	ϕ
AS	Yes	AS/1	-	-
	"	AS/2	-	-
	"	AS/A a	A + 10%	0.2105 V/p.s.i.g.
	"	AS/A b	A + 10%	0.2105 V/p.s.i.g.
	"	AS/A c	A - 10%	0.2105 V/p.s.i.g.
	"	AS/B	B - 10%	1.0 V/V
	"	AS/D*	D + 10%	1.0 V/V
DS	No	DS/1	-	-
	"	DS/2	-	-
	"	DS/3	-	-
	"	DS/4	-	-
	"	DS/A a	A + 10%	-0.516 V/p.s.i.g.
	"	DS/A b	A + 10%	"
	"	DS/A c	A + 10%	"
	"	DS/A d	A + 10%	"
	"	DS/A e	A - 10%	"
	"	DS/A f	A - 10%	"
	"	DS/A g	A - 10%	"
	"	DS/B a	B + 12.5%	-1.32 V/V
	"	DS/B b	B + 12.5%	"
	"	DS/C a	C + 10%	-0.7353 V/p.s.i.g.
	"	DS/C b	C + 10%	"
	"	DS/C c	C + 10%	"
	"	DS/C d	C - 10%	"
	"	DS/C e	C - 10%	"
	"	DS/C f	C - 10%	"

TABLE 3.15. Experiments performed on laboratory level control rig.

* denotes computer simulated error.

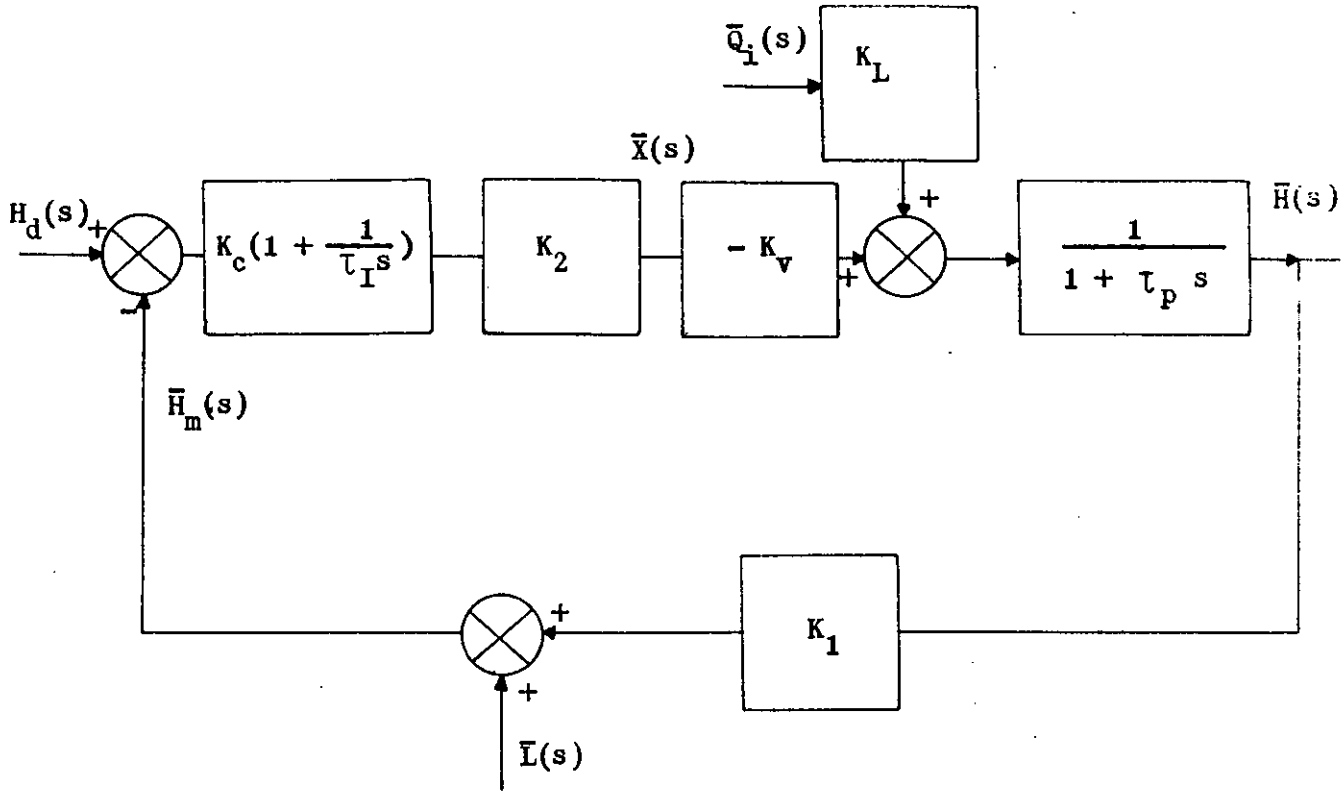


FIGURE 3.22 Block diagram for analogue setpoint control loop.

A malfunction in the Δ P/P transmitter is represented by a load disturbance $\bar{L}(s)$. Now the transfer function for the response of the control valve stem position to $\bar{L}(s)$ is:

$$\frac{\bar{X}(s)}{\bar{L}(s)} = \frac{-K_c\left(1 + \frac{1}{\tau_I s}\right) K_2}{1 + \frac{K_c(-K_v)K_1K_2}{(1 + \tau_p s)}\left(1 + \frac{1}{\tau_I s}\right)}$$

Defining $K_c(-K_v)\,K_1K_2 = K_F$ and rearranging gives:

$$\frac{\bar{X}(s)}{\bar{L}(s)} = \frac{-(K_2K_c\tau_I\tau_p\, s^2 + s(K_2K_c\tau_p + K_2K_c\tau_I) + K_2K_c)}{\tau_I\,\tau_p\, s^2 + s(\tau_I + \tau_I K_F) + K_F}$$

For a step change of magnitude L in $\bar{L}(s)$ then:

$$\bar{X}(s) = \frac{-L\,(K_2K_c\tau_I\tau_p\, s^2 + s(K_2K_c\tau_p + K_2K_c\tau_I) + K_2K_c)}{s(\tau_I\tau_p\, s^2 + s(\tau_I + \tau_I K_F) + K_F)}$$

Now the final value theorem states (56):

$$\lim_{t \to \infty} X(t) = \lim_{s \to 0} s\,\bar{X}(s)$$

So $$X(\infty) = -L\left(\frac{K_2 K_c}{K_F}\right)$$

Substituting the parameters of Table 3.3 yields:

$$X(\infty) = -0.2105\ L$$

hence $$\phi = -0.2105\ \text{V/p.s.i.g.}$$

The gain, ϕ, for a pure measurement error in the valve stem position or the P/I transmitter is simply 1.0 V/V.

The d.d.c. loop may be analysed in a similar manner to determine ϕ and the resulting values are summarised in Table 3.15.

3.17 Malfunction detection experimental results

Based upon the information generated by the state and loop security parameter (l.s.p.) estimators there are two types of check which indicate loop malfunction.

The first method is based upon the information generated by the Kalman filter. The filter innovation process was defined in equation (3.6.2.1). At a particular setpoint and load the process measurements were sampled and analysed by the Kalman filter, so that an innovation time series could be generated. This sequence may then be compared to an original malfunction free innovation sequence, corresponding to the same loop setpoint and load, and the changes examined. In fact, this type of check was adopted in the flow-control loop experiments where a residual process was monitored for indications of malfunction.

The methods of residual analysis discussed in section 2.6.2 and 2.6.3 may be applied to the filter innovation sequence, while some statistical tests of the time series were given by Mehra and Peschon (108). In these experiments inferences of loop security were derived by monitoring the innovation mean using a student's 't' test. As was pointed out in section 2.9, it was found that it was necessary only to monitor the valve demand innovation, and the results are presented in the same format as section 2.9, except that no innovation normalisation was performed.

The l.s.p. estimator forms the basis for the second method of malfunction detection. This estimator uses the innovations generated by the Kalman filter to estimate parameters which are indicative of the control loop security. Since a time series of bias estimates $\underline{\hat{b}}(k/k)$ are formed, a simple convenient visual display of the loop malfunction is possible.

The model formulations for the l.s.p. estimators in the analogue setpoint and d.d.c. loops were given in section 3.10. For the analogue setpoint control loop the l.s.p. $b_1(k)$ was indicative of malfunction in the P/I converter, while $b_2(k)$ arose because of an error in the control valve stem position measurement or malfunction in the actual feedback control loop instrumentation. The d.d.c. loop malfunctions were summarised by a single l.s.p. $b(k)$.

At a particular loop setpoint and load, the process measurements are analysed to yield estimates of the l.s.p.'s $\underline{\hat{b}}(k/k)$. These parameters then represent a characteristic of the control loop under the given process operating conditions. At some later date, with the control loop operating at the same setpoint and load, the l.s.p.'s may be re-estimated and compared visually, or by some other means, with

the original parameter estimates. Inferences may then be made concerning the control loop security.

3.17.1 Analogue setpoint control loop

The results corresponding to the experiments detailed in Table 3.15 are given in Table 3.16 and Figures 3.23 to 3.27.

These results show that the method of examining the innovation sequence mean is able to indicate loop malfunction. The runs AS/1 - AS/2 illustrate the consistency of innovations statistics when the

Run number	Valve demand innovation sequence statistics		Modulus of change in mean of valve demand innovation		Figure
	Mean	Variance	Actual change	Significant change at the 95% limit	
AS/1	0.97697×10^{-2}	0.101908×10^{-2}	-	-	3.23
AS/2	0.838387×10^{-2}	0.9370383×10^{-3}	0.138583×10^{-2}	0.272922×10^{-2}	3.23
AS/Aa	0.146417×10^{-1}	0.111302×10^{-2}	0.4872×10^{-2}	0.2850387×10^{-2}	3.24
AS/Ab	0.107003×10^{-1}	0.102374×10^{-2}	0.93233×10^{-3}	0.2789575×10^{-2}	3.24
AS/Ac	0.298211×10^{-2}	0.972764×10^{-3}	0.678759×10^{-2}	0.27662×10^{-2}	3.25
AS/B	0.16343×10^{-1}	0.949316×10^{-3}	0.65733×10^{-2}	0.273785×10^{-2}	3.26
AS/D*	0.18452×10^{-1}	0.115655×10^{-2}	0.86823×10^{-2}	0.287957×10^{-2}	3.27

TABLE 3.16 Behaviour of valve demand innovation for analogue setpoint control loop malfunction experiments.

control loop is malfunction free, while the other runs show the magnitude of the observed innovation mean shift compared to the significant shift at the 95% confidence limit.

In addition to the statistics of the innovations, the estimates of the l.s.p.'s reveal some interesting features.

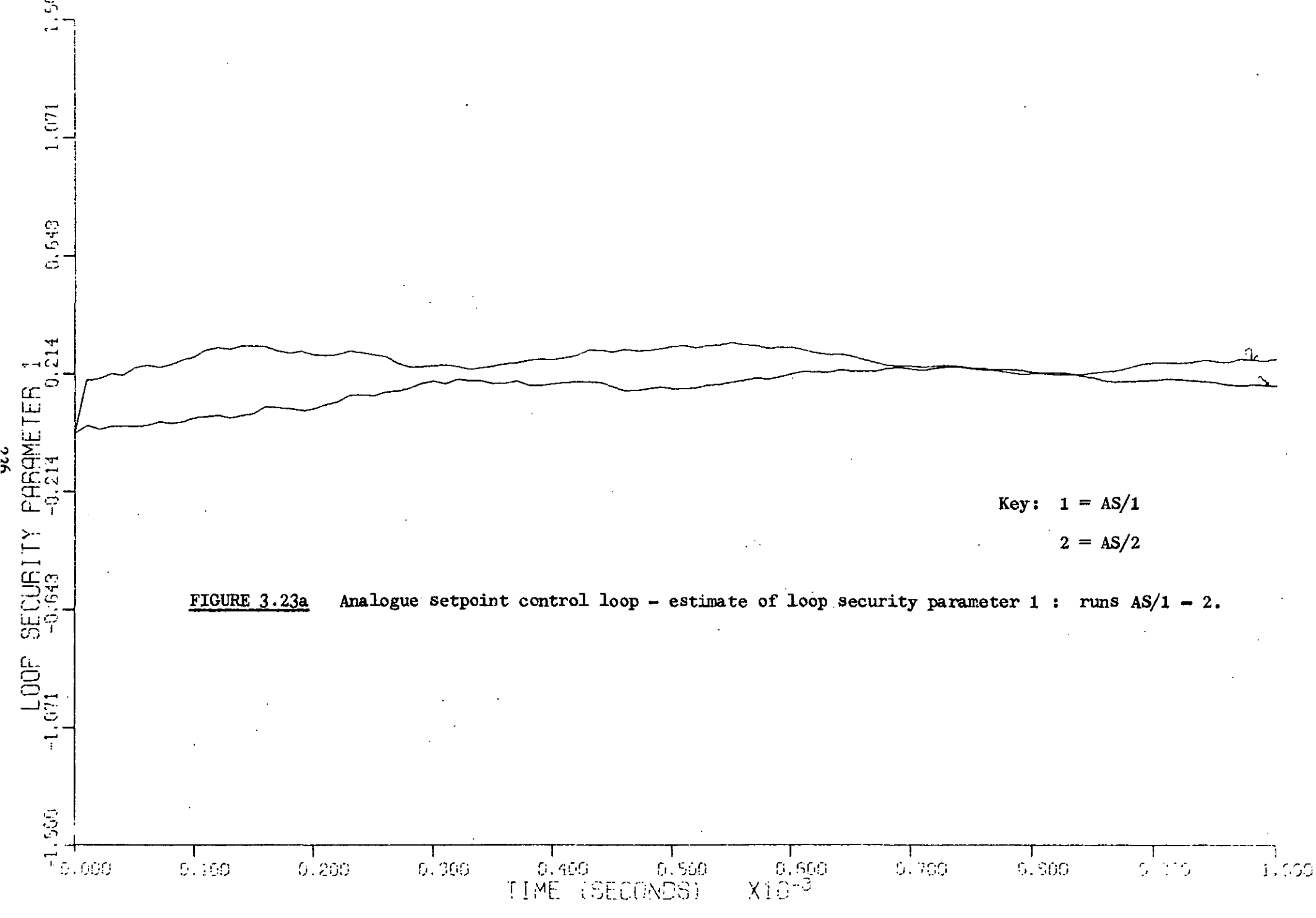


FIGURE 3.23a Analogue setpoint control loop - estimate of loop security parameter 1 : runs AS/1 - 2.

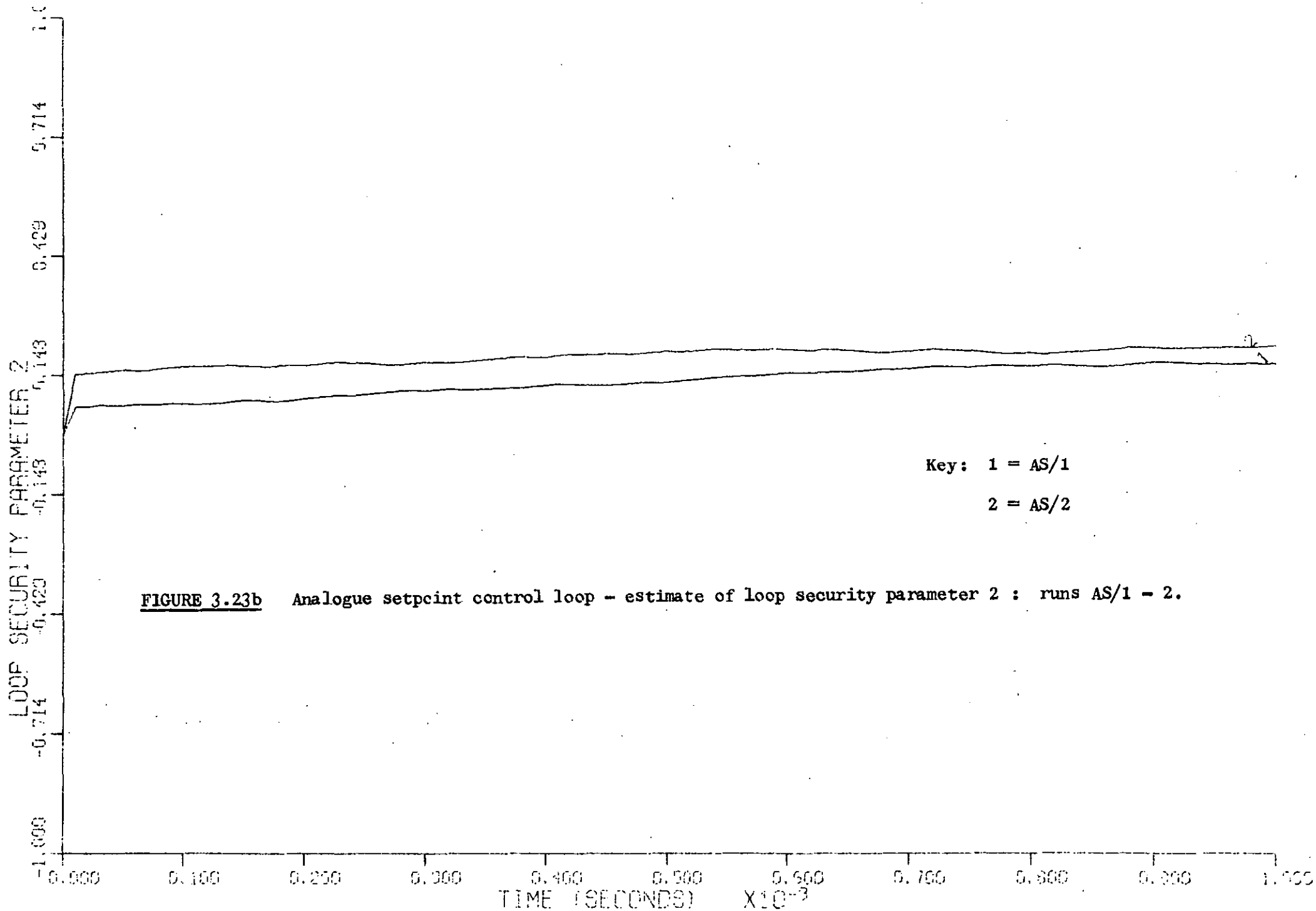


FIGURE 3.23b Analogue setpcint control loop - estimate of loop security parameter 2 : runs AS/1 - 2.

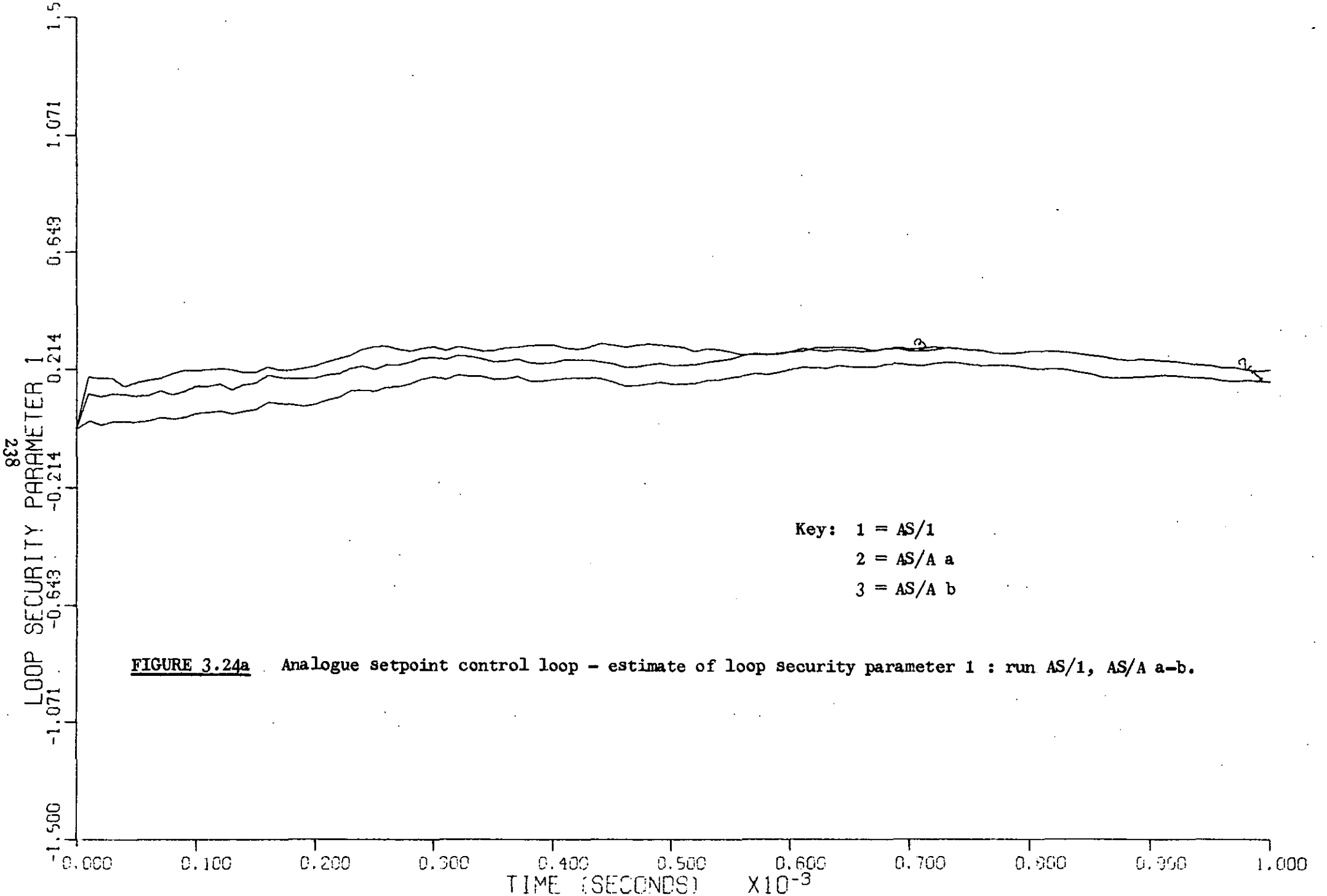


FIGURE 3.24a Analogue setpoint control loop - estimate of loop security parameter 1 : run AS/1, AS/A a-b.

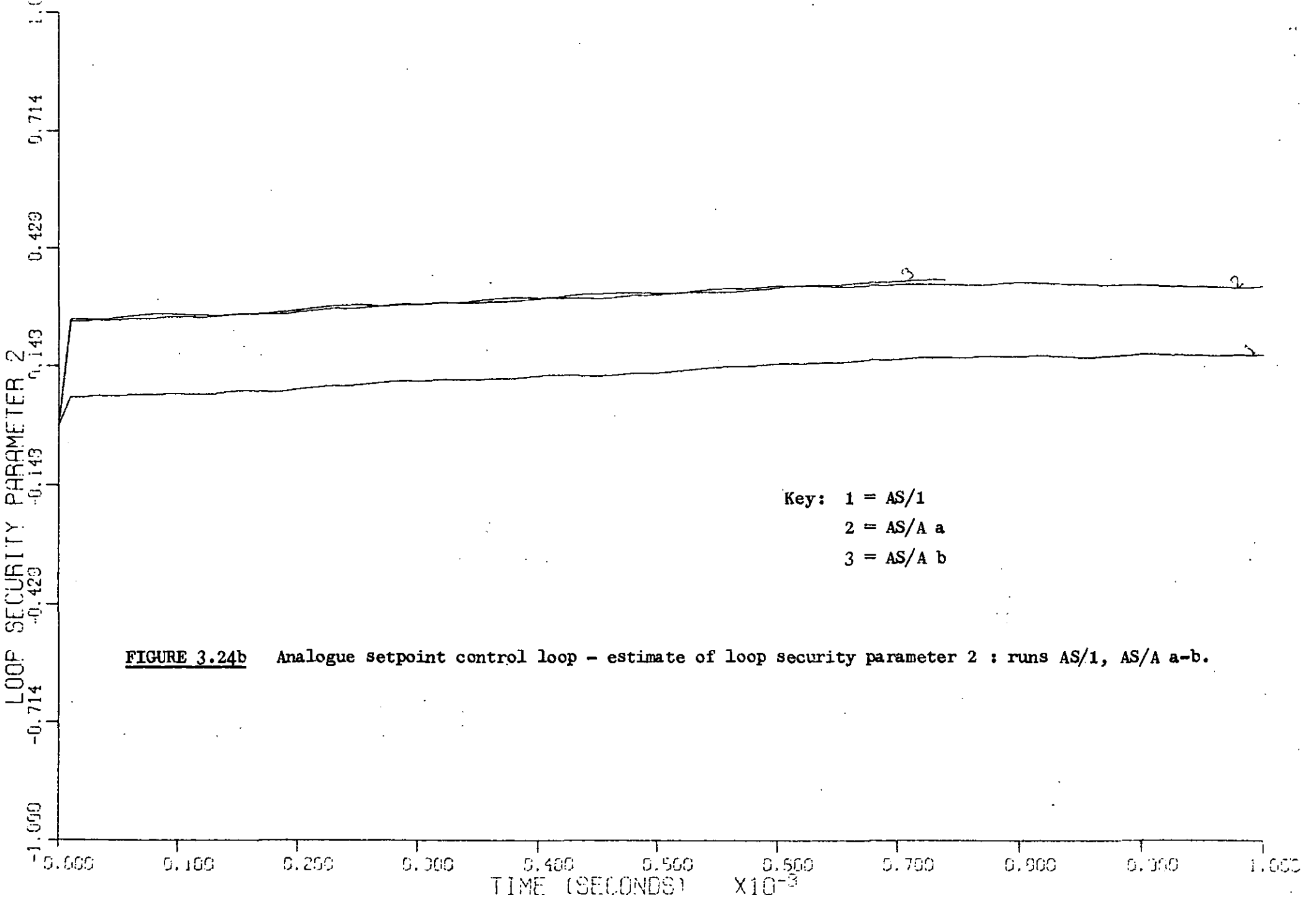
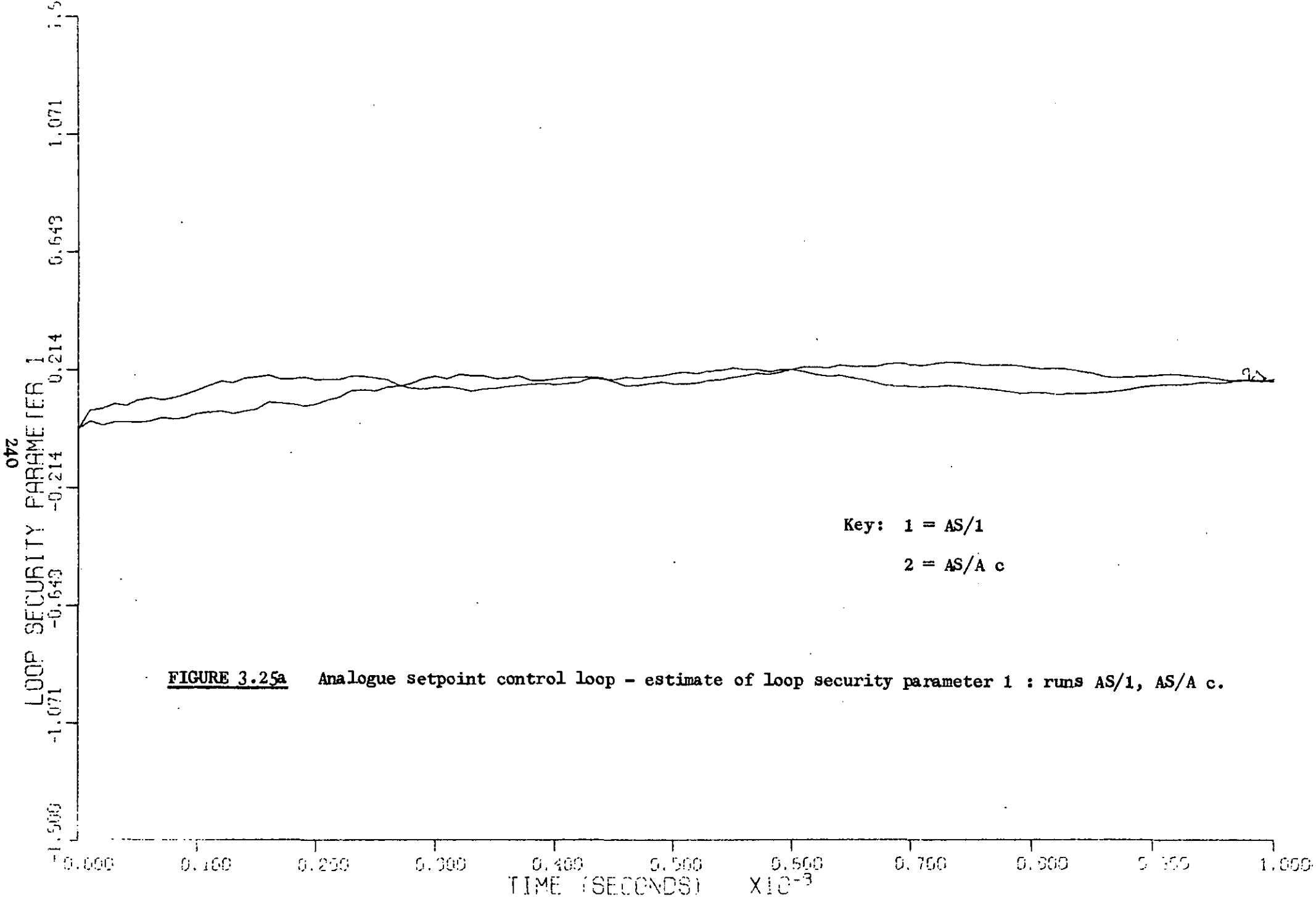


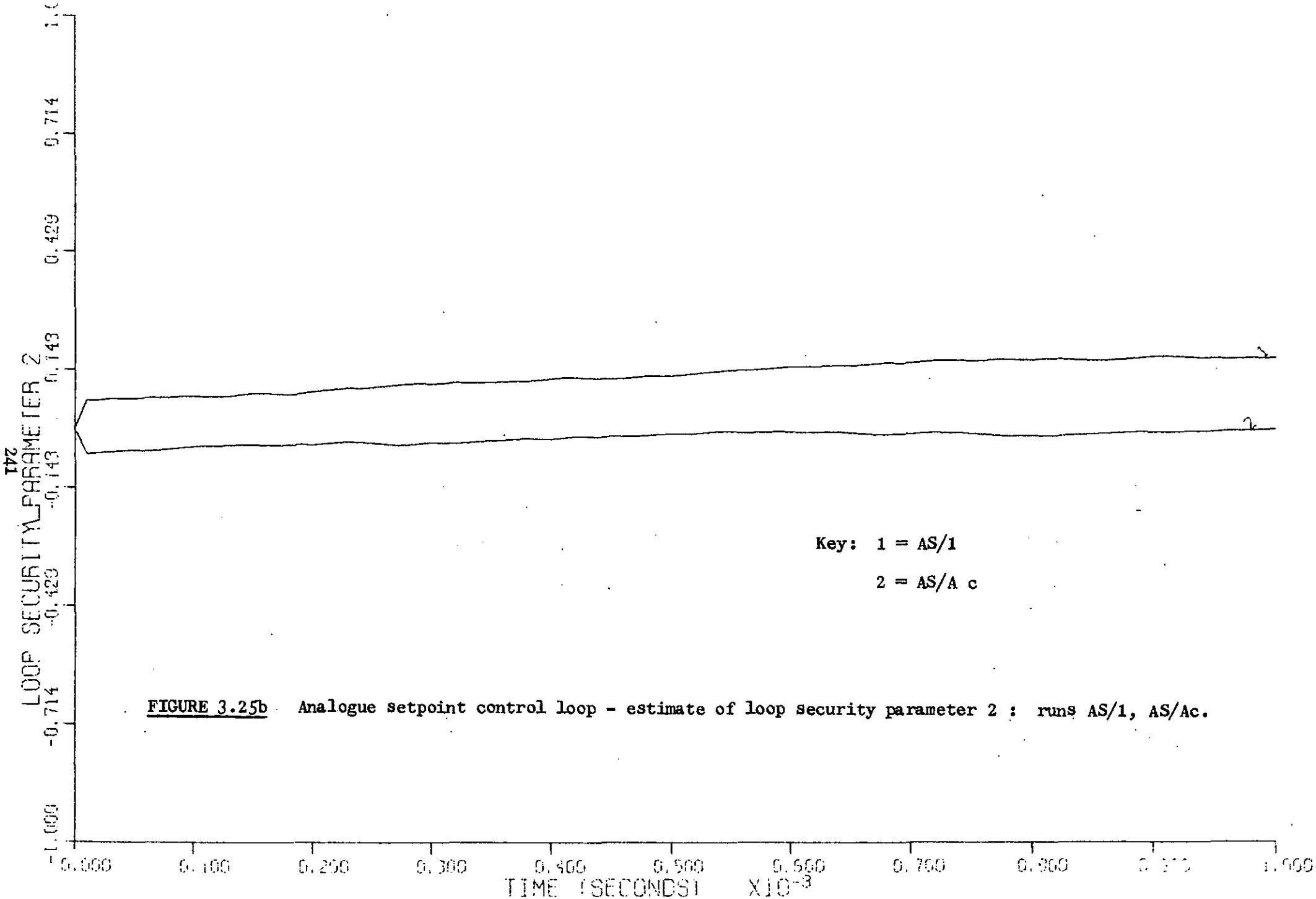
FIGURE 3.24b Analogue setpoint control loop - estimate of loop security parameter 2 : runs AS/1, AS/A a-b.



Key: 1 = AS/1

2 = AS/A c

FIGURE 3.25a Analogue setpoint control loop - estimate of loop security parameter 1 : runs AS/1, AS/A c.



Key: 1 = AS/1

2 = AS/A c

FIGURE 3.25b Analogue setpoint control loop - estimate of loop security parameter 2 : runs AS/1, AS/Ac.

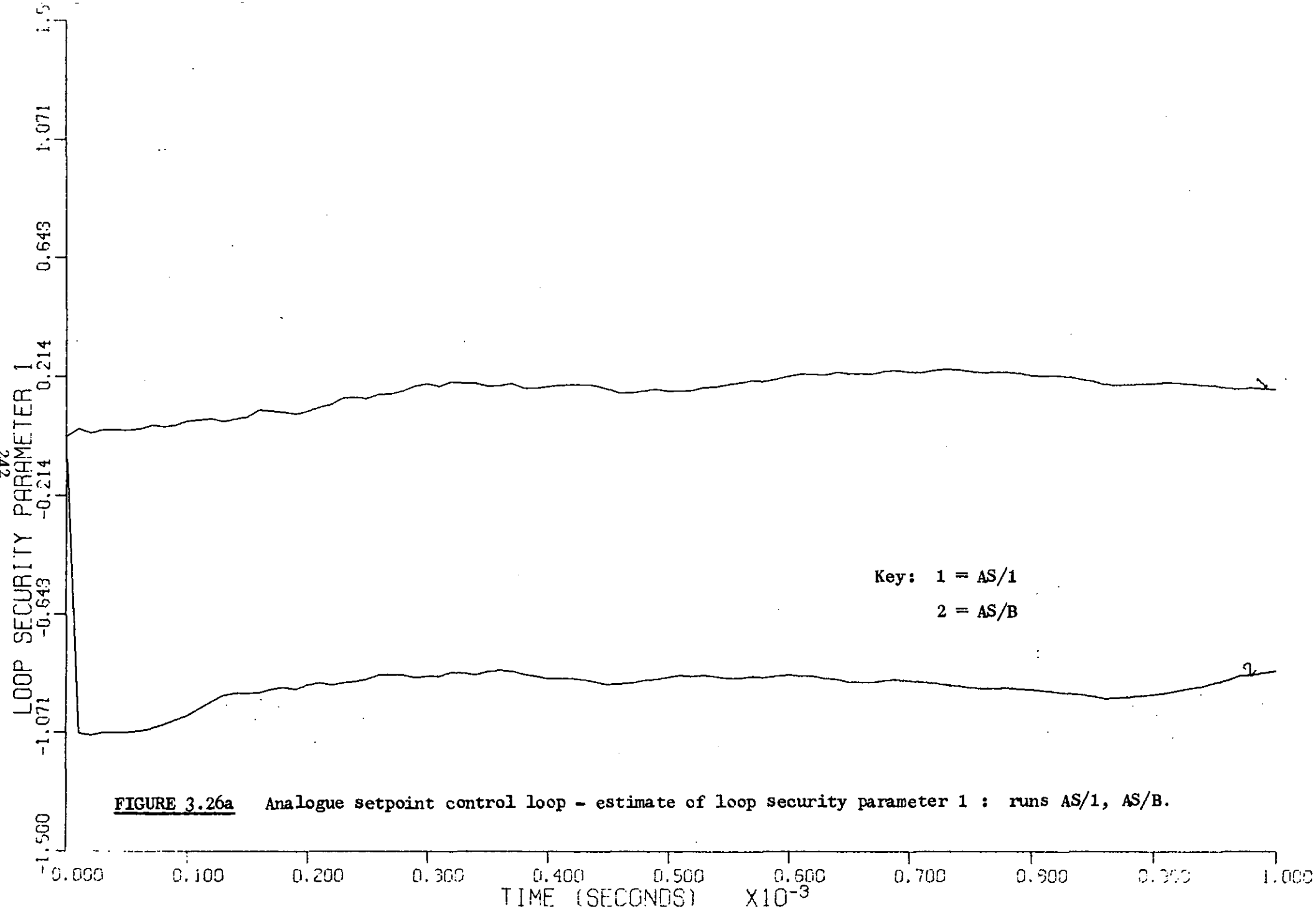
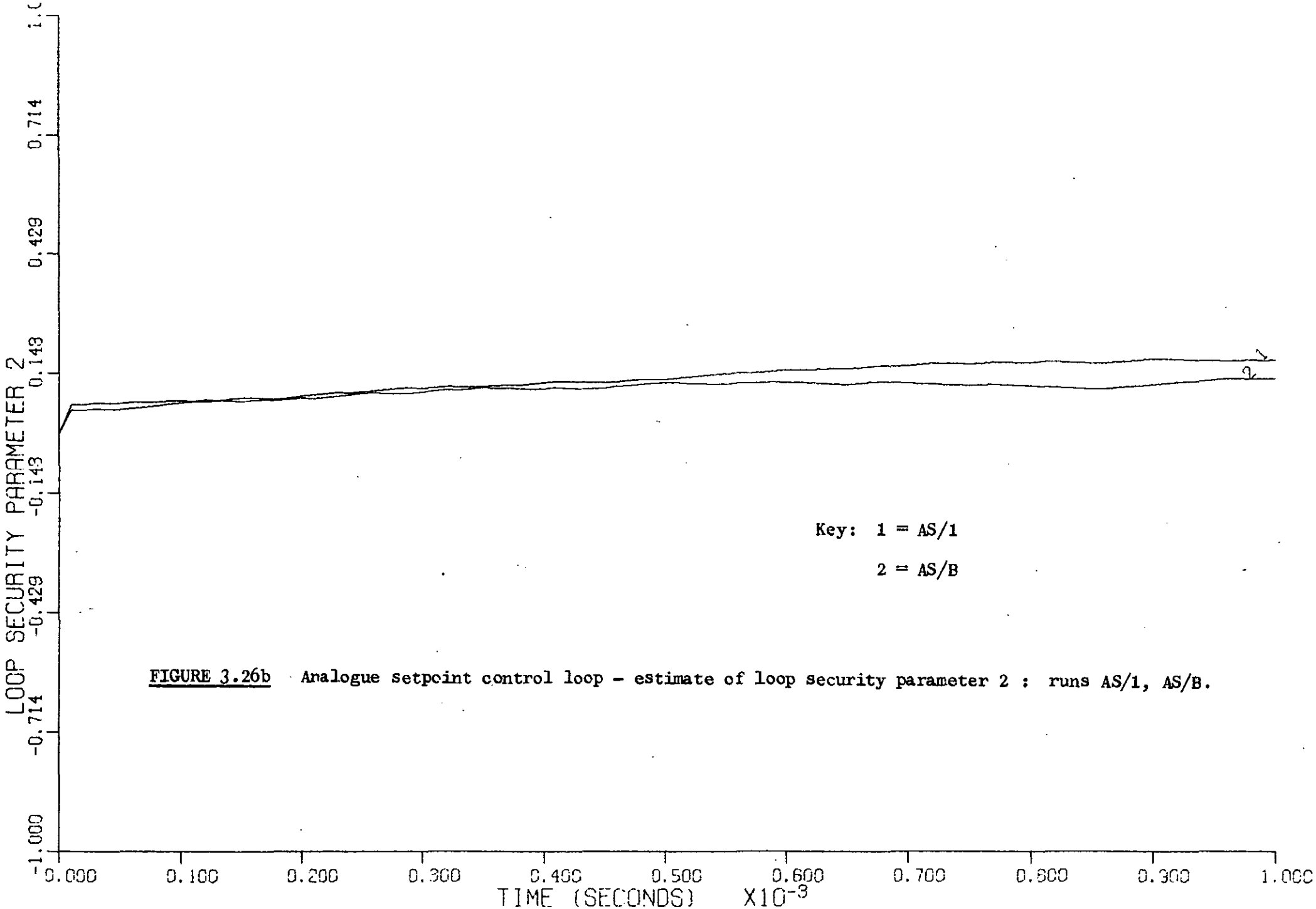


FIGURE 3.26a Analogue setpoint control loop - estimate of loop security parameter 1 : runs AS/1, AS/B.



Key: 1 = AS/1

2 = AS/B

FIGURE 3.26b Analogue setpoint control loop - estimate of loop security parameter 2 : runs AS/1, AS/B.

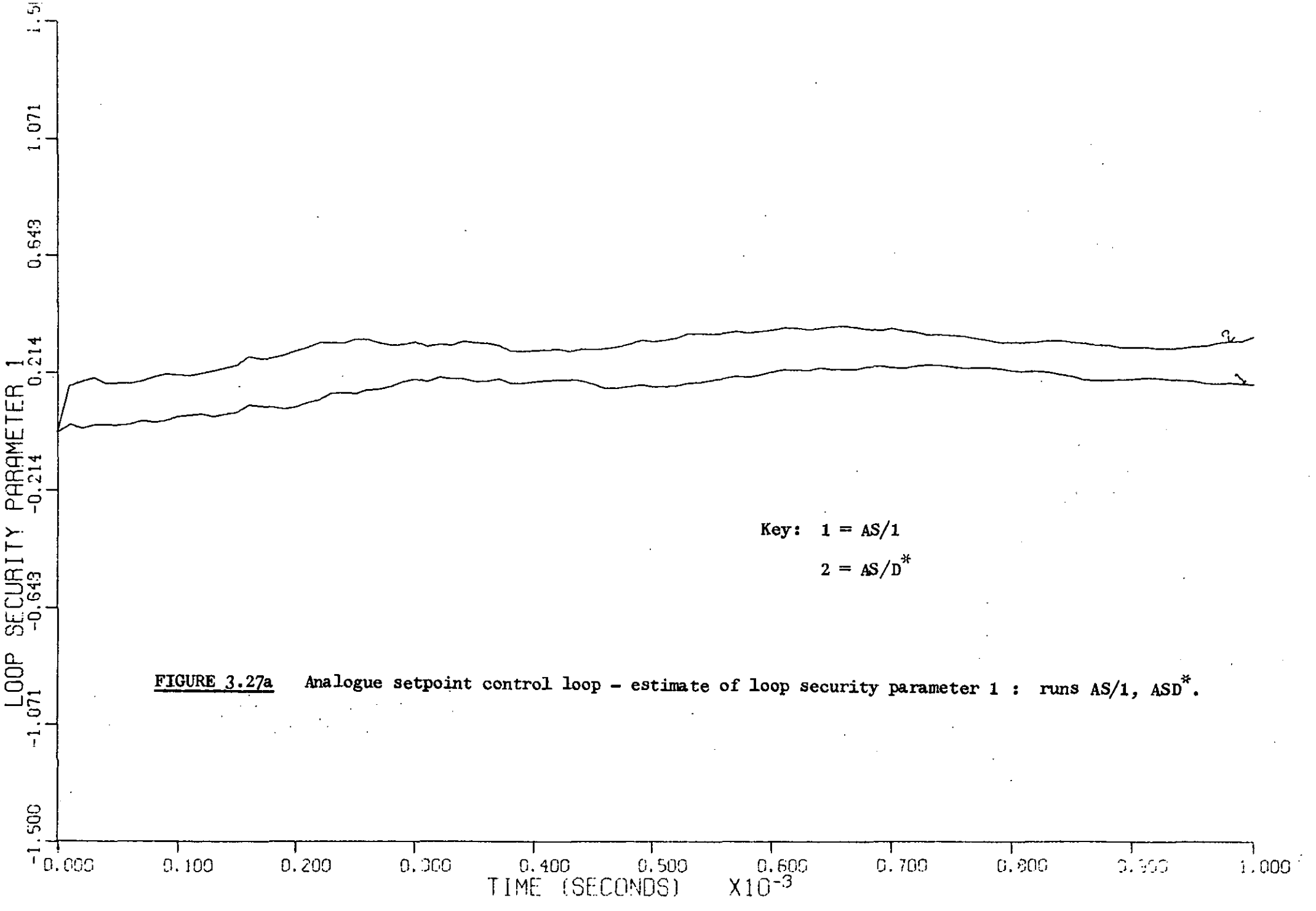


FIGURE 3.27a Analogue setpoint control loop - estimate of loop security parameter 1 : runs AS/1, ASD*.

