Control Loop Performance Monitoring

Christopher B. Lynch

B.A.Sc. University of Waterloo, Waterloo, 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL ENGINEERING

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December, 1992

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of _ELECTRICAL ENGINEERING_

The University of British Columbia
Vancouver, Canada

Date _DECEMBER 14, 1992_

Abstract

Control loop monitoring provides the opportunity to maintain high process performance. In this thesis, the regulation performance of a process is monitored by examining the time delay, static input-output relation, and the minimum achievable output variance of the process. The time delay estimate is obtained from the Fixed Model Variable Regressor Estimator and the static input-output relation is estimated using the Adaptive Nonlinear Modeller. The minimum achievable output variance of the process is determined by modelling the closed loop noise transfer function by a Laguerre network and finding the Laguerre gains of the network using Recursive Extended Least Squares estimation. An extensive simulation study was performed to examine the operation, usefulness and limitations of the monitoring tools. These simulations confirmed the operation and benefit of the monitoring methods, and defined their limitations. The monitoring tools were applied to data generated by industrial processes. The successful application to industrial processes highlighted the benefits obtainable by control loop monitoring. Due to the ease of applying the monitoring tools and the valuable insight they provide, these monitoring tools should be applied to any process where regulation performance is a concern.

Table of Contents

iv

List of Tables

List of Figures

ix

Acknowledgments

Chapter 1

Introduction

1.1 Background

A serious problem faced upon the completion of any engineering research is its acceptance and use by industry. This problem is very pronounced in the field of control engineering [7]. The reasons are numerous and the blame lies with both industry and the research community. The bottom line in industry is production, which dictates that processes and therefore controllers operate continuously, with as few disruptions as possible. This bottom line will allow promising control methods to pass by untried for fear of disturbing operation, even though process performance could be improved. Universities, however, must realize the needs and restrictions of industry and gauge research accordingly. Only once industry and universities work together will new research be integrated into industrial processes.

Control loop monitoring is a way to bring both sides closer together. Monitoring in its most basic form involves observing process operation, recording behavior and commenting on any unusual occurrence. It is very passive, in the sense that no changes are required to the setup of the loop, so process operation is not disturbed. Meanwhile, monitoring offers the research community an opportunity to use various identification and modelling techniques in an industrial environment. Thus industry is introduced to alternative methods and new research, while universities are given the opportunity to interact with industry. Successful results will demonstrate the usefulness of new methods and should help encourage its acceptance and possible use in control algorithms. Any unsuccessful results can be examined by researchers to determine problems, all without adversely affecting the process operation in any manner. Such cooperation would ultimately result in the fulfillment of the desires of both industry and the research community.

A pulp and paper mill can have thousands of control loops [8] which typically are the responsibility of one or two process control engineers. With this large number of loops, it is next to impossible to ensure that they all are operating satisfactorily and it is time consuming to investigate each loop

1

individually. This suggests that there is a requirement for methods that could detect which loops need the attention of the engineer [23]. The monitoring methods presented in this thesis can be used to flag loops whose performance has deteriorated. The cost of implementing control loop monitoring is small, as it requires only the addition of the software code to manipulate the process signals. It is this ease of implementation and the need of such methods in an industrial setting, that makes control loop monitoring an attractive way to improve process performance.

1.2 Motivation

An industry wide problem in the control of pulp and paper processes is that in practice many control principles are ignored. As a result, many loop controllers including low level single-input single-output controllers, are often poorly chosen and tuned [4]. If these low level processes operate poorly due to the control, the overall operation of the process will suffer. One way of improving the operation of these controllers is through monitoring the process.

The goal of control loop monitoring is to provide information about a process during closed loop operation. This information can be used to confirm if the process is operating as expected and also highlight any changes in the process. If the process is not operating as expected, then changes to the way it is controlled may be necessary to improve performance. By highlighting changes in the process over time, changes which often signal deteriorating operation, maintenance of the loop components can be undertaken to improve performance.

1.3 Contributions

Three properties of single-input single-output control loops are monitored. The time delay of the process is monitored using the Fixed Model Variable Regressor Estimator proposed by Elnaggar [12]. The static input-output relationship of the process is determined using the Adaptive Nonlinear Modeller proposed by Åström [3] but the relationship is estimated using the Recursive Least Squares method. Based on the work of Harris [17], the minimum achievable output variance of the control loop is monitored. In this work, the delay is not assumed known and Laguerre filters are used with Recursive Extended Least Squares estimation for process modelling.

2

1.4 Thesis Outline

The types of processes considered and the notation to be used are outlined in Chapter 2. As well, the purpose of monitoring the three properties chosen is discussed and the problems associated with closed loop monitoring are outlined. Chapter 3 describes the theory behind the Fixed Model Variable Regressor Estimator, lists its algorithm and presents some examples of its operation. The problems encountered when using the estimator on closed loop stochastic processes are discussed. The Adaptive Nonlinear Modeller used for the estimation of the static input-output relationship is studied in Chapter 4. The various components required for determining the minimum achievable output variance are presented in Chapter 5. This includes discussion of Laguerre filters, Recursive Extended Least Squares estimation and the method used to calculate the minimum achievable output variance. Chapter 6 contains numerous computer simulations of the various techniques studied. Practical aspects of the techniques are examined via computer simulation in order to gain further insight on their operation. Then, Chapter 7 presents the results when monitoring is used on industrial process data. Finally, in Chapter 8, conclusions are drawn and summarized based on previous results.

Chapter 2

Control Loop Monitoring

In this chapter, the foundation for the research is set, providing a starting point for the remaining chapters. First, the model used to describe an industrial process is given and so then the notation to be used is defined. Next, the assumed operating setup of the process is given, which is followed by a discussion on the importance of monitoring and the reason why the specific properties are monitored. Finally, this discussion includes benefits to be achieved from monitoring the chosen process properties. The importance of monitoring and the approach taken in this research will then have been presented.