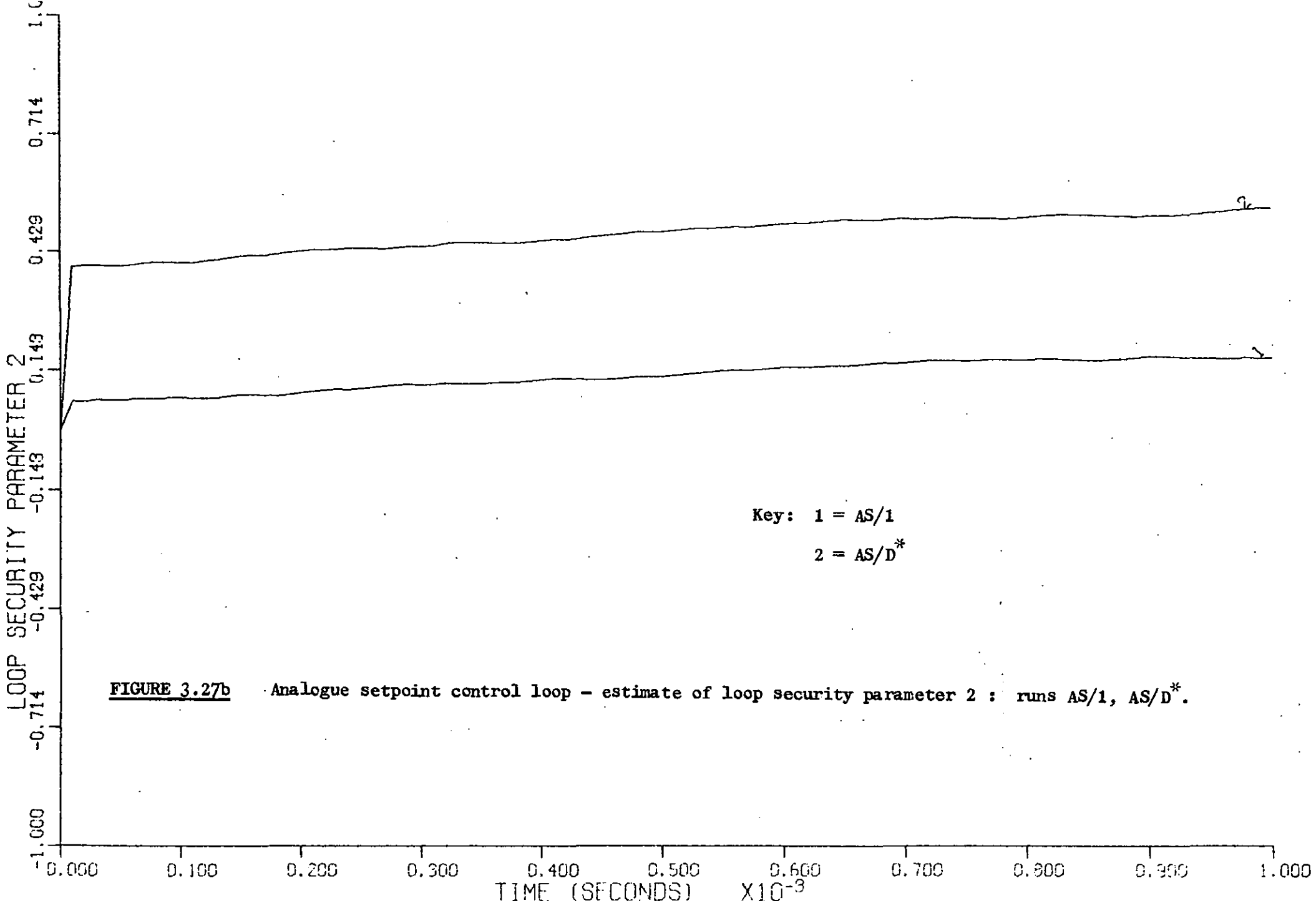


FIGURE 3.27b Analogue setpoint control loop – estimate of loop security parameter 2 : runs AS/1, AS/D*.

The estimates $\hat{b}_1(k/k)$ and $\hat{b}_2(k/k)$ for the malfunction free runs are shown in Figures 3.23a and b. The parameters are not zero, as may be expected for an error free loop, when the Kalman filter is operating optimally. There are two possible explanations for these non-zero parameter estimates.

Primarily the Kalman filter used to generate the innovation data base for malfunction detection was not optimal. The filter used in this study resulted from Mehra's adaptive estimator, run AM1/G/N, which was discussed in section 3.15. The suboptimality of the derived filter was indicated by the non-whiteness of the innovation sequence, as shown in Table 3.9.

The second possible source of error lies in the selection of the a priori information for the secondary l.s.p. estimator. It was pointed out in section 3.10 that to implement this secondary filter the designer must choose the a priori covariance matrix $P_b(0/0)$.

The effect of the choice of this matrix on the estimated value of $\underline{\hat{b}}(k/k)$ was investigated by repeating run AS/1 with various initial matrices $P_b(0/0)$. The results are given in Table 3.17, which indicates that $\underline{\hat{b}}(k/k)$ is relatively insensitive to $P_b(0/0)$ and so $P_b(0/0) = \text{diag}\ (10)$ was maintained throughout the studies on the analogue setpoint control runs.

Figures 3.23 to 3.27 confirm the postulations made in section 3.3 concerning the effect of malfunction on control loop operation and in particular the l.s.p. estimator is able to discriminate between some types of loop malfunction. For example, run AS/B, which has a -10% P/I transmitter zero error, results in the l.s.p. estimates shown in Figures 3.26 a and b. There is a large deviation between the currently

$P_b(0/0)$	$\hat{b}_1(k/k)$	$\hat{b}_2(k/k)$	$\hat{b}_1(k/k)$	$\hat{b}_2(k/k)$	$\hat{b}_1(k/k)$	$\hat{b}_2(k/k)$
k	500	500	750	750	1000	1000
$\begin{bmatrix} 10. & 0. \\ 0. & 10. \end{bmatrix}$	0.161142	0.12704	0.23368	0.164786	0.16998	0.172775
$\begin{bmatrix} 1. & 0. \\ 0. & 1. \end{bmatrix}$	0.160535	0.12694	0.23297	0.164656	0.16958	0.172655
$\begin{bmatrix} 0.1 & 0. \\ 0. & 0.1 \end{bmatrix}$	0.154707	0.125932	0.226108	0.163395	0.165641	0.171487
$\begin{bmatrix} 0.01 & 0. \\ 0. & 0.01 \end{bmatrix}$	0.113392	0.117948	0.17444	0.153157	0.134082	0.161781

TABLE 3.17. Effect of $P_b(0/0)$ on $\underline{\hat{b}}(k/k)$ in analogue setpoint control experiments.

estimated l.s.p. $\hat{b}_1(k/k)$ (denoted by curve 2 in Figure 3.26a), and the malfunction free characteristic (denoted by curve 1). Figure 3.26b shows that the corresponding change in the l.s.p. $\hat{b}_2(k/k)$ is small. This phenomenon is exactly that expected for a P/I transmitter malfunction as discussed in section 3.10.

The knowledge that a pure measurement error has occurred in the P/I transmitter and the estimated loop security parameter $\hat{b}_1(k/k)$ associated with it means that the measurement $y_1(k)$ can be corrected to eliminate the effect of the malfunction.

In section 3.10.1 the P/I malfunction was modelled as:

$$y_1(k) = \underline{h}_1 \underline{x}(k) + b_1(k)$$

Now the measurement $y_1(k)$ is biased because of the malfunction and so the true process measurement is:

$$y_1(k) - b_1(k)$$

The loop security parameter $\hat{b}_1(k/k)$ is an estimate of $b_1(k)$ and from Figure 3.26a it can be seen that the loop malfunction causes $\hat{b}_1(k/k)$ to deviate by -1.016 from the malfunction free estimate, thus the compensated process measurement becomes:

$$y_1(k) + 1.016$$

Now the actual error imposed on to the P/I transmitter measurement was -1.0V and so the malfunction compensation provides excellent estimates of the true process measurement.

This error compensation technique is particularly valuable when the measurement $y_1(k)$ is used in some other process performance evaluation computer program, because the compensated measurement may be used as a substitute for the real measurement until the P/I transmitter is repaired.

Examination of Figures 3.25 to 3.27 reveals the presence of a loop malfunction which causes the parameter $\hat{b}_2(k/k)$ to be large. Such a deviation could arise from either a pure measurement error in the control valve stem position or a Δ P/P transducer malfunction. The l.s.p. estimator is unable to discriminate between these two types of malfunction.

The actual shift in the estimates of the l.s.p.'s from the malfunction free characteristics may be compared with the expected deviations derived from a consideration of the loop malfunction gain ϕ. These results are summarised in Table 3.18.

Notice that the check based upon monitoring the innovation mean does not determine the malfunction for run AS/Ab as shown in Table 3.16. However, inspection of Figures 3.24 a and b, and Table 3.18

shows that the l.s.p. estimator does give an indication of this fault.

Summarising, these experiments on the analogue setpoint control loop have shown that the proposed malfunction detection method is able

Run number	ϕ	Expected deviation due to malfunction		Actual deviation due to malfunction	
		$\hat{b}_1$ (V)	$\hat{b}_2$ (V)	$\hat{b}_1$ (V)	$\hat{b}_2$ (V)
AS/1		-			
AS/Aa	0.2105 V/p.s.i.g.	0.	0.2526	0.045	0.166
AS/Ab	0.2105 V/p.s.i.g.	0.	0.2526	0.1128	0.18
AS/Ac	0.2105 V/p.s.i.g.	0.	-0.2526	0.00015	-0.172
AS/B	1.0 V/V	-1.0	0.	-1.016	-0.0441
AS/D*	1.0 V/V	0.	+0.5	0.175	0.362

TABLE 3.18 Expected and actual shifts of loop security parameters for analogue setpoint control experiments.

to expose errors. In particular, the method can discriminate between a P/I transmitter malfunction and a pure valve stem position measurement error or a Δ P/P transducer malfunction. However, the check is unable to differentiate between these two latter malfunctions.

3.17.2 Direct digital control loop

Table 3.19 details the statistics of the control valve demand signal innovation sequence corresponding to the experiments outlined in Table 3.15. Runs DS/1 - DS/4 represent malfunction free experiments and the results show the consistency of the innovation statistics. The remaining experiments demonstrate that malfunctions can be detected by examining the shift of the innovation mean from the malfunction free characteristics.

Run number	Valve demand innovation statistics		Modulus of change in mean of valve demand innovation		Figure
	Mean	Variance	Actual change	Significant change at 95% limit	
DS/1	0.155451×10^{-1}	0.321232×10^{-1}	-	-	3.29
DS/2	0.602586×10^{-2}	0.306119×10^{-1}	0.951924×10^{-2}	0.155242×10^{-1}	3.29
DS/3	0.390723×10^{-2}	0.31716×10^{-1}	0.1163787×10^{-1}	0.156602×10^{-1}	3.29
DS/4	0.274055×10^{-1}	0.315153×10^{-1}	0.118604×10^{-1}	0.156356×10^{-1}	3.29
DS/Aa	-0.277109×10^{-1}	0.309236×10^{-1}	0.43256×10^{-1}	0.155627×10^{-1}	3.30
DS/Ab	-0.226125×10^{-1}	0.310138×10^{-1}	0.381576×10^{-1}	0.155739×10^{-1}	3.30
DS/Ac	-0.242276×10^{-1}	0.319993×10^{-1}	0.397727×10^{-1}	0.15695×10^{-1}	3.30
DS/Ad	-0.240471×10^{-1}	0.312199×10^{-1}	0.395922×10^{-1}	0.155993×10^{-1}	-
DS/Ae	0.106938	0.385525×10^{-1}	0.913929×10^{-1}	0.164775×10^{-1}	3.31
DS/Af	0.898643×10^{-1}	0.340735×10^{-1}	0.743192×10^{-1}	0.159468×10^{-1}	3.31
DS/Ag	0.986288×10^{-1}	0.34666×10^{-1}	0.830837×10^{-1}	0.16018×10^{-1}	3.31
DS/Ba	-0.111814	0.516182×10^{-1}	0.1273591	0.17936×10^{-1}	3.32
DS/Bb	-0.105139	0.45693×10^{-1}	0.1206841	0.172898×10^{-1}	3.32
DS/Ca	-0.534161×10^{-1}	0.341493×10^{-1}	0.689612×10^{-1}	0.159559×10^{-1}	3.33
DS/Cb	-0.515355×10^{-1}	0.334092×10^{-1}	0.670806×10^{-1}	0.158666×10^{-1}	3.33
DS/Cc	-0.432993×10^{-1}	0.330094×10^{-1}	0.588444×10^{-1}	0.158181×10^{-1}	3.33
DS/Cd	0.148659	0.556092×10^{-1}	0.1331139	0.18358 $\times 10^{-1}$	3.34
DS/Ce	0.157395	0.556232×10^{-1}	0.1418499	0.183599×10^{-1}	3.34
DS/Cf	0.152682	0.655709×10^{-1}	0.1371369	0.193727×10^{-1}	3.34

TABLE 3.19. Behaviour of valve demand innovation for direct digital control loop malfunction experiments.

The malfunction detection method based upon the l.s.p. estimator was also used to check the control loop performance. At a particular loop setpoint and load the Kalman filter processed the measurements to yield an innovation sequence data base from which a l.s.p. estimate $\hat{b}(k/k)$ was determined. Figure 3.28 shows this parameter estimate for the four experiments on the malfunction free control loop DS/1 - DS/4.

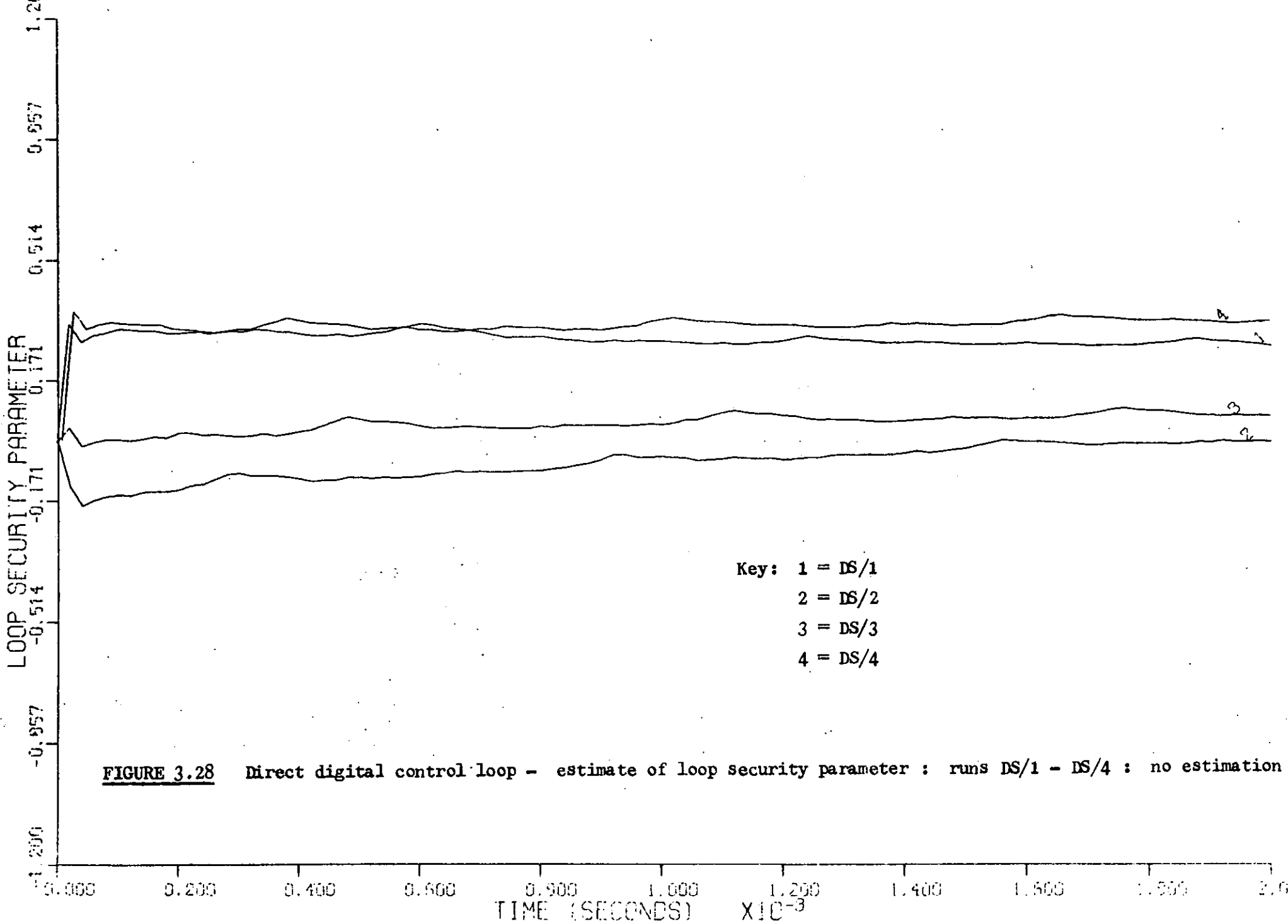


FIGURE 3.28 Direct digital control loop - estimate of loop security parameter : runs DS/1 - DS/4 : no estimation delay

This Figure illustrates that there is quite a wide spread of $\hat{b}(k/k)$ values which is unsatisfactory.

The effect of $P_b(0/0)$ on the quantity $\hat{b}(k/k)$ was investigated by examining run DS/1 for different a priori estimates of $P_b(0/0)$. The results are given in Table 3.20 which shows that $\hat{b}$ is insensitive to the choice of $P_b(0/0)$ and so $P_b(0/0) = 10$ was used throughout these experiments.

It is suggested that the reason for the inconsistency in the four malfunction free runs was due to the initial mechanisation of the l.s.p. estimator. At the beginning of the algorithm for runs DS/1, DS/2, DS/4, large values of the l.s.p. estimator gain $K_b(k)$ were associated with large innovations, thereby generating large estimates $\hat{b}(k/k)$. The matrix $K_b(k)$ soon decayed to its nominal value which was small. Now the basic loop security parameter algorithm is:

$$\hat{b}(k/k) = (I - K_b(k)\ S(k))\ \hat{b}(k-1/k-1) + K_b(k)\ \underline{\nu}\ (k)$$

This equation is basically a "smoother" and since $K_b(k)$ was small, its

$P_b(0/0)$	$\hat{b}(k/k)$	$\hat{b}(k/k)$	$\hat{b}(k/k)$
k	500	750	1000
[10]	0.280462	0.27064	0.267823
[0.01]	0.231668	0.229343	0.231403

TABLE 3.20 Effect of $P_b(0/0)$ on $\hat{b}(k/k)$ in d.d.c. experiments.

response is slow, thus if $\hat{b}(k/k)$ was initially large, it would decay to the true value very slowly.

This problem was overcome by delaying the beginning of the estimation

of $\hat{b}(k/k)$ until the matrices $K_b(k)$, $S(k)$ and $\underline{v}(k)$ had achieved their nominal values. This delay was chosen to be $k = 100$. The resulting estimates of $\hat{b}(k/k)$ for runs DS/1 - DS/4 are shown in Figure 3.29 from which it may be seen that reasonable consistency with a spread of 0.12, is achieved.

Figures 3.30 to 3.34 show how the parameter $\hat{b}(k/k)$ varies when the control loop is subject to malfunction while Table 3.21 compares the actual change of $\hat{b}(k/k)$ with the expected change calculated from the knowledge of ϕ.

Run number	ϕ	Expected deviation due to malfunction $\hat{b}$ (V)	Actual deviation due to malfunction $\hat{b}$ (V)
DS/1	-	-	-
DS/Aa	-0.5106 V/p.s.i.g.	- 0.619	- 0.284
DS/Ab	"	"	- 0.244
DS/Ac	"	"	- 0.246
DS/Ad	"	"	- 0.249
DS/Ae	"	0.619	0.556
DS/Af	"	"	0.462
DS/Ag	"	"	0.516
DS/Ba	-1.32 V/V	- 0.66	- 0.72
DS/Bb	"	"	- 0.673
DS/Ca	-0.7353 V/p.s.i.g.	- 0.882	- 0.392
DS/Cb	"	"	- 0.391
DS/Cc	"	"	- 0.357
DS/Cd	"	0.882	0.781
DS/Ce	"	"	0.823
DS/Cf	"	"	0.758

TABLE 3.21 Expected and actual shifts of loop security parameter for direct digital control experiments.

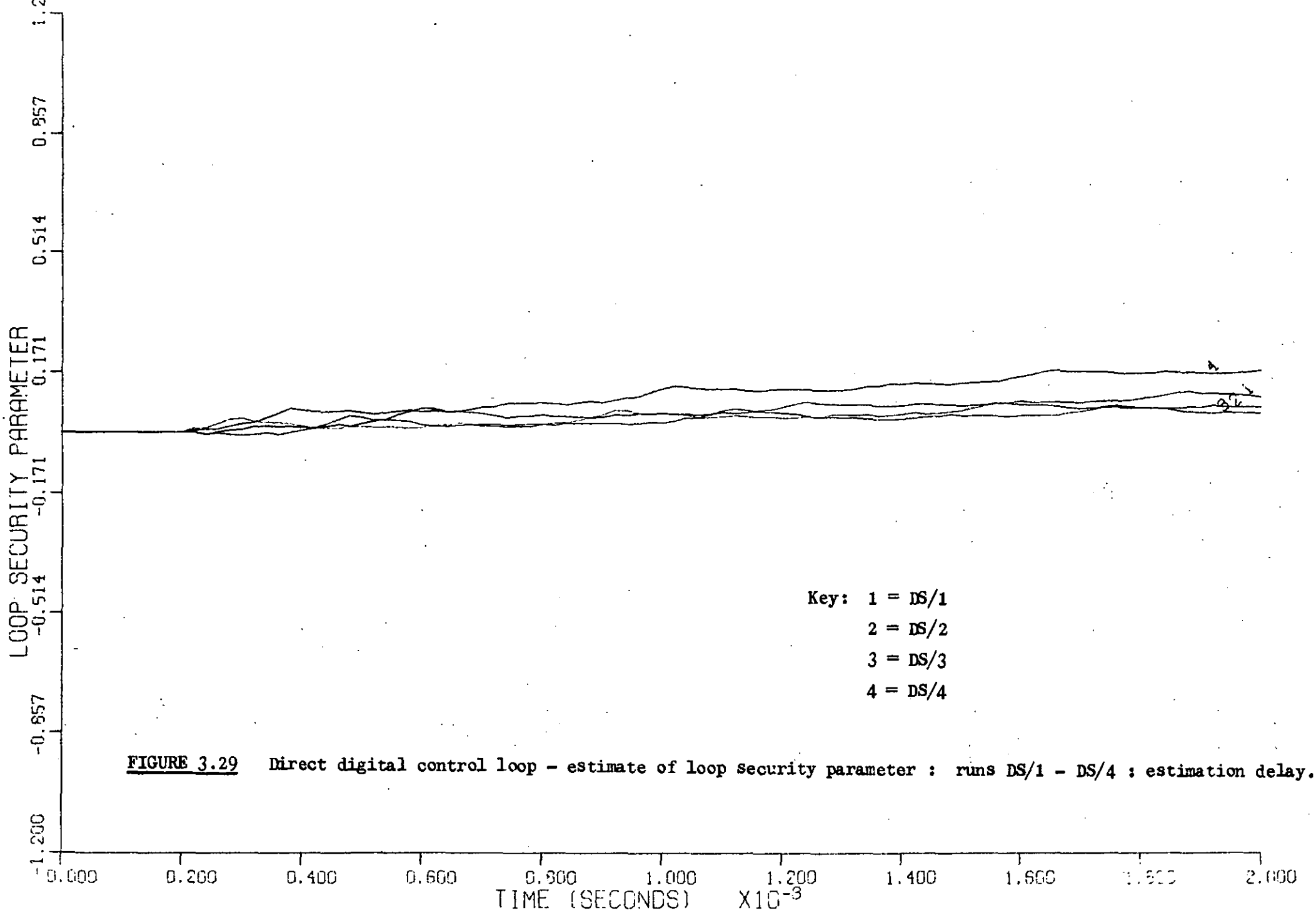


FIGURE 3.29 Direct digital control loop - estimate of loop security parameter : runs DS/1 - DS/4 : estimation delay.

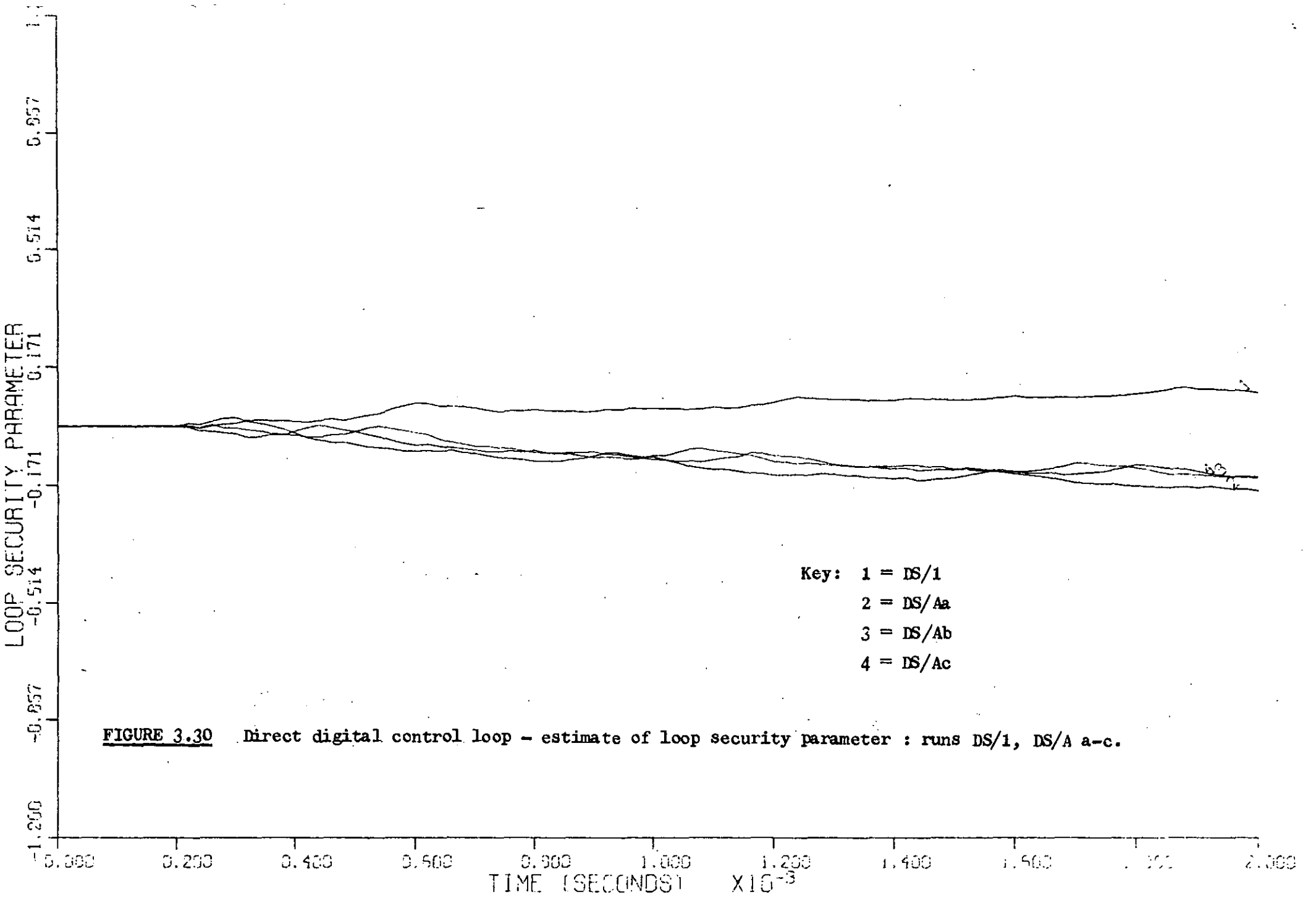


FIGURE 3.30 Direct digital control loop - estimate of loop security parameter : runs DS/1, DS/A a-c.

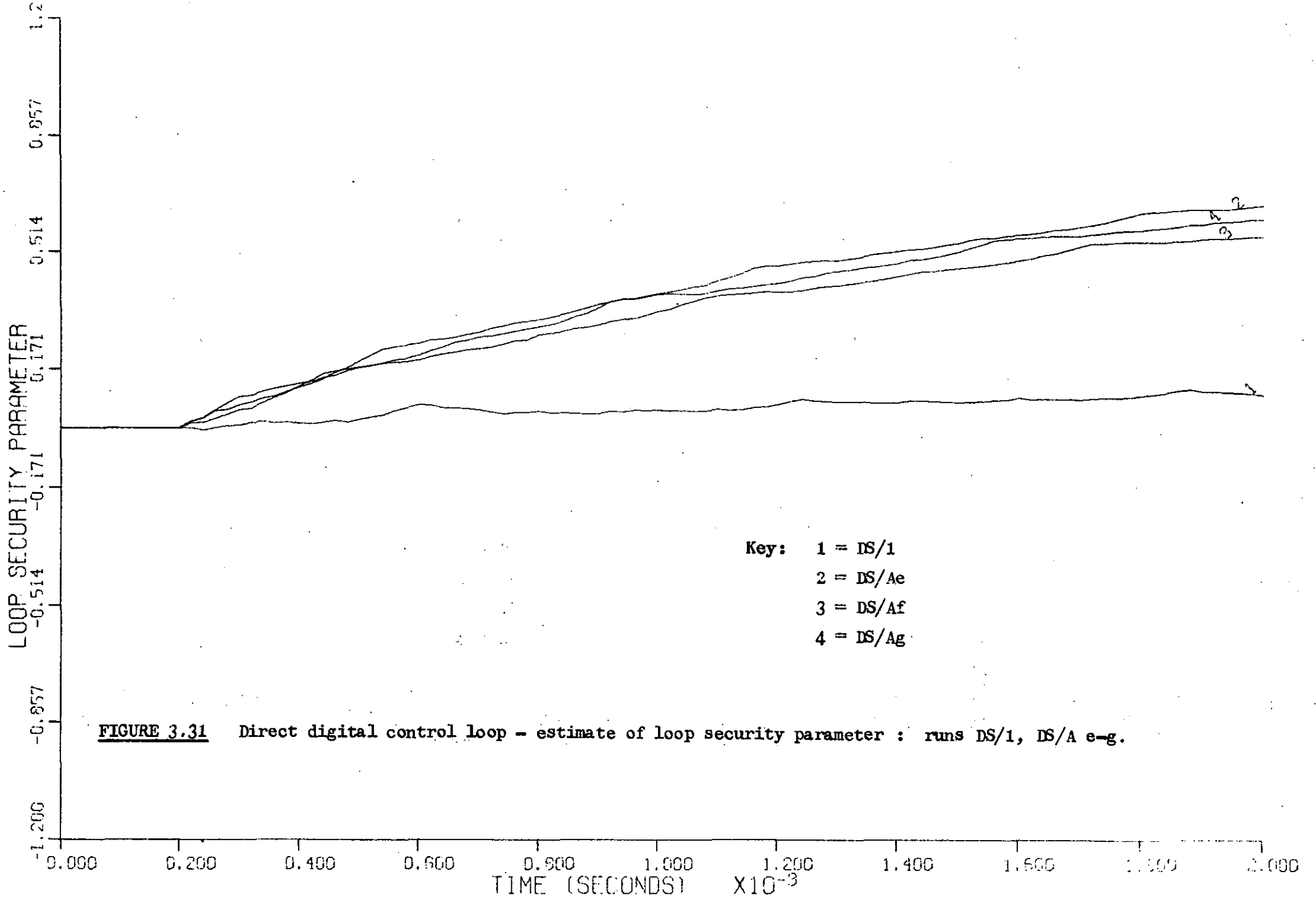


FIGURE 3.31 Direct digital control loop - estimate of loop security parameter : runs DS/1, DS/A e-g.

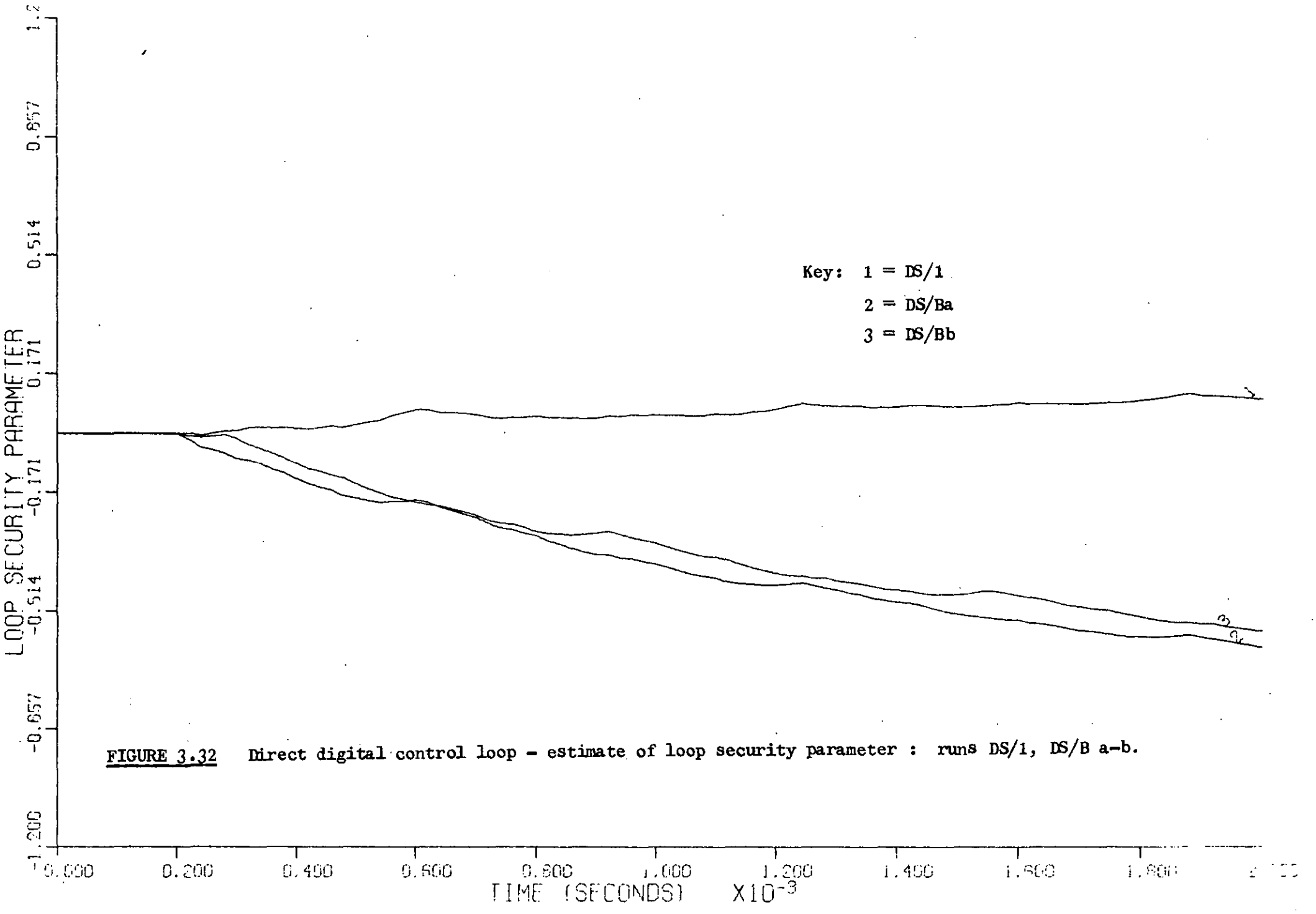


FIGURE 3.32 Direct digital control loop - estimate of loop security parameter : runs DS/1, DS/B a-b.

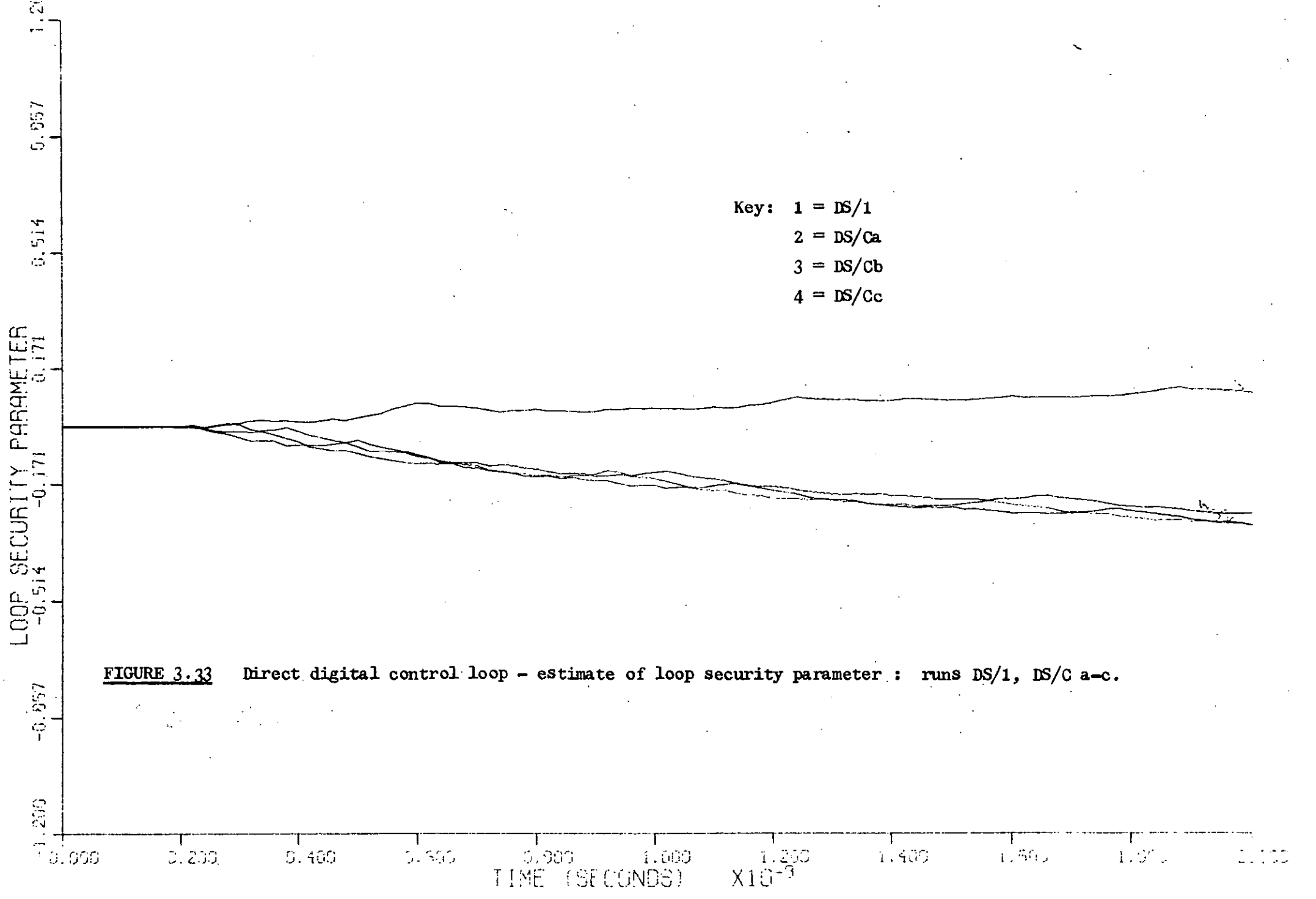


FIGURE 3.33 Direct digital control loop - estimate of loop security parameter : runs DS/1, DS/C a-c.

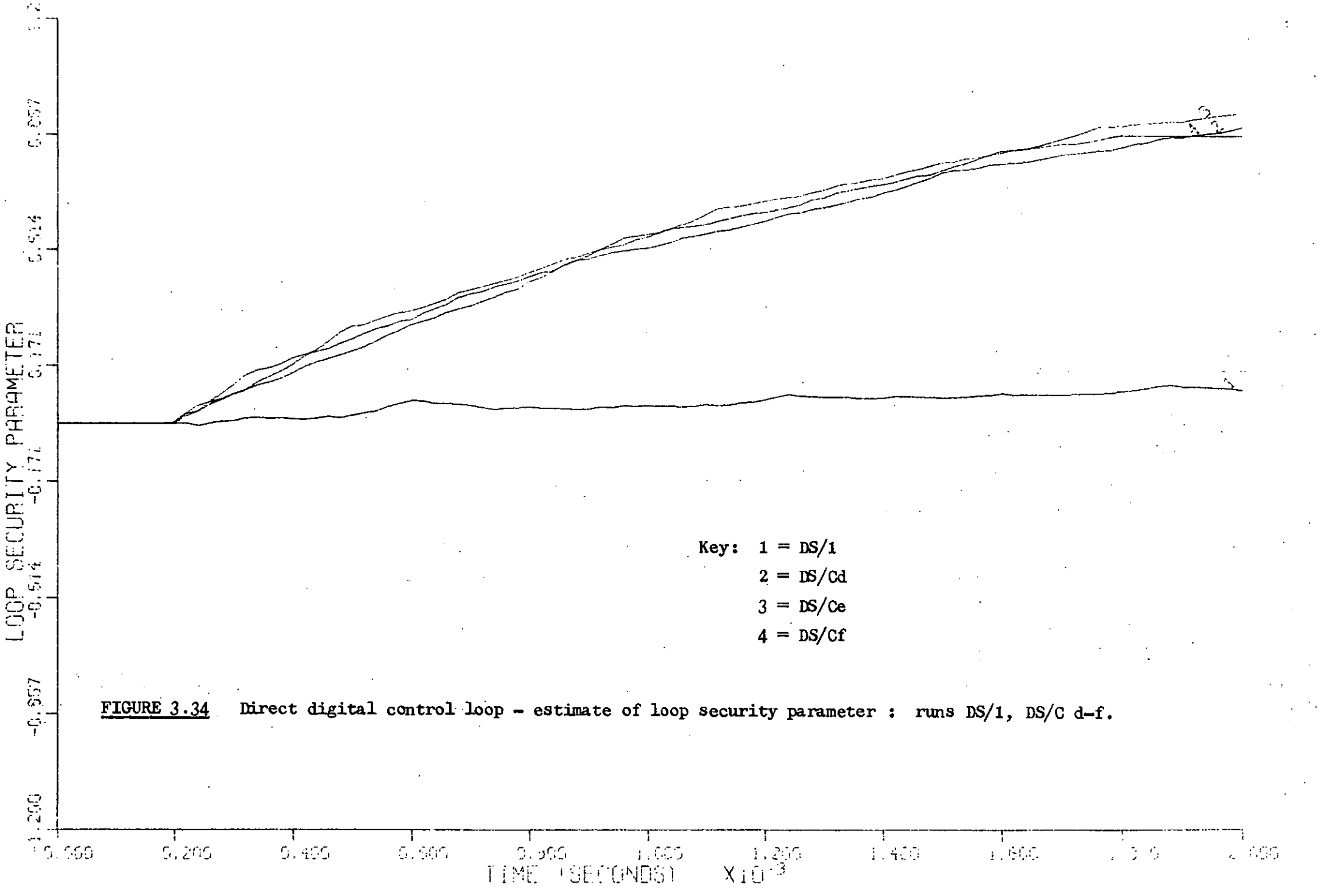


FIGURE 3.34 Direct digital control loop - estimate of loop security parameter : runs DS/1, DS/C d-f.

These results show that the two proposed techniques of malfunction detection can determine faults in a d.d.c. loop. However, the checks do not yield any diagnostic information concerning the source of the control loop malfunction.

3.18 Concluding remarks

A general method for detecting malfunction in an analogue setpoint or direct digital control loop has been proposed in this chapter.

The method assumes that the control loop can be represented by a linear, time invariant state variable model. The proposed checks can only be performed when the control loop is operating at a particular setpoint and nominal steady state value of process load as detailed in the algorithm specification.

The fundamental technique is based upon analysing closed control loop operation under conditions of malfunction. The feasibility of detecting a particular instrumentation malfunction in a given control loop has been formulated in terms of a malfunction gain, $\emptyset$, relating the change in magnitude of the control valve demand signal to the malfunction. As the magnitude of $\emptyset$ increases, the ease of malfunction detection increases. In fact, for some critical loops where small malfunction cannot be tolerated, it is suggested that $\emptyset$ may be used as a design criterion in specifying the control loop.

Two malfunction detection techniques have been developed and both can be used while the control is operating on-line. The two check methods are based upon an analysis of the available process measurements using a Kalman filter to provide the data base for loop security interrogation.

The Kalman filter used in the development of the proposed checks was constrained in this chapter to be "optimal", although this is not a necessary condition for the implementation of the malfunction detection schemes. The problem of designing an "optimal" Kalman filter, due to uncertain a priori information about the system models and process noise statistics has been considered and Mehra's adaptive estimator was used to overcome this difficulty.

The experimental results derived from a level control loop have shown that this adaptive technique fails to converge to an "optimal" filter when the true process measurement noise is small. In noisier systems, however, it has been found that the method is robust, consistent and provides good estimates of the system state in spite of poor a priori information.

The first proposed malfunction detection method is based upon the comparison of an original and current innovation sequence generated by the Kalman filter from a control loop operating at the same setpoint and load in each case. This check is applicable to both analogue setpoint and direct digital control loops. However, the information gained from this check is only that a malfunction exists in the control loop and no diagnosis of the cause is possible. This type of check was in fact used in the detection of malfunction in a flow control loop.

The second check method is based upon the estimation of various loop security parameters which are indicative of malfunction. These parameters are calculated from the innovation sequence, generated by the Kalman filter, and various system matrices estimated in Mehra's adaptive estimation scheme. This method provides a convenient and simple visual display of control loop security and does yield some diagnostic information.

The laboratory experiments on an analogue setpoint control loop have shown that this method is able to discriminate between a P/I transmitter malfunction and other control loop errors. When the Kalman filter is designed optimally, the resulting estimates of the loop security parameters may be used to compensate for this P/I transmitter malfunction, thereby creating a substitute corrected process measurement which is available to the computer. This is particularly important when this measurement is used in other computer programs for process evaluation, since incorrect measurements render such calculations useless.

However, these comments do not apply to a direct digital control loop and as in the first check method, the loop security parameter estimation yields no diagnostic information on loop malfunction.

The limitation of the proposed algorithms is the requirement that the tests should be performed at the same control loop setpoint and under the same process load each time an assessment of loop security is required.

The setpoint constraint is not a major drawback since, as discussed in the flow control loop experiments of Chapter 2, it is possible to store in the computer the means of the innovation sequence or the loop security parameter estimates for several different values of setpoint. Thus, when a malfunction check is performed the current control loop characteristics may be compared with the stored values at the given setpoint.

The problem of "large" changes in the nominal value of the load is more difficult.

If the load change is measured, or can be approximated, from other

process considerations, then this poses no problem since the basic state variable model may be extended to include this extra coefficient. For example, the state space representation of the analogue setpoint control loop used in this Chapter, assuming constant nominal process load, was

$$\underline{x}(k+1) = A\,\underline{x}(k) + \underline{\Gamma}\,Q_i(k)$$

$$\underline{y}(k) = H\,\underline{x}(k) + \underline{v}(k)$$

where Q_i was a deviation variable representing small random unmeasured load disturbances. Now if Q_i is subject to large changes, which are measured, the state variable model may be written as:

$$\underline{x}(k+1) = A\,\underline{x}(k) + \underline{\Gamma}\,Q_i(k) + \underline{\Gamma}\,w(k) \qquad (3.18.1)$$

$$\underline{y}(k) = H\,\underline{x}(k) + \underline{v}(k) \qquad (3.18.2)$$

In this formulation, $\underline{x}(k)$, A, $\underline{\Gamma}$, $\underline{y}(k)$, H, $\underline{v}(k)$ have the same meaning and values as defined in Table 3.4. The process load $Q_i(k)$ is measured at each sample interval and $w(k)$ is a zero mean Gaussian white noise sequence of covariance Q. $w(k)$ is included to account for uncertainty in measuring $Q_i(k)$ and other stochastic disturbances.

Equations (3.18.1) and (3.18.2) are suitable for use in the Kalman filter, c.f. equations (3.4.1) and (3.4.2), and the malfunction detection method may be used with this new formulation.

However, the case where the load is unmeasured and varies widely over short periods of time is not easily accounted for in the present method and it is suggested that this is an area for further research.

CHAPTER 4

THE ROLE OF MALFUNCTION DETECTION IN RELIABILITY, MAINTAINABILITY AND AVAILABILITY.

4.1 List of Symbols

A	random variable associated with real variable α	various
d_τ	mean system downtime per inspection interval	time units
E_1	random variable associated with e_1	-
e_1	event consisting of an observation of the monitor signal at time τ when it is healthy.	-
E_2	random variable associated with e_2	-
e_2	event consisting of an observation of the monitor signal at time τ when it is unhealthy.	-
$f_{A/E_2}(\alpha/e_2)$	conditional probability density function for the monitor signal α given event e_2	-
$f_{A/T_m,E_2}(\alpha/t_m,e_2)$	conditional probability density function for the monitor signal α given event e_2 and t_m	-
$f_{E_1}(e_1)$	probability density function for event e_1	-
$f_{E_2}(e_2)$	probability density function for event e_2	-
$f_T(t)$	equipment/system failure probability density function; designer's failure density function.	-
$f_T(t;\alpha)$	probability density function for time of failure when the monitor signal is α	-
$f_{T/T\geq\tau}(t/t\geq\tau)$	Conditional probability density function describing time of failure given survival up to time τ	-
$f_{T/T_m,E_1}(t/t_m, e_1)$	conditional probability density function for time to failure given t_m and event e_1	-
$f_{T/T_m,E_2}(t/t_m,e_2)$	conditional probability density function for time to failure given t_m and event e_2	-
$f_{T_m}(t_m)$	probability density function for time to initial malfunction.	-

$f_{T_m/E_1}(t_m/e_1)$ conditional probability density function for time to initial malfunction given event e_1 -

$f_{T_m/E_2}(t_m/e_2)$ conditional probability density function for time to initial malfunction given event e_2 -

$f_{T_m/A,E_2}(t_m/\alpha, e_2)$ conditional probability density function for time to initial malfunction t_m, given the monitor signal α and the event e_2 -

$f_\theta(\theta)$ probability density function for time to failure θ (operator's failure density function) -

$f_\theta(\theta; \alpha)$ probability density function for time of failure when the monitor signal is α -

$f_{\theta/T\geq\tau}(\theta/t\geq\tau)$ conditional probability density function for time of failure θ given survival up to time τ -

$f_{\theta/T_m=\tau}(\theta/t_m=\tau)$ conditional probability density function for time of failure θ given the time of observation equals the initial malfunction time (terminal failure density function) -

$f_{\theta/T_m,E_1}(\theta/t_m, e_1)$ conditional probability density function for time of failure θ given t_m and the event e_1 -

$f_{\theta/T_m,E_2}(\theta/t_m,e_2)$ conditional probability density function for time of failure θ given t_m and the event e_2 -

$f_X(x)$ probability density function -

$F(\tau)$ failure probability distribution ($= \int_0^\tau f_T(t)\,dt$) -

K constant in equation (4.7.1.5.6) -

k constant in equations (4.6.1.4), (4.6.1.5) and (4.7.1.1). -

k_1 constant in equation (4.7.1.1.1) -

k_2; k_3	constant in equation (4.7.1.2.1)	-
k_4	constant in equation (4.7.1.6.1)	-
m	mean time between failure ($^1/\lambda$):	time
	constant in equations (4.6.1.4), (4.6.1.5)	units
	and (4.7.1.1).	
m_τ	mean time between failure for an	time
	inspected system.	units
m_1	constant in equation (4.7.1.1.1)	-
m_2	constant in equation (4.7.1.2.1)	-
m_3	constant in equation (4.7.1.6.1)	-
P( )	probability.	-
R(t)	system reliability: reliability function on	-
	the designer's time scale: designer's	
	reliability.	
R(θ)	reliability function on the operator's time	-
	scale: operator's reliability.	
$R_D(t)$	designer's reliability.	-
$R_{oI}(\theta)$	operator's reliability I	-
$R_{oII}(\theta)$	operator's reliability II	-
$R_{oIII}(\theta)$	operator's reliability III	-
T	random variable associated with t.	time units
t	time : time on the designer's time scale.	" "
T_m	random variable associated with t_m	" "
t_m	time of initial malfunction measured on the	" "
	designer's time scale.	
X	random variable associated with x	-
x	dummy variable	-
z(t)	hazard rate on designer's time scale.	failures/ unit time

$z(t/t \geq \tau)$	hazard rate corresponding to failure density function $f_{T/T \geq \tau}(t/t \geq \tau)$	failure/ unit time
$z(t/t_m, e_1)$	hazard rate corresponding to failure density function $f_{T/T_m, E_1}(t/t_m, e_1)$	"
$z(t/t_m, e_2)$	hazard rate corresponding to failure density function $f_{T/T_m, E_2}(t/t_m, e_2)$	"

Greek letters

α	monitor signal	various
β	probability level	-
$\Gamma( )$	gamma function	-
Θ	time to failure on operator's time scale (measured from the time of monitor observation i.e. $\Theta = t - \tau$)	time units
θ	random variable associated with Θ	time units
λ	hazard (failure) rate (constant)	failures/ unit time
μ_T	mean of random variable T	time units
μ_{T_m}	mean of random variable T_m	time units
$\mu_{\theta/T_m=\tau}$	mean of random variable θ given that this is measured from the time of the initial malfunction.	" "
μ_x	mean of the dummy random variable X	-
τ	inspection frequency; time of observation of the equipment and/or of the monitor, measured on the designer's time scale.	time units
$\phi_{T,T_m,E_1}(t,t_m,e_1)$	joint probability density function of the	-

time to failure t, the time of the initial malfunction t_m and the event e_1 -

$\phi_{T,T_m,E_2}(t, t_m, e_2)$ joint probability density function of the time to failure t, the time of the initial malfunction t_m and the event e_2 -

$\phi_{T_m,A,E_2}(t_m,\alpha, e_2)$ joint probability density function of the time to initial malfunction, t_m, the monitor signal α and the event e_2 -

4.2 Introduction

As the chemical industry has developed, the stimulation of the economic rewards which may be obtained from rapid process commissioning, fewer major equipment failures, reduced maintenance costs and a higher on-stream production time has led to the increasing application and development of reliability engineering (3), (7), (116), (117).

The concepts of reliability engineering originated in the aircraft industry during World War II, and have now developed into a discipline with applications ranging from the aerospace to the nuclear reactor industry. Reliability engineering relies heavily upon probability theory for its development, and many of the fundamental concepts may be found in references (28), (118) -(121).

Paralleling this growth has been an interest in the monitoring of equipment to detect potential failures in advance. This may result in four advantages:

i) Dangerous process situations which may put personnel at risk may be avoided.

ii) Unscheduled process stoppages which disrupt production may be avoided.

iii) Unnecessary damage to equipment may be prevented by identifying the fault at an incipient stage.

iv) A rational approach to equipment maintenance may be employed since the forewarning of impending failure enables a planned scheduled repair to be performed rather than an emergency repair. Also the need to dismantle serviceable equipment for examination during the preventive maintenance period may be avoided.

The use of a monitoring device for any particular application requires a knowledge of the failure mechanisms of the system, the

operating characteristics of the healthy system and the optimum type of monitor for the given system. Edwards and Lees (1) have suggested that malfunction monitoring schemes can be based upon the condition or performance of the equipment. However, the former is more common and is referred to as condition monitoring.

Some methods of equipment condition monitoring have been reviewed by Dowson (11) and Trotter (12). The method of vibration analysis is well developed and widely used to monitor the state of rotating machinery (122), (123), (124). Other monitoring techniques, such as shock pulse testing (125), mechanical debris analysis (126) and acoustic measurements (22), (127), (128), (129) are less well developed, but their practical application is increasing. Typically these monitors produce a signal which is indicative, with some degree of uncertainty, that the equipment is healthy, unhealthy or failed.

The case for using a process computer to monitor and detect system malfunction was considered, in general terms, in Chapter 1. Initially, in formulating this project, it was decided to examine the role of equipment condition monitoring by considering the consequences of equipment inspection and malfunction detection on both the economic and reliability aspects of the system. Details of this study are given in section 4.4.

However, as this project developed, it became apparent that the precise way in which the condition monitor information was used to modify estimates of the system reliability, availability and maintainability was not adequately described by the existing reliability theory. Section 4.5 of this Chapter is a contribution to such a theory.

4.3 Some fundamental concepts and definitions

In performing any study in reliability engineering it is of prime importance to define explicitly what constitutes an equipment failure.

Lees (4) has suggested that the most common practical definition of instrument failure is that the instrument is not operating to the satisfaction of the process operator.

Green and Bourne (119) have formally defined failure as:

> "The condition of a component, equipment or system whereby a particular performance characteristic or number of performance characteristics, of such a device has moved outside the assessed specification range for that characteristic in such a way that the component, equipment or system can no longer perform adequately in the desired manner."

The point about these definitions is that the application of the equipment determines what constitutes a failure. This feature was discussed in Chapter 2, when the reliability of a flow control loop was discussed in terms of the reproducibility and absolute accuracy of the system.

Equipment failures may be further subdivided into revealed and unrevealed failures.

A revealed failure is defined in (119) as a failure of a component, equipment or system which is automatically brought to light on its occurrence.

An unrevealed failure (119) is a failure of a component, equipment or system which remains hidden until revealed by some thorough proof-testing procedure.

To complete this definition, the concept of proof testing is required (119).

"Proof testing is a method of ensuring that a component, equipment or system possesses all the required performance characteristics and is capable of responding to input conditions in the desired manner."

In terms of these definitions the malfunction detection algorithms developed in this thesis and (1) and (9) are system proof tests designed to detect unrevealed failures, particularly gradual and developing faults.

4.4 The effect of a periodic equipment inspection policy on some reliability performance indices

This section examines the effect of equipment condition monitoring on several aspects of system reliability.

There is an extensive literature on the design of optimal inspection and maintenance policies for stochastically failing equipment. Reviews of these schemes have been given by McCall (130), Berg and Epstein (131), and Jardine (132). In general, the design of such policies is based upon the optimisation of some formulated economic objective function. For example, the cost of equipment inspection in terms of labour, lost production, etc., may be balanced against the benefits of detecting failure, such as minimising the equipment repair time.

However in this section a much more fundamental approach is adopted. Possibly the simplest equipment inspection scheme of all is the periodic policy, where the state of the system is determined every τ time units. Now one of the suggested advantages of using a process

computer to monitor equipment condition was that the machine could perform very frequent tests or checks of the system state without incurring a higher probability of error in interpreting the results. This is contrary to the human operator's performance. Also, the use of a process computer to monitor equipment condition is relatively inexpensive, since often tests may be performed while the equipment is on-line.

In the following sections it is assumed that a process computer is used to periodically inspect and detect unrevealed faults in the system, and the effect of the inspection frequency, τ , on some features of the system reliability is examined.

It is assumed throughout the analysis that the computer is a perfect detector, i.e. if an unrevealed equipment failure is present, then the computer detects it with probability 1.

4.4.1 Mean downtime per inspection interval

Suppose that a system is operated continuously and is inspected for an unrevealed fault every τ time units. An example of such a system may be a control loop which is examined for malfunction using one of the techniques developed earlier. Obviously, the system can be in a failed state during some proportion of the time between monitorings without the process operator being aware of this.

Now let the system failure probability density function be $f_T(t)$, then the mean system downtime in the interval $(0,\tau)$ between inspections is (120):

$$d_\tau = \int_0^\tau (\tau - t)\, f_T(t)\, dt$$

d_τ represents the mean time from system failure until inspection (and

detection). Now assuming the repair time is significantly less than the time between monitorings, the mean downtime (unavailability) in a time period T can be found in the steady state as:

$$\sum_i d_{\tau_i} = \frac{T}{\tau} d_\tau$$

Clearly, d_τ is a function of τ, and an investigation of how this relationship varied with inspection interval for several equipment configurations was considered. Table 4.1 summarises the systems analysed and the assumptions made in deriving the system failure density function $f_T(t)$. The fundamental failure probability density function for a single equipment was assumed to be exponential with a hazard rate λ.

The relationship between d_τ and τ for various systems is shown in Figure 4.1. This shows that d_τ decreases as the inspection interval, τ, decreases, although the relationship is not linear and the reduction in downtime achieved becomes smaller as the inspection frequency increases. In particular, it can be seen that for system applications where an infrequent inspection policy is adopted, then a 2 out of 3 parallel redundant system will, on average, be in a failed condition for more time than a single equipment. Thus, in terms of the mean downtime per inspection interval, a single equipment offers better performance and also represents a smaller capital cost.

Configuration	$f_T(t)$	λ	d_τ	Assumptions
Single equipment	$\lambda e^{-\lambda t}$	0.001	$\tau - \frac{1}{\lambda}(1 - e^{-\lambda\tau})$	-
Two equipment standby redundant system.	$\lambda^2 t e^{-\lambda t}$	0.001	$e^{-\lambda\tau}(\tau + \frac{2}{\lambda}) + (\tau - \frac{2}{\lambda})$	i) Switch is perfect. ii) Equipments fail only in the on-line position. iii) System is failed when both equipments are failed. iv) Both equipments have the same failure rate and exponential failure probability density function.
Two equipment parallel redundant system.	$2\lambda(e^{-\lambda t} - e^{-2\lambda t})$	0.001	$e^{-\lambda\tau}(\frac{2}{\lambda} - \frac{1}{2\lambda} e^{-\lambda\tau}) + (\tau - \frac{3}{2\lambda})$	i) System is failed when both equipments are failed. ii) Both equipments have the same failure rate and exponential failure probability density function.
Two out of three parallel redundant system.	$6\lambda(e^{-2\lambda t} - e^{-3\lambda t})$	0.001	$\frac{1}{\lambda}(\frac{5}{6} + \frac{3}{2} e^{-2\lambda\tau} - \frac{2}{3} e^{-3\lambda\tau}) + \tau$	i) System is failed when two or more equipments are failed. ii) All equipments have the same failure rate and exponential failure probability density function.

TABLE 4.1. Effect of inspection interval on system downtime.

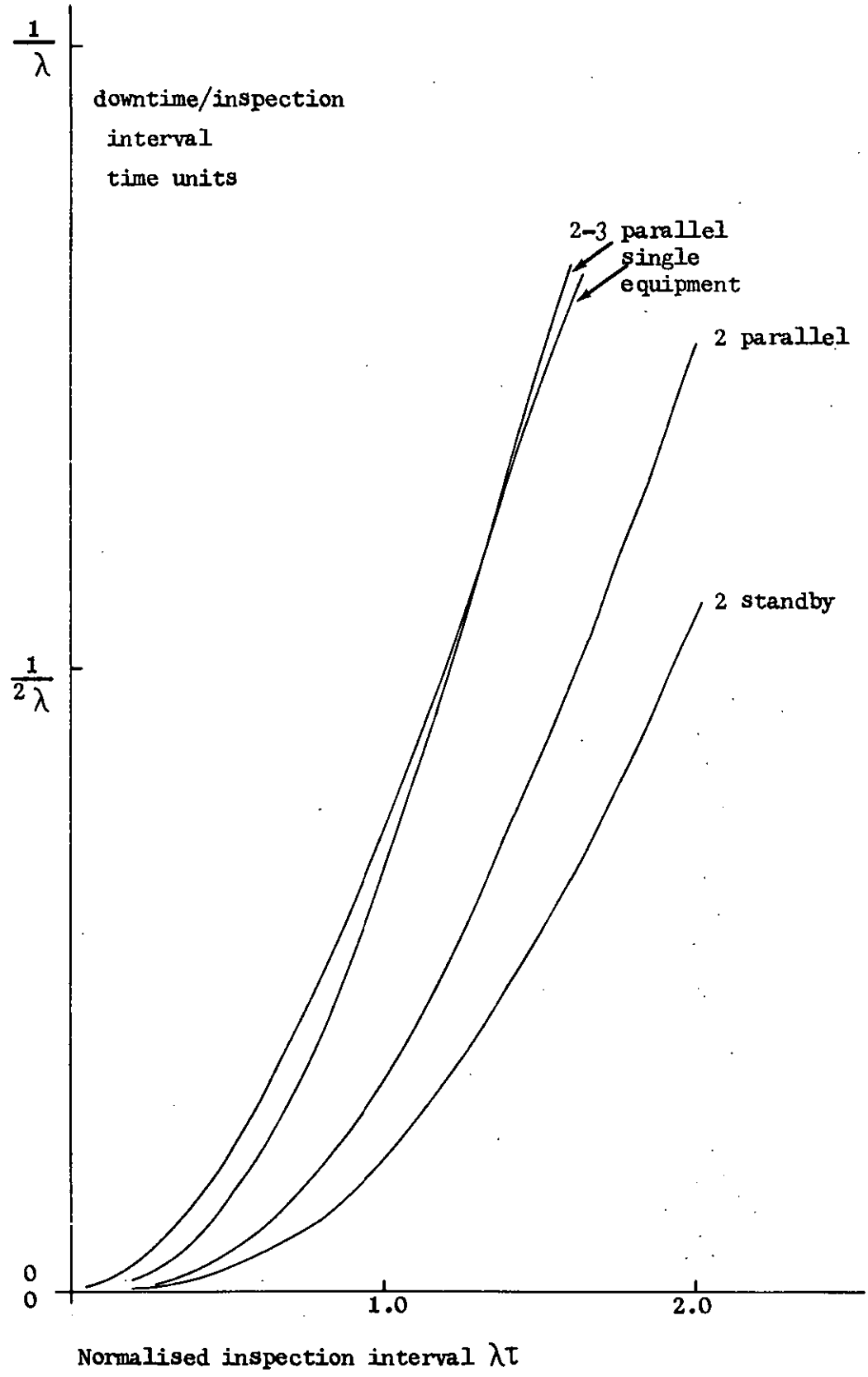


FIGURE 4.1 Effect of inspection interval on system downtime.

4.4.2 Mean time between failures of maintained redundant systems

The mean time between failure (MTBF) of a system is often used as a figure of merit in reliability engineering, and is defined as the expected time for a system to reach a failed state.

It may be shown (118) that the MTBF is given by integrating the reliability function, R(t), over the range $t = 0$ to ∞

$$m = \int_0^\infty R(t)\, dt$$

Thus on average any given redundant system will fail once every m time units if the failed components of a redundant system are not replaced until the complete system fails.

m may be increased by introducing a maintenance policy (130), (131). Suppose the adopted policy is one such that the system components are inspected periodically to determine failure, and on detection, failed components are repaired or replaced. If this policy is used then the system may be expected to fail less frequently than it would without inspections, because it is assumed that every new operating period after inspection begins with full system redundancy restored. Thus the MTBF for the inspected system, denoted m_τ, becomes longer than m and in theory it would become infinite if failed redundant components were immediately detected and replaced.

m_τ is a function of the inspection interval, τ, and Bazovsky (133) has shown that it is related to the system reliability, R (t), according to:

$$m_\tau = \frac{\int_0^\tau R(t)\, dt}{1 - R(\tau)} \qquad (4.4.2.1)$$

Under the adopted maintenance policy, the redundant system is inspected for failures in its components every τ time units. If the inspection reveals failed components, they are replaced or repaired so that full system redundancy is restored. If an inspection does not reveal failures, then assuming the system components have exponential failure probability laws, the system is considered renewed.

The inspected system MTBF, m_τ, is a function of the inspection frequency, τ, and the relations between them for several redundant systems were derived using equation (4.4.2.1) and are given in Table 4.2. The assumptions made in deriving Table 4.2 were the same as those described in Table 4.1.

Figure 4.2 shows how m_τ varies, as a function of τ for the systems

Configuration	$f_T(t)$	$R(t)$	m	m_τ
Two equipment standby redundant	$\lambda^2 t\, e^{-\lambda t}$	$e^{-\lambda t}(1+\lambda t)$	$\frac{2}{\lambda}$	$\frac{m - e^{-\lambda\tau}(m+\tau)}{1 - e^{-\lambda\tau}(1+\lambda\tau)}$
Two equipment parallel redundant	$2\lambda(e^{-2\lambda t} - e^{-\lambda t})$	$e^{-\lambda t}(2 - e^{-\lambda t})$	$\frac{3}{2\lambda}$	$\frac{m - e^{-\lambda\tau}(\frac{2}{\lambda} - \frac{e^{-\lambda\tau}}{2\lambda})}{1 - e^{-\lambda\tau}(2 - e^{-\lambda\tau})}$
Two out of three parallel redundant	$6\lambda(e^{-2\lambda t} - e^{-3\lambda t})$	$e^{-2\lambda t}(3 - 2e^{-\lambda t})$	$\frac{5}{6\lambda}$	$\frac{m - e^{-2\lambda\tau}(\frac{3}{2\lambda} - \frac{2e^{-\lambda\tau}}{3\lambda})}{1 - e^{-2\lambda\tau}(3 - 2e^{-\lambda\tau})}$

TABLE 4.2. Effect of inspection interval on redundant system mean time between failures

given in Table 4.2.

In each of the systems considered there is quite a significant improvement in MTBF as the inspection frequency increases (or τ decreases).

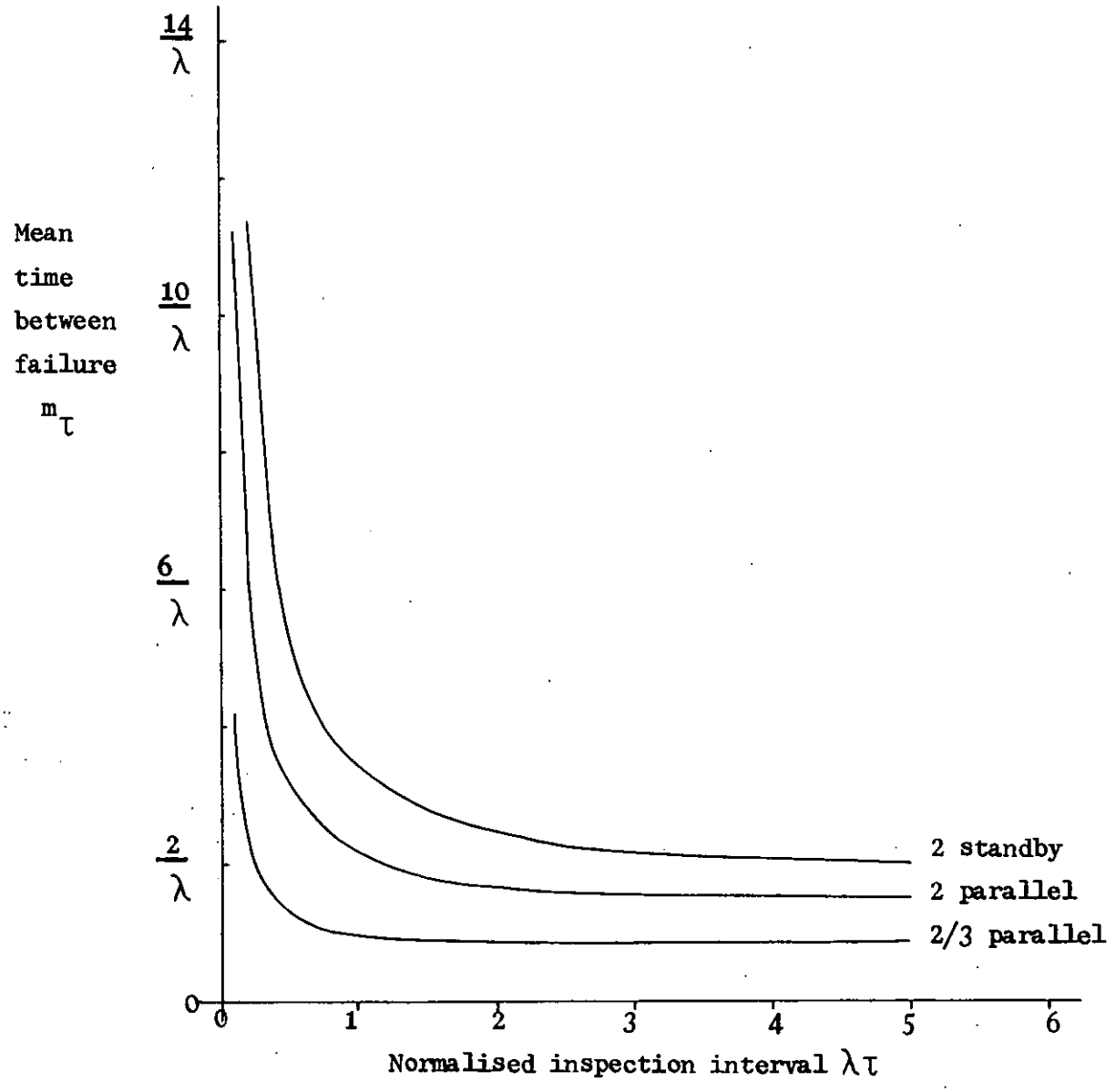


FIGURE 4.2 Effect of inspection interval on redundant system mean system between failure

In particular, for a two equipment standby redundant system, the MTBF can be doubled by inspecting the system once every $^{0.75}/_{\lambda}$ time units. For example, suppose $\lambda = 0.001$ then $m = 2000$. Now using the described inspection policy with $\tau = 750$, gives $m_{\tau} = 4000$.

4.4.3 Design and economics of maintained redundant systems

A designer is often called upon to specify a system to meet a certain MTBF criterion. There are a number of design alternatives available, ranging from a single equipment to redundant systems maintained in such a way as to minimise some cost function. The precise selection of the system is a function of many criteria. However, to illustrate the value of a malfunction detection policy, it is assumed that the designer has only two alternatives.

To meet the specified MTBF the designer may use an n parallel redundant system or a 2 parallel system employing a periodic inspection-maintenance policy with an inspection interval τ .

Section 4.4.2 has shown that by choosing the interval τ , the designer can achieve any desired system MTBF, and so make the 2 parallel system equivalent to any n parallel redundant system.

For example, the 2 parallel system with a periodic inspection policy has a MTBF given in Table 4.2 by

$$m_{\tau} = \frac{\frac{3}{2\lambda} - e^{-\lambda\tau}\left(\frac{2}{\lambda} - \frac{e^{-\lambda\tau}}{2\lambda}\right)}{1 - e^{-\lambda\tau}\left(2 - e^{-\lambda\tau}\right)}$$

Now suppose the design requirement is for a system to have a MTBF of not less than $\frac{11}{6\lambda}$. The designer may thus choose a 3 parallel redundant system or a 2 parallel system with a periodic inspection-maintenance policy where τ is determined by solving:-

$$\frac{11}{6\lambda} = \frac{\frac{3}{2\lambda} - e^{-\lambda\tau}\left(\frac{2}{\lambda} - \frac{e^{-\lambda\tau}}{2\lambda}\right)}{1 - e^{-\lambda\tau}\left(2 - e^{-\lambda\tau}\right)}$$

The solution of this equation yields the maximum value of the inspection interval τ_{max} which meets the design requirement. Any inspection interval τ less than τ_{max} consequently results in the 2 parallel system having a longer MTBF than the design specification.

In a similar manner, the value τ_{max} such that a 2 parallel system has an equivalent MTBF to any n parallel redundant system may be determined.

The results of such an analysis are shown in Figure 4.3. This Figure shows that the inspection frequencies needed to replace n parallel systems are modest. For example, a 7 parallel redundant system may be replaced by a 2 parallel system with a periodic inspection scheduled every $\frac{0.65}{\lambda}$ time units, where $\frac{1}{\lambda}$ is the MTBF of a single component.

Summarising, sections 4.4.1 to 4.4.3 have examined and quantified, in terms of some reliability performance indices, the consequences of employing a periodic malfunction detection policy to determine unrevealed failures in plant equipment and instruments.

This rather brief and tentative study has revealed that the consequences of monitoring can be profound and that, for the reliability performance indices considered, relatively infrequent inspections resulted in significant improvement. The indications are that computerised monitoring schemes, which may be implemented frequently, would enhance the overall system security and as such should be the subject of continuing effort so that basic system malfunction detection algorithms are available.

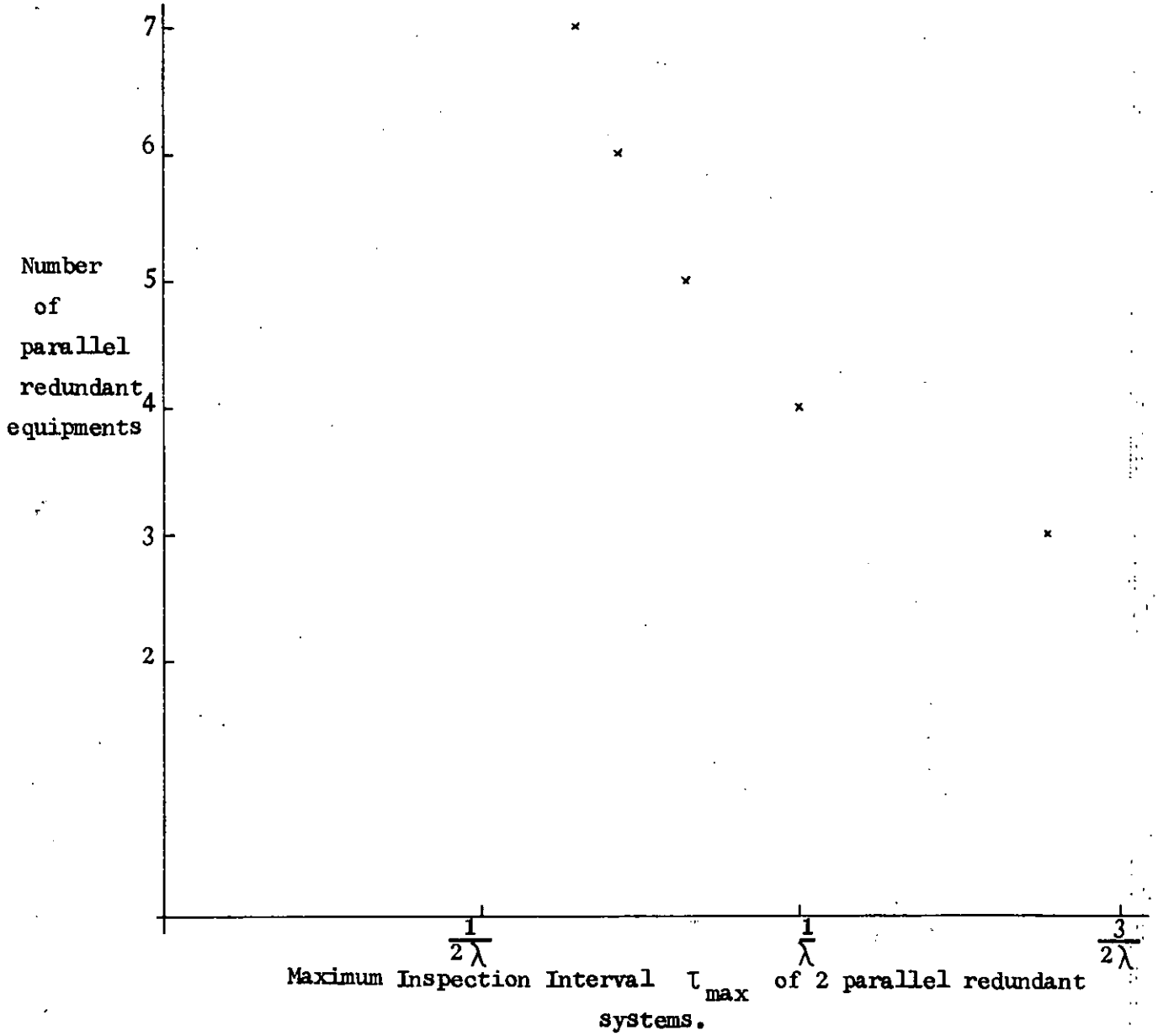


FIGURE 4.3 Equivalence of n parallel and inspected 2 parallel systems.

4.5 The effect of malfunction monitoring on reliability, maintainability and availability

Despite the voluminous literature on reliability and maintenance problems and policies (130), (131), the effect of equipment malfunction monitoring does not appear to have been adequately considered in terms of reliability theory.

Derman (134), Kolesar (135), Klein (136) and many others have represented a single equipment, subject to degradation, by a series of healthy, unhealthy and failed Markov states and subsequently derived maintenance policies designed to achieve the minimisation of some economic objective function. The proposed maintenance strategies assume that an equipment inspection procedure is capable of detecting which state the system is in at the observation time.

A more general Markov model was used by Satia and Lave (137) who assumed that the equipment state could not be observed directly, but only through a probabilistic observer. The extension of the Markov system degradation representation to redundant systems has not proved trivial, although Mine and Kawai (138) have considered two equipment systems.

The problem of a post-mortem failure diagnosis has been considered by Gross (139). It is assumed that a 2 component system fails, but whether component 1 or 2 is the cause is unknown. The repairman can use a diagnostic monitor which enables the failed component to be located. By balancing the additional cost incurred in system repair if the repairman makes the incorrect decision concerning the choice of component for repair against the cost of purchasing a diagnostic tool Gross determines the optimal amount to be spent in providing such a tool.

Polovko (140) and Jensen (141) have examined the problem of calculating the reliability of systems which are subject to gradual failure (as opposed to catastrophic failure) due to component parameter drift. In this case failure is defined when the system output does not meet some predetermined specification.

The problem of equipments deteriorating over a period of time in store has been discussed by Welker (142). The degradation process is characterised by some continuous parameter (for example temperature) which is discretised to form a series of Markov states. By using Markovian theory with this monitoring process an equipment replacement maintenance policy is derived.

However, in spite of this activity, there does not appear to be an adequate framework for the assessment and contribution of malfunction monitoring to reliability. In this work the approach adopted is that the reliability is a function of the state of knowledge of the system (143), and as such can be decomposed into the designer's and operator's reliability. The impact of malfunction monitoring on the operator's reliability is then considered.

The mathematical quantities and notation used in the following sections have been reviewed in Appendix I, while some problems encountered in the use of Markov models are given in Appendix V.

4.5.1 Malfunction monitoring

Some reviews of malfunction monitoring techniques were discussed in section 4.2. In general, equipment monitors are special, dedicated instruments which measure some distinctive feature of the system's condition or performance. However, a monitor need not be a special

instrument and a process computer examining the state of its measuring instruments may also be considered as a malfunction monitor.

The fundamental feature of these monitors is that they produce a signal which is indicative of the equipment health. This signal on which the monitor is based may be a direct reading or a derived function (for example, vibration measurements often involve spectral analysis). However, in each case it is assumed that the final monitor signal will resemble the function shown in Figure 4.4.

This curve has three regions: I, the equipment is healthy; II, the initial malfunction has developed at time t_m, so that the equipment is unhealthy and failing; III, the equipment has failed. Region II represents the failure characteristic of the equipment and is the time period in which incipient malfunction detection is possible.

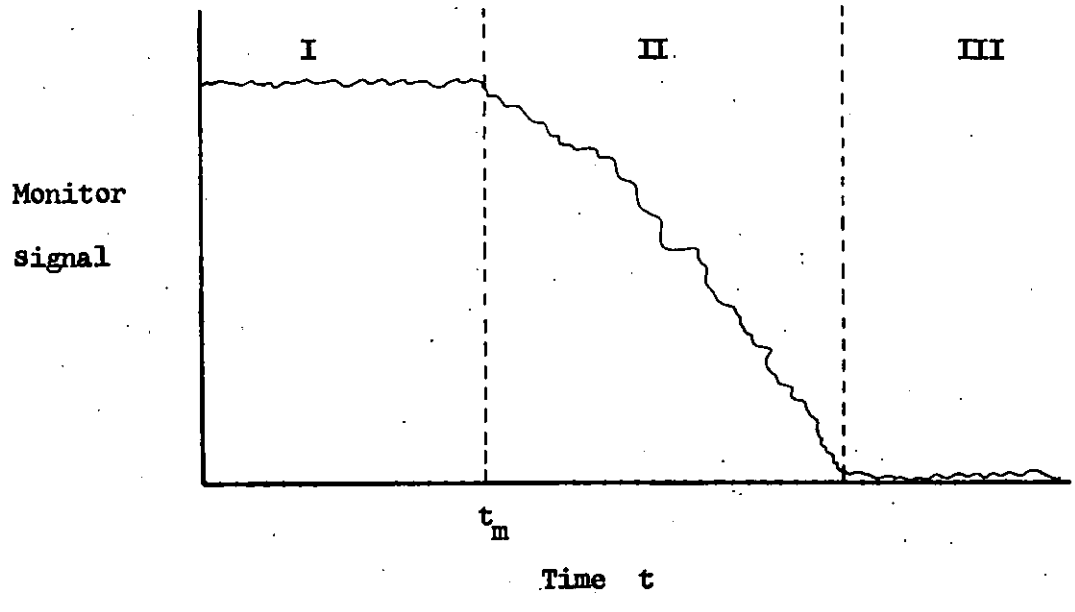


Figure 4.4 Typical signal from malfunction monitor.

4.5.2 Reliability, maintainability and availability

Before proceeding with the theoretical development, it is necessary to explicitly define what can and cannot be achieved by malfunction monitoring. As a vehicle for discussion a single equipment is considered since this illustrates most clearly the capabilities and limitations of malfunction monitoring.

4.5.2.1 Failure and reliability regimes

Failure rate may be regarded as a property of the equipment, although it may not be known. Reliability, however, is a probability which an engineer estimates. Now adopting the interpretation of Tribus (143) that probability is simply a numerical encoding of the state of knowledge about a system, then the concept of reliability may be decomposed into several different regimes.

The first reliability regime pertains to the plant designer's viewpoint of the equipment operation. The designer adopts a macroscopic view of the equipment operation and is concerned with the overall reliability of the equipment measured from the time of plant start-up. This time is denoted t. In formulating his estimate of reliability the designer considers the overall failure rate and characteristics of the equipment. This reliability is the conventional reliability, R(t), which is referred to here as the "designer's reliability".

The second reliability regime is that of the plant operator. The plant operator is much more concerned with the dynamic operation of the equipment than the designer since, after the plant start-up, he can inspect the equipment for failure at any particular time τ and thereby actively update his knowledge of equipment operation. For example,

the designer may estimate that the reliability of the equipment surviving τ time units from start up as $R(\tau) = \beta$. However, the operator can inspect the equipment at time τ and if it is operating his assessment of reliability is 1, while if it is failed it is 0. This is not the same reliability estimate as the designer's.

The time scale of the plant operator is measured from the time that he observes the state of the equipment and is denoted θ. Since the observation occurs at time τ on the first time scale t then:

$$\theta = t - \tau \qquad (4.5.2.1.1)$$

By redefining his time scale like this, the operator can make a running or dynamic estimate of the reliability, denoted $R(\theta)$. This is true whether or not the plant operator is practising malfunction monitoring, since the information gained at each inspection is simply that the equipment has not failed in the interval $(0, \tau)$.

This second reliability regime is referred to as the "operator's reliability".

If the plant operator is practising malfunction monitoring, however, his state of knowledge about the equipment operation is increased further by considering the monitor information. When the operator examines the monitor at time τ and records a healthy monitor signal, then he has three pieces of information. He knows the overall equipment failure rate and failure probability density function, the fact that the equipment has not failed in the interval $(0, \tau)$ and finally that the monitor signal is healthy and so immediate equipment failure is less likely. However, if the signal is unhealthy when he examines the monitor, he again knows the above information except that he now expects failure to be more likely in the near future.

Using these various information sets, it can be seen that there

are four possible reliability regimes and estimates to consider:

i) Designer's reliability;

ii) Operator's reliability I - no monitor;

iii) Operator's reliability II - monitor
- signal healthy;

iv) Operator's reliability III - monitor
- signal unhealthy.

These different estimates of the reliability of a single equipment do not in any way change its failure rate. However, they can be used to modify decisions in the operation and maintenance of the equipment.

4.5.2.2 Maintainability and availability

Several advantages of equipment monitoring were discussed in section 4.2. However, the aspect considered here is the reduction of repair time and the improvement of maintainability. There are several ways in which equipment repair may be improved by monitoring. These include avoiding severe equipment damage, permitting better repair strategies (such as the organisation and coordination of maintenance crews, and the location of the relevant equipment spares), and allowing better planning of process operation. If these benefits did materialise, then repair time and consequently maintainability of a single equipment may be a function of monitoring.

Since single equipment availability is also a function of repair time and failure rate, then it too may be a function of the monitoring activity.

However, as already stated, the reliability of a single equipment, in the conventional sense of the designer's reliability, is not affected by monitoring.

4.6 Reliability functions for a single equipment

4.6.1 Designer's reliability

This is the conventional reliability function. Denoting the overall probability failure density function by $f_T(t)$, then the reliability function is given by:

$$R_D(t) = 1 - \int_0^t f_T(t)\, dt \qquad (4.6.1.1)$$

If the designer uses an exponential failure density function for $f_T(t)$, then:

$$f_T(t) = \lambda \exp(-\lambda t) \qquad (4.6.1.2)$$

$$R_D(t) = \exp(-\lambda t) \qquad (4.6.1.3)$$

If a Weibull failure density function is used, then (118):

$$f_T(t) = k\, t^m \exp\left(-\frac{kt^{m+1}}{m+1}\right) \qquad (4.6.1.4)$$

$$R_D(t) = \exp\left(-\frac{kt^{m+1}}{m+1}\right) \qquad (4.6.1.5)$$

4.6.2 Operator's reliability I

In this case the operator does not have the benefit of a malfunction monitor, but he can inspect the equipment at any time τ and satisfy himself that the equipment has not failed. Thus at the inspection time τ the information available to the operator is:

i) The equipment has not failed in the interval $(0, \tau)$.

ii) The overall probability failure density function of the equipment.

He can use this information to condition the failure density function and pose the question: "What is the probability of failing at time t, given equipment survival up to time τ (i.e. $t > \tau$)?"