2.1 The Discrete Linear Model

The discrete linear model which is used to describe processes consists of two terms, one which represents the plant dynamics and another which accounts for the influence of the surroundings on the plant. These effects are denoted separately by a single term and combined together because the model used is linear. A block diagram form of such a model appears in Figure 2.1 and represents the plant and disturbance of a process. The equation:

$$y_o(t) = \frac{B(q^{-1})}{A_1(q^{-1})} q^{-k} u(t) \qquad (2.1)$$

completely describes the plant and its properties. The plant output is $y_o(t)$, its input $u(t)$, and the polynomials $A_1(q^{-1})$ and $B(q^{-1})$ contain information about its dynamics. The backward shift operator q^{-1} is used and is defined as $u(t-1) = q^{-1}u(t)$. The polynomial $A_1(q^{-1})$ is monic and of degree n, while the polynomial $B(q^{-1})$ is of degree m, where $m \geq n$. The plant dynamics are determined by the value of the roots of $A_1(q^{-1})$ and $B(q^{-1})$, known as the poles and zeroes of a plant respectively. The last piece of information about the plant is contained in k, which is the delay. For sampled-data processes, this delay is always greater than or equal to one, to ensure process causality.

The disturbance term represents the effect of the environment in which the plant exists and can take on many forms. It could be purely deterministic, such as randomly occurring steps changes, or it could be stochastic in nature. The Autoregressive–Integrated–Moving–Average (ARIMA) model

4
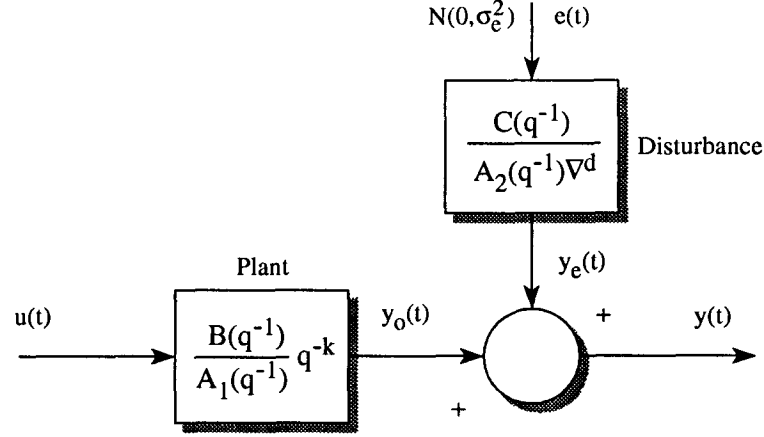
Figure 2.1. Block diagram representation of the discrete linear model

can be used to represent many different types of typical industrial disturbances. It makes up the second part of Figure 2.1 and the equation:

$$y_e(t) = \frac{C(q^{-1})}{A_2(q^{-1})\nabla^d} e(t) \tag{2.2}$$

contains information about the properties of the disturbances. Its output is $y_e(t)$ while the input $e(t)$ is a sequence of uncorrelated random variables with a zero mean and constant variance, which can be denoted as $N(0,\sigma_e^2)$. This input signal is then a white noise sequence with a normal distribution. The polynomials $C(q^{-1})$ and $A_2(q^{-1})$ have degree q and p respectively and like the plant polynomials, the roots of these polynomials determine the dynamics of the disturbance. It is also assumed that the roots of the $A_2(q^{-1})$ polynomial lie inside the unit circle. The operator, ∇, is defined as $\nabla = 1 - q^{-1}$ and the term ∇^d permits the mean of the disturbance output to change over time so that it can exhibit some nonstationary behavior. For example, when $d=1$ the disturbance will exhibit a random walk behavior where its mean varies over time. The specific order of the disturbance is denoted by ARIMA(p,d,q) where p, d, and q are as defined above and thus set the properties of the disturbance.

These are the two terms, when combined, that can be used to represent most industrial processes. The output of a process which combines the two terms is given by:

$$y(t) = \frac{B(q^{-1})}{A_1(q^{-1})} q^{-k} u(t) + \frac{C(q^{-1})}{A_2(q^{-1})\nabla^d} e(t) \tag{2.3}$$

which is made up of the effects of the plant and disturbance. The monitoring to be performed is not done on the process as it appears in Figure 2.1. This is because the processes to be monitored will be under feedback control.

2.2 Monitoring Closed Loop Processes

A goal of this process monitoring is to be able to extract information from the process successfully while it is operating under feedback control. This means that all estimation is performed under closed loop operation, which can cause some difficulties when it comes to estimation. These difficulties must be addressed to ensure successful monitoring and in future chapters, they will be taken into account as they become a concern.

Closed loop operation of a process can be represented by the block diagram in Figure 2.2. Using the definitions above, the plant transfer function is:

$$G_p(q^{-1}) = \frac{y_o(t)}{u(t)} = \frac{B(q^{-1})}{A_1(q^{-1})}q^{-k} = \frac{b_0 + b_1 q^{-1} + ...b_m q^{-m}}{1 + a_1 q^{-1} + ...a_u q^{-u}}q^{-k} \tag{2.4}$$

and the disturbance transfer function is:

$$G_d(q^{-1}) = \frac{y_e(t)}{e(t)} = \frac{C(q^{-1})}{\nabla^d A_2(q^{-1})} = \frac{1 + c_1 q^{-1} + ...c_q q^{-q}}{1 + a_1' q^{-1} + ...a_{p+d}' q^{-p-d}} \tag{2.5}$$

The final block to be defined is the controller, it has the form:

$$G_c(q^{-1}) = \frac{u(t)}{c(t)} = \frac{E(q^{-1})}{H(q^{-1})} = \frac{e_0 + e_1 q^{-1} + ...e_j q^{-j}}{h_0 + h_1 q^{-1} + ...h_i q^{-i}} \tag{2.6}$$