Mathematically, the conditional failure density function is described by (28):

$$f_{T/T\geq\tau}(t \,/\, t\geq\tau) = 0 \qquad 0\leq t<\tau \qquad (4.6.2.1)$$

$$= \frac{f_T(t)}{1-\int_0^{\tau} f_T(t)\,dt} \qquad \tau\leq t<\infty \qquad (4.6.2.2)$$

$$= \frac{f_T(t)}{1-F(\tau)} \qquad \tau\leq t<\infty \qquad (4.6.2.3)$$

Now setting $\theta = t-\tau$ and using the random variable transformation theorem (28),

$$f_{\theta/T\geq\tau}(\theta \,/\, t\geq\tau) = \frac{f_T(\theta+\tau)}{1-F(\tau)} \qquad 0<\theta\leq\infty \qquad (4.6.2.4)$$

and so the reliability becomes:

$$R_{oI}(\theta) = 1-\int_0^{\theta} f_{\theta/T\geq\tau}(\theta \,/\, t\geq\tau)\,d\theta \qquad (4.6.2.5)$$

$$= 1-\frac{1}{1-F(\tau)}\int_0^{\theta} f_T(\theta+\tau)\,d\theta \qquad (4.6.2.6)$$

When $f_T(t)$ is exponential then:

$$R_{oI}(\theta) = \exp(-\lambda\theta) \qquad (4.6.2.7)$$

If $f_T(t)$ is a Weibull function then:

$$R_{oI}(\theta) = 1-\frac{1}{\exp\left(-\frac{k\tau^{m+1}}{m+1}\right)}\int_0^{\theta} k(\theta+\tau)^m \exp\left(\frac{-k(\theta+\tau)^{m+1}}{m+1}\right)d\theta \qquad (4.6.2.8)$$

$R_{oI}(\theta)$ simply represents a "running" estimate of the reliability designed to include the fact that the equipment has not failed up to the current observation time.

4.6.3 Operator's reliability II

The operator has a malfunction monitor. At time τ he examines the monitor and records a healthy signal and so he has the following information:

i) The equipment has not failed in the interval (o, τ), i.e. $\tau < t \leq \infty$

ii) The monitor reading is healthy and so incipient failure has not begun, i.e. $\tau < t_m \leq \infty$

iii) The terminal probability failure density function. This density function represents the wear-out characteristic of the equipment and it is assumed that this can be derived from an analysis of the experimental failure data when the time of failure is measured from the first occurrence of the initial malfunction, i.e. $\tau = t_m$. This is denoted $f_{\theta/T_m=\tau}(\theta/t_m=\tau)$.

iv) The malfunction monitor is perfect, i.e. the instrument does not fail.

The reliability function may be derived as follows.

Let the event e_1 be defined as the facts that a monitor observation is made at time τ, the signal is healthy, $\tau < t \leq \infty$ and $\tau < t_m \leq \infty$. Now defining a joint probability density function for the random variables T, T_m and E_1 gives (28):

$$\phi_{T,Tm,E_1}(t,t_m,e_1) = f_{T/T_m,E_1}(t/t_m,e_1)\, f_{T_m/E_1}(t_m/e_1)\, f_{E_1}(e_1) \quad (4.6.3.1)$$

Integrating with respect to t_m and e_1 yields the marginal failure density function $f_T(t)$, i.e.

$$f_T(t) = \int_{\substack{\text{range} \\ T_m}} \int_{\substack{\text{range} \\ E_1}} f_{T/T_m,E_1}(t/t_m,e_1)\, f_{T_m/E_1}(t_m/e_1)\, f_{E_1}(e_1)\, dt_m\, de_1 \quad (4.6.3.2)$$

However, the event e_1 is a certain event and so:

$$f_{E_1}(e_1) = \delta(e_1) \qquad (4.6.3.3)$$

Then equation (4.6.3.2) becomes:

$$f_T(t) = \int_{\substack{\text{range} \\ T_m}} f_{T/T_m,E_1}(t/t_m, e_1)\, f_{T_m/E_1}(t_m/e_1)\, dt_m \qquad (4.6.3.4)$$

Now recalling that $\theta = t - \tau$ and using the transformation theorem (28) gives:

$$f_\theta(\theta) = \int_\tau^\infty f_{T/T_m,E_1}(\tau+\theta/t_m, e_1)\; f_{T_m/E_1}(t_m/e_1)\, d\, t_m \qquad (4.6.3.5)$$

where $f_{T/T_m,E_1}(\tau+\theta/t_m, e_1) = f_{\theta/T_m,E_1}(\theta/t_m, e_1)$ (4.6.3.6)

The density function for the random variable T_m, conditioned upon the event e_1, simply states that T_m cannot have a value in the interval (o, τ). Mathematically this is given by:

$$f_{T_m/E_1}(t_m/e_1) = \frac{f_{T_m}(t_m)}{1 - \int_o^\tau f_{T_m}(t_m)\, d\, t_m} \qquad \tau < t_m \leq \infty \qquad (4.6.3.7a)$$

$$= 0 \qquad o < t_m \leq \tau \qquad (4.6.3.7b)$$

The reliability function is determined from equation (4.6.3.5) as:

$$R_{oII}(\theta) = 1 - \int_o^\theta f_\theta(\theta)\, d\theta$$

or

$$R_{oII}(\theta) = 1 - \int_o^\theta \int_\tau^\infty f_{T/T_m,E_1}(\tau+\theta/t_m, e_1)\, f_{T_m/E_1}(t_m/e_1)\, dt_m\, d\theta \qquad (4.6.3.8)$$

4.6.4 Operator's reliability III

The operator has a monitor which he inspects at time τ, but this time the signal indicates an unhealthy equipment. Therefore at the observation time the following information is available to the operator:

i) The equipment has not failed in the interval $(0, \tau)$, i.e. $\tau < t \leq \infty$

ii) The monitor reading is unhealthy and has a particular value α at time τ, and consequently $0 < t_m \leq \tau$.

iii) The terminal failure density function $f_{\theta/T_m=\tau}(\theta/t_m=\tau)$.

iv) The probability density function describing the values of the monitor signal values α from the time of initial malfunction. It is assumed that this would be determined experimentally from the characteristics of the monitor.

v) The malfunction monitor is perfect, i.e. the instrument does not fail.

The derivation of the reliability proceeds as follows. The event e_2 is defined as the observation time τ, the signal is unhealthy, $0 < t_m \leq \tau$ and $\tau < t \leq \infty$. Then writing the joint probability density function for the random variables T, T_m and E_2 gives:

$$\phi_{T,T_m,E_2}(t, t_m, e_2) = f_{T/T_m,E_2}(t/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)\, f_{E_2}(e_2) \tag{4.6.4.1}$$

Integrating with respect to t_m and e_2 yields the marginal probability density function

$$f_T(t) = \int_{\substack{\text{range}\\T_m}} \int_{\substack{\text{range}\\E_2}} f_{T/T_m,E_2}(t/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)\, f_{E_2}(e_2)\, dt_m\, de_2 \tag{4.6.4.2}$$

But by definition:

$$f_{E_2}(e_2) = \delta(e_2) \qquad (4.6.4.3)$$

and so:

$$f_T(t) = \int_{\text{range } T_m} f_{T/T_m,E_2}(t/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)\, d\, t_m \qquad (4.6.4.4)$$

The conditional probability density function $f_{T_m/E_2}(t_m/e_2)$ is derived from the probability density function for the time of initial malfunction $f_{T_m}(t_m)$ given the event e_2, i.e. $0 < t_m \leq \tau$. This is given by:

$$f_{T_m/E_2}(t_m/e_2) = \frac{f_{T_m}(t_m)}{\int_0^{\tau} f_{T_m}(t_m)\, d\, t_m} \qquad 0 < t_m \leq \tau \qquad (4.6.4.5a)$$

$$= 0 \qquad \tau < t_m \leq \infty \qquad (4.6.4.5b)$$

In this context $f_{T_m/E_2}(t_m/e_2)$ is an a priori density function.

However, there is available further information in the form of the malfunction monitor signal α which may be used to modify the estimate of $f_{T_m/E_2}(t_m/e_2)$ by the application of Bayes' theorem (28).

The joint probability density function for the random variables T_m, A, E_2 may be written as:

$$\phi_{T_m,A,E_2}(t_m,\alpha,e_2) = f_{T_m/A,E_2}(t_m/\alpha,e_2)\, f_{A/E_2}(\alpha/e_2)\, f_{E_2}(e_2) \qquad (4.6.4.6)$$

$$= f_{A/T_m,E_2}(\alpha/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)\, f_{E_2}(e_2) \qquad (4.6.4.7)$$

Equating (4.6.4.6) and (4.6.4.7) yields Bayes' theorem:

$$f_{T_m/A,E_2}(t_m/\alpha,e_2) = \frac{f_{A/T_m,E_2}(\alpha/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)}{f_{A/E_2}(\alpha/e_2)} \qquad (4.6.4.8)$$

But by the extension rule (28), (143) $f_{A/E_2}(\alpha/e_2)$ may be written as:

$$f_{A/E_2}(\alpha/e_2) = \int_{\substack{\text{range} \\ T_m}} f_{A/T_m,E_2}(\alpha/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)\, dt_m \quad (4.6.4.9)$$

and so equation (4.6.4.8) becomes:

$$f_{T_m/A,E_2}(t_m/\alpha,e_2) = \frac{f_{A/T_m,E_2}(\alpha/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)}{\int_{\substack{\text{range} \\ T_m}} f_{A/T_m,E_2}(\alpha/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)\, dt_m} \quad (4.6.4.10)$$

The conditional probability density function $f_{T_m/A,E_2}(t_m/\alpha,e_2)$ is an a posteriori density function. This is simply the conditional density of T_m given the value α, and expresses the degree of belief of the location of the value of t_m given the result of observing the malfunction monitor.

Equation (4.6.4.10) is used in place of (4.6.4.5) in equation (4.6.4.4) to give:

$$f_T(t;\alpha) = \int_{\substack{\text{range} \\ T_m}} f_{T/T_m,E_2}(t/t_m,e_2)\, f_{T_m/A,E_2}(t_m/\alpha,e_2)\, d\, t_m \quad (4.6.4.11)$$

Changing the variable $\theta = t - \tau$ and using the transformation theorem (28) yields:

$$f_\theta(\theta;\alpha) = \int_0^\tau f_{T/T_m,E_2}(\tau+\theta/t_m,e_2)\, f_{T_m/A,E_2}(t_m/\alpha,e_2)\, d\, t_m \quad (4.6.4.12)$$

The reliability function then becomes:

$$R_{oIII}(\theta) = 1 - \int_0^\theta f_\theta(\theta;\alpha)\, d\theta$$

$$= 1 - \int_0^{\theta} \int_0^{\tau} f_{T/T_m,E_2}(\tau + \theta/t_m, e_2)\, f_{T_m/A,E_2}(t_m/\alpha, e_2)\, d\, t_m\, d\theta$$

(4.6.4.13)

4.7 Illustrative example

4.7.1 Basic probability density functions

The expressions for the designer's and operator's reliability were given in equations (4.6.1.1), (4.6.2.5), (4.6.3.8) and (4.6.4.13). Their use will be illustrated, but first the probability density functions used in the previous material will be considered, and typical density functions chosen for the illustrative example.

The main probability density function used is the Weibull function which is (118):

$$f_X(x) = k\, x^m \exp\left(-\frac{k\, x^{m+1}}{m+1}\right) \qquad 0 < x \leq \infty \quad (4.7.1.1)$$

and $$\mu_X = \frac{\Gamma\left(\frac{1}{m+1}\right)}{m+1\left(\frac{k}{m+1}\right)^{1/m+1}}$$

This density function is very flexible and by appropriate choice of m and k can be made to approximate other standard probability density functions. When $m = 0$ the exponential distribution is obtained, and when $m = 1$ the distribution becomes the Rayleigh distribution. Thus the parameter m determines the shape of the distribution while k does not affect the shape but serves as a scaling factor.

4.7.1.1 Overall failure density function $f_T(t)$

It is assumed that the designer knows from an analysis of the equipment failure data the overall failure density function, which is represented in this case by the Weibull function:

$$f_T(t) = k_1 t^{m_1} \exp\left(\frac{-k_1 t^{m_1+1}}{m_1+1}\right) \qquad (4.7.1.1.1)$$

The corresponding hazard function is

$$z(t) = k_1 t^{m_1}$$

4.7.1.2 Failure density functions $f_{T/T_m,E_1}(t/t_m,e_1)$, $f_{\Theta/T_m,E_1}(\Theta/t_m,e_1)$

For the purpose of this analysis it is assumed that the operator knows the equipment terminal failure density function. This is determined on monitored equipment by analysing the data on failures when the time of failure is measured from the initial malfunction time t_m. In this example it is assumed that the terminal failure hazard rate is given by:

$$z(t) = k_2 + k_3 (t - t_m)^{m_2} \qquad t > t_m \qquad (4.7.1.2.1)$$

Generally the failure density function is related to the hazard rate according to (118):

$$f_T(t) = z(t) \exp\left(-\int_0^t z(t)\,dt\right) \qquad (4.7.1.2.2)$$

Thus using equation (4.7.1.2.1) in (4.7.1.2.2) the corresponding terminal failure density function is found as:

$$f_T(t) = 0 \qquad 0 < t \leq t_m \qquad (4.7.1.2.3a)$$

$$= (k_2 + k_3(t - t_m)^{m_2}) \exp\left\{-k_2(t - t_m) - \frac{k_3(t - t_m)^{m_2+1}}{m_2+1}\right\}$$

$$t_m < t \leq \infty \qquad (4.7.1.2.3b)$$

Setting $\Theta = t - \tau$ and noting $t_m = \tau$ then the transformation theorem yields the terminal failure density function

$$f_{\theta/T_m=\tau}(\theta/t_m=\tau) = (k_2 + k_3\,\theta^{m_2}) \exp\left\{ - k_2\theta - \frac{k_3\,\theta^{m_2+1}}{m_2+1} \right\}$$

$$0 < \theta \leq \infty \quad (4.7.1.2.4)$$

Now consider the event e_1, which is an observation at time τ with the monitor signal healthy.

The operator's equipment failure model is described as follows. He knows the equipment has not failed in the time interval $(0, \tau)$, and that it cannot fail until the initial malfunction has occurred. After t_m the equipment fails according to the terminal probability failure density function. Thus the hazard rate is:

$$z(t) = 0 \qquad 0 < t \leq \tau \quad (4.7.1.2.5a)$$

$$= 0 \qquad \tau < t \leq t_m \quad (4.7.1.2.5b)$$

$$= k_2 + k_3\,(t - t_m)^{m_2} \qquad t_m < t \leq \infty \quad (4.7.1.2.5c)$$

and the corresponding failure density function is:

$$f_T(t) = 0 \qquad 0 < t \leq \tau \quad (4.7.1.2.6a)$$

$$= 0 \qquad \tau < t \leq t_m \quad (4.7.1.2.6b)$$

$$= (k_2 + k_3\,(t - t_m)^{m_2}) \exp\left\{ - k_2(t - t_m) - \frac{k_3(t - t_m)^{m_2+1}}{m_2+1} \right\}$$

$$t_m < t \leq \infty \quad (4.7.1.2.6c)$$

Now the failure density function conditioned on the time t_m and on the event e_1 is determined using an equation corresponding to (4.6.3.7), i.e.

$$f_{T/T_m,E_1}(t/t_m,e_1) = \frac{f_T(t)}{1 - \int_0^{\tau} f_T(t)\,dt}$$

so

$$f_{T/T_m,E_1}(t/t_m,e_1) = 0 \qquad 0 < t \leq t_m \quad (4.7.1.2.7a)$$

$$= (k_2 + k_3(t - t_m)^{m_2}) \exp \left\{ - k_2(t - t_m) - \frac{k_3(t - t_m)^{m_2+1}}{m_2 + 1} \right\}$$

$$t_m < t \leq \infty \quad (4.7.1.2.7b)$$

Now setting $\theta = t - \tau$ and using the random variable transformation theorem (28):

$$f_{\theta/T_m,E_1}(\theta/t_m,e_1) = 0 \qquad \theta < 0 \qquad (4.7.1.2.8a)$$

$$= 0 \qquad 0 < \theta < t_m - \tau \qquad (4.7.1.2.8b)$$

$$= (k_2 + k_3(\tau + \theta - t_m)^{m_2}) \exp \left\{ - k_2(\tau + \theta - t_m) - \frac{k_3(\tau + \theta - t_m)^{m_2+1}}{m_2 + 1} \right\}$$

$$t_m - \tau < \theta \leq \infty \quad (4.7.1.2.8c)$$

4.7.1.3 Failure density functions $f_{T/T_m,E_2}(t/t_m,e_2)$ $f_{\theta/T_m,E_2}(\theta/t_m,e_2)$

The development of these density functions follows the same procedure as above. The event e_2 is a monitor inspection at time τ resulting in an unhealthy signal.

The operator thus formulates a failure model which informs him that the equipment has not failed up to time τ, that the initial malfunction has occurred at some time $0 < t_m \leq \tau$ and that the equipment will subsequently fail according to the terminal probability failure density function. For this model the hazard rate becomes:

$$z(t) = 0 \qquad 0 < t \leq t_m \qquad (4.7.1.3.1a)$$

$$= k_2 + k_3(t - t_m)^{m_2} \qquad t_m < t \leq \tau \qquad (4.7.1.3.1b)$$

$$= k_2 + k_3(t - t_m)^{m_2} \qquad \tau < t \leq \infty \qquad (4.7.1.3.1c)$$

The corresponding failure density function is:

$$f_T(t) = 0 \qquad 0 < t \leq t_m \qquad (4.7.1.3.2a)$$

$$= (k_2 + k_3(t - t_m)^{m_2}) \exp\left\{ - k_2(t - t_m) - \frac{k_3(t - t_m)^{m_2+1}}{m_2 + 1} \right\}$$

$$t_m < t \leq \tau \qquad (4.7.1.3.2b)$$

$$= (k_2 + k_3(t - t_m)^{m_2}) \exp\left\{ -k_2(t - t_m) - \frac{k_3(t - t_m)^{m_2+1}}{m_2 + 1} \right\}$$

$$\tau < t \leq \infty \qquad (4.7.1.3.2c)$$

Now conditioning the failure density function on t_m and event e_2 gives

$$f_{T/T_m, E_2}(t/t_m, e_2) = \frac{f_T(t)}{1 - \int_0^{\tau} f_T(t)\, dt}$$

$$= 0 \qquad 0 < t \leq \tau \qquad (4.7.1.3.3a)$$

$$= \frac{(k_2 + k_3(t - t_m)^{m_2}) \exp\left\{ - k_2(t - t_m) - \frac{k_3(t-t_m)^{m_2+1}}{m_2 + 1} \right\}}{\exp\left\{ - k_2(\tau - t_m) - \frac{k_3(\tau - t_m)^{m_2+1}}{m_2 + 1} \right\}}$$

$$\tau < t \leq \infty \qquad (4.7.1.3.3b)$$

The variable is changed according to $\theta = t - \tau$ which yields (28):

$$f_{\theta/T_m, E_2}(\theta/t_m, e_2) = 0 \qquad \theta \leq 0 \qquad (4.7.1.3.4a)$$

$$= \frac{(k_2 + k_3(\tau + \theta - t_m)^{m_2}) \exp\left\{ -k_2\theta - \frac{k_3(\tau + \theta - t_m)^{m_2+1}}{m_2 + 1} \right\}}{\exp\left\{ - \frac{k_3(\tau - t_m)^{m_2+1}}{m_2 + 1} \right\}}$$

$$0 < \theta \leq \infty \qquad (4.7.1.3.4b)$$

4.7.1.4 Summary of failure density functions

Sketches of the shapes of typical failure density functions for the four reliability regimes are shown in Figure 4.5. Figure 4.5a shows the designer's failure density function calculated from equation (4.7.1.1.1) with values $m_1 = 0$ and 3. Figure 4.5b gives the operator's failure density function $f_{T/T\geq\tau}(t/t\geq\tau)$, corresponding to operator's reliability II, which is calculated from equations (4.6.2.3) and (4.7.1.1.1) This function is similar to Figure 4.5a except for the modification due to τ .

The failure density function $f_{T/T_m,E_1}(t/t_m,e_1)$ for a particular value of t_m envisaged by the operator, corresponding to operator's reliability II, is shown in Figure 4.5c. This is calculated from equation (4.7.1.2.3) with an assumed value $m_2 = 3$.

Figure 4.5d shows $f_{T/T_m,E_2}(t/t_m,e_2)$ for operator's reliability III, which again involves a modification by truncation at the observation time τ . The density is determined from equation (4.7.1.3.3) with the same value of m_2 as above, and t_m is a particular value envisaged by the operator for the purpose of the sketch.

Figure 4.6 shows the corresponding hazard rates.

4.7.1.5 Density functions $f_{T_m}(t_m)$, $f_{T_m/E_1}(t_m/e_1)$, $f_{T_m/E_2}(t_m/e_2)$

The analysis of section 4.6 assumed that the operator knew the overall probability density function for failure when time was measured from the beginning of the equipment's life ($t = 0$), $f_T(t)$, and the terminal failure density when the failure time was measured from the time of the initial malfunction ($t = t_m$, $\theta = 0$), $f_{\theta/T_m=\tau}(\theta/t_m = \tau)$.

However, there is a third probability density function which is

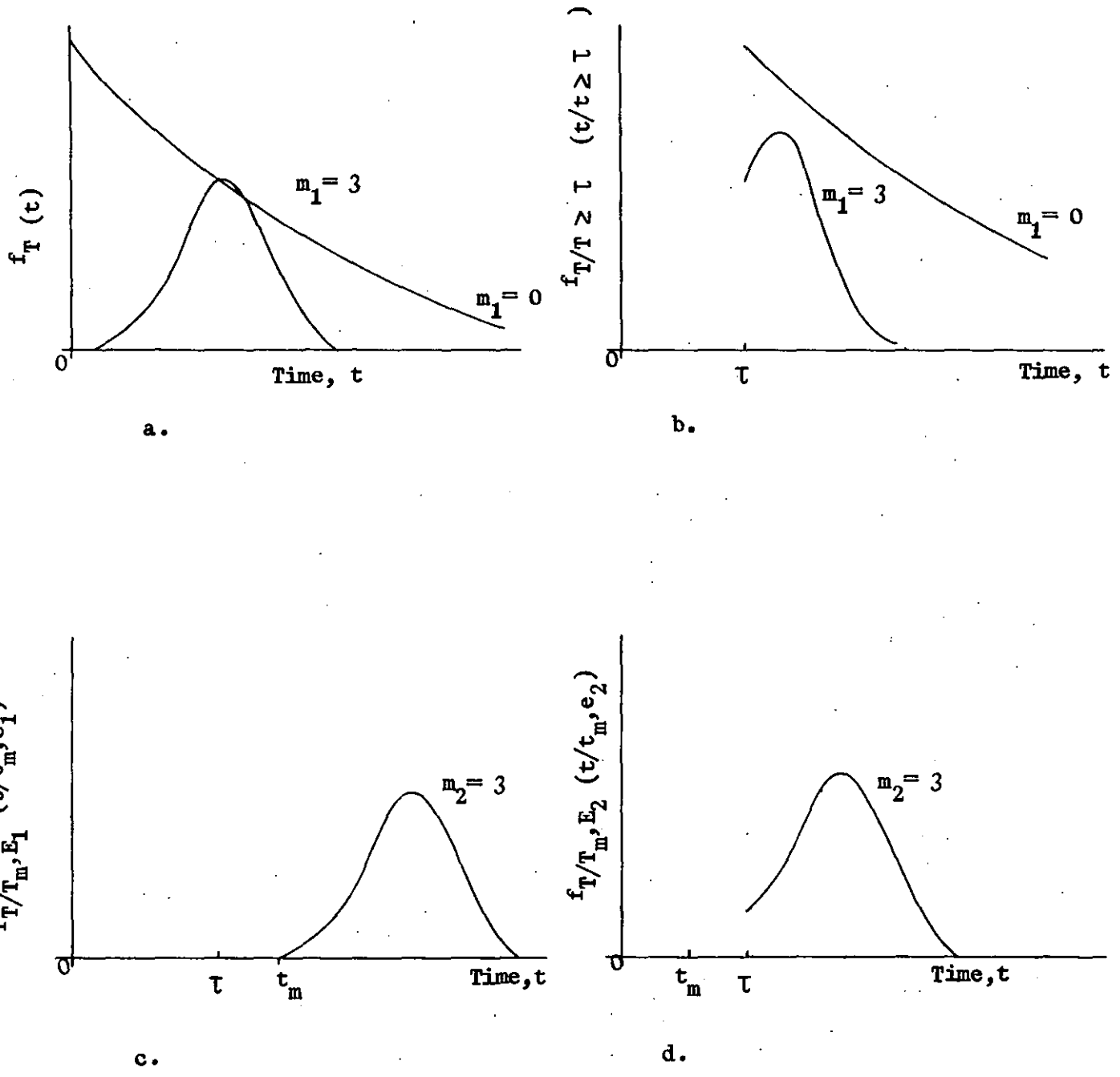
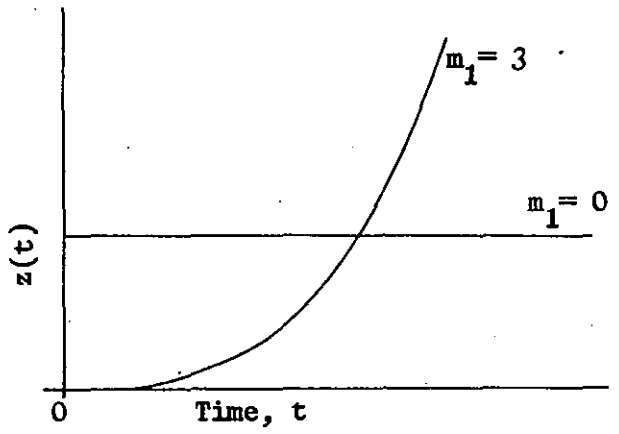
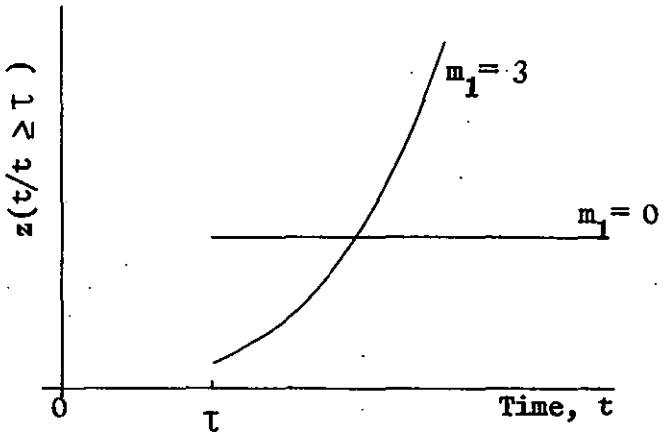


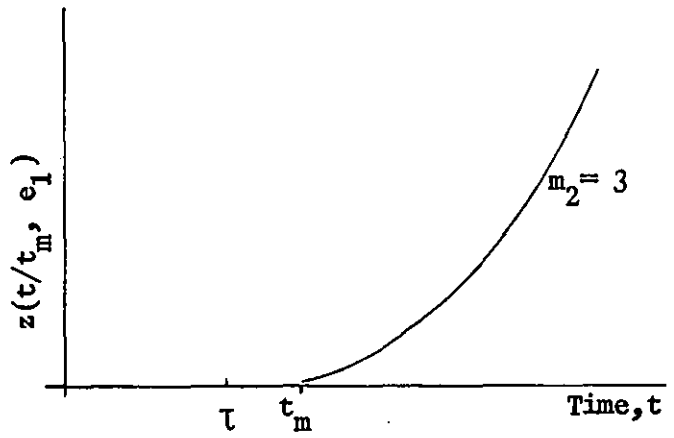
FIGURE 4.5 Typical failure density functions.



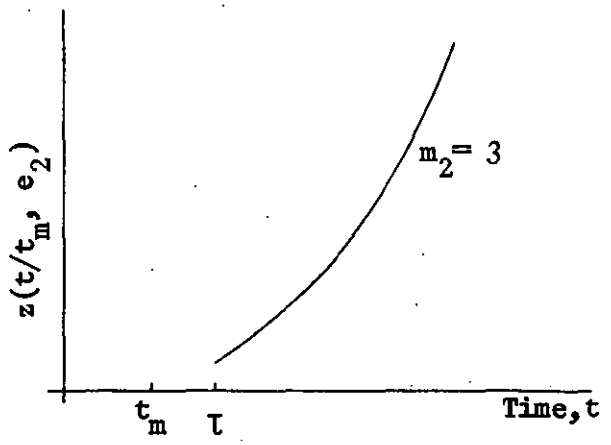
a.



b.



c.



d.

FIGURE 4.6 Typical hazard rates.

needed. This is the density function describing the time of initial malfunction, $f_{T_m}(t_m)$, when time is measured from the beginning of the equipment life ($t = 0$).

In practice this may be available from an analysis of the experimental data, however, it is assumed that this is unlikely and therefore must be derived.

Suppose that the observation time τ is made equal to the initial malfunction time t_m, then by definition

$$\theta = t - t_m \qquad (4.7.1.5.1)$$

or

$$t_m = t - \theta \qquad (4.7.1.5.2)$$

Now $f_T(t)$ and $f_{\theta/T_m=\tau}$ $(\theta/t_m = \tau)$ are given and so $f_{T_m}(t_m)$ may be determined from the transformation theorem as (28):

$$f_{T_m}(t_m) = \int_{\text{range } T} f_T(t)\ f_{\theta/T_m=\tau}\ (t - t_m/t_m = \tau)dt \qquad (4.7.1.5.3)$$

The range of t and θ are both 0 to ∞ and so examination of equation (4.7.1.5.2) reveals that in theory negative values of t_m are possible. However, this contradicts intuition since it is impossible for the equipment to malfunction before it has started to operate, i.e. t_m cannot be less than 0, although t_m can equal 0.

Now for $t_m \geq 0$ equation (4.7.1.5.3) becomes

$$f_{T_m}(t_m) = \int_{t_m}^{\infty} f_T(t)\ f_{\theta/T_m=\tau}(t - t_m/t_m = \tau)\ dt \qquad (4.7.1.5.4)$$

and for $t_m < 0$, theoretically

$$f_{T_m}(t_m) = \int_{0}^{\infty} f_T(t)\ f_{\theta/T_m=\tau}(t - t_m/t_m = \tau)\ dt \qquad (4.7.1.5.5)$$

Since t_m cannot be less than 0, then a density function is formed where the probability of obtaining $t_m < 0$ is located, in the form of an impulse, at the origin, i.e.,

$$P(t_m<0) = \int_{-\infty}^{0} \int_{0}^{\infty} f_T(t)\, f_{\Theta/T_m=\tau}\,(t - t_m/t_m=\tau)\, dt dt_m \qquad (4.7.1.5.6a)$$

$$= K\ \delta\ (t_m = 0) \qquad (4.7.1.5.6b)$$

The resulting probability density function $f_{T_m}(t_m)$ is therefore a mixture of both an impulse at the origin and a continuous function, which is given by substituting equations (4.7.1.1.1) and (4.7.1.2.4) into (4.7.1.5.4) and (4.7.1.5.6b) respectively, i.e.,

$$f_{T_m}(t_m) = \int_{t_m}^{\infty} k_1 t^{m_1} \exp\left\{\frac{-k_1 t^{m_1+1}}{m_1+1}\right\} \left((k_2 + k_3(t - t_m)^{m_2}) x \right.$$
$$\left. \exp\left\{-k_2(t - t_m) - \frac{k_3(t - t_m)^{m_2+1}}{m_2+1}\right\}\right) dt$$
$$0< t_m \leq \infty \qquad (4.7.1.5.7a)$$

$$P(t_m<0) = \int_{-\infty}^{0} \int_{0}^{\infty} k_1 t^{m_1} \exp\left\{\frac{-k_1 t^{m_1+1}}{m_1+1}\right\} \left((k_2 + k_3(t - t_m)^{m_2}) x\right.$$
$$\left. \exp\left\{-k_2(t - t_m) - \frac{k_3(t - t_m)^{m_2+1}}{m_2+1}\right\}\right) dt\, dt_m$$

$$= K\ \delta\ (t_m = 0) \qquad (4.7.1.5.7b)$$

Then the failure density functions conditioned upon the events e_1 and e_2 are given by equations (4.6.3.7) and (4.6.4.5) respectively, i.e.:

$$f_{T_m/E_1}(t_m/e_1) = \frac{f_{T_m}(t_m)}{1 - \int_0^{\tau} f_{T_m}(t_m)\, dt_m} \qquad (4.7.1.5.8)$$

$$f_{T_m/E_2}(t_m/e_2) = \frac{f_{T_m}(t_m)}{\int_0^{\tau} f_{T_m}(t_m)\, dt_m} \qquad (4.7.1.5.9)$$

4.7.1.6 Density functions $f_{A/T_m,E_2}(\alpha/t_m,e_2)$, $f_{T_m/A,E_2}(t_m/\alpha, e_2)$

The density function $f_{A/T_m,E_2}(\alpha/t_m,e_2)$ described the location of A, given a particular observation time τ and initial malfunction time $T_m = t_m$. This density is dependent upon the characteristics of the malfunction monitor and the equipment failure characteristic as shown in Figure 4.4.

In this study it is assumed that $f_{A/T_m,E_2}(\alpha/t_m,e_2)$ is given by a Weibull-type distribution. However, in this case the monitor variable is normalised to be in the range $0 \leq \alpha \leq 1$ and so the Weibull density function of equation (4.7.1.1) is transformed using the relation (28):

$$\alpha = \frac{x}{x+1}$$

Hence

$$f_{A/T_m,E_2}(\alpha/t_m,e_2) = \frac{k_4}{(1-\alpha)^2}\left(\frac{\alpha}{\alpha-1}\right)^{m_3} \exp\left\{\frac{-k_4}{m_3+1}\left(\frac{\alpha}{\alpha-1}\right)^{m_3+1}\right\}$$

$$0 \leq \alpha \leq 1 \qquad (4.7.1.6.1)$$

The a posteriori conditional probability density function for the time to initial malfunction $f_{T_m/A,E_2}(t_m/\alpha, e_2)$ is calculated from equation (4.6.4.10) using (4.7.1.5.9) and (4.7.1.6.1).

4.7.2 Reliability estimates for a single equipment

The probability density functions described in section 4.7.1 are now used to estimate single equipment reliability for the four regimes discussed earlier.

The parameters used in this example are given in Table 4.3.

i	k_i	m_i
1	0.04	0
2	0.02	1
3	0.1615	1
4	$(\tau - t_m)$	-
μ_T	25	
$\mu_{\theta/T_m=\tau}$	3	
μ_{T_m}	21.8	
τ	25	
α	0.25; 0.75	

TABLE 4.3. Numerical values of the parameters used in the illustrative example.

4.7.2.1 A priori density functions

The overall equipment failure characteristic $f_T(t)$, which the designer knows, is the exponential density function ($m_1 = 0$) with a mean $\mu_T = 25$. This is shown in Figure 4.7.

To determine the operator's reliabilities, it is assumed that the operator has a priori information in the form of experimentally determined probability density functions $f_{\theta/T_m=\tau}(\theta/t_m=\tau)$, $f_{A/T_m,E_2}(\alpha/t_m,e_2)$ and the calculated function $f_{T_m}(t_m)$. These were given in equations

(4.7.1.2.4), (4.7.1.6.1) and (4.7.1.5.7) and using the parameters of Table 4.3 are shown in Figures 4.8, 4.9 and 4.10.

Figure 4.9 was derived by assuming $k_4 = (\tau - t_m)$, i.e. the scaling factor increases linearly as the difference between the observation time τ and the time of initial malfunction detection t_m. Normally it is expected that in practice this density function would be available from an experimental analysis of the malfunction monitor signal characteristics.

Notice in Figure 4.10 that $f_{T_m}(t_m)$ is a discontinuous function consisting of an impulse at the origin and a continuous function.

The ratio of the means of the total equipment life (μ_T) to the unhealthy period ($\mu_{\theta/T_m=\tau}$) is 25 : 3.

4.7.2.2 Designer's reliability

The designer's reliability is simply calculated from the overall equipment failure probability density function using equation (4.6.1.1) and is actually given by equation (4.6.1.5) with $k = k_1$ and $m = m_1$. This reliability is shown in Figure 4.11.

4.7.2.3 Operator's reliability I

The operator's failure density function I $f_{\theta/T\geq\tau}(\theta/t\geq\tau)$ and reliability I are calculated from equations (4.6.2.4) and (4.6.2.5) respectively and are shown in Figures 4.12a and b.

Since the fundamental overall equipment failure characteristic is exponential, then observation at time τ contributes nothing to the assessment of reliability and the operator's reliability I is the same as the designer's except the time scale is different. This is a basic

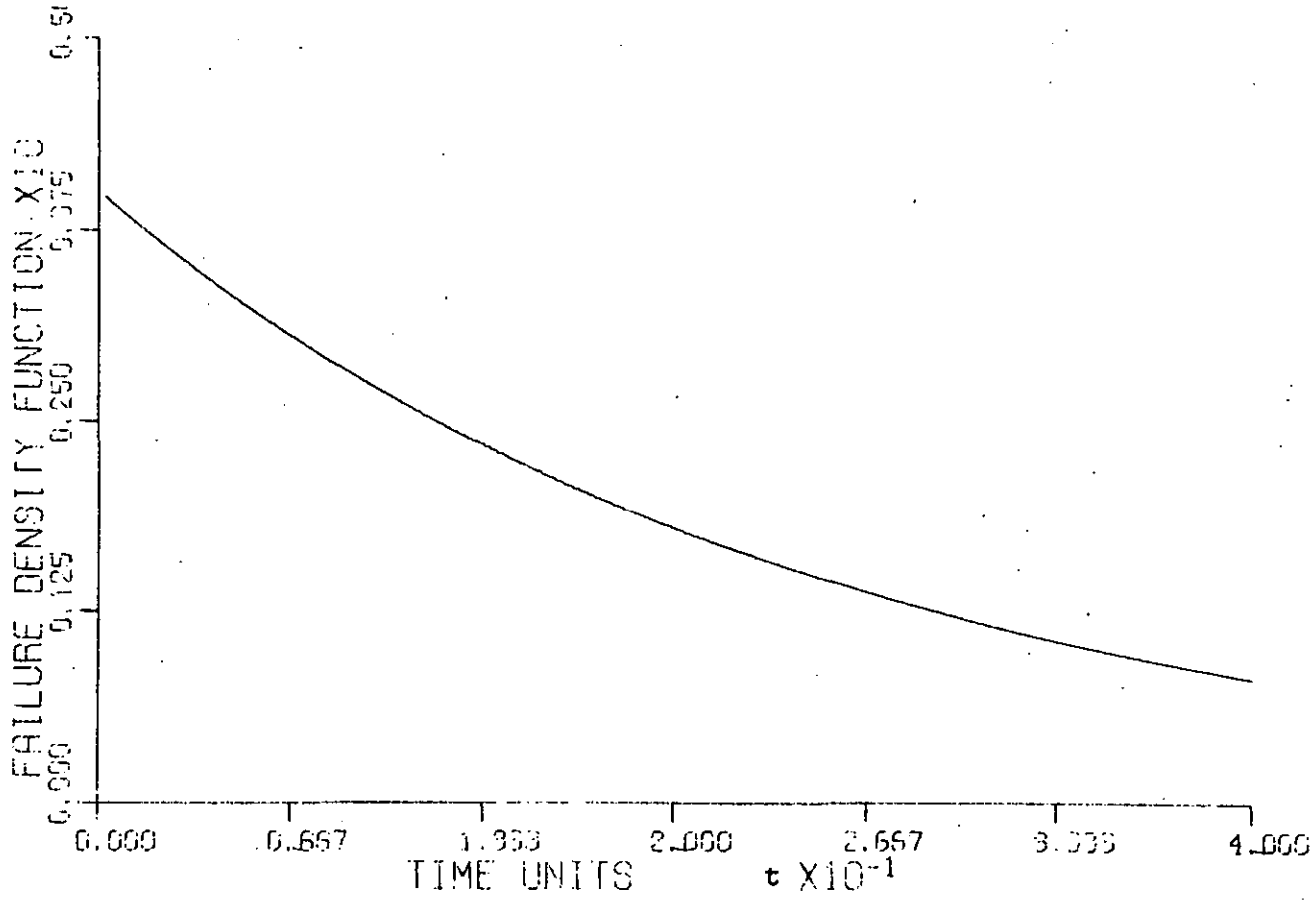


FIGURE 4.7 Illustrative example: A priori failure density function $f_T(t)$

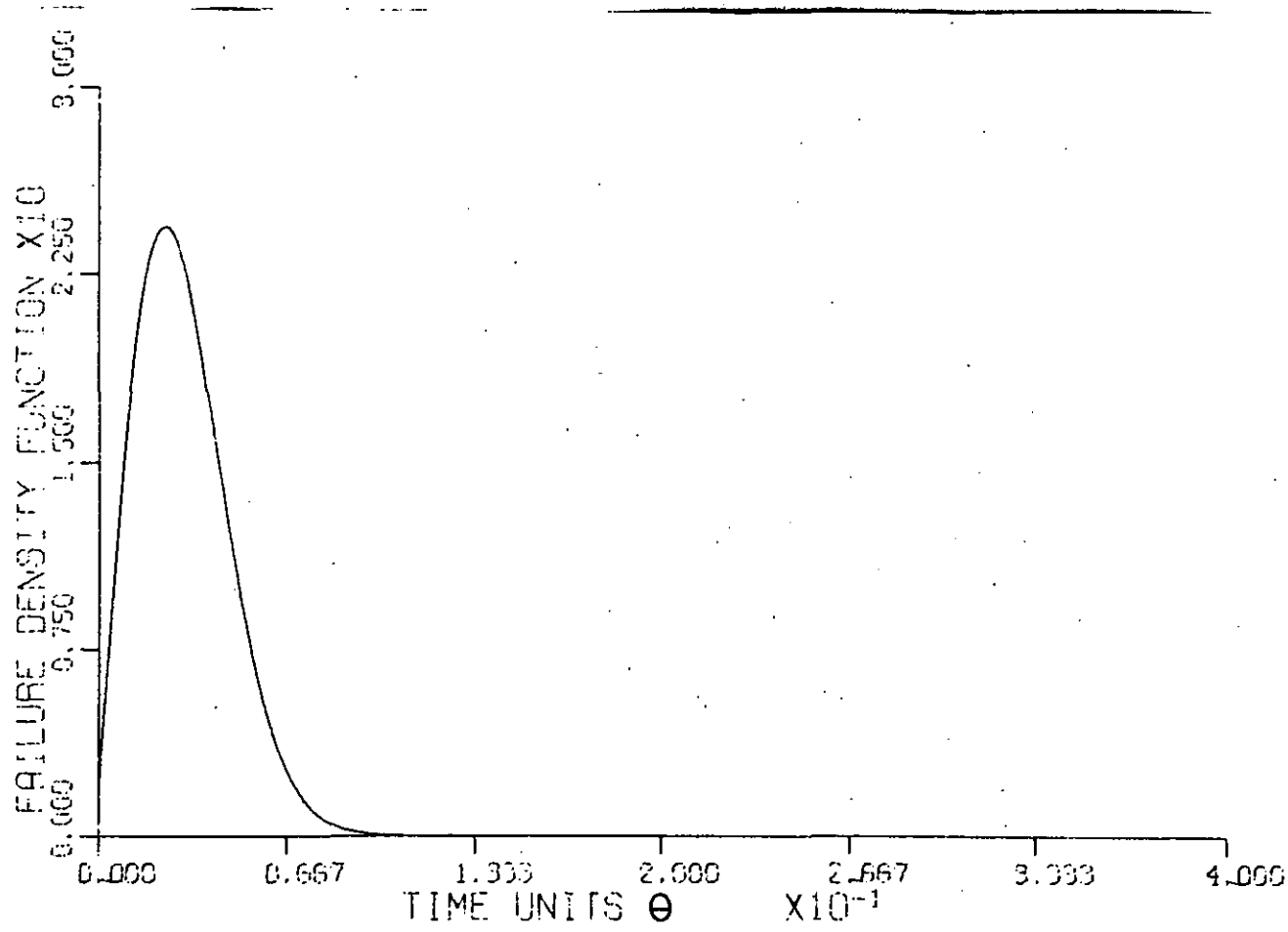


FIGURE 4.8 Illustrative example: A priori terminal failure density function

$$f_{\Theta/T_m=\tau}(\Theta/t_m = \tau)$$

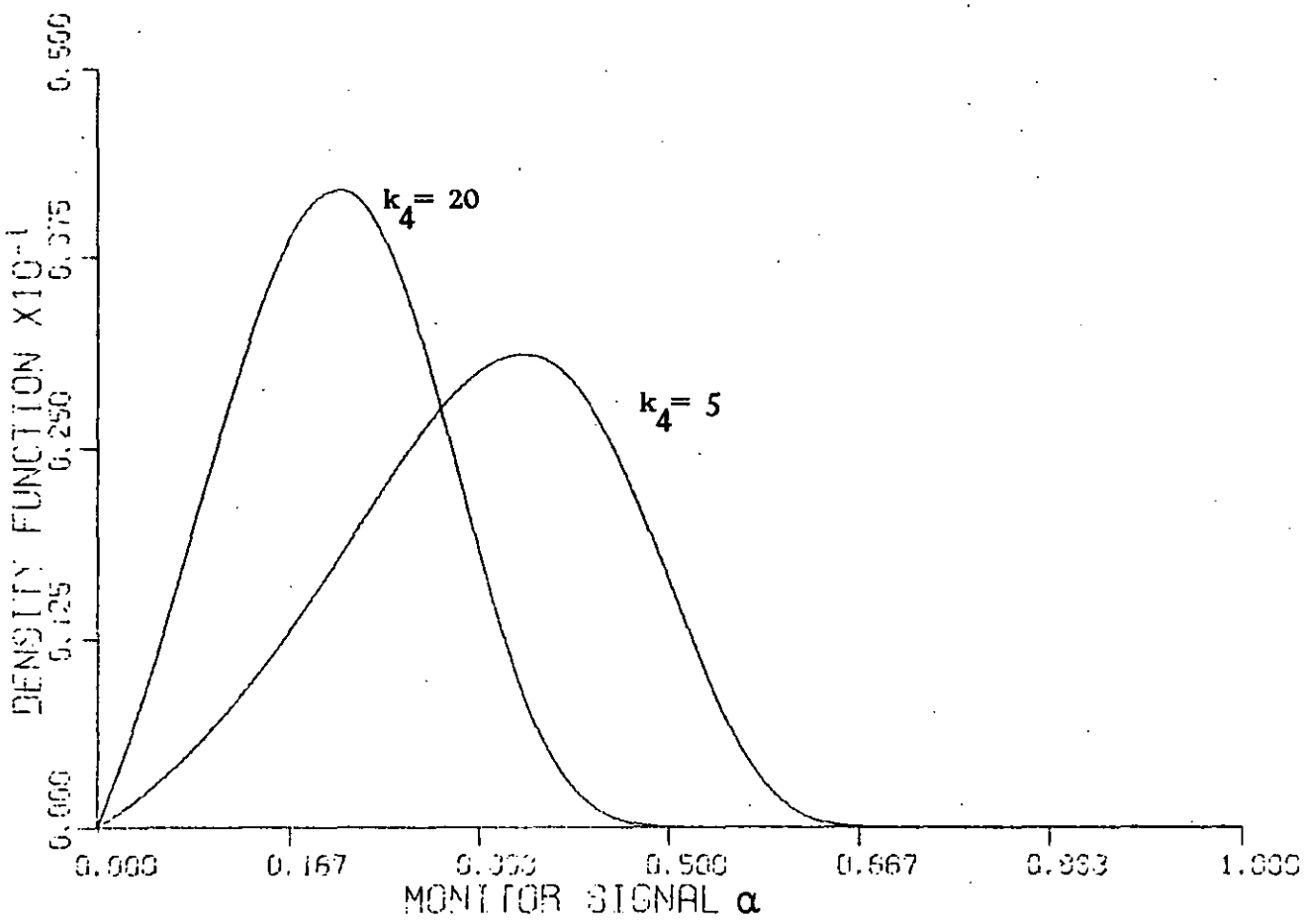


FIGURE 4.9 Illustrative example: A priori density function $f_{A/T_m,E_2}(\alpha/t_m,e_2)$