and is chosen so that the plant output response performs in some desired manner.
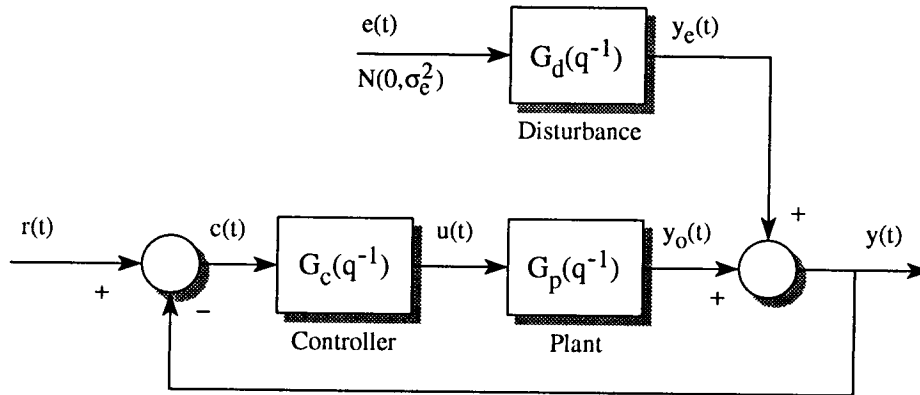
Figure 2.2. Block diagram representation of closed loop process
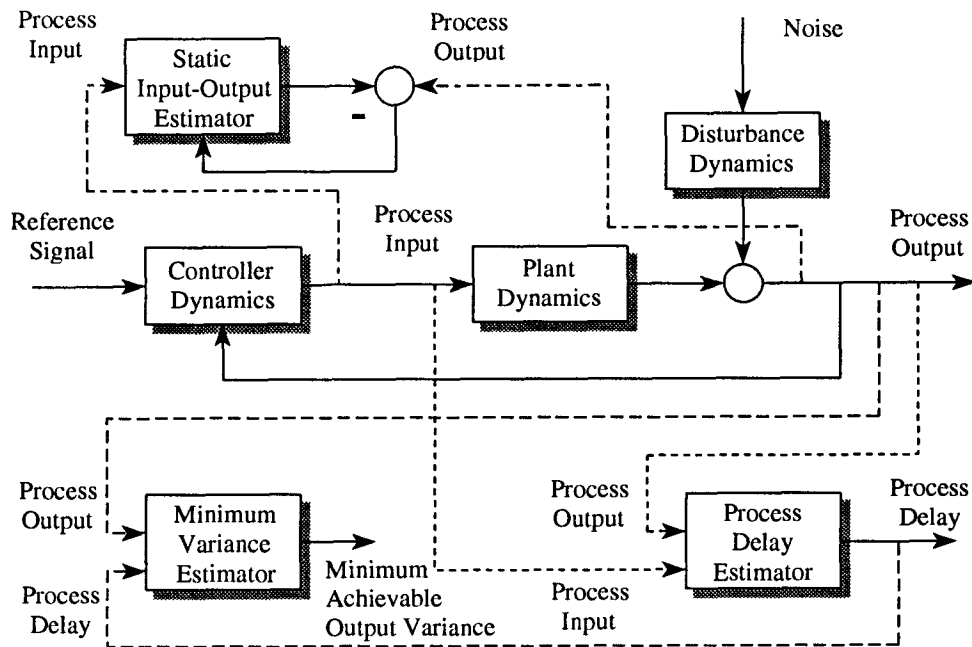
Figure 2.3. Block diagram of the monitoring scheme

2.3 Purpose of Monitoring

The successful operation of a process requires the inclusion of some sort of performance monitoring and/or fault detection. There are many insights that can be obtained by monitoring the process, all which will aid in choosing a control scheme for the process or provide a better understanding of the operation of the process. Monitoring will help determine if the process is operating as expected, which is important for successful control. The detection of changes in the process over time will be noticed by monitoring, which will allow preventive measures to be taken before the changes become too serious. The cumulation of the insights provided by monitoring will be used to indicate if and how performance can be improved.

There are different properties of a process that can be monitored to gain knowledge of process performance. When the following three attributes of a process are monitored, namely the time delay, static characteristics, and output variance, many insights will be provided which will aid in the successful operation of a process. These properties can be used separately or together to serve this purpose. These process properties were chosen because of an interest in monitoring the operation of processes in regulation mode. Examining the process output variance and determining

7

the minimum achievable output variance becomes very important for regulation loops. Knowledge of the time delay and static characteristic will help in determining and interpreting the minimum achievable output variance level. They can also stand alone as monitoring tools. The block diagram of the monitoring scheme appears in Figure 2.3, it showing the way information is extracted from a process. It is obtained using only the available input and output signals. The benefits obtained from monitoring these properties will now be discussed, emphasizing the value of successful monitoring.

2.3.1 Time Delay

Monitoring the time delay of a process serves several purposes. Most importantly, knowledge of the true time delay provided by monitoring will ensure that the controller can compensate for the effect of the delay. Also, monitoring can indicate if changes in the process are required for improved performance. Lastly, as changes in the delay can be noted as they occur, the reason for the change can be determined and corrected if possible.

The knowledge of the correct time delay of a process is very important for successful control of any industrial process. Many processes have long and often varying time delays, which makes proper control difficult. This is because the delay adds a phase contribution to the process frequency response which slows the output response and must be compensated for by a controller. Monitoring the process to obtain the delay can allow the controller to be designed correctly, so the performance of the process can be improved.

Typically, processes with delay are controlled by one of two approaches. One approach uses a type of controller, such as a Dahlin, a Deadbeat, or a Minimum Variance controller, that we shall call dead-time controllers, whose design requires the knowledge of the exact delay. If the controller uses an incorrect delay value, the resulting process performance is poor or even unstable. Monitoring the time delay of the process allows confirmation of correct controller design and allows the controller to be changed on-line if the delay varies, avoiding the associated possible poor performance.

The other approach involves the use of Proportional-plus-Integral (PI) and Proportional-plus-Integral-plus-Derivative (PID) controllers. These controllers ignore the exact time delay in their design. The performance of these controllers is therefore not as sensitive to variation in delay as

the dead-time controllers, so they are often used to control processes with unknown time delay. The reason for using these controllers lies in the fact that the performance of dead-time controllers deteriorates or even becomes unstable if the controller does not use the correct delay. However, the performance of PI and PID controllers is worse than the dead-time controllers when the time delay is known. By monitoring the process and obtaining the correct time delay, dead-time controllers can be correctly designed, thus replacing PI and PID controllers, in these situations. The process can then enjoy the superior performance of the dead-time controllers.

There are situations where the control scheme is not able to improve performance to a satisfactory level due to limitations imposed by the process itself. The information obtained by monitoring the process may reveal that changes in the process are necessary to further improve performance. If, for example, monitoring the time delay reveals that the time delay changes frequently or is very long, changes in the process may be required to correct this so performance can be improved. It may be necessary to consider moving sensors to shorten delay or perform other measures which affect the length or variability of the time delay. If the results of these changes make the delay shorter or more constant, then the process performance will be improved.

A final benefit which arises from monitoring is that changes in the delay will be noticed. Once a change is noted, the process operation can be investigated to determine the cause of the change. Knowledge about the cause of the change or the degradation in process behavior will provide insight into the important aspects of the process. If these aspects are watched and kept from changing, the overall operation of the process can be improved. This discussion has shown a few advantages of monitoring the process time delay.