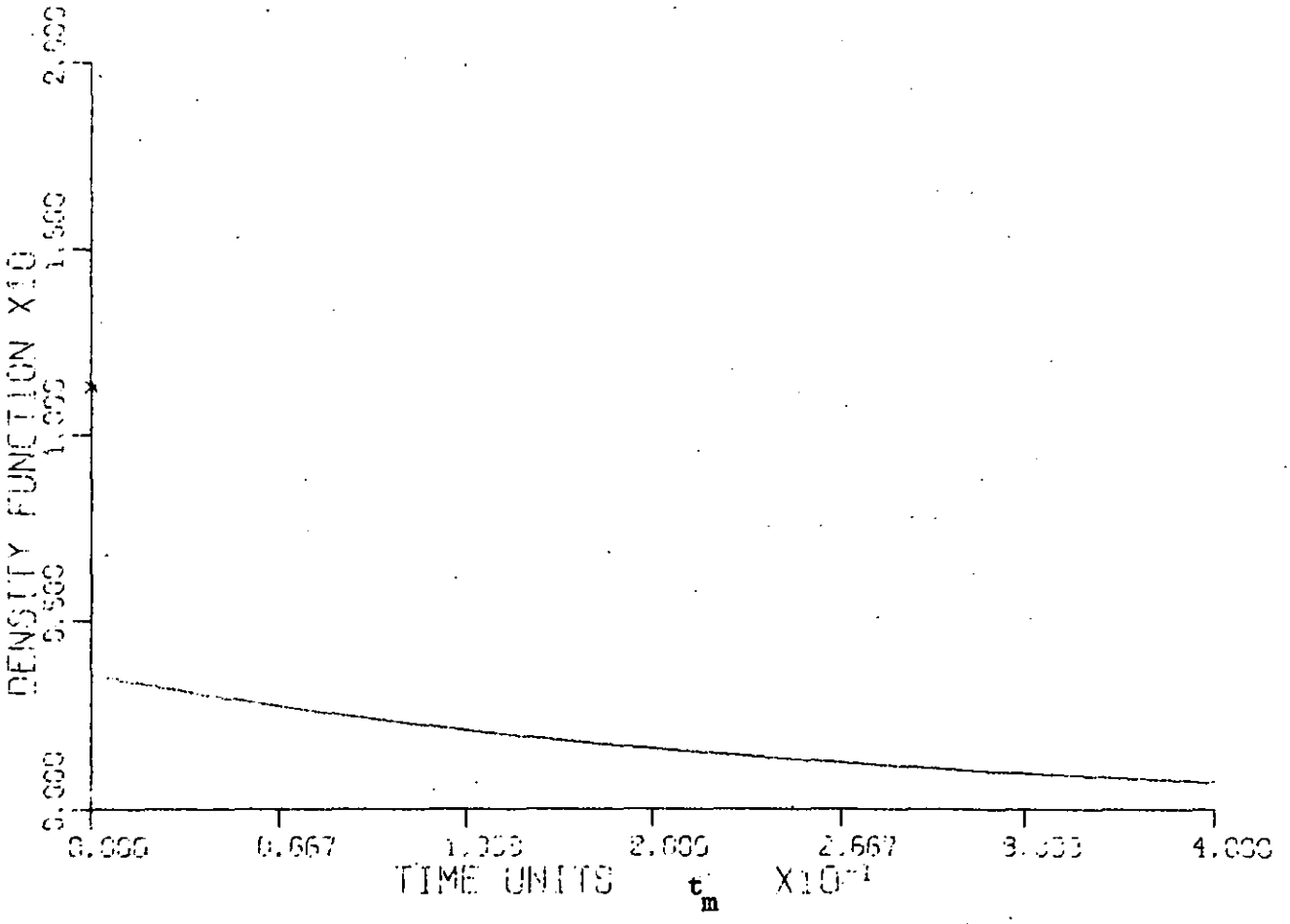


FIGURE 4.10 Illustrative example: A priori density function $f_{T_m}(t_m)$

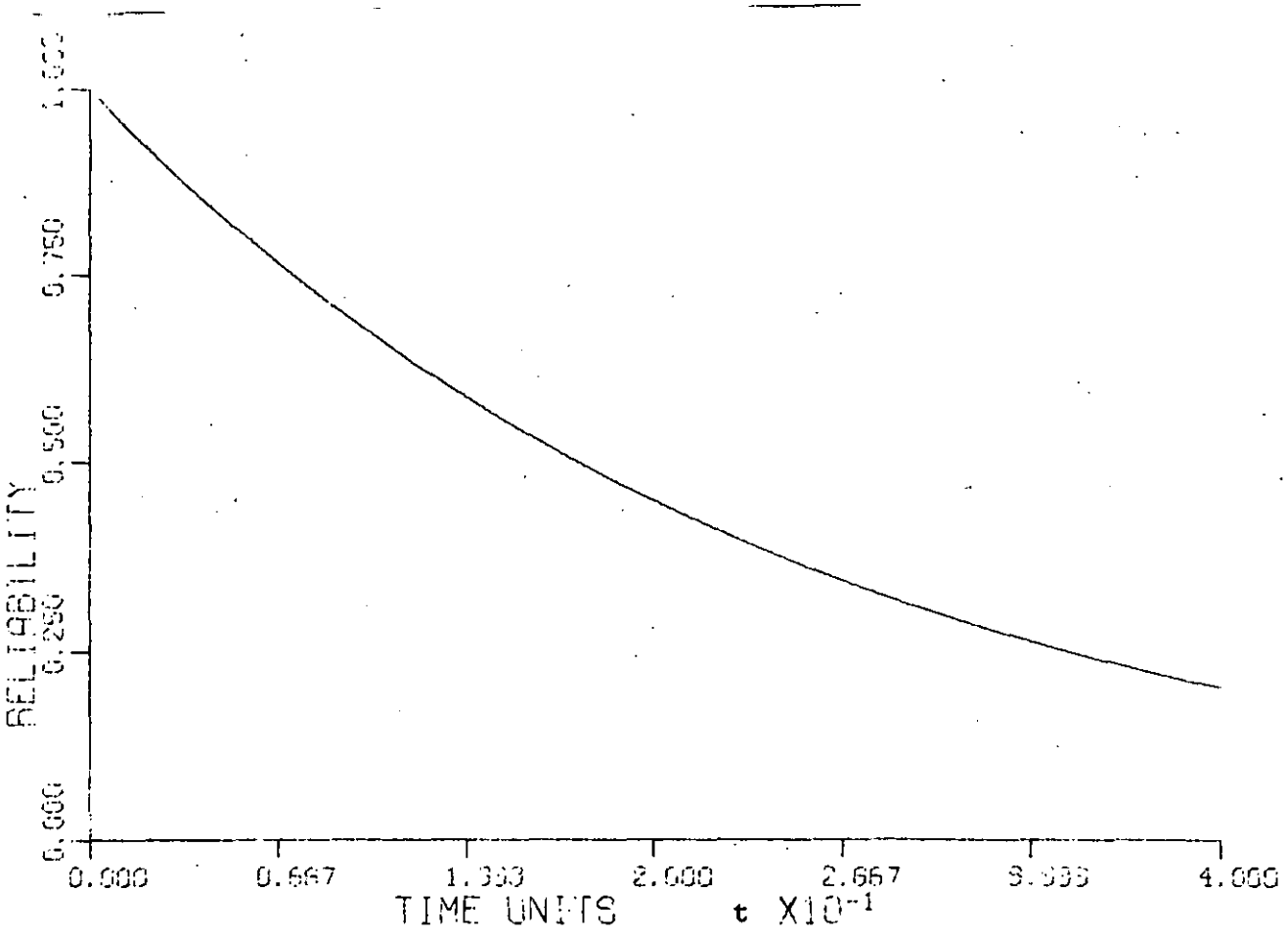
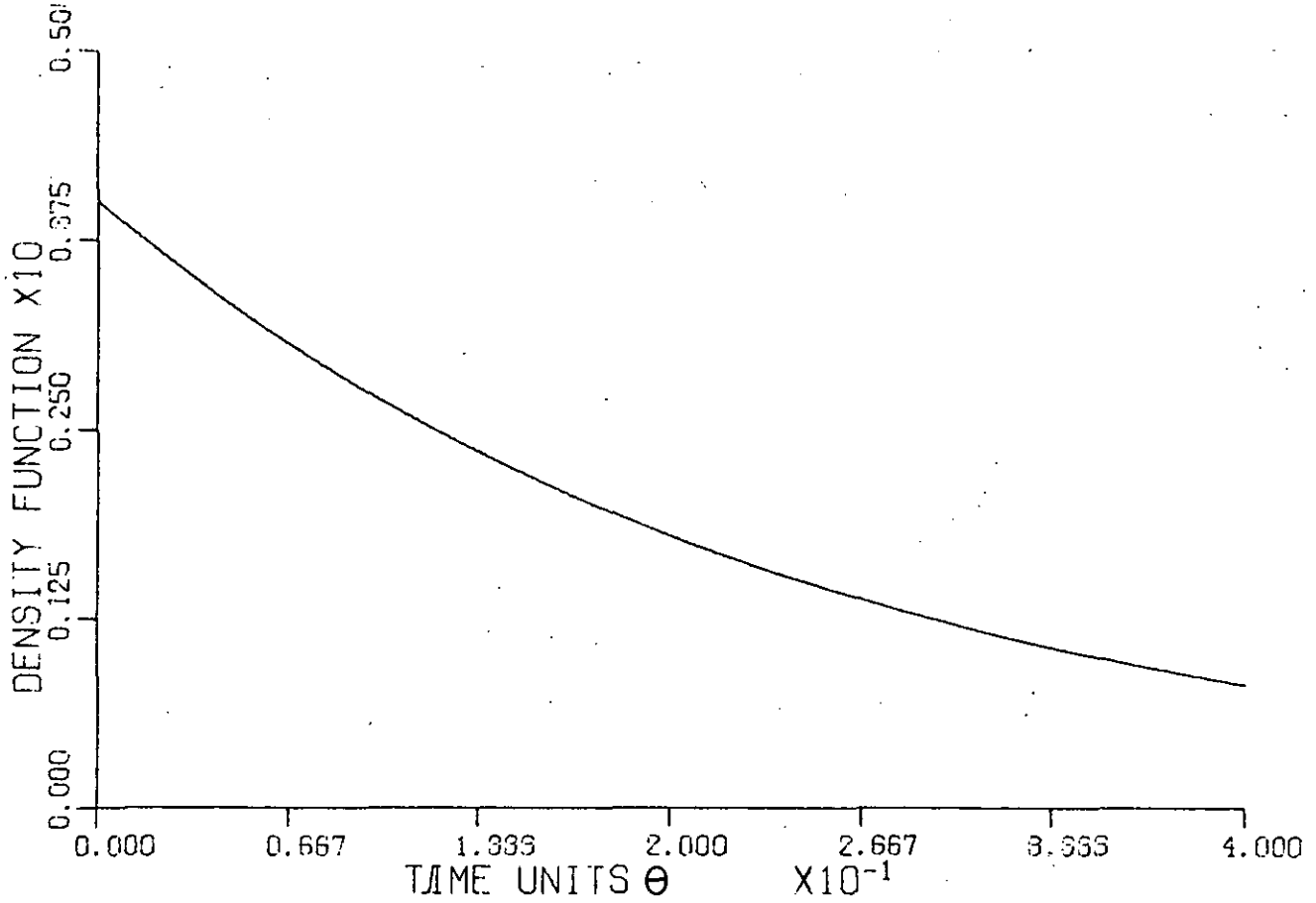
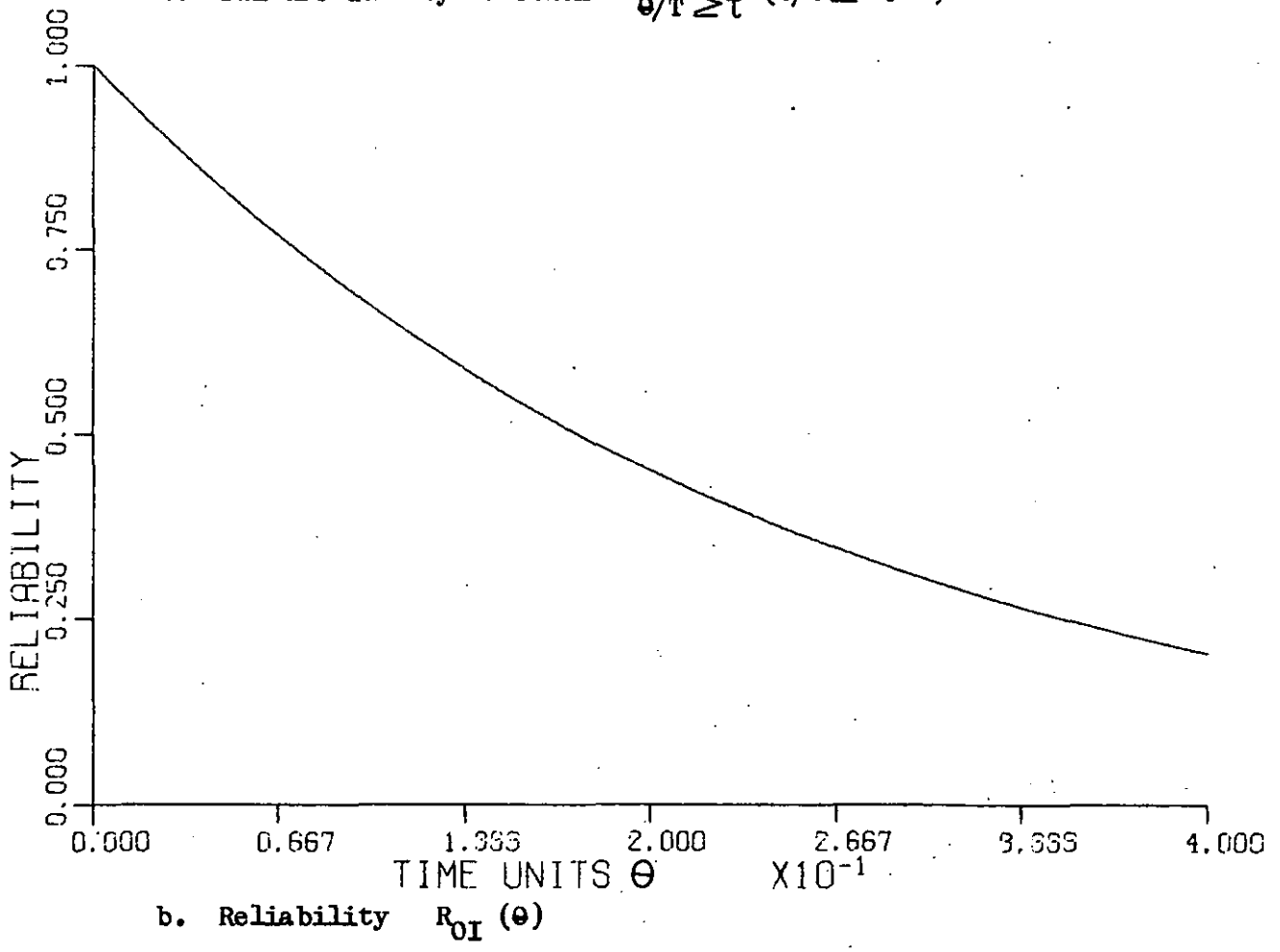


FIGURE 4.11 Illustrative example: Designer's reliability function, $R_D(t)$



a. Failure density function $f_{\theta/T \geq \tau}$ $(\theta/t \geq \tau)$



b. Reliability R_{0I} (θ)

FIGURE 4.12 Illustrative example: Operator's failure density function and reliability I

property of the exponential probability distribution (118).

4.7.2.4 Operator's reliability II

This failure density function is calculated from equation (4.6.3.5). The density functions $f_{\Theta/T_m, E_1}(\Theta/t_m, e_1)$ and $f_{T_m/E_1}(t_m/e_1)$ which are required for this integration were calculated from equations (4.7.1.2.8) and (4.7.1.5.8) respectively.

The operators reliability II $R_{oII}(\Theta)$ was subsequently determined by equation (4.6.3.8).

This calculation procedure is summarised as a flow diagram in Figure 4.13. The integration limits of equation (4.6.3.5) are changed from τ to ∞ to τ to $\tau + \Theta$ in Figure 4.13, which may be verified by inspection from equations (4.7.1.2.8).

Figure 4.14 shows the density function $f_{T_m/E_1}(t_m/e_1)$ while the resulting operator's failure density function $f_{\Theta}(\Theta)$ and his reliability II, $R_{oII}(\Theta)$ are shown in Figures 4.15a and b.

Figure 4.15b also shows the operator's reliability I and reliability III if the time of initial malfunction t_m is the current observation time τ (which is calculated by integrating equation (4.7.1.2.4)). This Figure illustrates that operator's reliability II is significantly higher than operator's reliability I in the immediate future (i.e. Θ small). This is the expected result since there is a malfunction monitor which is showing a healthy signal and therefore there is a greater probability that equipment failure will not occur in the immediate future. Operator's reliability III is much lower than operator's reliability I or II. Again this is the expected result since the monitor signal has changed from

Equation Number

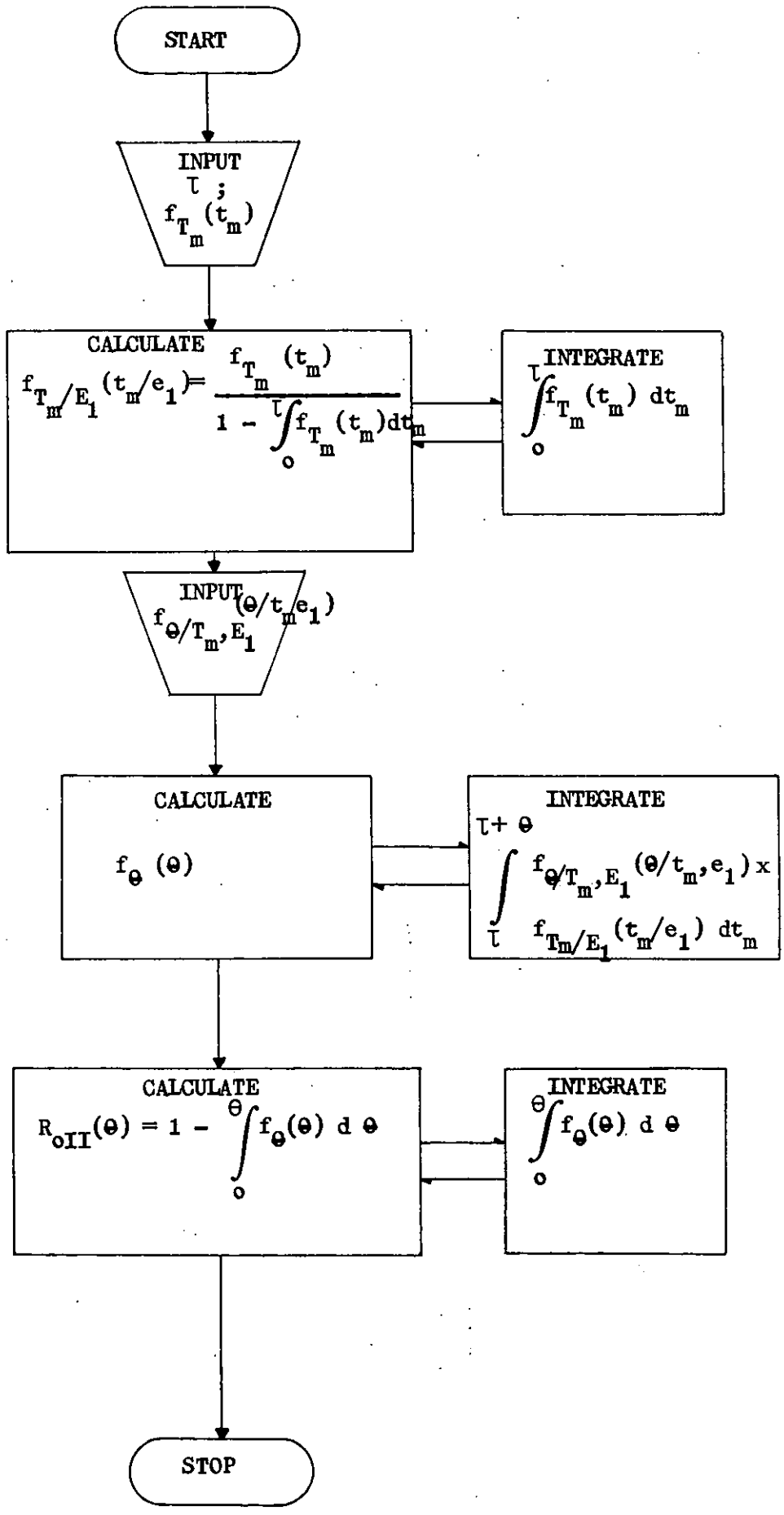


FIGURE 4.13 Flow diagram of operator's reliability II calculation.

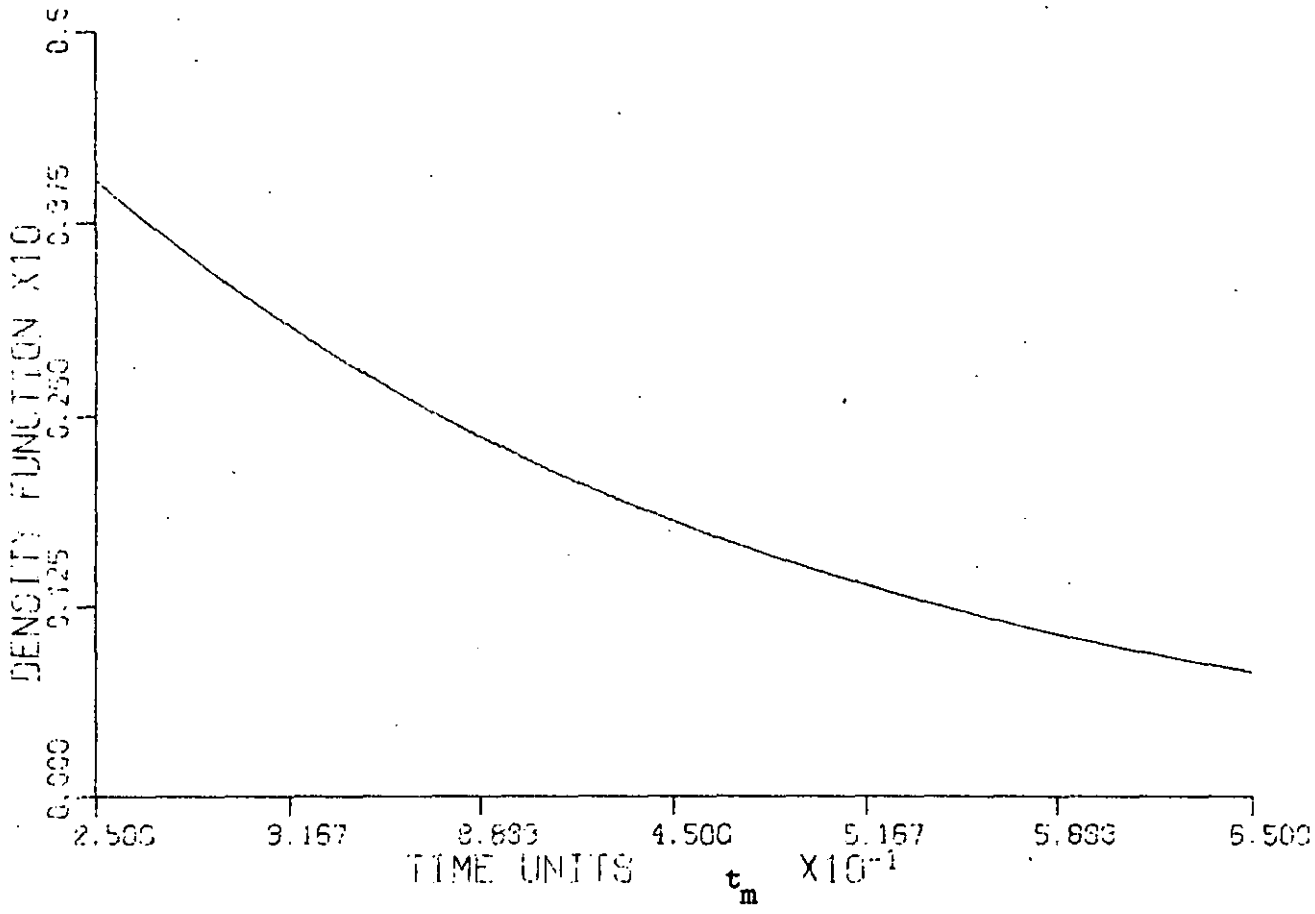
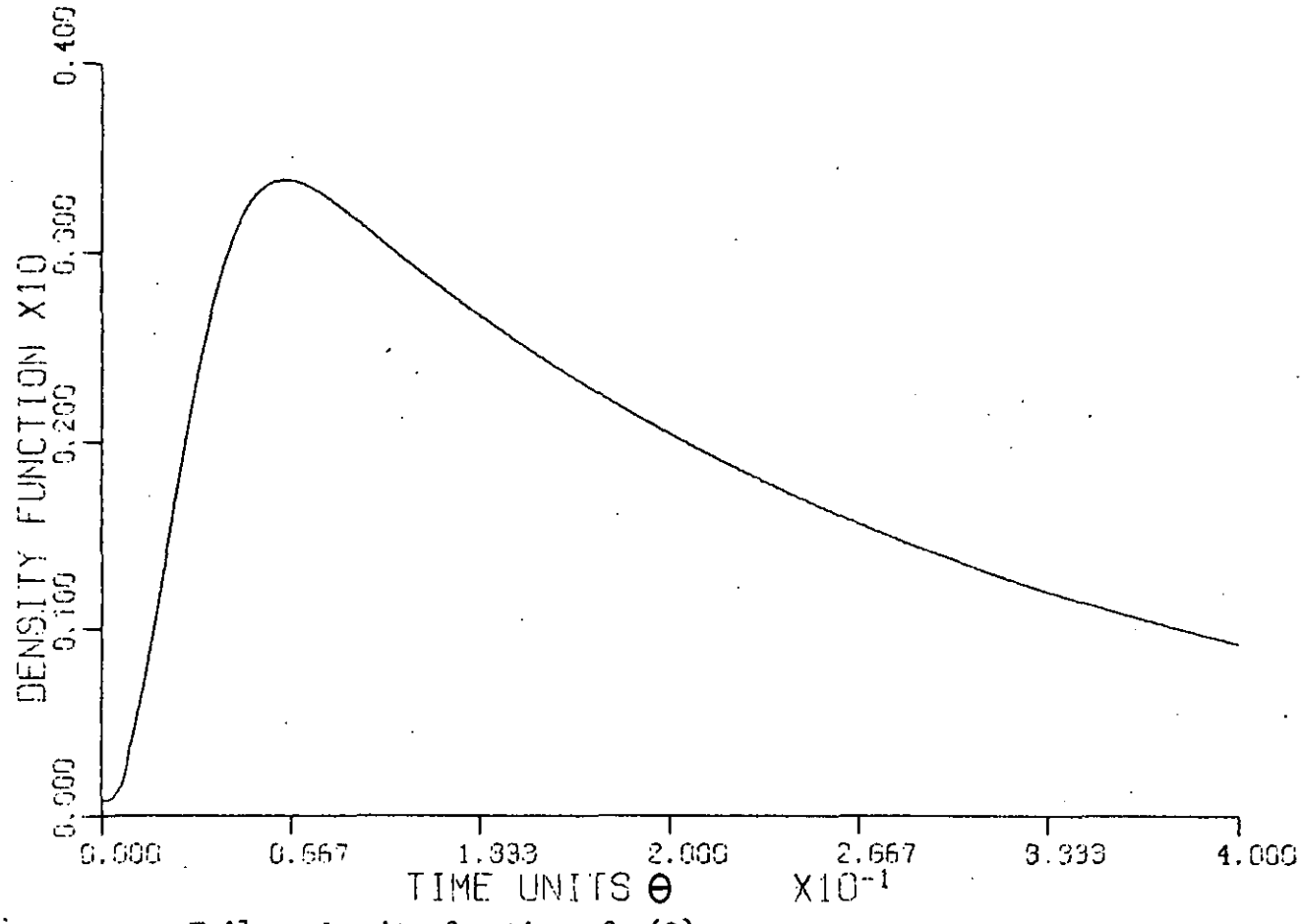
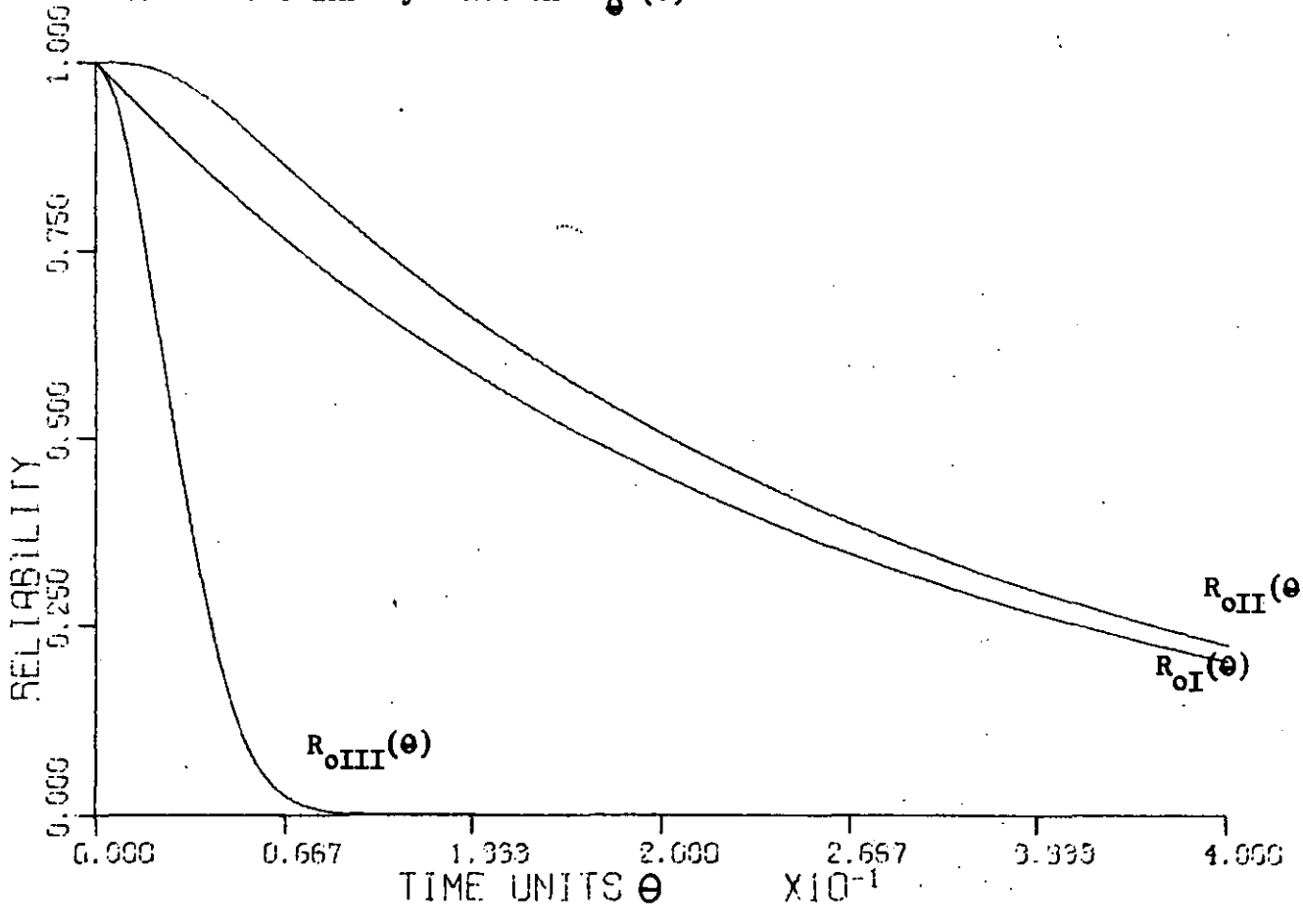


FIGURE 4.14 Illustrative example: Density function $f_{T_m/E_1}(t_m/e_1)$



a. Failure density function $f_{\theta}(\theta)$



b. Reliability $R_{oII}(\theta)$

FIGURE 4.15 Illustrative example: Operator's failure density function and reliability II.

healthy to unhealthy thereby indicating that the equipment has started to wear out and so there is a much greater probability that the equipment will fail in the near future.

4.7.2.5 Operator's reliability III

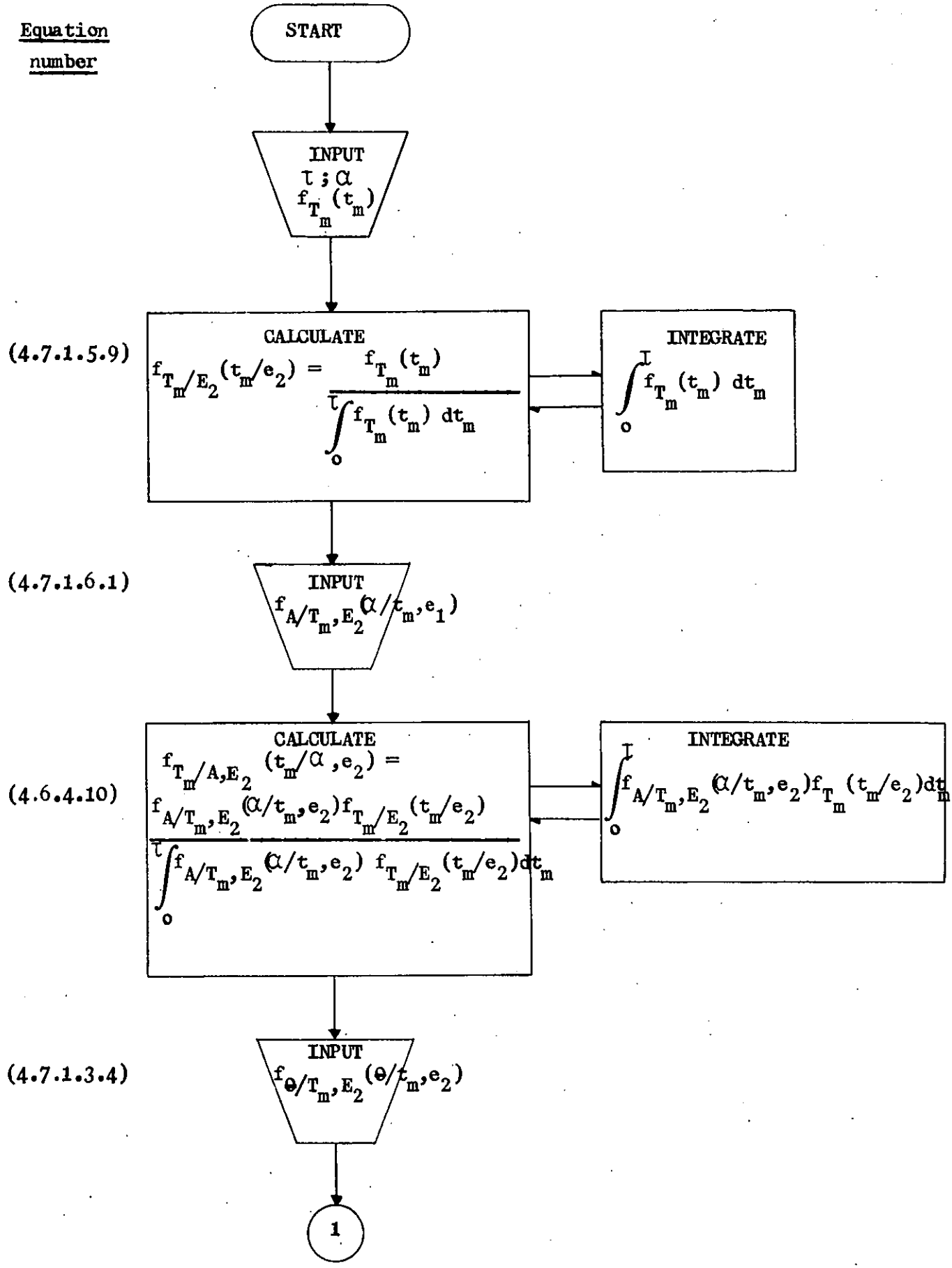
The calculation procedure is outlined in Figure 4.16.

The a priori information input was described above. This first stage in the calculation is the determination of $f_{T_m/E_2}(t_m/e_2)$ which is shown in Figure 4.17 for $\tau = 25$. The a posteriori density function $f_{T_m/A,E_2}(t_m/\alpha, E_2)$ is then calculated from equation (4.6.4.10) and is shown in Figures 4.18a and b for $\alpha = 0.25$ and 0.75.

The operator's failure density function III, $f_\theta(\theta;\alpha)$ and his reliability function $R_{oIII}(\theta)$ are then determined from equations (4.6.4.12) and (4.6.4.13), and are illustrated in Figures 4.19a and b, and Figure 4.20 respectively. Curves are shown for the same parameter values as in Figure 4.18.

Also shown in Figure 4.20 for comparison are the operator's reliability I and reliability III if the time of initial malfunction is the current observation time. These additional curves were also illustrated in Figure 4.15b.

The same comments apply to the comparison of the reliability curves in Figure 4.20 as were made in section 4.7.2.4. In particular the operator's reliability III estimate when $\alpha = 0.25$ is lower than that for $\alpha = 0.75$. This is expected since a reading $\alpha = 0.25$ means that the initial equipment malfunction occurred further in the past from τ than that indicated by $\alpha = 0.75$. Therefore the equipment has been "wearing-out" longer and so has a higher probability of failing in the immediate future.


Equation number
START
INPUT
τ ; α
$f_{T_m}(t_m)$
(4.7.1.5.9)
CALCULATE
$f_{T_m/E_2}(t_m/e_2) = \frac{f_{T_m}(t_m)}{\int_0^{\tau} f_{T_m}(t_m)\,dt_m}$
INTEGRATE
$\int_0^{\tau} f_{T_m}(t_m)\,dt_m$
(4.7.1.6.1)
INPUT
$f_{A/T_m,E_2}(\alpha/t_m,e_1)$
(4.6.4.10)
CALCULATE
$f_{T_m/A,E_2}(t_m/\alpha,e_2) = \frac{f_{A/T_m,E_2}(\alpha/t_m,e_2) f_{T_m/E_2}(t_m/e_2)}{\int_0^{\tau} f_{A/T_m,E_2}(\alpha/t_m,e_2)\, f_{T_m/E_2}(t_m/e_2)\,dt_m}$
INTEGRATE
$\int_0^{\tau} f_{A/T_m,E_2}(\alpha/t_m,e_2) f_{T_m}(t_m/e_2)\,dt_m$
(4.7.1.3.4)
INPUT
$f_{\theta/T_m,E_2}(\theta/t_m,e_2)$
1

Equation number

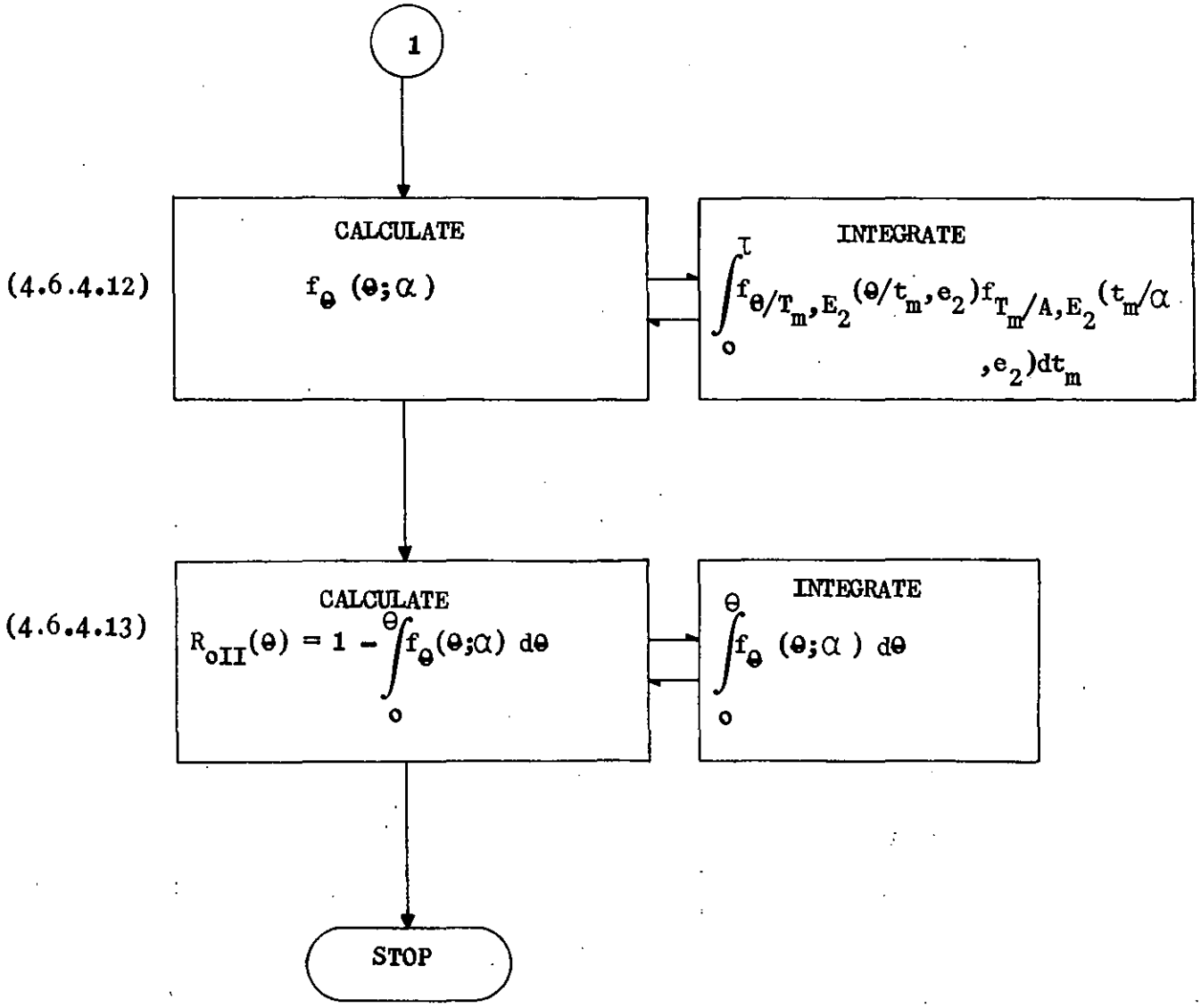


FIGURE 4.16 Flow diagram of operator's reliability III calculation.