2.3.2 Static Characteristics

A major assumption made when controllers are designed, is that the process to be controlled is linear. Because no process is truly linear, it would be informative to know the degree of nonlinearity present in the process. By monitoring the static input-output relationship, the degree of nonlinearity will be found and can be examined. If it differs a great deal from what was expected then steps can be taken to take accommodate this difference and improve control performance.

9

Also, nonlinearities can develop over time, often as a result of deterioration in actuator operation, such as control valves. If this nonlinearity is detected, then maintenance can be performed to return the actuator operation to normal. Often there will be nonlinearities present that were not expected when a control strategy was chosen. If this occurs, then the control strategy must be reworked to account for the nonlinearity. It must also be realized that achievable performance will be limited by any nonlinearities.

Finally, control performance will deteriorate if sensors or actuators are poorly sized. For example, an undersized actuator will most likely exhibit saturation which in turn will limit the process response time to changes in reference or disturbance upsets. Conversely, an oversized actuator may not be capable of making fine adjustments required to keep the output very close to the reference. Similar problems will be experienced with sensors as well. Poorly chosen sensors may be driven outside their linear region, or may not have a resolution fine enough to detect small, yet undesirable changes. Problems such as these can be determined from estimating the static input-output relationship.

2.3.3 Minimum Achievable Output Variance

It has been shown [7] that the variability of a control loop output can often be lower when the process is in open loop or manual operation as opposed to feedback control. Sometimes, this situation results from poorly maintained control valves and sensors, but often it arises from poorly tuned PI controllers which cause much of the variability themselves. Bialkowski [7] estimates that the output variance of processes in a pulp and paper mill can be reduced by as much as thirty to fifty percent by improved loop tuning, controller selection and equipment maintenance. In industries such as pulp and paper, where a small variation in the final product is an important goal, this reduced variability can result in major economic savings.

It has been determined by Åström [2] that the minimum variance in an output is achievable through the use of a minimum variance controller. The output variance when this minimum variance controller is used is the minimum achievable output variance obtainable by feedback control. A method of calculating this minimum achievable output variance level for a process under any linear

control was determined by Harris [17]. There are several reasons why this minimum achievable output variance level should be estimated and monitored.

The ratio of output variance to minimum achievable output variance can be used to assess the performance of the control algorithm which is employed. If this ratio is large, it suggests that alternative controller parameters be chosen. This retuning of the controller may result in a smaller ratio. It may be found that no choice of parameters for the present controller will reduce the output variance to the desired level. A new control algorithm may then be required to further reduce the ratio. In the case of processes with a large time delay, a dead-time controller should be employed.

It is just as likely that the minimum achievable output variance determined is not satisfactory. No feedback controller can reduce the output variance below this level, so some other steps must be taken to reduce the output variance. Modelling of the disturbance can be done so that feedforward control can be employed and this control scheme will reduce the minimum achievable output variance. Another way of reducing the minimum achievable output variance is by decreasing the time delay. Thus the cause of the delay must be found so that one can be determined if it can be reduced. These are a few uses of the minimum achievable output variance once it is known.

There are several more process properties which must be estimated in order to determine the minimum achievable output variance. It is only natural that these properties are monitored, because the cause of any change in the minimum achievable output variance level can be determined by examining the estimation of these properties. Simulations will be used to study different changes in a process and the effect these changes have on the calculated minimum achievable output variance. In practice, many of these changes will suggest that maintenance be performed on the process, which will help reduce the output variability and operation stoppages.

Now that the process properties to be monitored have been chosen, the tools required to determine them will be examined. In turn, the following three chapters examine the steps necessary to determine the time delay, static relationship and the minimum achievable output variance.

Chapter 3

Time Delay Estimation

There is a great advantage in estimating the time delay of the process and there have been various algorithms suggested recently to provide that estimate [14][12]. The chosen method, Fixed Model Variable Regressor Estimation, is described in this chapter and the steps required to implement the scheme and estimate the delay are listed. Examples are then given to demonstrate the operation of the estimator. This is followed by an examination of the estimator to the operating conditions under which monitoring is to be performed.

3.1 Fixed Model Variable Regression Estimator

The Fixed Model Variable Regression Estimator (FMVRE) was proposed by Elnaggar [12] as a suitable method for estimating the delay of industrial processes. The FMVRE has advantages over other delay estimation techniques as it is very straightforward to implement and it converges very quickly. The FMVRE is decoupled from the estimation of any other process parameters and this aids in its speed of convergence and ease of implementation.

In the derivation of the FMVRE, a fixed auxiliary model is used to represent the process. This auxiliary model is chosen as a first-order plus delay, since most industrial processes are overdamped and can be approximated by such a model. If the process is of a higher order, it is shown in [12] that this auxiliary model is still suitable to estimate the delay.

3.1.1 Derivation of FMVRE

For a plant represented by a sampled-data linear model described earlier, the output is:

$$y(t) = q^{-k}\frac{B(q^{-1})}{A_1(q^{-1})}u(t) = q^{-k}\frac{b_0 + ... + b_m q^{-m}}{1 + a_1 q^{-1} + ... + a_n q^{-n}}u(t) \qquad (3.7)$$

where the delay is k and is always larger than or equal to one. When this delay is known only to lie within a range $[k_{min}, k_{max}]$, it can be correctly estimated, $\hat{k}$, if the cost function:

$$J(\hat{k}) = E\left\{\varepsilon^2(\hat{k})\right\} \qquad (3.8)$$

12

is minimized, where E is the expectation function. The prediction error ε is defined as:

$$\varepsilon(\hat{k}) = y(t) - \hat{y}(t, \hat{k}) \tag{3.9}$$

where $\hat{y}(t, \hat{k})$ is the output determined for every delay in the range of interest, as given by a chosen auxiliary model. This model is chosen to be first-order, for the reason given above, so its output is:

$$\hat{y}(t, \hat{k}) = ay(t - 1) + bu(t - \hat{k}) \tag{3.10}$$

and the cost function can be minimized once a and b are chosen. This minimization is not easy to do yet, so the squared prediction error is evaluated as it will simplify finding the delay that minimizes the cost function. The squared prediction error is given as:

$$\begin{aligned}
\varepsilon^2(\hat{k}) &= \left[y(t) - ay(t-1) - bu(t - \hat{k}) \right]^2 \\
&= y^2(t) + a^2 y^2(t-1) + b^2 u^2(t - \hat{k}) - 2ay(t)y(t-1) - \\
&\quad 2b\left[y(t)u(t - \hat{k}) - ay(t-1)u(t - \hat{k}) \right]
\end{aligned} \tag{3.11}$$

Before taking the expectation of this expression, which is required to evaluate the cost function, the definitions of autocorrelation and cross-correlation are introduced:

$$\begin{aligned}
r_y(\tau) &= E\{y(t)y(t \pm \tau)\} \\
r_{uy}(\tau) &= E\{u(t)y(t + \tau)\} \\
&= E\{y(t)u(t - \tau)\} \\
&= r_{yu}(-\tau)
\end{aligned} \tag{3.12}$$

Using these definitions, the expectation of the squared prediction error is rewritten as:

$$\begin{aligned}
E\left\{\varepsilon^2(\hat{k})\right\} &= r_y(0) + a^2 r_y(0) + b^2 r_u(0) - 2ar_y(1) - 2b\left[r_{uy}(\hat{k}) - ar_{uy}(\hat{k} - 1) \right] \\
&= \qquad\qquad E_0 \qquad\qquad\qquad - 2bE_1
\end{aligned} \tag{3.13}$$

Upon examination of the terms that make up E_0 and E_1, it can be seen that E_0 does not depend upon the delay, whereas E_1 does. Thus to minimize $J(\hat{k})$, E_1 must be maximized if $b > 0$, or minimized if $b < 0$. This means that for processes with a positive gain E_1 is to be maximized and then for processes with a negative gain E_1 is to be minimized to determine the correct delay. Now E_1 is given as:

$$E_1 = r_{uy}(\hat{k}) - ar_{uy}(\hat{k} - 1) \tag{3.14}$$

which still requires that the auxiliary pole a is chosen. It was shown in Elnaggar [12] that the choice of a=1 gives unbiased estimation of delay for low order processes and good delay approximation for higher order processes. Thus:

$$
\begin{aligned}
E_1 &= r_{uy}(\hat{k}) - r_{uy}(\hat{k} - 1) \\
&= E\Big\{ [y(t) - y(t - 1)]u(t - \hat{k}) \Big\} \\
&= E\Big\{ \triangle y(t)u(t - \hat{k}) \Big\} \\
&= r_{u\triangle y}(\hat{k})
\end{aligned}
\tag{3.15}
$$

An advantage of choosing a=1 becomes apparent by the appearance of $\triangle y$. It means delay estimation will not be affected by step changes in either the reference signal or disturbances. This is how the process time delay is determined and now the FMVRE algorithm will be outlined followed by two examples demonstrating its operation.

3.1.2 FMVRE Algorithm

The FMVRE algorithm is very simple to implement and is outlined below. These steps include the important equations required to implement the algorithm:

1. Update E_1 for every delay in the interval of interest. A forgetting factor, λ, can be used as shown to track time-varying time delay.

$$
E_1(t, k_i) = \lambda E_1(t - 1, k_i) + [y(t) - y(t - 1)]u(t - k_i) \quad \forall \quad k_i \epsilon [k_{min}, k_{max}]
\tag{3.16}
$$

2. For positive process gain, choose the delay that maximizes E_1.

$$
E_1(t, \hat{k}(t)) = \max\{E_1(t, k_i)\} \quad \forall \quad k_i \epsilon [k_{min}, k_{max}]
\tag{3.17}
$$

3. Return to step 1 for the next sampled data.

3.1.3 Examples

The following two examples demonstrate the operation of FMVRE and confirm its quick convergence and its ability to quickly track changes in the delay. A first-order process plus delay was used for these examples:

$$
G(s) = \frac{e^{-T_d s}}{\tau s + 1}
\tag{3.18}
$$

14

Now after zero-order hold sampling, with a sample time of T seconds, the process has the following discrete form :

$$G(q^{-1}) = \frac{b\ q^{-k}}{1 + aq^{-1}} \qquad (3.19)$$

where:

$$a = -e^{-\frac{T}{\tau}}$$

$$b = 1 - e^{-\frac{T}{\tau}} \qquad (3.20)$$

$$k = \frac{T_d}{T}$$

For these simulations a white input signal, N(0,1) was used for the estimation of the delay.

EXAMPLE 1:

The parameters of the process were chosen as follows:

Continous Time	Discrete Time (T=1sec)
τ=2.5 T_d=4	a=–0.67 b=0.33 k=5

The input and output data were run through the FMVRE and the results are shown in Figure 3.4. It can be seen from Figure 3.4(a) that the estimate converges to the correct delay after about six samples and the plot of the E_1 function confirms that the delay is definitely five.
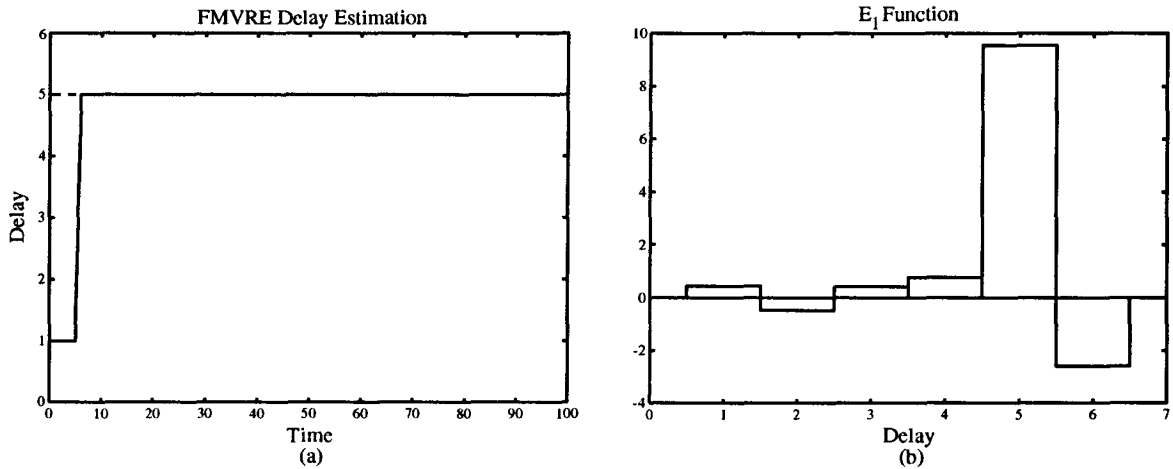
Figure 3.4: Results of FMVRE time delay estimation with (a) presenting the estimation over time while (b) shows the E_1 function at sample one hundred