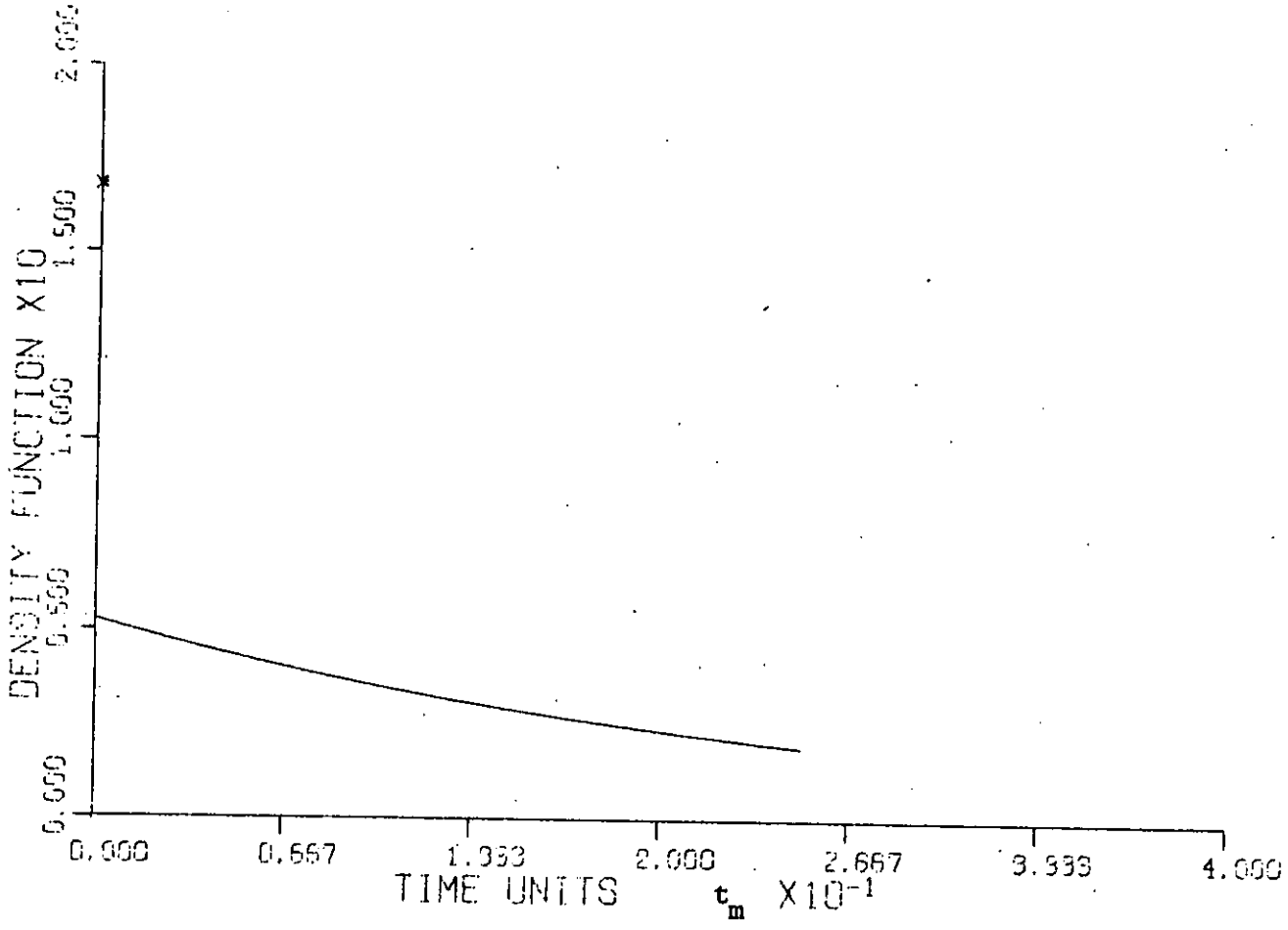
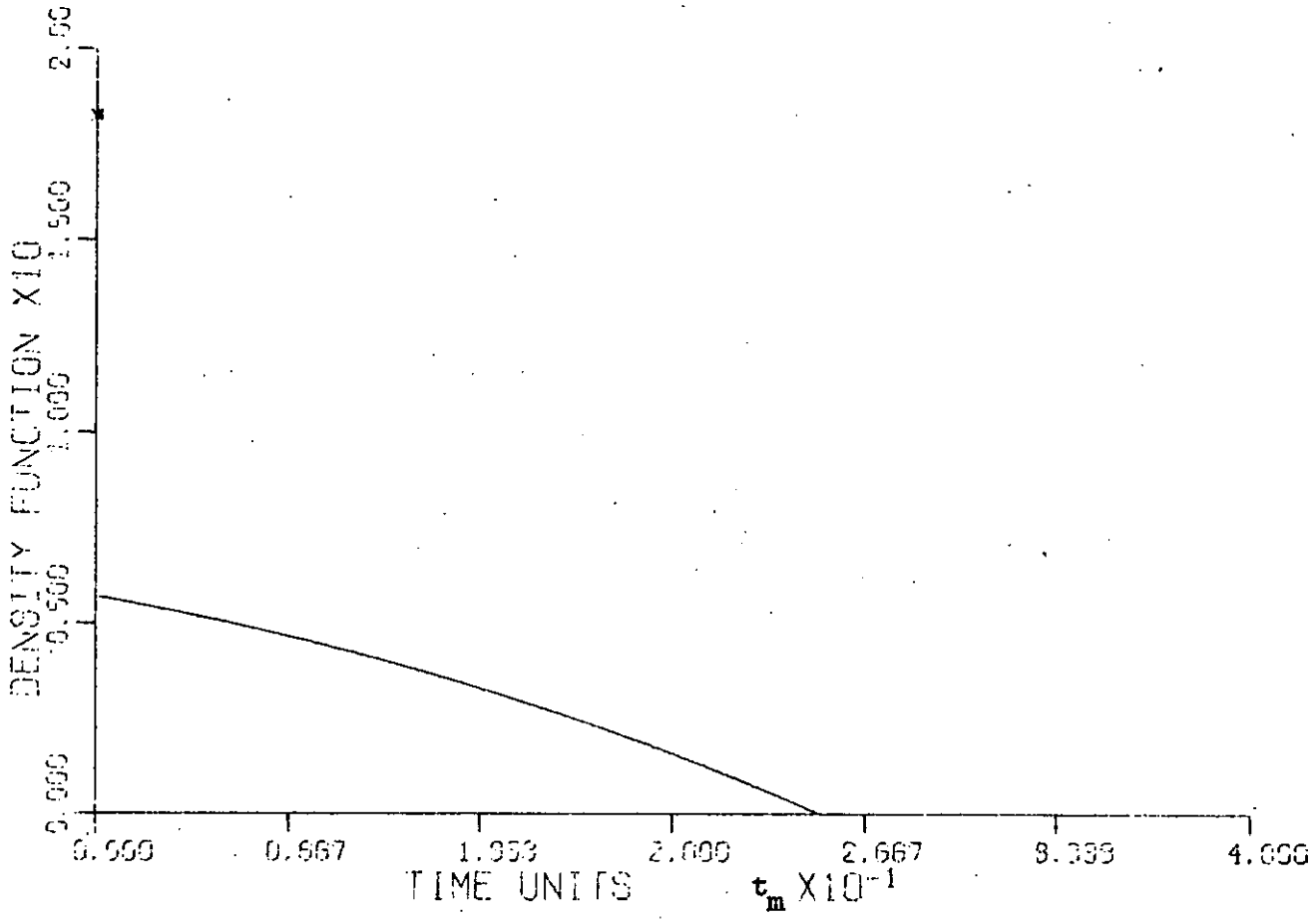
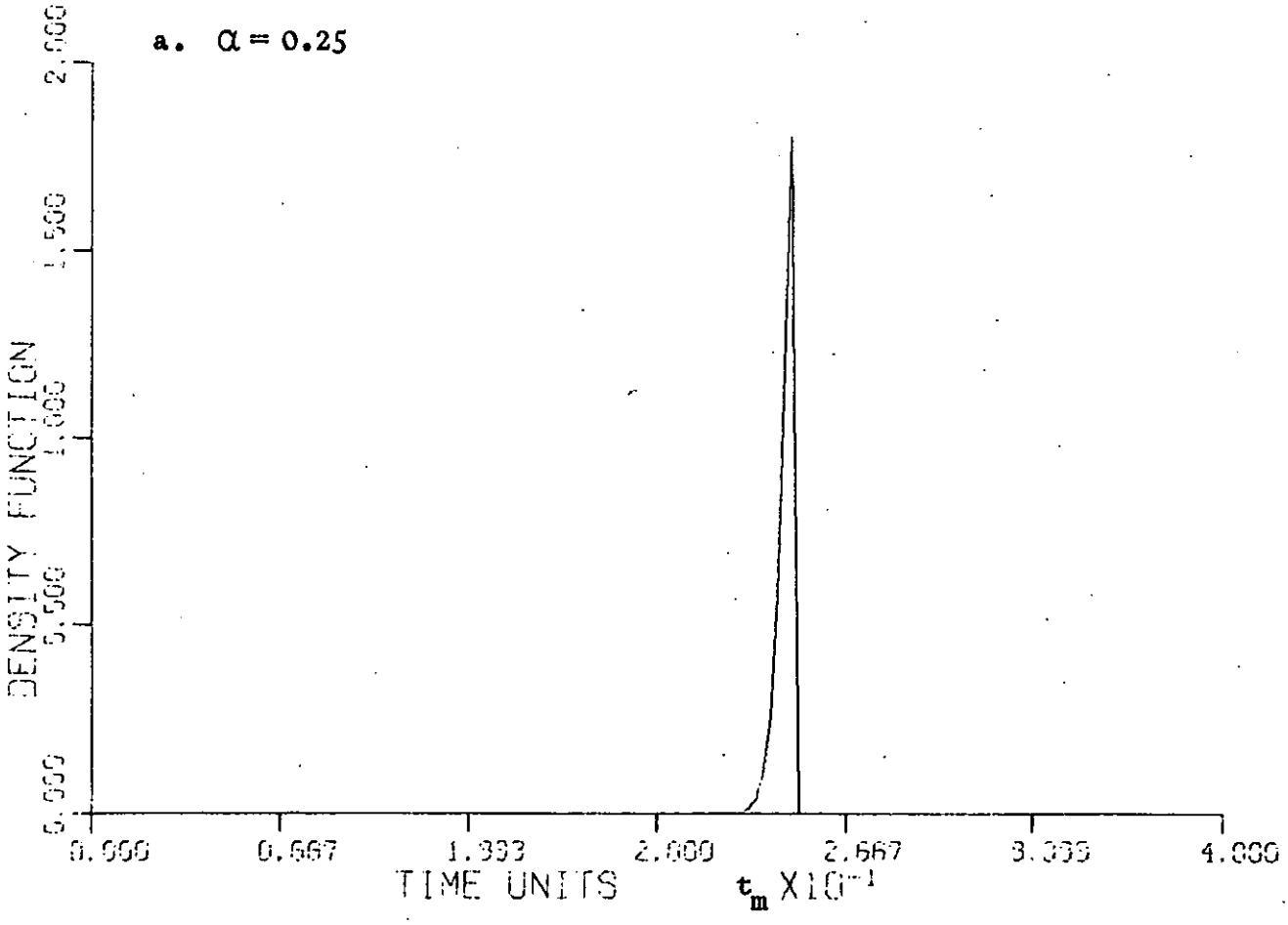


FIGURE 4.17 Illustrative example: Density function $f_{T_m/E_2}(t_m/e_2)$.

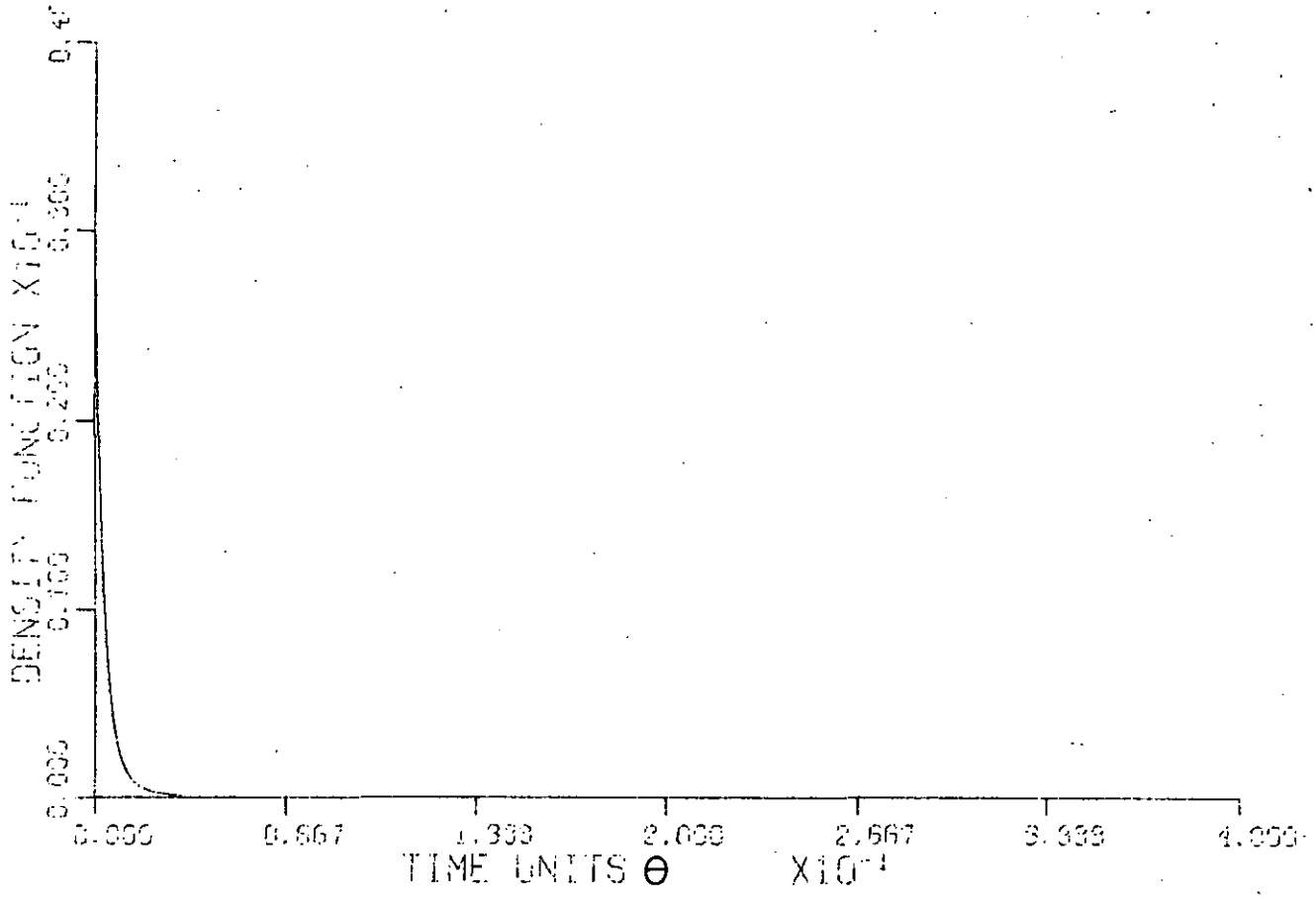


a. $\alpha = 0.25$

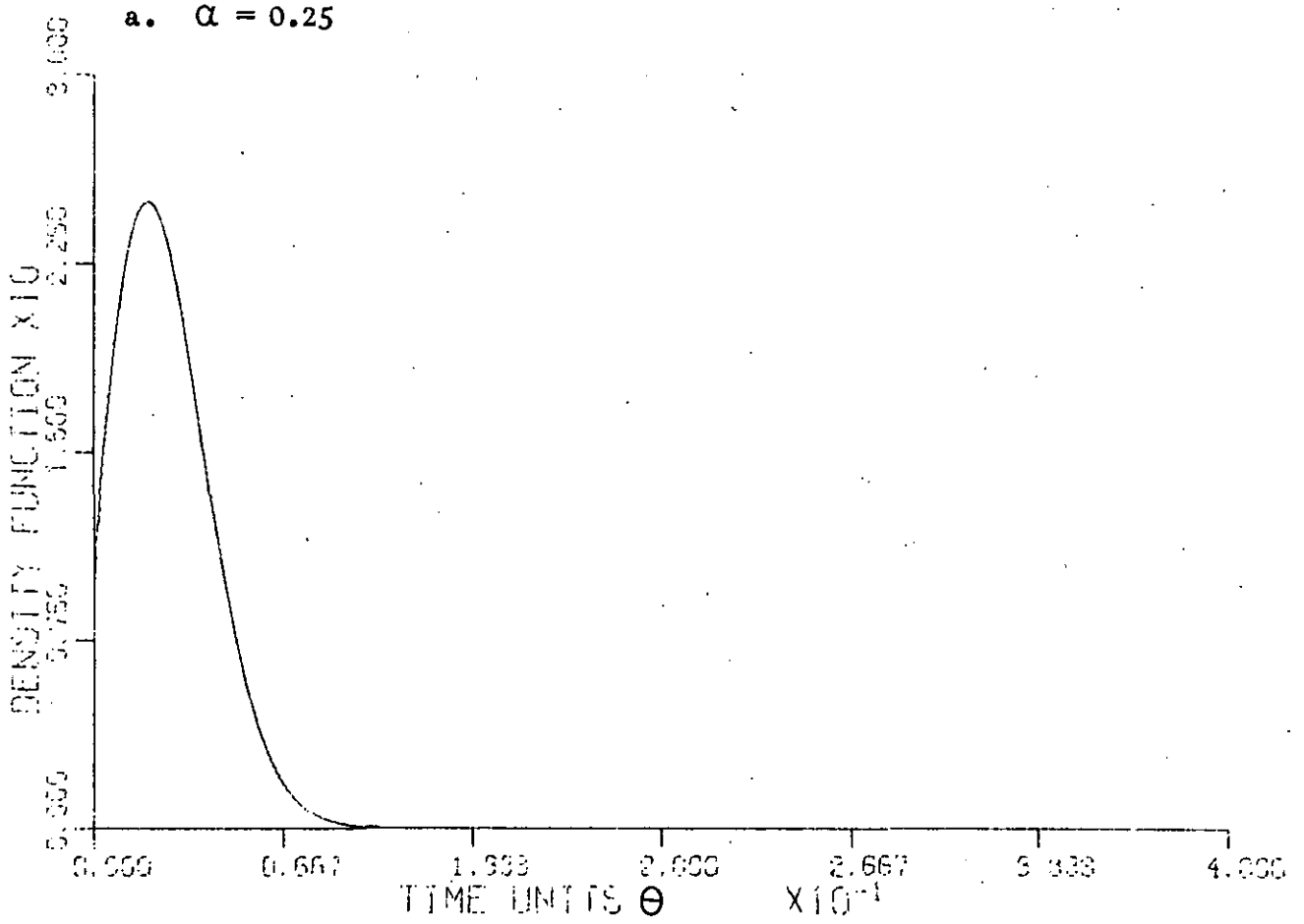


b. $\alpha = 0.75$

FIGURE 4.18 Illustrative example: A posteriori density function $f_{T_m/A,E_2}(t_m/\alpha,e_2)$.



a. $\alpha = 0.25$



b. $\alpha = 0.75$

FIGURE 4.19 Illustrative example: Operator's failure density function $f_\Theta(\theta;\alpha)$.

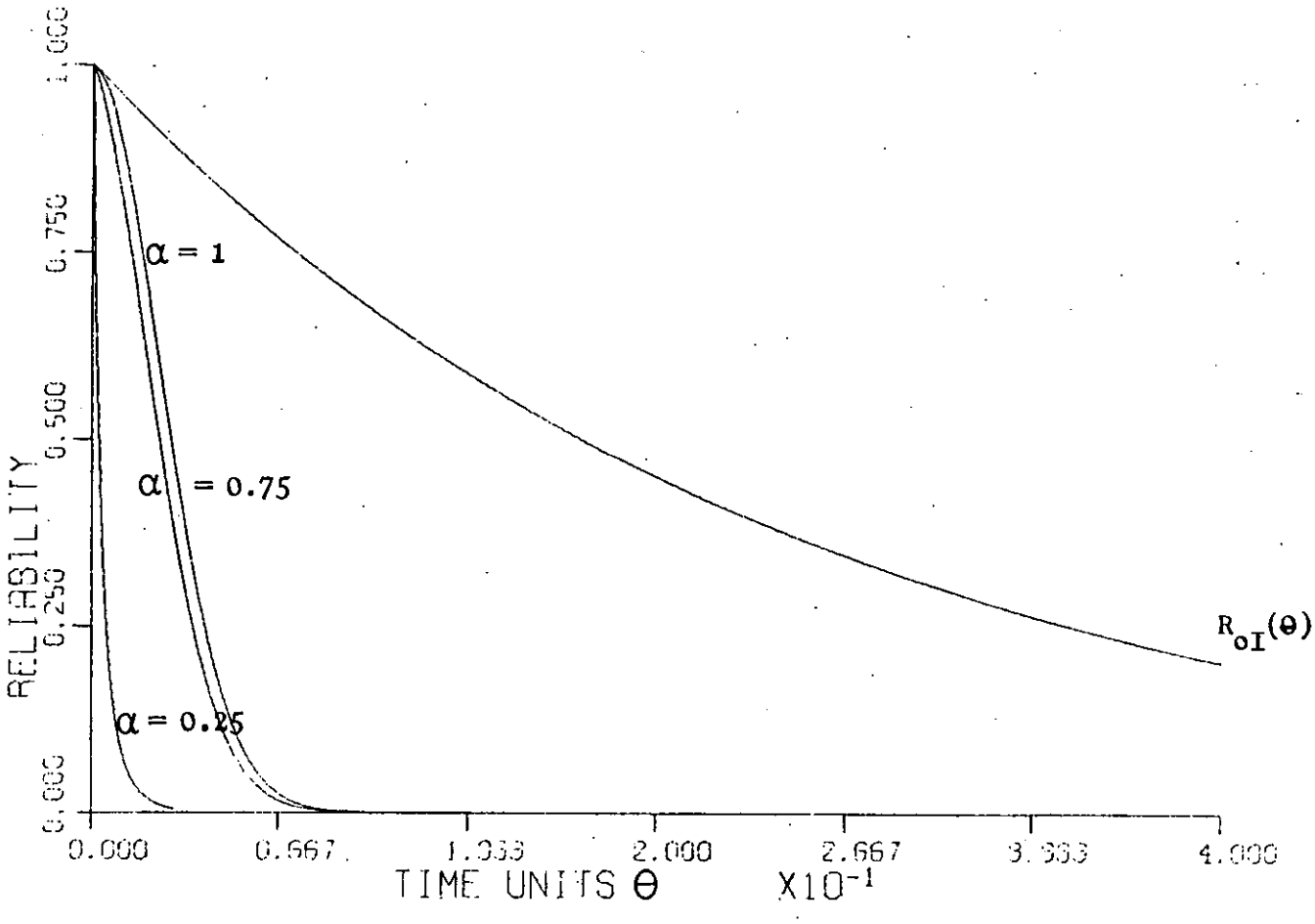


FIGURE 4.20 Illustrative example: Operator's reliability III.

4.7.2.6 Summary of reliability functions

The designer's reliability and the operator's reliability I - III are shown together in Figure 4.21, which illustrates how they are related. The curves are self explanatory.

4.8 Applications of the reliability functions

The immediate application of the operator's reliability functions with malfunction monitoring is as an operational tool to obtain updated estimates of the reliability of single equipments and systems. However, it is anticipated that the knowledge of the functions could be used to determine a "real time" maintenance strategy which involves continuous decision-making, after the detection of the initial malfunction.

Another application is as a design tool to obtain estimates of reliability, maintainability and availability of systems as a function of the particular maintenance policy adopted when a malfunction is detected. Such a study may in fact suggest the optimal type of action to take when a malfunction is shown on the monitor.

Although the theory developed in this study has emphasised the condition monitoring of machinery and malfunction detection in instruments, it is potentially applicable to a diverse range of monitoring activities from non-destructive testing to medical work. For example, in non-destructive testing the ideas may be applied to the monitoring of crack growth in materials to determine reliability functions.

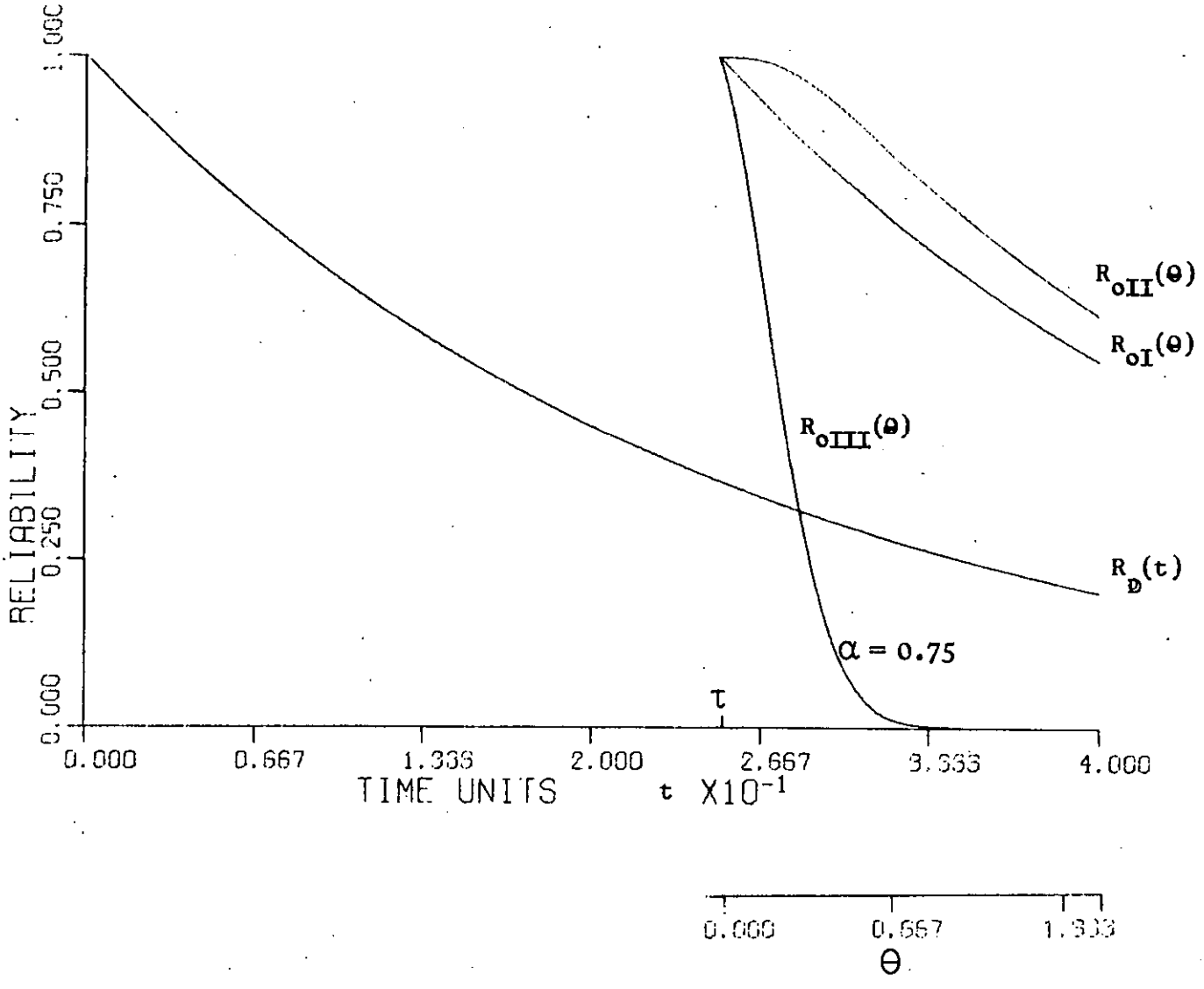


FIGURE 4.21 Illustrative example: Designer's and Operator's reliabilities.

4.9 Concluding remarks

The effect of equipment malfunction monitoring on some aspects of system security has been considered. A periodic inspection policy was used to determine unrevealed equipment failures and the relationship between inspection frequency and several reliability performance indices was examined. It was found, for the criteria considered, that the system security increased as the inspection frequency increased and in particular significant improvements of performance could be achieved by adopting relatively infrequent inspection policies.

The second part of this Chapter was concerned with the relation between malfunction monitoring and reliability, maintainability and availability. The reliability of a system has been categorised into four regimes depending upon whether malfunction monitoring is carried out or not and the type of information obtained from a monitor.

Expressions have been derived for the reliability function of a single equipment with and without malfunction monitoring from the time of observation of the equipment state. The information required to use these expressions is the conventional reliability function together with the probability density functions for the time to failure and for the monitor signal after the initial occurrence of the malfunction.

These reliability expressions may be used as an operational tool to give "running" estimates of the reliability of the equipment or as a design tool to obtain advance estimates of the equipment reliability, maintainability and availability. In this latter case it is necessary to specify maintenance policies in which the monitoring is an integral part. The determination of such maintenance policies and their relation to the information gained from a malfunction monitor provides a basis for further research.

CHAPTER 5

CONCLUSIONS

This thesis has examined ways in which a process computer can be used to improve the overall security of chemical plant by detecting instrument and equipment malfunction. To this end the first part of this study developed methods by which a process computer can be used to detect malfunction in its own instrumentation and control loops. The second study in this thesis assumed the existence of malfunction detection techniques and examined the role of malfunction monitoring in improving process security. These two aspects will now be discussed in some detail.

Chapter 2 has described a control valve position - flow check which may be implemented on a process computer to detect malfunction in a flow control loop. The fundamental idea of the proposed detection algorithm is that for loops in which there is a constant system flow pressure drop characteristic, there exists a unique relationship between the control valve position and the flowrate. This means that knowledge of the valve position implies a measurement of the flowrate, which is additional to the usual primary flowmeter measurement. Thus the control loop has measurement redundancy, thereby providing a foundation for a malfunction detection algorithm.

Two types of check have been developed. In the first a control valve characteristic is obtained of flowrate vs valve position. This requires that the valve be moved over its entire range of travel, thereby permitting the original and current characteristic to be compared. This full valve characteristic comparison enables the source of the malfunction to be determined in some cases.

The second check method assumes that control valve "stroking" is inadmissible and so uses a state estimator to process the measurements

to yield a data base from which the control loop security is inferred. The actual check is based upon the comparison of the original and current residuals generated from the state estimator. No diagnostic information concerning the source of the malfunction can be obtained from this check.

In formulating this check the state estimates were not constrained to be particularly accurate, and so no dynamic modelling of the control loop was necessary. The sole purpose of the state estimator is to condition the process measurements into a convenient form for loop security interrogation.

The proposed methods do not require additional process instrumentation, but exploit the capability of the computer to process, store and display information. The computer storage and time requirements for the checks are modest and since most malfunctions appear not to occur very suddenly the algorithms can be executed at quite infrequent intervals and at low computer priority.

The malfunction detection algorithms have been tested by extensive laboratory trials on an experimental rig which used industrial control equipment. In addition some industrial experiments have been performed. These revealed that models based upon process design manuals were adequate to implement the proposed checks and no difficulty was experienced in tracking the operation of d.d.c. loops or calculating the values of residuals. However, during the period of experimentation control loop malfunction did not occur, although a simulated fault was detected by the algorithm.

The major assumption in the derivation of this position-flow check is that the system flow-pressure drop characteristic is relatively

constant. If this criterion is not completely met then the level of detectable malfunction will be increased. The adaptation of the method to control loops which do not satisfy this assumption of constant system flow-pressure drop characteristic is an area for further research. However, in spite of this it is considered that the flow control loop malfunction detection techniques are at a stage where more comprehensive industrial tests will be most profitable in determining the value and success of the methods.

In principle, there appears to be no reason why the techniques of Chapter 2 should not be generalised to other control loops. All that is required of the loop is that there should be a constant load and a known relationship between the controlled variable and the valve demand signal, thereby creating measurement redundancy. The success of the method will then depend upon the characteristics of the control loop such as the dynamics. However, in practice the relationship between the controlled variable and the valve demand signal is unknown and so a more general malfunction detection technique is needed.

Chapter 3 has developed a general malfunction detection method to encompass all control loops using a Kalman filter. This method was adopted so that tests could be performed on operating control loops. Also it was desired to generate optimal estimates of the control loop state vector when the system was without malfunction, thereby increasing the process operator's knowledge of the loop operation.

The control loops are modelled by linear time - invariant transfer functions and it is assumed that the load entering the loop is constant at some nominal value although some stochastic deviations are accounted for. At a given setpoint and load there is an expected control valve demand signal. Deviations from this expected value are used to

determine the loop malfunction by directly estimating various loop security parameters or examining changes in the statistical properties of the resulting filter innovation sequence.

The implementation of the malfunction detection method has been separated into two problems; the first being the ordinary Kalman filter and the second being Friedland's filter to estimate the loop security parameters. The mechanisation of the primary Kalman filter is fraught with uncertainty in the system models, noise statistics and initial conditions. These problems were overcome by using the innovation correlation technique of Mehra to directly estimate the optimal Kalman gain. The combination of Mehra's and Friedland's algorithms reduces the uncertainty problem to the specification of one covariance matrix.

The malfunction detection method has been tested on a laboratory level control rig, where experiments on both analogue setpoint and direct digital control loops have been performed. It was found that Mehra's adaptive estimator failed to converge to the optimal state estimator if the true process measurement noise covariance matrix was small. However, in "noisier" systems the technique proved to be robust and consistent.

The loop security parameter estimator is successful in estimating control loop malfunction and in particular is able to diagnose and estimate the magnitude of a measurement instrument fault in an analogue setpoint control loop. Thus a correction can be made to the measurement and hence a substitute measurement is available for use in other computer programs. The experiments on the d.d.c. loops provide no diagnostic information.

The second check method which is based upon examining the statistics of the Kalman filter innovation sequence also indicates control loop malfunction, but no diagnostic information is obtained.

The main limitation of the proposed check is the need for a relatively constant nominal load. If this is not true, but the load is measured or can be approximated, e.g. from the overall process throughput, then the method can still be used simply by including this "measurement" in the system model as discussed in section 3.18. For the case of an approximated or estimated unmeasured load the level of detectable malfunction will depend upon the malfunction gain, ϕ, and upon how well the load can be estimated. For example, errors in estimating the load may cause a loop security parameter to change by x%. Hence the malfunction has to be of such a magnitude that the resultant change in the loop security parameter is greater than x%.

The case of unmeasured load disturbances which cannot be approximated and which occur relatively frequently is more difficult to account for and the methods will fail to detect malfunction under these circumstances.

As in the control valve position-flowmeter the modification of the method to cope with large unmeasured load changes is a topic for further exploration.

The methods of malfunction detection developed in Chapters 2 and 3 have been based upon the premise that in a given control loop, which is operating at a particular setpoint and load, there is a unique relationship between the controlled variable and the control valve demand signal. This relationship has been exploited to detect malfunction. The technique used in this thesis was to relate the control loop measurements through a state estimator mathematical model to yield a data base from

which inferences on loop security could be made.

The precise form of the state estimator used for malfunction detection differed in Chapters 2 and 3.

The flow control loop check in Chapter 2 assumed that there was a known steady state relationship between the system flowrate (controlled variable) and the control valve demand signal, i.e. the control valve characteristic. This relationship was additional to the usual equation describing the primary flowmeter measurement. Now a state estimator was used to process measurements of the valve demand and flowmeter signals to estimate the system flowrate $\hat{Q}$, thereby creating a set of residuals from which loop security could be inferred.

The philosophy adopted in designing the flow control loop estimator was that the simplest technique which detected malfunction was the best. It was therefore considered unnecessary to model the dynamics of the loop and a steady state model was assumed, although some dynamic effects were accounted for by expanding the steady state estimator into a tracking steady state estimator. The resultant state estimate, $\hat{Q}$, was therefore not necessarily optimal.

The malfunctions in the control loop were detected by examining the difference between the assumed state estimator model and the actual process under conditions of no malfunction and malfunction.

The techniques of Chapter 3 used the same fundamental ideas as Chapter 2, i.e. the concept of relating the control valve demand signal and the controlled variable, except that the more general case of state estimation was considered.

In Chapter 3 it was assumed that the steady state measurement equation relating the valve demand signal to the tank height was

unknown. The design of the state estimator was also constrained to yield optimal state estimates and therefore the state estimator was based upon a dynamic model of the closed loop. A Kalman filter state estimator was used to relate the resultant process measurements and hence generate a residual from which loop malfunction could be inferred exactly as was done in Chapter 2.

This discussion has highlighted the common and differing features of the proposed malfunction detection algorithms suggested in Chapters 2 and 3. The choice of method in any particular application will depend upon:

i) The degree of information known.
   For example, the application of the tracking steady state estimator to the level control loop requires the characteristic of the tank height versus the control valve demand signal. Usually this will not be available but could be determined on-line during normal process operation.

ii) The quality of the information required.
   If optimal state estimates are required when the control loop operates in a malfunction free condition, then it is unlikely that the tracking steady state estimator would be adequate. Also the diagnostic information on malfunction which is generated from the Kalman filter may be important in some applications.

iii) The characteristics of the control loop.

iv) The available effort for technique implementation.
   The use of the Kalman filter method for detecting control loop malfunction is likely to involve more work than the tracking steady state estimator because of the modelling and computer programming involved.

Many of these features may be explored by examining the Kalman filter method applied to the flow control loop, while the tracking state estimator's performance may be tested by detecting malfunction in the level control loop. These comparisons could be performed on the existing laboratory rigs.

The final part of this thesis has assumed the existence of malfunction detection algorithms or malfunction monitors and has examined the effect of monitoring on equipment reliability, availability and maintainability.

The details of this study are given in Chapter 4.

Initially, it was assumed that a periodic inspection policy was used to detect unrevealed equipment malfunctions and the relationship between the inspection frequency and several reliability performance indices was examined. This analysis shows that the system security increases as the inspection frequency increases and in particular significant improvements can be achieved by adopting relatively infrequent inspection policies.

The second part of Chapter 4 has considered how the information obtained from a malfunction monitor can be used to update estimates of reliability, availability and maintainability. The system reliability is viewed as a numerical encoding of the process operator's state of knowledge of the system rather than an inherent system characteristic. Using this concept, the reliability of a system has been categorised into four regimes depending upon whether malfunction monitoring is performed or not, and the type of information obtained from a monitor.

Reliability expressions for a single equipment have been derived for these four regimes based upon the time at which a monitor inspection

occurs (if monitoring is practised) and the indicated monitor signal. These expressions are derived from probability density functions and the information required is the conventional reliability function together with the probability density functions for the time to failure and for the monitor signal after the occurrence of the initial malfunction.

These reliability expressions provide an operational tool to give "running" estimates of equipment reliability. It is envisaged that these renewed estimates may provide a basis from which real time decisions concerning equipment maintenance policies may be made. However, precisely how this information can be incorporated into the decision making process is not trivial.

For example, if a process engineer inspects a malfunction monitor which indicates a failing equipment, then he can use one of the derived reliability expressions to estimate the probability of equipment survival in the subsequent time intervals. He is then faced with a decision either to let the equipment continue operating, thereby risking severe equipment damage, but performing a planned repair, or to shut the equipment down immediately and perform an emergency repair. The operator's decision must balance the relative advantages and disadvantages of these criteria. It is suggested that the determination of "real time" maintenance policies is an area for further study.

6.0 REFERENCES

(1) Edwards, E., and Lees, F.P., "Man and computer in process control", Inst. Chem. Engrs., 1973.

(2) Allen, D.H., "Economic aspects of plant reliability", presented at the "Design for Reliability" Symp., Inst. Chem. Engrs., N.W. Branch, Univ. of Manch., Inst. Sci. and Tech., 12 April, 1973.

(3) Buffham, B.A., Freshwater, D.C., and Lees, F.P., "Reliability engineering - a rational approach for minimising losses", Inst. Chem. Eng. "Major Loss Prevention in the Process Industries", p. 87, 1971.

(4) Lees, F.P., "The reliability of instrumentation", Chemy. Ind. (in press).

(5) Anon, "Valve malfunctions in nuclear power plant", Nucl. Safety, 15, No. 3, p. 316, 1974.

(6) Emel'yanov, I. Ya., et al., "Evaluating the reliability of the Beloyarsk atomic power station's pumps on the basis of operating data", Sov. At. Energy, 33, No. 1, p. 637, 1972.

(7) King, C.F., and Rudd, D.F., "Design and maintenance of economically failure tolerant processes", AIChE Jnl., 18, No.2, p. 257, 1972.

(8) Skala, V., "Improving instrument service factors", Instrum. Tech., 21, No. 10, p. 27, 1974.

(9) Anyakora, S.N., "Malfunction of process instruments and its detection using a process control computer", Ph.D. Thesis, Loughborough Univ. of Techn., Leics., 1971.

(10) Takada, T., "A study of an engineering system for predictive maintenance"presented at the First Terotechnology Conf. of Inst. Plant Engrs., Bristol, 10 May, 1973.

(11) Dowson, D., "Monitoring, pts. 1 and 2", Tribology, 3, No. 3, p. 138, 1970.

(12) Trotter, J.A., "New tools for modern maintenance", Chem. Eng., Deskbook Issue, p. 87, Feb. 26, 1973.

(13) Lees, F.P., "Principles of the improvement of plant integrity using a process control computer", Measmt. Control, 5, No. 8, p. T118, 1972.

(14) Whitman, K.A., "Computer monitoring and surveillance of process equipment", Instrum. Tech. 19, No. 7, p. 50, 1972.

(15) Cumming, P. J., "Computer and telemetry systems in Offshore Oil production", presented at the Technology Offshore (North Sea) Conf., Aberdeen, 30-31, July, 1974.

(16) Hawickhorst, W., "The improvement of engineered safeguards availability by means of computerised inspection techniques", paper presented at "The Development and Application of Reliability Techniques to Nuclear Plant", CSNI Specialist Meeting, Liverpool, 8 April, 1974.

(17) Damon, D.L., "Application of mini-computers to testing of safety systems", 1972. Nuclear Science Symp. and Nuclear Power Systems Symp., Miami, Florida, 6-8 Dec., 1972, reprinted in IEEE Trans. Nucl. Sci., NS-20, p. 753, 1973.

(18) Lewis, J.B., "Automatic fault sensing and isolation - an important computer technique", Automation, 16, No. 2, p. 94, 1969.

(19) Hoyte, D.W., "The checking and calibration of process instruments on the plant with a digital computer", in "Industrial Measurement Techniques for On-line Computers", London, IEE Conf. Pub. No. 43, p. 54, 1968.

(20) Fraade, D. J., "Computer calibration of a process", Inst. Control Syst., 38, No. 12, p. 93, 1966.

(21) Thompson, A., "Operating experience with direct digital control", IFAC., Ed. Miller, W.E., "Digital Computer Applications to Process Control", Stockholm, p. 55, 1964.

(22) Burton, E.J., "Acoustics diagnostics and nuclear power plants", J. Brit. Nucl. Energy, 13, No. 2, p. 183, 1974.

(23) Ohsawa, Y., Kato, K., and Oyamada, R., "Experiments of anomaly detection system of a reactor core", Nucl. Tech., 23, No. 1, p. 5, 1974.

(24) Compton, M.R., "Real time software operating system for a computer controlled acoustic emission flow monitor", Mater. Eval., 31, No. 7, p.121, 1973.

(25) Izumi, M., and Iida, H., "Application of on-line digital noise analysis to reactor diagnosis in JMTR", Jnl. Nucl. Sci. and Tech., 10, No. 4, p. 227, 1973.

(26) Anyakora, S.N., and Lees, F.P., "Detection of instrument malfunction by the process operator", Chem. Engr., 264, p. 304, Aug., 1972.

(27) Deutsch, R., "System analysis techniques", Prentice-Hall, Englewood Cliff, N.J., 1969.

(28) Papoulis, A., "Probability, random variables and stochastic processes", McGraw-Hill, 1965.

(29) Lees, F.P., "Some data on the failure modes of instruments in the chemical plant environment", Chem. Engr., 277, p. 418, Sept., 1973.

(30) Anon, "Valve malfunctions in nuclear power plants", Nucl. Safety, 15, No. 3, p. 316, 1974.

(31) Stiles, G.F., "Cavitation in control valves", Control and Instrum., 47, No. 4, p. 20, 1974.

(32) Davidson, D.P., "Environmental influences on pressure transducer performance", Measmt. and Control, 8, No. 7, p. 273, 1975.