15

EXAMPLE 2:

This example illustrates the ability of the FMVRE to adapt to changes in the process delay. For this simulation, the following parameters were used:

Period	Continous Time	Discrete Time (T=1sec)
0-100	τ=2.5 T$_d$=4	a=–0.67 b=0.33 k=5
101-200	τ=3 T$_d$=1	a=–0.72 b=0.28 k=2
201-300	τ=2 T$_d$=3	a=–0.61 b=0.39 k=4

and it should be noted that the plant for the first one hundred samples is identical to the plant used in Example 1. The results of delay estimation are in Figure 3.5 and include four different values of the forgetting factor. These results show the change in delay convergence, as it varies from ninety
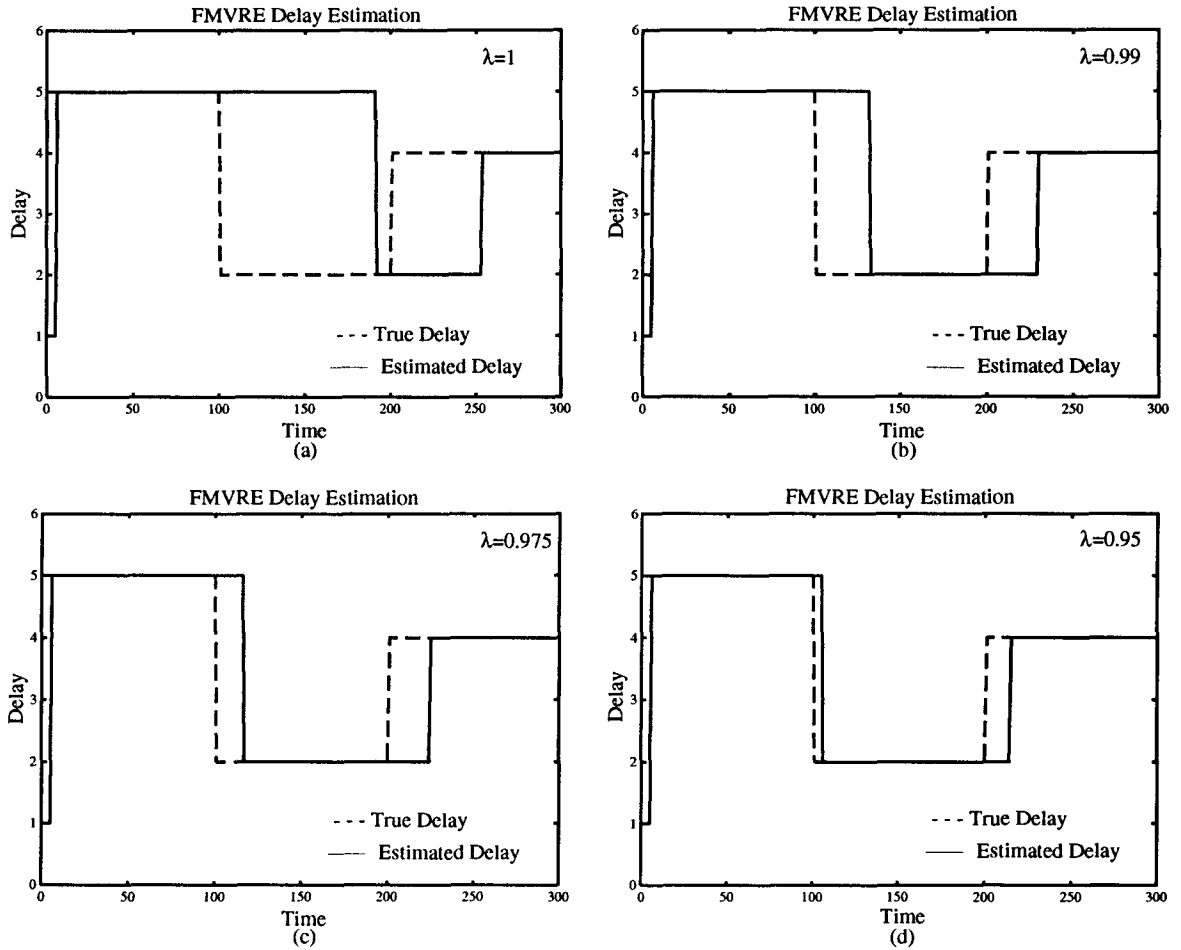
Figure 3.5: Delay estimation with varying forgetting factor, with (a) λ=1, (b) λ=0.99, (c) λ=0.975, (d) λ=0.95

16

samples in Figure 3.5(a) for a forgetting factor of one, down to fifteen samples for a forgetting factor of 0.95 shown in Figure 3.5(d). The reason for this can be understood by examining Figure 3.4(b), which shows the E_1 function at sample one hundred. The difference between the E_1 function at five, the correct delay at sample one hundred and E_1 at two, the next delay, is quite large and must be made up before the estimate of the delay will change. This time can be shortened by decreasing the forgetting factor as it reduces the weighting of previous values of E_1. This increases the weight of each new sample and will keep the difference between two and five in the E_1 function smaller, so the estimate of the delay will change quicker. Care must be taken so the forgetting factor is not made so small that the estimation changes too rapidly.

3.2 Estimation of Delay in Control Loops

Up until now, the time delay of the plant has been determined while the plant has operated in open loop and with no disturbances. To be of use for monitoring, the FMVRE must estimate the delay correctly when the plant is under feedback control and subject to stochastic disturbances. To ensure this is possible, the FMVRE operation is investigated with the addition of a stochastic disturbance and then with the process under closed loop control. It can then be determined if the delay estimation becomes biased under any circumstances.

3.2.1 Effect of Stochastic Disturbances

Before treating closed loop operation, the effect of adding a stochastic disturbance will be examined. The form of such a disturbance has been discussed and a stochastic process was shown in Figure 2.2, where the output was given by:

$$y(t) = \frac{B(q^{-1})}{A_1(q^{-1})}q^{-k}u(t) + \frac{C(q^{-1})}{A_2(q^{-1})\nabla^d}e(t)$$
$$= y_o(t) + y_e(t)$$

(3.21)

which consists of output due to the white noise as well as output due to the input. The effect of this extra component on correct delay estimation must be examined.