(33) Kuehn, D.R., and Davidson, H., "Computer control: II Mathematics of control". Chem. Eng. Progr., 57, No. 6, p.44, 1961.

(34) Clementson, A.T., "Statistical determination of plant yields", Brit. Chem. Engng., 8, No. 8, p. 564, 1963.

(35) Ripps, D.L., "Adjustment of experimental data", Chem. Eng. Progr. Symp. Ser., 61, No. 55, p. 8, 1965.

(36) Nogita, S., "Statistical test and adjustment of process data", Ind. Eng. Chem. Proc. Des. Dev., 11, No. 2, p. 197, 1972.

(37) Nogita, S., and Uchiyama, Y., "Multibalance measurement of flow systems", Proc. 1973 Ann. Reliab. and Maintainab. Symp., Philadelphia, Pa., p. 590, 23-25, Jan. 1973.

(38) Dwyer, P., "Linear computations", Wiley, 1960.

(39) Flum, H.H., and Fraade, D.J., "Control computer enables operators to use today's instrumentation tools", Oil and Gas Jnl., 64, No. 38, p. 115, 1966.

(40) Garden, M., "Learning techniques adapt D.D.C. for valve actuators", Instrum. Tech. 15, No. 1, p. 39, 1968.

(41) Barton, A.D., Jenkins, D.G., Lees, F.P., and Murtagh, R.W., "Commissioning and operation of a direct digital control computer on a cement plant", Measmt. and Control, 3, No. 5, p. T77, 1970.

(42) Anyakora, S.N., and Lees, F.P., "The detection of malfunction using a process computer: simple noise power techniques for instrument malfunction", in "The use of Digital Computer in Measurement", London IEE, Conf. Pub. No. 103, p. 35, 1973.

(43) Shinskey, F.C., "Process control systems", McGraw-Hill, 1967.

(44) Buckley, P.S., "Techniques of process control", Wiley, 1964.

(45) Sage, A.P., and Melsa, J.L., "Estimation theory with applications to communications and control", McGraw-Hill, 1971.

(46) Kalman, R.E., "A new approach to linear filtering and prediction problems", Trans. A.S.M.E., Jnl. Basic Engng., 82, ser. D, p.35, 1960.

(47) Schweppe, F.C., et al., "Power system static state estimation", Part I, II, III, IEEE. Trans. on Power App. and Systems, PAS-89, No. 1, p. 120, 1970.

(48) Larson, R.E., Tinney, W.F., and Peschon, J., "State estimation in power systems, Part I Theory and feasibility", IEEE, Trans. on Power App. and Systems, PAS-89, No. 3, p. 345, 1970.

(49) Kendall, M.G., and Stuart, A., "The advanced theory of statistics, vol. 2", Griffin, 1961.

(50) Masiello, R.D., and Schweppe, F.C., "A tracking static state estimator", IEEE. Trans. on Power App. and Systems, PAS-90, No. 3, p. 1025, 1971.

(51) Draper, N.R., and Smith, H., "Applied regression analysis", Wiley, 1966.

(52) Volk, W., "Applied statistics for engineers", McGraw-Hill, 1958.

(53) Walker, H.M., and Lev, J., "Statistical inference", Holt and Co., 1953.

(54) British Standards Institute, "B.S.1042 Code for flow measurement, part I: orifice plates, nozzles and venturi pipes", 1943 and "Part II: Guide to the effects of departure from the methods of part I", 1965, London.

(55) Ceagalske, N.N., "Transient vs. frequency response in analysing chemical process systems", Chem. Eng. Progr. Symp. Ser. 57, No.36, p. 53, 1961.

(56) Coughanowr, D.R., and Koppel, L.B., "Process systems analysis and control", McGraw-Hill, 1965.

(57) Jazwinski, A.H., "Stochastic processes and filtering theory", Academic Press, 1970.

(58) Kailath, T., "An innovations approach to least squares estimation, Part I: Linear filtering with additive white noise", IEEE Trans. Auto. Control, AC-13, No. 6, p.646, 1968.

(59) Bryson, A.E., and Johansen, D.E., "Linear filtering for time varying systems using measurements containing coloured noise." IEEE Trans. Auto. Control, AC-10, No. 2, p. 198, 1965.

(60) Bryson, A.E., and Henrikson, L.J., "Estimation using sampled data containing sequentially correlated noise", AIAA Guidance, Control and Flight Dynamics Conf., Huntsville, Alabama, Paper 67-541, 1967.

(61) Schlee, F.H., Standish, C.J., and Toda, N.F., "Divergence in the Kalman filter", AIAA Jnl., 5, No. 6, p. 1114, 1967.

(62) Fitzgerald, R.J., "Error divergence in optimal filtering problems", 2nd IFAC Symp. on Auto. Contr. in Space, Vienna, Aust., 1967.

(63) Heffes, H., "The effect of erroneous models on the Kalman filter response", IEEE Trans. Auto. Control, AC-11, No. 3, p.541, 1966.

(64) Coggan, G.C., and Noton, A.R.M., "Discrete-time sequential state and parameter estimation in chemical engineering", Trans. Inst. Chem. Engrs., 48, p. T255, 1970.

(65) Goldmann, S.F., and Sargent, R.W.H., "Applications of linear estimation theory to chemical processes: a feasibility study", Chem. Eng. Sci., 26, No. 10, p. 1535, 1971.

(66) Tarn, T.J., and Zaborsky, J., "A practical non-diverging filter", AIAA Jnl., 8, No. 6, p. 1127, 1970.

(67) Joffe, B.L., and Sargent, R.W.H., "The design of an on-line control scheme for a tubular catalytic reactor", Trans. Inst. Chem. Engrs., 50, No. 3, p. 270, 1972.

(68) Seinfeld, J.H., Gavalas, G.R., and Hwang Myung, "Control of non-linear stochastic systems", Ind. and Eng. Chem. Fund., 8, No. 2, p. 257, 1969.

(69) Seinfeld, J.H., "Optimal stochastic control of non-linear systems", AIChE Jnl., 16, No. 6, p.1016, 1970.

(70) Wells, C.H., "Application of modern estimation and identification techniques to chemical processes", AIChE Jnl., 17, No. 4, p.966, 1971.

(71) Hamilton, J.C., Seborg, D.E., and Fisher, D.G., "An experimental evaluation of Kalman filtering", AIChE Jnl, 19, No. 5, p. 901, 1973.

(72) Seborg, D.E., Fisher, D.G., and Hamilton, J.C., "An experimental evaluation of state estimation in multivariable control systems", Automatica, 11, No.4, p.351, 1975.

(73) Coggan, G.C., and Wilson, J.A., "Approximate models and the elimination of bias in on-line estimation of industrial processes", The British Computer Soc. Symp. "On-line Computer Methods Relevant to Chem. Engng." Univ. of Nottingham, p.136, 1971.

(74) Vakil, H.B., Michelsen, M.L., and Foss, A.S., "Fixed bed reactor control with state estimation", Ind. Eng. Chem. Fund., 12, No.3, p. 328, 1973.

(75) Wells, C.H., and Larson, R.E., "Application of combined optimum control and estimation theory to d.d.c." Proc. IEEE, 58, No. 1, p. 16, 1970.

(76) Gustavsson, I., "Survey of applications of identification in chemical and physical processes", Automatica, 11, No.1, p.3, 1975.

(77) Choquette, P., Noton, A.R.M., and Watson, C.A.G., "Remote computer control of an industrial process", Proc. IEEE, 58, No.1, p.10, 1970.

(78) Wells, C.H., and Wismer, D.A., "Advances in process control applications", in "Advances in Contr. Syst., Theory and Appl." C. T. Leondes (Ed.), Academic Press, 1971.

(79) Thé, G., "Kalman filter divergence due to process noise decoupling", Proc. IEE, 121, No. 6, p. 525, 1974.

(80) Sastry, V.A., and Vetter, W.J., "A paper making wet-end dynamics model and parameter identification by iterative filtering", in "Industrial Appl. of Dynamic Modelling", Durham, IEE Conf. Pub. No. 57, 1969.

(81) King, R.P., "On line digital computer control of slurry conditioning in mineral flotation", Automatica, 10, No.1, p.5, 1974.

(82) Godbole, S.S., "Application of Kalman filtering technique to nuclear reactors", 1972 Nuclear Science Symp. and Nuclear Power Symp., Miami, Florida, 6-8 Dec., 1972, reprinted in IEEE Trans. Nucl. Sci., NS-20, p. 661, 1973.

(83) Venerus, J.C., and Bullock, T.E., "Estimation of dynamic reactivity using digital Kalman filtering", Nucl. Sci. Eng. 40, No. 2, p. 199, 1970.

(84) Shinohara, Y., and Oguma, R., "Estimation of time varying reactivity using a method of non-linear filtering", Nucl. Sci. Eng., 52, No. 1, p. 76, 1973.

(85) Sorenson, H.W., "On the error behaviour in linear minimum variance estimation problems", IEEE Trans. Auto. Control, AC-12, No.5, p. 557, 1967.

(86) Berkovec, J.W., "A method for the validation of Kalman filter models", Joint Auto. Control Conf., Univ. of Colorado, Boulder, Colorado, p.488, Aug. 5-7, 1969.

(87) Mehra, R.K., "On the identification of variances and adaptive Kalman filtering", IEEE Trans. Auto. Control, AC-15, No. 2, p. 175, 1970.

(88) Tompetrini, K., "Adaptive filtering techniques and premium-protection approach for performance evaluation", Ph.D. Thesis, Stevens Inst. of Tech., U.S.A., 1973.

(89) Mehra, R.K., "Approaches to adaptive filtering", IEEE Trans. Auto. Control, AC-17, No. 5, p. 693, 1972.

(90) Pearson, J.O., "Estimation of uncertain systems", in "Control and Dynamic syst.; Advances in Theory and Appl." C.T. Leondes (Ed.), Academic Press, 1973.

(91) Weiss, I.M., "A survey of discrete Kalman-Bucy filtering with unknown noise covariances", AIAA Guidance, Control and Flight Mechanics Conf., Santa Barbara, Calif., Paper 70-955, 1970.

(92) Schmidt, S.F., Weinburg, J.D., and Lukesh, J.S., "Case study of Kalman filtering in the C-5 aircraft navigation systems", Case studies in system control, IEEE Group on Auto. Control, 1968 Joint Auto Control Conf., Ann Arbor, Mich., 1968.

(93) Crump, N.D., "Estimation of discrete signals containing a non-random component ", Ph.D. Thesis, Georgia Inst. of Tech., U.S.A., 1969.

(94) Quigley, A.L.C., "An approach to the control of divergence in Kalman filter algorithms", Int. Jnl. Control, 17, No.4, p.741, 1973.

(95) Sriyananda, H., "A simple method for the control of divergence in Kalman filter algorithms", Int. Jnl. Control, 16, No.6, p.1101, 1972.

(96) Magill, D.T., "Optimal adaptive estimation of sampled stochastic processes", IEEE Trans. Auto. Control, AC-10, No.4, p.434, 1965.

(97) Sims, F.L., Lainiotis, D.G., and Magill, D.T., "Recursive algorithm for the calculation of the adaptive filter weighting coefficients", IEEE Trans. Auto. Control, AC-14, No. 2, p.215, 1969.

(98) Smith, G.L., "Sequential estimation of observation error variances in a trajectory estimation problem", AIAA Jnl., 5, No.11, p.1964, 1967.

(99) Abramson Jr., P.D., "Simultaneous estimation of the state and noise statistics in linear dynamical systems", NASA Tech. Rpt. R-332, March 1970.

(100) Shellenbarger, J.C., "Estimation of covariance parameters for an adaptive filter", Nat. Electr. Conf. Proc., 22, p.698, 1966.

(101) Sage, A.P., and Wakefield, C.D., "Maximum likelihood identification of time varying and random system parameters", Int. J. Control, 16, No.1, p.81, 1972.

(102) Jazwinski, A.H., "Adaptive filtering", Automatica, 5, No.4, p.475, 1969.

(103) Carew, B., and Bélanger, P., "Identification of optimum filter steady state gain for systems with unknown noise covariances", IEEE Trans. Auto. Control, AC-18, No.6, p.582, 1973.

(104) Neethling, C., and Young, P., "Comments on Identification of optimum steady state gain for systems with unknown noise covariances", IEEE Trans. Auto. Control, AC-19, No.5, p.623, 1974.

(105) Bélanger, P., "Estimation of noise covariance matrices for a linear time-varying stochastic process", Automatica, 10, No.3, p. 267, 1974.

(106) Godbole, S.S., "Kalman filtering with no a priori information about noise–white noise case". Proc. 1973 IEEE Decision and Control Conf., paper WA1-2, p.9, 1973.

(107) Godbole, S.S., "Kalman filtering with no a priori information about noise – white noise case: identification of covariances", IEEE Trans. Auto. Control, AC-19, No.5, p.561, 1974.

(108) Mehra, R.K., and Peschon, J., "An innovations approach to fault detection and diagnosis in dynamic systems", Automatica, 7, No.5, p. 637, 1971.

(109) Friedland, B., "Treatment of bias in recursive filtering", IEEE Trans. Auto. Control, AC-14, No.4, p. 359, 1969.

(110) Tacker, E.C., "Linear filtering in the presence of time-varying bias", IEEE Trans. Auto. Control, AC-17, No.6, p.828, 1972.

(111) Tanaka, A., "Parallel computation in linear discrete filtering", IEEE Trans. Auto. Control, AC-20, No.4, p. 573, 1975.

(112) Davis, M.H.A., "The application of non-linear filtering to fault detection in linear systems", IEEE Trans. Auto. Control, AC-20, No.2, p.257, 1975.

(113) "Optimum controller settings", Taylor Transcope Controller Instructions, 1B404, Issue 3, Taylor Inst. Co., Herts, Eng.

(114) Cadzow, J.A., and Martens, H.R., "Discrete time and computer control systems", Prentice-Hall, 1970.

(115) Wiberg, D.M., "State space and linear systems", McGraw-Hill, 1971.

(116) Jenkins, G.M., Ottley, D.J., and Packer, S.H., "A systems study of the effect of unreliability of a petrochemical complex", Jnl. Syst. Eng., 2, No.1, p.65, 1971.

(117) Houston, D.E.L., "New approaches to the safety problem", Inst. Chem. Eng. Symp. Ser. No.34, "Major loss prevention in the process industries", p.210, 1971.

(118) Shooman, M., "Probabilistic reliability: an engineering approach", McGraw-Hill, 1968.

(119) Green, A.E., and Bourne, A.J., "Reliability technology", Wiley Interscience, 1972.

(120) Sandler, G.H., "System reliability engineering", Prentice-Hall, 1963.

(121) Larson, H.J., "Introduction to probability theory and statistical inference", Wiley, 1969.

(122) Biheller, J.H., "Vibration-aid in diagnosing pump troubles", Power, 115, No.5, p.102, 1971.

(123) Maten, S., "Program machine maintenance by measuring vibration velocity", Hydroc. Proc., 49, No. 9, p.291, 1970.

(124) Tustin, W., "Measurement and analysis of machinery", Chem. Eng. Progr., 67, No.6, p.62, 1971.

(125) Boto, P.A., and Fernlund, I., "Shock pulse measurement of bearings", Aircr. Eng., 42, No.12, p. 22, 1970.

(126) Drost, D.G., "Analysis of lube-oil pinpoints engine trouble spots", Oil and Gas Jnl., 70, No.23, p.70, 1972.

(127) Collacott, R.A., "Sonic monitoring of plain bearings subject to seizure", Tribol. Int., 8, No.3, p.123, 1975.

(128) Deeprose, W.M., and McNulty, P.J., "Cavitation noise in process pumps", Chem. and Proc. Eng., 53, No.5, p.46, 1972.

(129) Sebestyén, Gy., Fáy, A., Csemniczky, J., "Measurements of cavitation characteristics of a pump connected with measurement of noise", Acta. Tech. Acad. Sci. Hung., 66, No.4, p.305, 1969.

(130) McCall, J., "Maintenance policies for stochastically failing equipment: a survey", Managmt. Sci., 11, No.5, p.493, 1965.

(131) Berg, M., and Epstein, B., "Grouping of maintenance policies", Proc. NATO Conf. on "Reliab. testing and reliab. eval.", The Hague, Neth., p. I-C-1, 4-8, Sept., 1972.

(132) Jardine, A.K.S., "Maintenance, replacement and reliability", Pitman, 1973.

(133) Bazovsky, I., "Reliability theory and practice", Prentice-Hall, 1961.

(134) Derman, C., "On optimal replacement rules when changes of state are Markovian", in "Mathematical Optimisation Techn.", R. Bellman (Ed.), Rand Corp., 1963.

(135) Kolesar, P., "Minimum cost replacement under Markovian deterioration", Managmt. Sci., 12, No.9, p.694, 1966.

(136) Klein, M., "Inspection-maintenance-replacement schedules under Markovian deterioration", Managmt. Sci., 9, No.1, p.25, 1962.

(137) Satia, J.K., and Lave, R.E., "Markovian decision processes with probabilistic observation of states", Managmt. Sci., 20, No.1, p.1, 1973.

(138) Mine, H., and Kawai, H., "An optimal maintenance policy for a 2-unit parallel system with degraded states", IEEE Trans. Reliab., R-23, No.2, p.81, 1974.

(139) Gross, A.J., "Minimisation of misclassification of component failures in a 2-component system", IEEE Trans. Reliab., R-19, No.3, p.120, 1970.

(140) Polovko, A.M., "Fundamentals of reliability theory", Academic Press, 1968.

(141) Jensen, F., "The computation of yield and drift reliability of electronic circuits", Microelect. Reliab., 11, No.2, p.139, 1972.

(142) Welker, E.L., "Some statistical techniques useful in system ageing studies", Proc. 1973 Ann. Reliab. and Maintainab. Symp., Philadelphia, Pa., p.10, 23-25, Jan. 1973.

(143) Tribus, M., "Rational descriptions, decisions and designs, Pergamon Press, 1969.

APPENDIX I.

NOMENCLATURE

A.I.1 List of Symbols

A	matrix	-
$\underline{a}$	vector	-
$\underline{a}_i$	i^{th} vector	-
a_{ij}	ij^{th} element of A	-
$a(\ )$	parameter	-
B	matrix	-
b	scalar variable	-
C	matrix	-
c_{ij}	ij^{th} cofactor of A	-
$E(\ )$	expectation	-
$E(\ /\ )$	conditional expectation	-
$F(x)$	probability distribution function for random variable X	-
$f_X(x)$	probability density function describing random variable X	-
$f_{X_1,X_2}(x_1,x_2)$	joint probability density function describing random variables X_1 and X_2	-
$f_{X_1/X_2}(x_1/x_2)$	conditional probability density function describing X_1 given $X_2 = x_2$	-
k	discrete time counter	-
n	matrix dimension	-
m	matrix dimension	-
$P(\ )$	probability	-
$p_X(x)$	probability function for discrete random variable X	-
$p_{X_1,X_2}(x_1,x_2)$	joint probability function describing discrete random variables X_1 and X_2	-

$p_{X_1/X_2}(x_1/x_2)$	conditional probability function describing discrete random variable X_1 given $X_2 = x_2$	-
r	matrix dimension	-
t_o	initial time	time units
t_k	time at k^{th} sampling interval	time units
t	time	" "
Δ t	sampling interval	" "
X	random variable associated with the real variable x	-
$\underline{X}$	random vector associated with the real vector $\underline{x}$	-
$\underline{x}$	vector	-
x_i	i^{th} element of $\underline{x}$	-
x	real variable	-
z_i	i^{th} scalar variable	-

Greek letters

μ_x	expected value of random variable X	-
$\mu_{\underline{x}}$	expected value of random vector $\underline{X}$	-
σ^2_x	variance of random variable X	-

Superscripts

T	matrix transpose
- 1	matrix inverse
##	matrix pseudoinverse

Subscripts

o	initial value
k	k^{th} sampling interval

A.I.2 Nomenclature

In this Appendix the notation and mathematical quantities used in this thesis are defined. Although most of the concepts used are standard, they are included here to avoid confusion in terminology.

Several estimation problems considered here will be restricted to those containing discrete time dynamic and measurement models. In other words, the dynamics and measurements are treated at fixed time increments which take on only integer values, i.e. $a(k) = a(t_k)$ where $t_k = t_o + k\Delta t$ with t_o the initial time treated in the problem and Δt the time increment used. If the parameter $a(k)$ does not vary with time, then the subscript k is omitted. If the state dynamics are characterised by a difference equation with constant coefficients, the model is said to be stationary.

When the dynamics and measurement models are written matrix-vector notation is used. Vectors are denoted by lower-case letters with an underscore, i.e.

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \qquad \text{(A.I.2.1)}$$

where x_i, $i = 1, \ldots, n$ are scalar quantities and are referred to as the components of the vector. If $n = 1$, the quantity x is a scalar and the letter is not underscored. Matrices are denoted by upper-case letters, i.e.

$$A = \begin{bmatrix} a_{11} & a_{12} \cdots & a_{1m} \\ a_{21} & & \\ \vdots & & \\ a_{n1} & a_{n2} \cdots & a_{nm} \end{bmatrix} \qquad \text{(A.I.2.2)}$$

where the elements of A, a_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, m$ are scalars and n is the number of rows and m is the number of columns. If $n = m = 1$, A is a scalar quantity.

The transpose of a matrix is denoted by a superscript T, i.e.

$$A^T = \begin{bmatrix} a_{11} & a_{21} \cdots\cdots\cdots\cdots\cdots a_{n1} \\ a_{12} & \\ \vdots & \\ a_{1m} & a_{nm} \end{bmatrix} \qquad \text{(A.I.2.3)}$$

Throughout this thesis several matrix operations are referred to or used. These include the determinant of a matrix, the trace of a matrix, the inverse of a matrix and the pseudoinverse of a matrix. The determinant of a matrix, A, is denoted by det(A) and is definable for square matrices only, i.e., the number of rows equals the number of columns. The determinant of A is given by:

$$\det(A) = a_{11}c_{11} + a_{12}c_{12} + \ldots a_{1n}c_{1n} \qquad \text{(A.I.2.4)}$$

where A is given by equation (A.I.2.2) with $m = n$ and c_{1i} is the $1i^{th}$ cofactor and is the determinant of the submatrix formed by striking out the first row and the i^{th} column multiplied by $(-1)^{i+1}$.

The trace of a square matrix is defined as the sum of the diagonal elements, i.e.,

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii} \qquad \text{(A.I.2.5)}$$

The inverse of a matrix, A, is formed as follows:

$$(A^{-1})_{ij} = \frac{1}{\det(A)} (c_{ji}) \qquad \text{(A.I.2.6)}$$

where c_{ji} is the ji^{th} cofactor of A and $(A^{-1})_{ij}$ is the ij^{th} element of A^{-1}. From equation (A.I.2.6) it is obvious that the inverse of a

matrix exists only if it is square and its determinant is nonzero.

For matrices which are not square or those which have zero valued determinants, the concept of a pseudoinverse (27) can be applied. Before this concept can be defined, the concepts of linear independence and matrix rank must first be discussed. A set of n-vectors, i.e. each vector contains n components, $\underline{a}_1, \ldots, \underline{a}_n$, is said to be linearly independent if the following vector equation

$$z_1 \underline{a}_1 + \ldots + z_n \underline{a}_n = \underline{0} \qquad \text{(A.I.2.7)}$$

implies that each scalar $z_1, \ldots, z_n$ is zero. The rank of a matrix is then defined as the maximum number of linearly independent rows or columns of the matrix. For an n x n matrix, a determinant of zero implies a rank less than n.

Returning to the definition of the pseudoinverse of a m x n matrix, A of rank r, denoted by $A^{\#\#}$, then:

$$A^{\#\#} = C^T (C C^T)^{-1} (B^T B)^{-1} B^T \qquad \text{(A.I.2.8)}$$

where

$$A = B C$$

and B is m x r, C is r x n and both B and C are of rank r. From equation (A.I.2.8) it is obvious that the pseudoinverse of a m x n matrix is an n x m matrix. A complete discussion of the matrix pseudoinverse can be found in Deutsch (27).

Another matrix property that is referred to is that of a positive definite matrix. A real, symmetric ($a_{ij} = a_{ji}$, $i,j = 1, \ldots, n$) matrix A is said to be positive definite if the quantity b,

$$b = \underline{x}^T A \underline{x} \qquad \text{(A.I.2.9)}$$

is non-negative for all $\underline{x} \neq \underline{0}$ and zero only if $\underline{x} = \underline{0}$. Similarly, A is positive semidefinite if b can be zero for some $\underline{x} \neq \underline{0}$.

The remainder of this Appendix deals with some basic definitions associated with random processes. Initially the concept of a random variable must be defined. Simply, a random variable X is a function whose values depend upon the outcome of a chance event. Here capital letters are used to denote random variables while lower case letters stand for particular values in the range of the random variable.

Random variables may be discrete or continuous. A random variable X is said to be discrete if its range forms a discrete (countable) set of real numbers, while X is continuous if its range forms a continuous (uncountable) set of real numbers and the probability of X equalling any single value in its range is zero.

If X is a discrete random variable then the probability function for X is a function of the real variable x and is denoted:

$$p_X(x) = P(X = x) \qquad \text{for all real } x$$

The distribution function for a random variable X is defined and denoted as:

$$F(x) = P(X \leq x)$$

Hence F(x) gives the total accumulation of probability for X equalling any number less than or equal to x.

If X is a discrete random variable, then:

$$F(x) = \sum_{x} p_X(x)$$

When X is a continuous random variable then the concept of a probability density function is valuable. The probability density function is denoted by $f_X(x)$ and is related to the distribution function according to

$$f_X(x) = \frac{d}{dx} F(x)$$

and $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)\, dx$$

The expected or mean value of a random variable X is given by:

$$\mu_x = E(X) = \sum_{\substack{\text{range} \\ \text{of } X}} x\, p_X(x) \quad : \text{discrete } X$$

$$\mu_x = E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \quad : \text{continuous } X$$

The variance of a random variable is denoted by:

$$\sigma_x^2 = E\left((X - \mu_x)^2\right)$$

A random vector, denoted $\underline{X}$, is a vector whose components are each a random variable. The above notation for one-dimensional random variables may be extended to the n dimensional case and the joint distribution and density functions are given by:

$$F(\underline{x}) = P(X_1 \leq x_1;\ X_2 \leq x_2;\ \ldots,\ X_n \leq x_n)$$

$$F(\underline{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\underline{X}}(\underline{x})\, dx_1 \ldots dx_n$$

$$f_{\underline{X}}(x) = d\,F(\underline{x}) / d\underline{x}$$

Knowledge of the n dimensional density function $f_{\underline{X}}(\underline{x})$ enables the one dimensional density function $f_{X_1}(x_1)$ to be determined. This is called the marginal density function for X_1 and is found by integrating $f_{\underline{X}}(\underline{x})$ over x_2 to x_n, i.e.

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x})\, dx_2\, dx_3 \ldots dx_n$$

The expected or mean value of a random vector is given by:

$$\mu_{\underline{x}} = E(\underline{X}) = \int_{-\infty}^{\infty} \underline{x}\, f_{\underline{X}}(\underline{x})\, d\underline{x}$$

A measure of how $X_1, \ldots, X_n$ vary together is called the covariance matrix and is defined by:

$$\text{Cov}\ (\underline{X}) = E\ ((\underline{x} - E\ (\underline{x}))\ (\underline{x} - E\ (\underline{x}))^T)$$

One of the most important concepts in the application of probability theory is that of independent random variables. n random variables are said to be independent if the joint density function or probability function will factor into the product of the marginals, i.e.,

$$p_{\underline{X}}\ (\underline{x}) = p_{X_1}(x_1)\ p_{X_2}(x_2) \ \ldots\ p_{X_n}(x_n) : \text{discrete}$$

$$f_{\underline{X}}\ (\underline{x}) = f_{X_1}(x_1)\ f_{X_2}(x_2) \ \ldots\ f_{X_n}(x_n) : \text{continuous}$$

Another statement which is true for the random variables $X_1, \ \ldots\ X_n$ to be independent may be written in terms of the joint distribution function, i.e.,

$$F\ (\underline{x}) = F\ (x_1)\ F\ (x_2) \ \ldots\ F\ (x_n)$$

If the above statements are not true then the random variables are dependent, and it is necessary to define a conditional probability function. Suppose X_1 and X_2 are jointly discrete random variables, then the conditional probability function for X_1, given $X_2 = x_2$ is defined to be:

$$p_{X_1/X_2}\ (x_1/x_2) = \frac{p_{X_1,X_2}\ (x_1,x_2)}{p_{X_1}(x_1)}$$

If X_1 and X_2 are continuous random variables, then the conditional density for X_1, given $X_2 = x_2$ is given by

$$f_{X_1/X_2}(x_1/x_2) = \frac{f_{X_1,X_2}\ (x_1,x_2)}{f_{X_1}(x_1)}$$

Of particular interest in estimation theory is the concept of a random or stochastic process. The estimation theory used in this work is limited to discrete random processes. A discrete random process is a

sequence of random variables which are indexed according to their time of occurrence, i.e. X_k is the random variable at the k^{th} time interval.

A sequence of random variables may be described by the probability distributions and density functions discussed above.

A parameter of particular importance is the expected value of a random variable or vector at the $k+1^{th}$ time increment, $\underline{X}_{k+1}$, given that the sequence of random variables $\underline{X}_1$ to $\underline{X}_k$ have occurred. The conditional expectation is defined by:

$$E(\underline{X}_{k+1}/\underline{X}_1, \ldots, \underline{X}_k) = \frac{\int_{-\infty}^{\infty} \underline{x}_{k+1} \, f_{\underline{X}} (\underline{x}_1, \ldots, \underline{x}_{k+1}) \, d\underline{x}_{k+1}}{f_{\underline{X}} (\underline{x}_1, \ldots, \underline{x}_k)}$$

Further information and definitions for random processes may be found in Papoulis (28).

APPENDIX II

PROCESS DATA VALIDATION CHECKS

A.II.1 List of symbols

A	constraint matrix	-
a_{ij}	ij^{th} element of A	-
$\underline{b}$	vector	-
C_p	specific heat	chu/lb.mole
D	square matrix	-
d_{ij}	ij^{th} element of D	-
D_B	distillation column product	lb.mole/hr
e_i	relative error	-
$e(D)$	matrix of error coefficients in D	-
$e(d_{ij})$	ij^{th} element of e(D)	-
$\underline{e}(\underline{g})$	vector of error coefficients in $\underline{g}$	-
$e(g_i)$	i^{th} element of $e(\underline{g})$	various
$e(g_i)_{g_j}$	error in solution g_i when matrix error coefficients are chosen to maximise g_j	various
E	solution vector error matrix	-
F_A	distillation column feed	lb.mole/hr
$\underline{f}$	vector	-
$f_j(\underline{y})$	non-linear constraint equation	-
$\underline{g}$	vector	-
H_F; H_{W_A}; H_{D_B}; H_1; H_2	enthalpy	chu/lb.mole
I	unit matrix	-
k	fraction	-
L	square matrix	-
l_{ij}	ij^{th} element of L	-
l	number of constraint equations	-

L_A	bottom product flowrate from column B	lb.mole/hr
L_B	reflux flowrate	lb.mole/hr
$\underline{m}$	vector of process measurements	-
$\underline{m}_i$	i^{th} element of $\underline{m}$	various
m	number of process measurements	-
Q_c	condenser heat load	chu/hr
Q_L	distillation column heat loss	chu/hr
Q_R	reboiler heat load	chu/hr
q	q line	-
S	diagonal matrix	-
s_{ij}	ij^{th} element of S	-
T	temperature of column B overhead liquid	^oC
T^L_B	temperature of column B subcooled overhead liquid	^oC
T^V_B	temperature of column B overhead vapour	^oC
V_A	overhead vapour flowrate from column A	lb.mole/hr
V'_A	vapour flowrate below feed point in column A	" "
V_B	overhead vapour flowrate from column B	" "
W_A	bottoms product flowrate from column A	" "
W_B	bottoms product flowrate from column B	" "
$\underline{y}$	vector of predicted process measurements	-
y_i	i^{th} element of $\underline{y}$	various

Greek

$\underline{\lambda}$	vector of Lagrange multipliers	-
λ	latent heat	chu/lb.mole
σ	measurement noise standard deviation	various
ϕ	sum of squares cost function	-
$\underline{\psi}$	vector of constraint equations	-

Subscripts

i	variable i
A	distillation column A
B	distillation column B
D_B	overhead product from column B
F	feed to column A
V_A	overhead vapour from column A
V_B	overhead vapour from column B
W_A	bottoms product from column A

Superscripts

L	liquid
T	transpose
V	vapour
-1	inverse
$'$	stripping section of column A

A.II.2 Sensitivity analysis

Section 2.1 reviewed methods due to Ripps (35) and Nogita (36) for detecting instrument malfunction when the process measurements were related by linear equations.

A least squares criterion was proposed to predict a consistent set of data $\underline{y}$ from a set of measured data $\underline{m}$. This was written:

$$\phi = \sum_{i=1}^{m} \left(\frac{y_i - m_i}{\sigma_i} \right)^2 \qquad \text{(A.II.2.1)}$$

It was assumed that the measurements were related by linear relationships yielding the constraint equations:

$$\psi_j = \sum_{i=1}^{m} a_{ji}\, y_i = 0 \qquad j = 1,2,\ldots.l \qquad \text{(A.II.2.2)}$$

Now equation (A.II.2.1) may be minimised subject to the constraints (A.II.2.2) by introducing Lagrange multipliers resulting in the solution:

$$\begin{bmatrix} \underline{\lambda} \\ \underline{y} \end{bmatrix} = \begin{bmatrix} 0 & A \\ A^T & S \end{bmatrix}^{-1} \begin{bmatrix} \underline{0} \\ \underline{b} \end{bmatrix} \qquad \text{(A.II.2.3)}$$

where: $\underline{\lambda}$ is an (lx1) vector of Lagrange multipliers

$\underline{y}$ is a (mx1) vector of the estimates of the process measurements

A is an (lxm) matrix of constraint equation coefficients

S is a (mxm) matrix with

$$s_{ii} = \frac{2}{\sigma_i^2} \qquad s_{ij} = 0 \qquad i \neq j$$

$\underline{b}$ is a (mx1) vector with $b_{i1} = \frac{2m_i}{\sigma_i^2}$

The solution given by equation (A.II.2.3) is valid only if the constraint coefficients a_{ji} are known with certainty or are assumed as such. However, in formulating the constraint equations (A.II.2.2) often

a_{ji} will be subject to uncertainty, as for example in the case of enthalpy coefficients in heat balances. It is then necessary to examine the sensitivity of the solution vector to this uncertainty in the coefficients a_{ji} and determine whether this invalidates the malfunction detection algorithm.

Equation (A.II.2.3) is simply the solution of a set of linear equations. In order to develop a sensitivity analysis a general set of linear equations may be written as:

$$D\underline{g} = \underline{f} \qquad \text{(A.II.2.4)}$$

where $D = \begin{bmatrix} 0 & A \\ A^T & S \end{bmatrix}$

$$\underline{g} = \begin{bmatrix} \underline{\lambda} \\ \underline{Y} \end{bmatrix}$$

$$\underline{f} = \begin{bmatrix} \underline{0} \\ \underline{b} \end{bmatrix}$$

If e(D) denotes the matrix of errors in the coefficients of matrix D then Dwyer (38) shows that the change in the solution vector $\underline{e}(\underline{g})$ is given by:

$$\underline{e}(\underline{g}) = D^{-1} (- e(D) D^{-1} \underline{f}) \qquad \text{(A.II.2.5)}$$

Now usually the actual value of e(D) is unknown but often a bound on each error is definable. In such cases it is possible to determine the maximum errors associated with each element g_i of the solution vector $\underline{g}$. Recall that $\underline{g} = D^{-1}\underline{f}$ and so equation (A.II.2.5) may be written:

$$\underline{e}\ (\underline{g}) = D^{-1} (- e(D)\ \underline{g}) \qquad \text{(A.II.2.6)}$$

For a two dimensional problem this equation written in full becomes:

$$\begin{bmatrix} e\ (g_1) \\ e\ (g_2) \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} -e\ (d_{11}) & -e(d_{12}) \\ -e\ (d_{21}) & -e(d_{22}) \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$$

where the matrix $L = D^{-1}$. Multiplying out gives:

$$e(g_1) = -l_{11}\, e(d_{11})\, g_1 - l_{11}\, e(d_{12})\, g_2 - l_{12}\, e(d_{21})\, g_1 - l_{12}\, e(d_{22})\, g_2 \qquad \text{(A.II.2.7)}$$

$$e(g_2) = -l_{21}\, e(d_{11})\, g_1 - l_{21}\, e(d_{12})\, g_2 - l_{22}\, e(d_{21})\, g_1 - l_{22}\, e(d_{22})\, g_2 \qquad \text{(A.II.2.8)}$$

Now by inspection of equation (A.II.2.7) it can be seen that the error in g_1, $e(g_1)$, may be maximised by choosing the signs of $e(d_{11})$, $e(d_{12})$, $e(d_{21})$ and $e(d_{22})$ to be identical with those of $-l_{11}g_1$, $-l_{11}g_2$, $-l_{12}g_1$ and $-l_{12}g_2$ respectively. This error is denoted as $e(g_1)_{g_1}$. When these values $e(d_{ij})$ are substituted into equation (A.II.2.8) they result in a value of $e(g_2)$, denoted $e(g_2)_{g_1}$, which corresponds to the maximum error $e(g_1)_{g_1}$. By similar reasoning, the values of $e(d_{ij})$ may be chosen to maximise the error in g_2 by examining equation (A.II.2.8) and choosing appropriate error signs. This results in an error $e(g_2)_{g_2}$ and a corresponding error in g_1, denoted $e(g_1)_{g_2}$, found by substituting the chosen errors $e(d_{ij})$ into equation (A.II.2.7).

These solution vector errors $e(g_i)_{g_j}$ may be combined to give a matrix of error coefficients of the general form:

$$E = \begin{bmatrix} e(g_1)_{g_1} & e(g_2)_{g_1} & \cdots & e(g_m)_{g_1} \\ e(g_1)_{g_2} & e(g_2)_{g_2} & & \vdots \\ \vdots & \vdots & & \vdots \\ e(g_1)_{g_m} & \vdots & & e(g_m)_{g_m} \end{bmatrix} \qquad \text{(A.II.2.9)}$$

A.II.3 Illustrative example: Distillation column mass and heat balance

This sensitivity analysis was applied to the data validation method of Ripps and Nogita by considering a mass and heat balance around an industrial distillation column situated at Works A.

A schematic diagram of the column is shown in Figure A.II.1.

The following notation is adopted in Figure A.II.1:

F^*_A = feed to column A

V_A = overhead vapour flowrate from column A

W^*_A = bottoms product from column A

Q_R^* = reboiler heat load for column A

V_B = overhead vapour flowrate from column B

L^*_B = reflux flowrate to column B

D^*_B = column B product flowrate

Q_C = column B condenser heat load

$W^*_B = L_A$ = bottom product flowrate from column B

Q_L = heat loss from both columns.

The measured process variables are denoted by an asterisk. Now the constraint equations relating the process measurements are formulated by considering material and heat balances around the column as follows:

1) Overall Mass Balance:

$$F_A = D_B + W_A \qquad \text{(A.II.3.1)}$$

2) Overall Heat Balance:

$$F_A H_F + Q_R = W_A H_{W_A} + D_B H_{D_B} + Q_C + Q_L \qquad \text{(A.II.3.2)}$$

where H_F = enthalpy of the feed

H_{W_A} = enthalpy of the liquid W_A

H_{D_B} = " " " " D_B

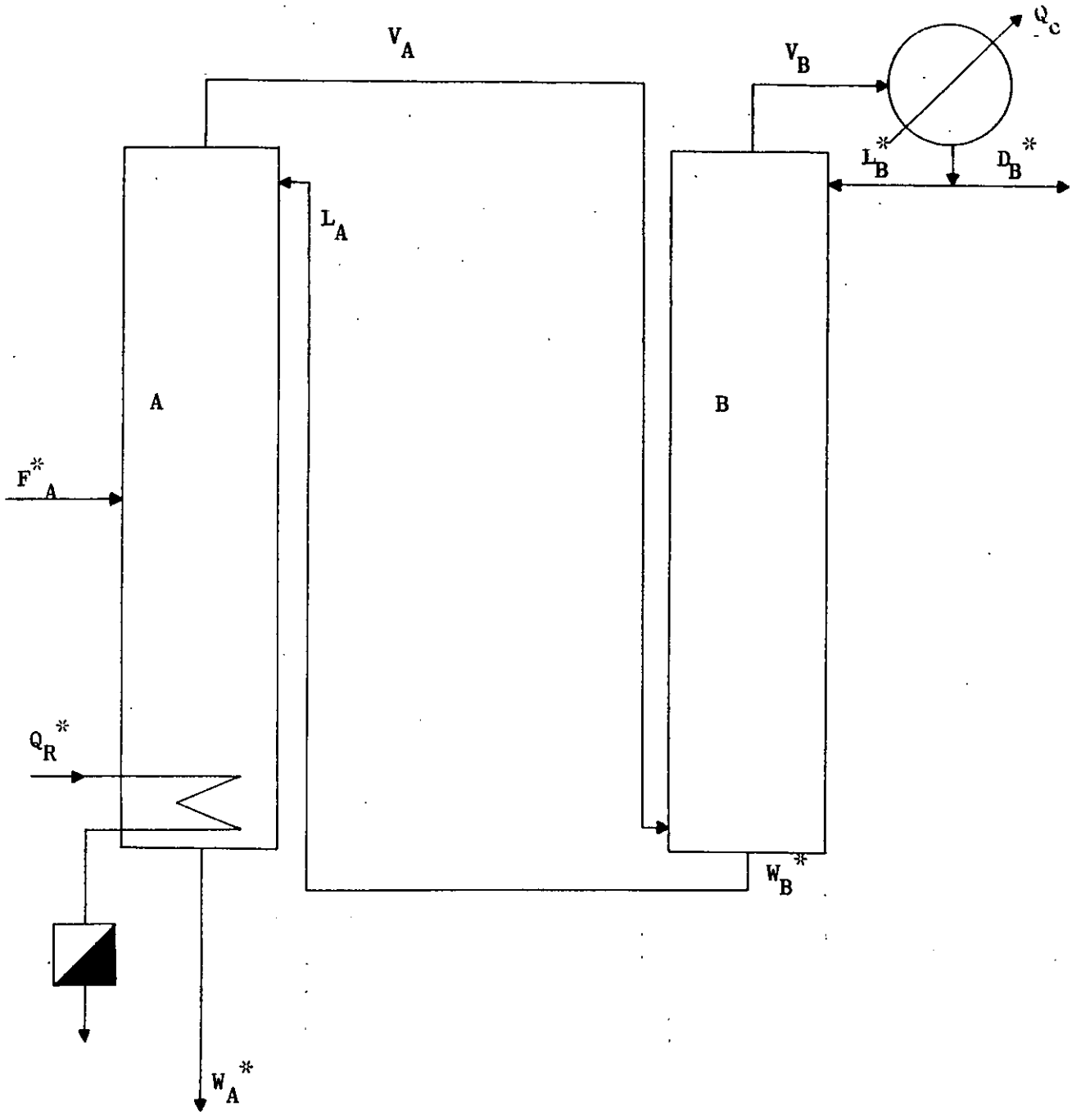


FIGURE A.II.1 Schematic illustration of distillation column.

* denotes a process measurement.

Now Q_C may be derived by considering a heat balance over the condenser, i.e.,

$$Q_C = V_B \lambda_{V_B} + V_B C_{p_{V_B}} (T_B^V - T) + V_B C_{p_L} (T - T_B^L)$$

where λ_{V_B} = latent heat of vaporisation of vapour V_B

$C_{p_{V_B}}$ = specific heat of vapour V_B

C_{p_L} = specific heat of condensed vapour V_B

T = boiling point of column B overhead liquid

T_B^V = temperature of vapour V_B

T_B^L = temperature of column B subcooled overhead liquid.

However, $V_B = L_B + D_B$ and

writing $C_{p_{V_B}} (T_B^V - T)$ as H_1 and $C_{p_L} (T - T_B^L)$ as H_2, then equation (A.II.3.2) becomes:

$$F_A H_F + Q_R = W_A H_{W_A} + D_B(H_{D_B} + \lambda_{V_B} + H_1 + H_2) + L_B(\lambda_{V_B} + H_1 + H_2) + Q_L \qquad \text{(A.II.3.3)}$$

3) Mass Balance on the vapour:

$$V_B = V_A' + (1 - q) F_A - \frac{Q_L}{\lambda_{V_B}} - \frac{L_B H_{D_B}}{\lambda_{V_B}}$$

$$\text{or } L_B + D_B = \frac{Q_R}{\lambda_{V_A'}} + (1 - q) F_A - \frac{Q_L}{\lambda_{V_B}} - \frac{L_B H_{D_B}}{\lambda_{V_B}} \qquad \text{(A.II.3.4)}$$

where $q = \dfrac{\text{heat of vaporise 1 mole of feed } F_A}{\text{molar latent heat of the feed } F_A}$

V_A' = vapour below the feed point in column A

$\lambda_{V_A'}$ = latent heat of vaporisation of vapour V_A'

4) Liquid balance on column B:

$$L_A = L_B + \frac{Q_L}{k\lambda_{V_B}} + \frac{L_B H_{D_B}}{\lambda_{V_B}}$$

but $L_A = W_B$

so $$W_B = L_B + \frac{Q_L}{k\lambda_{V_B}} + \frac{L_B H_{D_B}}{\lambda_{V_B}} \qquad \text{(A.II.3.5)}$$

where k is a constant denoting the proportion of the total heat loss from column B.