The cross-correlation between the input and output is given by:

$$r_{uy}(t) = r_{uy_o}(t) + r_{uy_e}(t)$$

(3.22)

17

and since the delay is determined by evaluating this cross-correlation, the effect of adding the $r_{uy_e}(t)$ term is important. This cross-correlation can be rewritten once the $y_e(t)$ is expressed as a sum of noise terms up to time t:

$$y_e(t) = w_0 e(t) + w_1 e(t-1) + w_2 e(t-2) + ...$$
(3.23)

Then the input-disturbance cross-correlation can be written as:

$$r_{uy_e}(t) = w_0 r_{ue}(t) + w_1 r_{ue}(t-1) + w_2 r_{ue}(t-2) + ...$$
$$= 0$$
(3.24)

This is because in open loop operation, the input signal and noise signal are not correlated. This means that the delay that maximizes the E_1 function is unbiased in the presence of process noise if the noise is uncorrelated with the process input. This is the case with an open loop process and this property of the FMVRE will be illustrated by an example.

EXAMPLE

The first-order plant plus delay is used once again, though this time with the addition of a stochastic disturbance. Figure 3.6 shows the process used for this example, which has white, N(0,1), uncorrelated sequences as the input and noise signals.

FMVRE delay estimation was used to estimate the delay and the results are shown in Figure 3.7. From these results, it can be seen that the delay estimation was not biased by the addition of the stochastic disturbance, as the E_1 function has a maximum at four. Since this test was performed on a computer, it is possible to access the signals $y_0(t)$ and $y_e(t)$ so that their contributions to the E_1 function can be studied. Figure 3.8(a) shows the E_1 function when the signals $u(t)$ and $y_0(t)$ are used and Figure 3.8(b) is the result of using $u(t)$ and $y_e(t)$. The E_1 function resulting from the signals $u(t)$
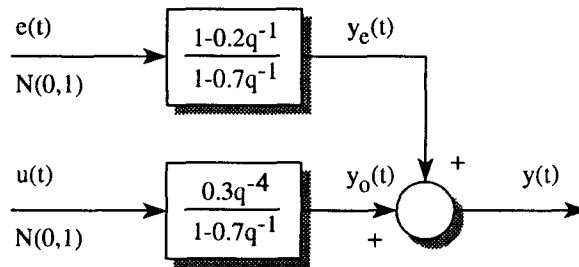
Figure 3.6: The stochastic process used for the example
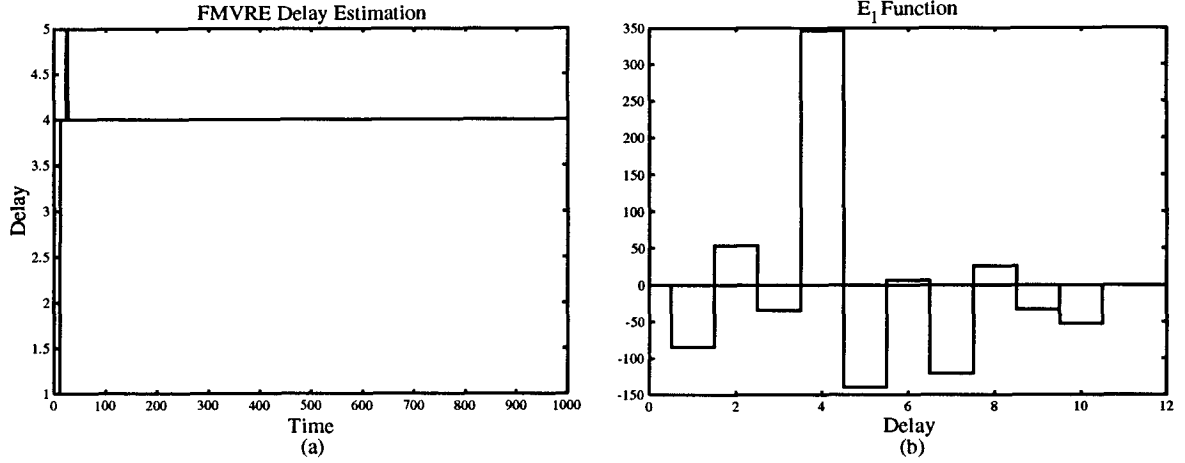
Figure 3.7: Results of FMVRE time delay estimation with (a) presenting the estimation over time while (b) shows the E_1 function at sample one thousand

and $y_o(t)$ all but matches the E_1 function in Figure 3.7(b) while the E_1 function for the signals $u(t)$ and $y_e(t)$ is essentially zero. This example then highlights the fact that when the process input and noise signal are uncorrelated, as in open loop operation, the delay estimation is unbiased.

3.2.2 Closed Loop Estimation

Attention is now turned to the process configuration of Figure 2.2, which has the process under feedback control. FMVRE delay estimation will be performed on this setup and the possibility of biased delay estimation will be investigated. It will be shown that the possibility of biased delay does exist and examples to demonstrate its seriousness will be given in Chapter 6.
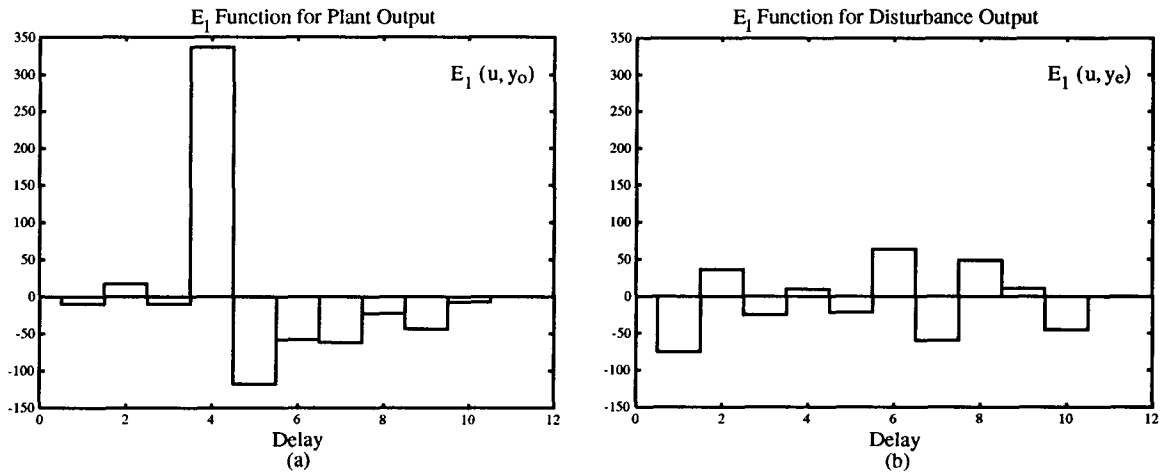
Figure 3.8: E_1 functions for (a) the input and plant output and (b) the input and disturbance output