The constraint equations (A.II.3.1), (A.II.3.3), (A.II.3.4) and (A.II.3.5) may be written in the matrix form of equation (A.II.2.1) as

$$\underline{\Psi} = A\,\underline{Y} = \underline{0} \qquad \text{(A.II.3.6)}$$

where $A =$

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ H_F & -H_{W_A} & 1 & 0 & -(\lambda_{V_B}+H_1+H_2) & -(H_{D_B}+\lambda_{V_B}+H_1+H_2) & -1 \\ -(1-q) & 0 & \frac{-1}{\lambda_{V'_A}} & 0 & \left(1+\frac{H_{D_B}}{\lambda_{V_B}}\right) & 1 & \frac{1}{\lambda_{V_B}} \\ 0 & 0 & 0 & 1 & -\left(1+\frac{H_{D_B}}{\lambda_{V_B}}\right) & 0 & -\frac{1}{k\lambda_{V_B}} \end{bmatrix}$$

and $$\underline{Y} = \begin{bmatrix} F_A \\ W_A \\ Q_R \\ W_B \\ L_B \\ D_B \\ Q_L \end{bmatrix}$$

The important feature of these constraint equations is that many of the elements of matrix A are enthalpy terms which themselves are functions of temperature and pressure process measurements and therefore subject to some uncertainty.

In order to examine the effect of this uncertainty, the mass and heat balance was closed by choosing the following parameters.

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 1188.49 & -1584.55 & 1 & 0 & -4901.9 & -6089.29 & -1 \\ -1.34 & 0 & -0.2203\times10^{-3} & 0 & 1.2465 & 1 & 0.2073\times10^{-3} \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}$$

and a vector of true measurements:

$$\underline{y} = \begin{bmatrix} 194.712 & \text{lb.mole/hr} \\ 18.1 & \text{" "} \\ 0.138785 \times 10^{8} & \text{chu/hr} \\ 1934.72 & \text{lb.mole/hr} \\ 1934.72 & \text{" "} \\ 176.612 & \text{" "} \\ 0.352202 \times 10^{7} & \text{chu/hr} \end{bmatrix}$$

To demonstrate the use of Nogita's serial elimination algorithm several simulations were performed.

The parameters of the constraint matrix A were chosen to be correct and assumed constant, while the measurement noise standard deviations were selected to be approximately 5% of nominal measurements. This results in a vector of measurement standard deviations of:

$$\begin{bmatrix} 10 \\ 1. \\ 0.5 \times 10^{6} \\ 100 \\ 100 \\ 10 \\ 0.5 \times 10^{5} \end{bmatrix}$$

The measurement correlation matrix was assumed to be the unit matrix I, i.e.

$$\begin{bmatrix} 1. & & & 0. \\ & 1. & & \\ & & 1. & \\ 0. & & & 1. \end{bmatrix}$$

The experiments performed are given in Table A.II.1, where the measurement errors are defined in terms of a relative error:

$$e_i = \frac{y_i - m_i}{\sigma_i}$$

Simulation number	Predicted measurements in error	Predicted y_i	True value	Relative error	
				True	Found
M1	-	194.71	194.712	0	0
		18.1	18.1	0	0
		0.138786×10^8	0.138785×10^8	0	0
		1934.73	1934.72	0	0
		1934.73	1934.72	0	0
		176.611	176.612	0	0
		0.352202×10^7	0.352202×10^7	0	0
M2	1, 4, 5	194.473	194.712	- 6	- 6.023
		18.6	18.1	-0.5	0
		0.138655×10^8	0.138785×10^8	-0.003	-0.029
		1932.73	1934.72	0.2	0.18
		1932.73	1934.72	0.2	0.18
		175.87	176.612	0.1	0.027
		0.35222×10^7	0.352202×10^7	0	0
M3	1, 2, 4	194.8	194.712	- 6	- 6
		19.79	18.1	- 6	- 4.31
		0.138756×10^8	0.138785×10^8	-0.003	-0.0088
		1935.58	1934.72	0	0.0088
		1935.58	1934.72	0	0.0088
		175.01	176.612	0.1612	0
		0.352204×10^7	0.352202×10^7	0	0

TABLE A.II.1. Detection of measurement data inconsistencies on a distillation column.

The results presented in Table A.II.1 illustrate the power of Nogita's technique for detecting errors in process measurements, when the constraint equation coefficients are deterministically known. In the simulations M2 and M3, the method makes it clear which process measurements contain gross errors.

However, when there is uncertainty in the constraint equation coefficients a sensitivity analysis should be performed. Consider again the constraint equations defined in equation (A.II.3.6). Many of the A coefficients are enthalpies or latent heats which were estimated from process measurements of temperature and pressure. Of course these measurements themselves are subject to errors (both small random and gross errors). It was determined that a $2^{o}C$ temperature measurement error resulted in the following coefficient errors:

$$|\delta H| = 50 \quad \text{chu/lb.mole}$$

$$\left|\delta \frac{1}{\lambda}\right| = 0.000005 \quad \text{lb.mole/chu}$$

$$|\delta q| = 0.05$$

The sensitivity of the solution of equation (A.II.2.3) to these coefficients errors was determined by implementing equation (A.II.2.6) and examining the maximum errors in the solution vector using the method discussed in deriving equations (A.II.2.7) and (A.II.2.8). The results of two sensitivity analyses are given in Tables A.II.2 and A.II.3, and are presented as a matrix of solution vector errors, E, corresponding to the predicted measurement vector $\underline{y}$.

Table A.II.2 shows the E matrix when the enthalpy coefficients H_F, H_{W_A} and H_{D_B} of the constraint equations can each have an error of 50 chu/lb.mole.

1.63	0.015	-67516.3	-11.7	-11.7	1.62	1390
1.63	0.015	-67516.3	-11.7	-11.7	1.62	1390
-1.63	-0.015	67516.3	11.7	11.7	-1.62	-1390
-1.63	-0.015	67516.3	11.7	11.7	-1.62	-1390
-1.63	-0.015	67516.3	11.7	11.7	-1.62	-1390
1.63	0.015	-67516.3	-11.7	-11.7	1.62	1390
1.63	0.015	-67516.3	-11.7	-11.7	1.62	1390

TABLE A.II.2 Matrix of measurement error coefficients: H_F, H_{W_A}, H_{D_B} constraint coefficients in error.

This shows that the solution vector $\underline{y}$ is quite insensitive to the enthalpy errors, since the maximum solution vector errors are well within the measurement standard deviations. Hence the measurement error achieved in calculating enthalpies from process temperatures does not particularly hinder the operation of the gross measurement detection algorithm.

Table A.II.3, however, details the solution vector error matrix E when the constraint coefficient $^1/\lambda_{V_A}$, has an error of 0.000005 lb.mole/chu. The results show that the maximum errors in the solution vector are now quite significant. For example, the predicted measurement y_1 can have an apparent error of 25 lb.mole/hr which is 2.5 measurement standard deviations. These solution vector errors would manifest themselves as gross measurement errors when a set of heat and mass balance data are analysed using Nogita's data consistency check and hence yield an incorrect solution.

This feature is illustrated by examining the check method when the constraint equation coefficient $^1/\lambda_{V_A}$, has an "unknown" error of 0.000005.

$$\begin{bmatrix} 25.11 & 0.24 & -0.84 \times 10^6 & -201 & -201 & 24.9 & 0.2 \times 10^5 \\ 25.11 & 0.24 & -0.84 \times 10^6 & -201 & -201 & 24.9 & 0.2 \times 10^5 \\ -25.11 & -0.24 & 0.84 \times 10^6 & 201 & 201 & -24.9 & -0.2 \times 10^5 \\ -25.11 & -0.24 & 0.84 \times 10^6 & 201 & 201 & -24.9 & -0.2 \times 10^5 \\ -25.11 & -0.24 & 0.84 \times 10^6 & 201 & 201 & -24.9 & -0.2 \times 10^5 \\ 25.11 & 0.24 & -0.84 \times 10^6 & -201 & -201 & 24.9 & -0.2 \times 10^5 \\ 25.11 & 0.24 & -0.84 \times 10^6 & -201 & -201 & 24.9 & -0.2 \times 10^5 \end{bmatrix}$$

TABLE A.II.3 Matrix of measurement error coefficients: $^1/\lambda_{V_A}$, constraint coefficient in error.

The constraint equations, measurement standard deviations and parameters have been given earlier.

Table A.II.4 shows the results of two simulations.

In simulation S1 there were no process measurement errors. Nogita's algorithm knows nothing of the constraint equation uncertainty and processes the measurements in the usual manner. The final result is a suggestion that the measurements Q_R and Q_L contain gross errors, although the predicted value of the heat loss Q_L, is negative (i.e., there is a heat input to the distillation column), which is nonsense. However, this type of solution is not altogether valueless since it may be able to use such results as an inference of which constraint equation coefficient is in error.

The case of a gross measurement error and an uncertain constraint equation coefficient $^1/\lambda_{V_A}$, is more difficult to analyse and the simulation S2 in Table A.II.4 details the results of Nogita's algorithm. The method suggests measurements 1, 2 and 4 are in error, which is correct with respect to measurement y_1. However, the predicted measurements $\underline{y}$ are nonsense and would inspire little confidence in these suggestions.

Simulation number	Predicted measurements in error	Predicted y_i	True value	Relative error	
				True	Found
S1	3,7	194.7	194.712	0	0
		18.1	18.1	0	0
		0.100238×10^8	0.138785×10^8	0	-7.71
		1934.69	1934.72	0	0
		1934.69	1934.72	0	0
		176.6	176.612	0	0
		-3.3255×10^5	0.352202×10^7	0	-77.1
S2	1,2,4	146.43	194.712	-6	-10.83
		- 29.19	18.1	-0.5	-47.8
		0.138681×10^8	0.138785×10^8	-0.003	-0.024
		1937.36	1934.72	0	0.0026
		1937.36	1934.72	-0.003	0.0024
		175.622	176.612	0.1	0.0022
		0.352214×10^7	0.352202×10^7	0	0.0024

TABLE A.II.4 Detection of measurement data inconsistencies on a distillation column; subject to constraint equation uncertainty.

These results demonstrate that care is needed in making assumptions about the nature of the constraint equation coefficients. If these coefficients are not deterministic, then the particular application of the detection method will determine to what degree the performance is degraded. The sensitivity check derived earlier provides an insight into this phenomenon.

A.II.4 Process data validation check when the constraint equation coefficients are uncertain

The problem of uncertain constraint equation coefficients may be solved by treating these coefficients as additional process measurements with a corresponding measurement error standard deviation. In this case

the least squares criterion of equation (A.II.2.1), i.e.,

$$\phi = \sum_{i=1}^{m} \left(\frac{y_i - m_i}{\sigma_i} \right)^2 \qquad \text{(A.II.4.1)}$$

can still be minimised, but the constraint equations are now non linear, i.e., equation (A.II.2.2) becomes:

$$\psi_j = f_j(\underline{y}) = 0 \qquad j = 1,2,\ldots l \qquad \text{(A.II.4.2)}$$

The minimisation of equation (A.II.4.1) subject to the non linear constraints of equation (A.II.4.2) may be solved directly as an optimisation problem by using standard techniques, e.g. penalty function optimisation, or the problem may be transformed into the solution of a set of non linear simultaneous equations by introducing Lagrange multipliers. This latter approach is illustrated here.

In this example, the mass and heat balances around the distillation column are considered again, except that the uncertain coefficient $^1/\lambda_{V_A}$, is treated as a further process measurement, y_8, to be estimated. The set of constraint equations which are equivalent to equation (A.II.3.6) become :

$$\psi_1 = a_{11} y_1 - a_{12} y_2 - a_{16} y_6 = 0 \qquad \text{(A.II.4.3)}$$

$$\psi_2 = a_{21} y_1 - a_{22} y_2 + a_{23} y_3 - a_{25} y_5 - a_{26} y_6 - a_{27} y_7 = 0 \qquad \text{(A.II.4.4)}$$

$$\psi_3 = -a_{31} y_1 - y_3 y_8 + a_{35} y_5 + a_{36} y_6 + a_{37} y_7 = 0 \qquad \text{(A.II.4.5)}$$

$$\psi_4 = a_{44} y_4 - a_{45} y_5 - a_{47} y_7 = 0 \qquad \text{(A.II.4.6)}$$

and the vector of measurements to be predicted is :

$$\underline{y} = \begin{bmatrix} F_A \\ W_A \\ Q_R \\ W_B \\ L_B \\ D_B \\ Q_L \\ 1/\lambda_{V_A'} \end{bmatrix}$$

Now equation (A.II.1.1) is minimised subject to equation (A.II.4.2) by introducing the Lagrangian multipliers λ_1 to λ_4 which results in the following set of non linear simultaneous equations :

$$\frac{2}{\sigma_1^2}(y_1 - m_1) + a_{11}\lambda_1 + a_{21}\lambda_2 - a_{31}\lambda_3 = 0 \qquad \text{(A.II.4.7)}$$

$$\frac{2}{\sigma_2^2}(y_2 - m_2) - a_{12}\lambda_1 - a_{22}\lambda_2 = 0 \qquad \text{(A.II.4.8)}$$

$$\frac{2}{\sigma_3^2}(y_3 - m_3) + a_{23}\lambda_2 - y_8\lambda_3 = 0 \qquad \text{(A.II.4.9)}$$

$$\frac{2}{\sigma_4^2}(y_4 - m_4) + a_{44}\lambda_4 = 0 \qquad \text{(A.II.4.10)}$$

$$\frac{2}{\sigma_5^2}(y_5 - m_5) - a_{25}\lambda_2 + a_{35}\lambda_3 - a_{45}\lambda_4 = 0 \qquad \text{(A.II.4.11)}$$

$$\frac{2}{\sigma_6^2}(y_6 - m_6) - a_{16}\lambda_1 - a_{26}\lambda_2 + a_{36}\lambda_3 = 0 \qquad \text{(A.II.4.12)}$$

$$\frac{2}{\sigma_7^2}(y_7 - m_7) - a_{27}\lambda_2 + a_{37}\lambda_3 - a_{47}\lambda_4 = 0 \qquad \text{(A.II.4.13)}$$

$$\frac{2}{\sigma_8^2}(y_8 - m_8) - y_3\lambda_3 = 0 \qquad \text{(A.II.4.14)}$$

Given a set of process measurements m_i, measurement standard deviations σ_i, $i = 1, 8$ and the coefficients a_{ij} defined earlier, equations

(A.II.4.3) to (A.II.4.14) can be solved for the vector of unknown $(\underline{\lambda}, \underline{y})^T$.

The detection of gross measurement errors may be performed using Ripp's suggestion of discarding elements from the basic least squares criterion in equation (A.II.4.1), i.e.,

$$\phi = \sum_{i=1}^{m} \delta_i \left(\frac{y_i - m_i}{\sigma_i} \right)^2 \qquad \text{(A.II.4.15)}$$

where $\delta_i = 0$ if element i is discarded

$= 1$ otherwise.

The consequence of this new least squares criterion in equation (A.II.4.15) is simply to modify the first term in equations (A.II.4.7) to (A.II.4.14) to $\frac{2\delta_i}{\sigma_i^2}(y_i - m_i)$. Gross measurement errors are then determined by examining ϕ when measurements are discarded, and the minimum ϕ usually coincides with the gross measurement errors.

This technique is illustrated by using the same parameters as before with $\sigma_8 = 0.000005$, and repeating simulations S1 and S2, denoted NS1 and NS2 respectively. The set of non linear equations (A.II.4.13) to (A.II.4.14) were solved using Powell's method for minimising a sum of squares without calculating the derivatives.

Table A.II.5 shows how the least squares criterion ϕ varies as measurements are discarded from the problem formulation for runs NS1 and NS2.

The predicted data set corresponding to the minimum ϕ in each run is presented in Table A.II.6. Run NS1 shows that the method is able to cope with the uncertain constraint equation parameter $^1/\lambda_{V_A'}$ by estimating it as well as the other process measurements. Run NS2 contains both a gross measurement error in m_1 and an error in $^1/\lambda_{V_A'}$. The results

Simulation NS1		Simulation NS2	
Discarded measurements	ϕ	Discarded measurements	ϕ
0	0.8566	0	20.633
1	0.8387	1	0.820
2	1.3776	2	4.452
3	0.8526	3	20.495
4	0.8464	4	46.228
5	0.8464	5	20.630
6	0.8390	6	4.727
7	1.1063	7	20.379
8	0.000233	8	18.264
1 ; 8	0.000232	1 ; 2	0.00216
2 ; 8	0.486	1 ; 3	0.817
3 ; 8	0.13×10^{-14}	1 ; 4	0.810
4 ; 8	0.000176	1 ; 5	0.811
5 ; 8	0.000176	1 ; 6	0.113
6 ; 8	0.000232	1 ; 7	0.795
7 ; 8	0.139×10^{-18}	1 ; 8	0.879×10^{-4}

TABLE A.II.5 Variation of ϕ in non-linear distillation column problem.

Simulation number	Predicted measurements in error	Predicted y_i	True value	Relative error	
				True	Found
NS1	7 ; 8	194.712	194.712	0	0
		18.1	18.1	0	0
		0.138879×10^{8}	0.138785×10^{8}	-0.0188	0
		1934.72	1934.72	0	0
		1934.72	1934.72	0	0
		176.612	176.612	0	0
		0.353135×10^{7}	0.352202×10^{7}	0	0.186
		0.2203×10^{-3}	0.2203×10^{-3}	- 1	- 1
NS2	1 ; 8	194.207	194.712	- 6	- 6.05
		18.6	18.1	- 0.5	0
		0.138763×10^{8}	0.138785×10^{8}	- 0.003	- 0.0074
		1935.21	1934.72	0	0.0051
		1935.21	1934.72	- 0.003	0.0021
		175.607	176.612	0.1	0
		0.352206×10^{7}	0.352202×10^{7}	0	0
		0.22036×10^{-3}	0.2203×10^{-3}	- 1	-0.934

TABLE A.II.6 Detection of measurement data inconsistencies; non-linear distillation problem.

illustrate that the detection method determines these errors and also predicts a consistent, accurate data set. The results presented for runs NS1 and NS2 in Table A.II.6 are a vast improvement over those of Table A.II.4 where the uncertainty in $^{1}/\lambda_{V_A}$, was ignored.

This Appendix has considered a process measurement validation method which may be used to detect gross measurement errors. The method was applied to a typical distillation column and the effect of uncertainty in the process measurement linear constraint equation coefficients was examined. A method has been suggested whereby the sensitivity of the technique to these uncertainties may be examined, and it is suggested that this analysis should be performed to test the feasibility of the method in any particular application.

If this analysis reveals critical constraint coefficients the measurement validation check may still be performed by considering the uncertain coefficients as additional process measurements. This formulation results in an optimisation, subject to non-linear constraints, problem. The optimisation was solved here by using Lagrange multipliers to transform the problem into the solution a set of non-linear simultaneous equations, which were solved using Powell's method. The serial elimination method of Ripps was used to detect malfunction and simulations demonstrated the power of the technique.

APPENDIX III

MEHRA'S INNOVATION CORRELATION ADAPTIVE ESTIMATOR

This Appendix describes Mehra's algorithm (87) for adapting the Kalman filter to cope with uncertainty in a priori system information. The nomenclature is the same as that used in Chapter 3.

Mehra considers the system model described by:

$$\underline{x}(k+1) = A\,\underline{x}(k) + \Gamma\,\underline{w}(k) \qquad \text{(A.III.1)}$$

$$\underline{y}(k) = H\,\underline{x}(k) + \underline{v}(k) \qquad \text{(A.III.2)}$$

The noise sequences $\{\underline{w}(k)\}$ and $\{\underline{v}(k)\}$ are assumed to be zero mean Gaussian, white noise sequences with covariances.

$$E(\underline{w}(k)\,\underline{w}^T(j)) = Q\,\delta(k,j)$$

$$E(\underline{v}(k)\,\underline{v}^T(j)) = R\,\delta(k,j)$$

$$E(\underline{w}(k)\,\underline{v}^T(j)) = 0 \qquad \text{for all } k,j \text{ and } Q \text{ and } R \text{ are}$$

bounded positive definite matrices.

The system considered above is also assumed to be completely controllable and observable.

Now Mehra examines the steady state Kalman filter, as guaranteed by the constraint of system controllability and observability, for which the filtering equations are:

$$\hat{\underline{x}}(k/k) = \hat{\underline{x}}(k/k-1) + K(\underline{y}(k) - H\,\hat{\underline{x}}(k/k-1)) \qquad \text{(A.III.3)}$$

$$K = PH^T(HPH^T + R)^{-1} \qquad \text{(A.III.4)}$$

$$P = A(I - KH)\,P\,(I - KH)^T A^T + AKRK^T A^T + \Gamma Q \Gamma^T \qquad \text{(A.III.5)}$$

In equations (A.III.4) and (A.III.5) Q and R represent approximations to the true noise covariance matrices. For this case, the covariance matrix of the innovation sequence is given by:

$$C_k = E(\underline{V}(i+k)\,\underline{V}^T(i))$$

$$= HPH^T + R \qquad k = 0$$

$$= H(A(I - KH))^{k-1} A(PH^T - K C_0) \qquad k > 0 \qquad \text{(A.III.6)}$$

Since the autocorrelation function is not a function of time (i), then the innovation sequence is stationary, and by virtue of linearity Mehra shows it is also Gaussian.

The testing scheme proposed by Mehra is based upon the criterion that the innovation sequence be white for the Kalman filter to be optimal. This means that if C_k is estimated then the filter is optimal if $\hat{C}_k$ is approximately zero for $k > 0$. The estimate of C_k is formed as:

$$\hat{C}_k = \frac{1}{N} \sum_{i=1}^{N-k} \underline{\nu}(i+k)\,\underline{\nu}^T(i) \qquad \text{(A.III.7)}$$

with N, the number of sample points assumed to be large.

The elements of the matrix of the normalized autocorrelation coefficients are estimated by:

$$(\hat{\rho}_k)_{ij} = \frac{(\hat{C}_k)_{ij}}{((\hat{C}_o)_{ii}\,(\hat{C}_o)_{jj})^{\frac{1}{2}}}$$

where $(\hat{\rho}_k)_{ij}$ is the ij^{th} element of the matrix $\hat{\rho}_k$. It can be shown that the probability distribution for $(\hat{\rho}_k)_{ii}$ is asymptotically Normal. Thus the 95% confidence limits for $(\hat{\rho}_k)_{ii}$ for $k > 0$ are $\pm (1.96/\sqrt{N})$. The test for filter optimality therefore becomes:

> Compute $(\hat{\rho}_k)_{ii}$, $k > 0$, for N sample points. If less than 5% of the k values fall outside the band formed by the 95% confidence limits, the sequence $\underline{\nu}(i)$ is white and the filter is optimal.

If the above test is not passed, this indicates that the Kalman gain used in the filter equation is incorrect. To correct the filter, Mehra generates new estimates for Q and R. However, it is first necessary to estimate PH^T.

Now, using equation (A.III.6), Mehra shows that an estimate of

PH^T is given by

$$\widehat{PH}^T = K\,\hat{C}_o + L^{\#\#} \begin{bmatrix} \hat{C}_1 \\ \vdots \\ \hat{C}_n \end{bmatrix} \qquad \text{(A.III.8)}$$

where $L^{\#\#}$ is the pseudo inverse of

$$L = \begin{bmatrix} HA \\ HA(I - KH)\,A \\ \vdots \\ H\,(A\,(I - KH))^{n-1}\,A \end{bmatrix}$$

and $L^{\#\#} = (L^T L)^{-1} L^T$

The estimate of R is given by:

$$\hat{R} = \hat{C}_o - H\,(\widehat{PH}^T) \qquad \text{(A.III.9)}$$

To estimate Q, equations (A.III.9) and (A.III.5) are used, although the latter equation cannot be used directly since an estimate of P is not available. Equation (A.III.5) may be rewritten as:

$$P = APA^T + M + \Gamma Q \Gamma^T \qquad \text{(A.III.10)}$$

where $M = A(KC_oK^T - KHP - PH^TK^T)\,A^T$

Now the expression for P is substituted into the right hand side of equation (A.III.10). The process is repeated n-1 times and the following set of equations is generated:

$$\sum_{j=o}^{k-1} A^j \Gamma Q \Gamma^T (A^j)^T = P - A^k\,P(A^k)^T - \sum_{j=o}^{k-1} A^j\,M(A^j)^T$$

$$k=1,\ldots,n \qquad \text{(A.III.11)}$$

If equation (A.II.11) is premultiplied by H and postmultiplied by $(A^{-k})^T H^T$, the result is:

$$\sum_{j=o}^{k-1} HA^j \Gamma Q \Gamma^T (A^{j-k})^T H^T = HP (A^{-k})^T H^T - HA^k P H^T - \sum_{j=o}^{k-1} HA^j M (A^{j-k})^T H^T$$

$$k=1,\ldots,n \qquad \text{(A.III.12)}$$

The right hand side of equation (A.III.12) can be computed if PH and C_o are known. If the estimates of $\hat{PH}^T$ and $\hat{C}_o$ are used in equation (A.III.12) then nm equations will be available to solve for the q^2 unknown elements of Q. Clearly if $q^2 > nm$, a unique solution cannot be found. If the $q^2 < nm$, a solution can be obtained but first a linearly dependent subset of equations (A.III.12) must be chosen. The Kalman filter may then be adapted using the computed values of $\hat{Q}$ and $\hat{R}$.

The adaptation scheme derived above is only applicable if Q has nm or fewer unknown elements. If Q has more than nm unknowns or unknown structure, then Mehra shows it is still possible to estimate an optimal Kalman gain $\hat{K}_{opt}$.

In his paper Mehra describes two related iterative procedures for calculating $\hat{K}_{opt}$. He proves that, for steady state filtering, the successive error covariance matrices (obtained at each iteration) converge, i.e. $P_{j+1} < P_j$ and $P_j > 0$, and in conjunction the corresponding sequence of Kalman gain K_j converge to some K_{opt}, i.e.,

$$\lim_{j \to \infty} K_j = K_{opt}$$

One gain estimation scheme proceeds as follows.

i) Let K_o be initial or a priori gain of the filter. Obtain an estimate of K_1 using equations (A.III.4) and (A.III.8)

$$\hat{K}_1 = K_o + L^{\#} \begin{bmatrix} \hat{C}_1 \\ \hline \vdots \\ \hline \hat{C}_n \end{bmatrix} \hat{C}_o^{-1} \qquad \text{(A.III.13)}$$

and obtain $\hat{P_1 H^T}$ and $\hat{R}$ from equations (A.III.8) and (A.III.9).

ii) Define $\delta P_1 = P_2 - P_1$ and then obtain an estimate:

$$\hat{\delta P_1} = A(I - \hat{K}_1 H)\ \hat{\delta P_1}(I - \hat{K}_1 H)^T A^T - A(\hat{K}_1 - K_o)\ \hat{C}_o(\hat{K}_1 - K_o)^T A^T$$

where P_j $j = 1, 2$ are the steady state error covariance matrices using K_o and $\hat{K}_1$ for the filter gains.

$\hat{\delta P_1}$ is the estimate of δP_1 using $\hat{K}_1$

iii) Obtain $\hat{P_2 H^T}$ and $\hat{K}_2$ as follows:

$$\hat{P_2 H^T} = \hat{P_1 H^T} + \hat{\delta P_1}\ H^T$$

$$\hat{K}_2 = \hat{P_2 H^T}\ (H\ \hat{P_2 H^T} + \hat{R}\ )^{-1}$$

iv) Repeat steps ii) and iii) until $\left\| \hat{\delta P_i} \right\|$ or $\left\| \hat{K}_i - \hat{K}_{i-1} \right\|$ become small with respect to $\left\| \hat{P}_i \right\|$ or $\left\| \hat{K}_i \right\|$. $\|.\|$ refers to a suitable matrix norm.

An alternative way of determining $\hat{K}_2$ would be to refilter the data sequence ($\mathbf{Y}(k)$) using $\hat{K}_1$ and then using equation (A.III.13) obtain a new $\hat{K}_2$. This procedure may then be repeated until convergence is obtained.

This second measurement refiltering method was used in the present work.

APPENDIX IV

FRIEDLAND'S BIAS ESTIMATOR

The nomenclature used in this Appendix has been defined in Chapter 3.

In the application of the Kalman filter an accurate model of the process dynamics and observation is required. If this is not possible, then unknown parameters may be added to the original state vector. The filter then estimates these unknown parameters or bias terms as well as the original state vector. Although this method is reasonably effective when the number of bias terms is relatively small, a problem arises when the number of bias terms is comparable to the number of state variables of the original problem because of the larger problem dimension. Friedland's paper (109) was thus motivated by the need for a method whereby the numerical inaccuracies introduced by computations with large vectors and matrices could be avoided.

Although Friedland considers linear continuous time filtering, only the discrete time case will be reviewed here.

Friedland writes the system models as:

$$\underline{x}\,(k+1) = A(k+1,\,k)\,\underline{x}(k) + A_b(k+1,k)\,\underline{b}(k) + \underline{w}(k) \qquad \text{(A.IV.1)}$$

$$\underline{b}\,(k+1) = \underline{b}\,(k) \qquad \text{(A.IV.2)}$$

$$\underline{y}\,(k) = H(k)\,\underline{x}(k) + H_b(k)\,\underline{b}\,(k) + \underline{v}\,(k) \qquad \text{(A.IV.3)}$$

The matrices A_b and H_b determine how the components of the bias vector enter into the dynamics and observations respectively and represent the general case.

Now an augmented state variable model may be written from equations (A.IV.1) - (A.IV.3) as:

$$\underline{z}\,(k+1) = F(k+1,k)\,\underline{z}\,(k) + G\,\underline{w}\,(k)$$

$$\underline{y}\,(k) = L(k)\,\underline{z}\,(k) + \underline{v}\,(k)$$

where $$\underline{z}\,(k) = \begin{bmatrix} \underline{x}\,(k) \\ \hline \underline{b}\,(k) \end{bmatrix}$$

$$F(k+1,k) = \left[\begin{array}{c|c} A\ (k+1,\ k) & A_b\ (k+1,\ k) \\ \hline 0 & I \end{array}\right]$$

$$G = \begin{bmatrix} I \\ \hline 0 \end{bmatrix}$$

$$L(k) = \left[H\ (k) \;\vdots\; H_b\ (k)\right]$$

Direct application of the Kalman filter of section 3.4 results in:

$$\hat{\underline{z}}(k/k) = F(k,k-1)\ \hat{\underline{z}}(k-1/k-1) + K(k)(\underline{y}(k) - L(k)\ F(k,k-1)\ \hat{\underline{z}}(k-1/k-1))$$

$$K(k) = P(k/k-1)\ L^T(k)\ (L(k)\ P(k/k-1)\ L^T(k) + R(k))^{-1}$$

$$P(k/k-1) = F(k,k-1)P\ (k-1/k-1)F^T(k,k-1) + G\ Q\ (k-1)\ G^T \qquad \text{(A.IV.4)}$$

The error covariance matrix $P(k/k-1)$ may be partitioned as:

$$P(k/k-1) = \left[\begin{array}{c|c} P_x\ (k/k-1) & P_{xb}\ (k/k-1) \\ \hline P_{xb}^T(k/k-1) & P_b\ (k/k-1) \end{array}\right]$$

where:

$P_x\ (k/k-1)$ = variance of the original state $\underline{x}$

$P_b\ (k/k-1)$ = variance of the bias $\underline{b}$

$P_{xb}(k/k-1)$ = covariance of $\underline{x}$ and $\underline{b}$

Now by partitioning of the other matrices in a similar fashion and using a transformation of the variance equation (A.IV.4), Friedland derives the following recursive algorithm for estimating the bias $\underline{b}(k)$. The subscript x refers to matrices calculated by ordinary bias-free Kalman filtering.

$$U(o) = 0 \ : \ M(o) = P_b(o) \ : \ \underline{b}\ (o) = 0$$

$$S(k) = H(k)\ U(k) + H_b\ (k)$$

$$V(k) = U(k) - K_x\ (k)\ S(k)$$

$$M(k+1) = M(k) - M(k)\, S^T(k)\, (H(k)\, P_x\, (k/k-1)\, H^T(k) + R\,(k) + S(k)\, M(k)\, S^T\,(k))^{-1}\, S(k)\, M(k)$$

$$K_b\,(k) = M(k+1)\, (V^T(k)\, H^T\,(k) + H_b(k))\, R^{-1}\,(k)$$

$$\hat{\underline{b}}(k/k) = (I - K_b(k)\, S(k))\, \hat{\underline{b}}(k-1/k-1) + K_b(k)\, (\underline{y}(k) - H(k)\, \hat{\underline{x}}(k/k-1))$$

$$U(k+1) = A(k+1,k)\, V(k) + A_b(k+1,\, k)$$

In this algorithm $K_x(k)$, $P_x(k/k-1)$ and $\hat{\underline{x}}(k/k-1)$ are calculated from the usual Kalman filter as if there were no bias $\underline{b}(k)$ present.

Thus the result of Friedland's variance transformation is that the problem of estimating the state $\underline{x}$ in the presence of a constant but unknown bias $\underline{b}$ is decomposed into two separate tasks. The first part is the standard Kalman filter algorithm that generates an estimate of the state vector $\hat{\underline{x}}(k/k)$ as if there were no bias present. The second filter due to Friedland estimates the bias vector $\hat{\underline{b}}(k/k)$ from the innovation sequence generated by the "bias-free" Kalman filter.

Finally, Friedland shows that the "bias-free" estimate of the state $\hat{\underline{x}}(k/k)$ and the estimated bias $\hat{\underline{b}}(k/k)$ may be combined to form an optimum state estimate $\hat{\underline{x}}'(k/k)$ according to:

$$\hat{\underline{x}}'(k/k) = \hat{\underline{x}}(k/k) + V(k)\, \hat{\underline{b}}(k/k)$$

Recent extensions of Friedland's technique to cover time varying bias and randomly varying bias estimation have been given in references (110), (111).

APPENDIX V

MARKOV RELIABILITY MODELS

A.V.1 List of symbols

k_1) k_2)	constants in hazard rate equation	$\frac{\text{failures}}{(\text{unit time})^2}$
$P_{s_i}(t)$	probability of being in state s_i at time t.	-
P_{ij}	transition probability from state i to j.	-
R(t)	system reliability.	-
s_i	i^{th} system state.	-
T	random variable associated with time	time units
t	time	" "
Δt	time increment	" "
t_m	time on time scale t	" "
X	random variable associated with the system state.	-
z(t)	hazard rate	$\frac{\text{failure}}{\text{unit time}}$

Greek letters

λ	failure rate	$\frac{\text{failure}}{\text{unit time}}$

Subscripts

0, 1, 2	state 0, 1 and 2
i	state i
j	state j

A.V.2 Markov processes (118), (120)

Markov models play a central role in reliability theory. Markov models are a function of two random variables, the state of the system X and the time of observation T. Reliability theory is usually concerned with the discrete-state continuous time Markov process.

A Markov model is defined by a set of probabilities P_{ij} which define the probability of transition from any state i to any state j. The important feature of a Markov process is that the future states of the process depend only on its immediate past history, i.e. the transition probability P_{ij} depends only on states i and j and is completely independent of all past states except the last one, state i.

In order to formulate a Markov model it is first necessary to define all the mutually exclusive states of the system. For example, in a system composed of a single equipment there are two possible states: $s_o = X_1$, the element is good, $s_1 = \bar{X}_1$, the element is failed. The states of the system at time $t = 0$ are called initial states and those representing a final state are called final states. The set of Markov state equations describes the probabilistic transitions from the initial to the final states and the transition probabilities must obey the following rules:

i) The probability of transition in time Δt from one state to another is given by $z(t)\,\Delta t$ where $z(t)$ is the hazard associated with the two states in question.

ii) The probabilities of more than one transition in time Δt are negligible.

Although Markov models have been extensively used some conceptual difficulties were encountered in the course of this work.

Consider a single equipment with an overall constant hazard rate λ_o failure/unit time.

Defining s_o and s_1 as good and failed states, $P_{s_o}(t + \Delta t)$ and $P_{s_1}(t + \Delta t)$ as the probability of being in states s_o and s_1 at time $t + \Delta t$, and $P_{s_o}(t)$ and $P_{s_1}(t)$ as the probability of being in state s_o and s_1 at time t then the Markov state equations are:

$$P_{s_o}(t + \Delta t) = (1 - \lambda_o \Delta t)\, P_{s_o}(t) \qquad \text{(A.V.2.1)}$$

$$P_{s_1}(t + \Delta t) = \lambda_o \Delta t\, P_{s_o}(t) + P_{s_1}(t) \qquad \text{(A.V.2.2)}$$

These equations yield, in the limit, the first order differential equations

$$\frac{dP_{s_o}}{dt}(t) + \lambda_o P_{s_o}(t) = 0 \qquad \text{(A.V.2.3)}$$

$$\frac{dP_{s_1}}{dt}(t) = \lambda_o P_{s_o}(t) \qquad \text{(A.V.2.4)}$$

Equations (A.V.2.3) and (A.V.2.4) may be solved using Laplace transforms with the initial conditions $(P_{s_o}(o),\ P_{s_1}(o))$ to yield:

$$P_{s_o}(t) = P_{s_o}(o)\, e^{-\lambda_o t} \qquad \text{(A.V.2.5)}$$

and $$P_{s_1}(t) = 1 - P_{s_o}(o)\, e^{-\lambda_o t} \qquad \text{(A.V.2.6)}$$

Now the reliability of the single equipment is the probability of being in state s_o i.e.,

$$R(t) = P_{s_o}(t) = P_{s_o}(o)\, e^{-\lambda_o t} \qquad \text{(A.V.2.7)}$$

Suppose the single equipment is now considered to have 3 states, i.e., s_o is the good state, s_1 is an unhealthy state (equivalent to a degrading equipment) and s_2 is the failed state. Then the Markov

graph becomes (118):

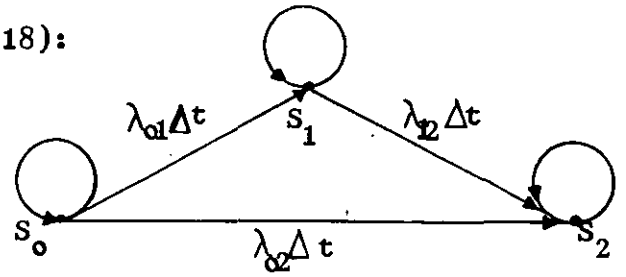


FIGURE A.V.1 Markov graph for an equipment with a single degraded state.

In this process $\lambda_{o1}\,\Delta t$ is the probability of the equipment moving from a good to a degraded state, $\lambda_{12}\,\Delta t$ is the probability of moving from the degraded to the failed state and $\lambda_{o2}\Delta t$ is the probability of the equipment failing directly.

The Markov state equations are:

$$P_{s_o}(t+\Delta t) = (1 - (\lambda_{o_2} + \lambda_{o_1})\Delta t)\, P_{s_o}(t) \qquad \text{(A.V.2.8)}$$

$$P_{s_1}(t+\Delta t) = \lambda_{o_1}\Delta t\, P_{s_o}(t) + (1 - \lambda_{12}\Delta t)\, P_{s_1}(t) \qquad \text{(A.V.2.9)}$$

$$P_{s_2}(t+\Delta t) = \lambda_{o_2}\Delta t\, P_{s_o}(t) + \lambda_{12}\Delta t\, P_{s_1}(t) + P_{s_2}(t) \qquad \text{(A.V.2.10)}$$

Solving the resultant differential equations, subject to the initial conditions $P_{s_o}(o)$, $P_{s_1}(o)$ and $P_{s_2}(o)$ gives:

$$P_{s_o}(t) = P_{s_o}(o)\,(1 - e^{-(\lambda_{o1}+\lambda_{o2})t}) \qquad \text{(A.V.2.11)}$$

$$P_{s_1}(t) = P_{s_1}(o)\, e^{-\lambda_{12}t} + \frac{P_{s_o}(o)\lambda_{o_1}}{(-\lambda_{12}+\lambda_{o1}+\lambda_{o2})}\left(e^{-\lambda_{12}t} - e^{-(\lambda_{o1}+\lambda_{o2})t}\right) \qquad \text{(A.V.2.12)}$$

$$P_{s_2}(t) = 1 - \sum_{i=o}^{1} P_{s_i}(t) \qquad \text{(A.V.2.13)}$$

The equipment reliability may be defined as the probability of not being in state s_2, i.e.

$$R(t) = 1 - P_{s_2}(t) = P_{s_o}(t) + P_{s_1}(t)$$

$$= P_{s_o}(o)\,(1 - e^{-(\lambda_{o1}+\lambda_{o2})t}) + P_{s_1}(o)\,e^{-\lambda_{12}t}$$

$$+ \frac{P_{s_o}(o)\,\lambda_{o1}}{(-\lambda_{12}+\lambda_{o1}+\lambda_{o2})}\left(e^{-\lambda_{12}t} - e^{-(\lambda_{o1}+\lambda_{o2})t}\right) \qquad \text{(A.V.2.14)}$$

Now the three state single equipment model is characterised by the hazard rates λ_{o1}, λ_{o2}, λ_{12} but how these coefficients are specified is not trivial. Intuitively, it may be postulated that λ_{o2} represents the number of failures/unit time which do not fail through a degraded state. Also it may be expected that $\lambda_{12} > \lambda_{o1}$ or λ_{o2} since this represents the failure rate of a degraded equipment and represents a wearing out phase of the equipment.

However, the fact that the single equipment has been represented by a three state Markov model should not affect the overall equipment failure rate and reliability as given in equation (A.V.2.7), thus equation (A.V.2.14) should reduce to this.

Unfortunately this is impossible because the 3 state formulation is fundamentally a different model. For example, an analogy is two stirred tanks in series. No matter how the time constants are chosen for the tanks, the dynamic response will never be the same as a single stirred tank although the responses can be very close. By appropriate selection of λ_{o1}, λ_{o2} and λ_{12} it is possible to achieve the same overall equipment failure rate λ_o but the reliability functions will never be quite equal (except when $\lambda_o = \lambda_{o1} + \lambda_{o2}$ and $\lambda_{12} = \infty$).

Thus the introduction of additional Markov states to represent the degradation of an equipment yields a model which does not satisfy intuitive conceptions of single equipment reliability.

In the above formulations, it was assumed that the equipment hazard rates were constants. Although the Markov formulation can handle time varying hazard rates (118) of the type shown in Figure A.V.2a, the method breaks down for the hazard shown in Figure A.V.2b.

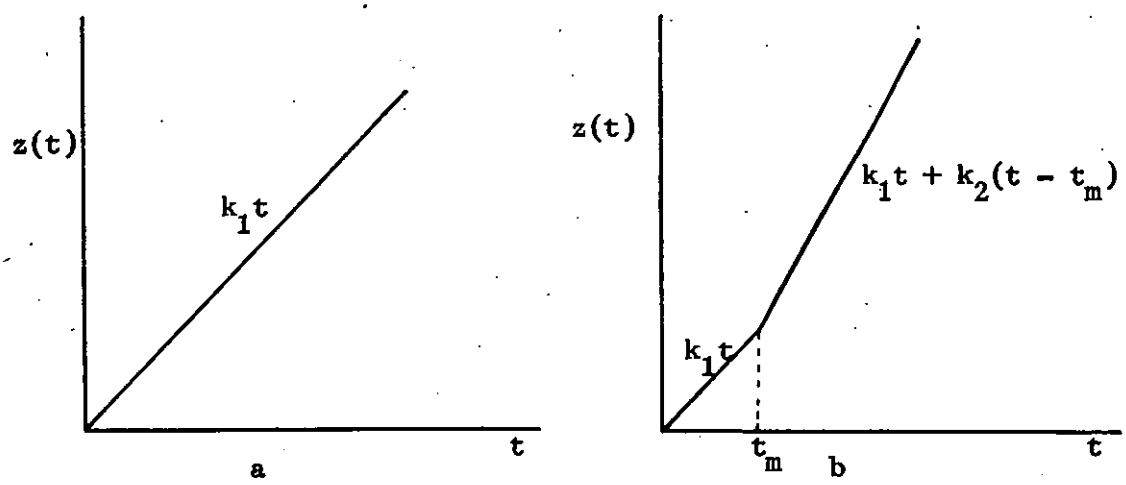


FIGURE A.V.2 Time dependent hazard models.

For the model shown in Figure A.V.2b, the hazard for $t > t_m$ is $k_1 t + k_2(t - t_m)$, obviously a function of t_m. Thus the state probabilities will be functions of t and t_m and the technique is no longer directly applicable. Shooman (118) has discussed this problem in some detail, and suggests that in such situations the analysis should be performed using a joint density function or compound events approach. The joint density function method is particularly convenient for incorporating information from a malfunction monitor into the analysis, whereas the use of such information with a Markov process is not so well defined.