19

In the majority of cases the process will be assumed to be operating with a constant reference signal, so any effect of changing the reference signal will be ignored. Now return to the cross-correlation of the input and output which was found previously:

$$
\begin{aligned}
r_{uy}(t) &= r_{uy_o}(t) + r_{uy_e}(t) \\
&= r_{uy_o}(t) + w_0 r_{ue}(t) + w_1 r_{ue}(t-1) + w_2 r_{ue}(t-2) + ...
\end{aligned}
\tag{3.25}
$$

By adding feedback to the process, the input and noise signal will no longer be independent and thus the E_1 function will have contribution from $r_{uy_e}(t)$. The amount of this contribution will depend on the noise filter and controller and may result in biased delay estimation of the delay. This problem will be examined further in Chapter 6 to determine its extent and to explore possible solutions.

So, now that the FMVRE algorithm has been introduced and its operation demonstrated, attention will be turned to the problem of estimating the static input-output relationship and the minimum achievable output variance. Further simulations to investigate the operation of the FMVRE are saved until Chapter 6 so that the other tools can be introduced.

Chapter 4

Static Characteristics

It is possible to identify the static relationship between the input and output of a process while it is operating in feedback control. The shape of this relationship will provide additional understanding of the process operation. The Adaptive Nonlinear Modeller represents the static relationship by using a table of values and an interpolation formula. After the operation of the estimator is explained, its algorithm will be outlined. Then an example of the Adaptive Nonlinear Modeller operation will be given to exhibit its functionality. The application of the estimator to the closed loop regulation problem is briefly discussed and while simulations are saved for a future chapter.

4.1 Adaptive Nonlinear Modeller

The Adaptive Nonlinear Modeller (ANM) was proposed by Åström [3] as a multiple use tool to model the static input-output relationship of a process. The ANM described here uses Recursive Least Squares estimation to identify this relationship avoiding some difficulties of the integration method used by Åström. A static input-output relationship can be described by:

$$y = f(u) \tag{4.26}$$

where $f(\cdot)$ is simply a function which relates the process input, u, to the process output, y. The ANM represents this function with a table of values and an interpolation formula. The table consists of
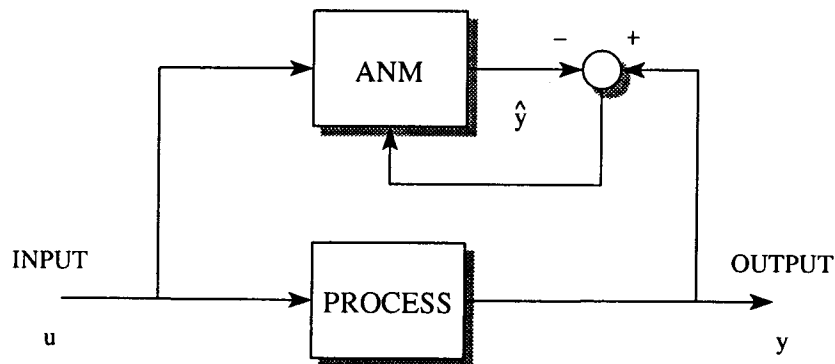
Figure 4.9: Block diagram of the Adaptive Nonlinear Modeller

discrete values of the input over the interval of interest along with the corresponding output for each discrete input. The role of the interpolation formula is to provide output when the input is between interval values. Estimation is required to identify the output values for each discrete input value in the table, from sampled data. As an aid in the discussion, Figure 4.9 presents the structure of the ANM and the following notation is defined:

u_k – a discrete input value in the table

y – the process output during estimation

$\hat{y}$ – the Adaptive Nonlinear Modeller (ANM) output

$\hat{f}_k$ – the estimated value in the table corresponding to input u_k

$f(u_k)$ – the process output for the input u_k

Using this notation, the input interval is $[u_1, u_n]$ and each entry in the table is simply a pair of numbers $\{u_k, f(u_k)\}$. The interpolation formula then uses the table values to determine the output for any arbitrary input which lies between two entries in the table, $u_{k+1} \geq u \geq u_k$. Linear interpolation is a very simple and straightforward formula to use and because of this is used by the ANM for interpolation. The linear interpolation formula which determines the output for an input is given by:

$$y = f(u) = \frac{u_{k+1} - u}{u_{k+1} - u_k} f(u_k) + \frac{u - u_k}{u_{k+1} - u_k} f(u_{k+1}) \tag{4.27}$$

The remaining problem is to estimate the values of $f(u_k)$ for the table. These entries will be determined by using Recursive Least Squares (RLS) estimation.

In order to use RLS estimation, the model must be placed into a linear regression form. The output for any input in the interval $[u_1, u_n]$ can be determined using the table of values along with the interpolation formula and is given as:

$$y = a_1 f(u_1) + a_2 f(u_2) + \dots + a_{n-1} f(u_{n-1}) + a_n f(u_n) \tag{4.28}$$

All but two of the a_i's will be zero, as the input will lie in only one interval $[u_k, u_{k+1}]$. The linear interpolation formula will determine the value of the two nonzero a_i's as:

$$a_k = \frac{u_{k-1} - u}{u_{k+1} - u_k} \quad \text{and} \quad a_{k+1} = \frac{u - u_k}{u_{k+1} - u_k} \tag{4.29}$$

22

For every sampled data, it is not known which a_i's are zero, so in general the regression vector is defined as:

$$\Phi(t) = [\, a_1 \quad a_2 \quad a_3 \quad \ldots \quad a_u \,]^T \tag{4.30}$$

and the parameter vector is:

$$\theta = [\, f(u_1) \quad f(u_2) \quad f(u_3) \quad \ldots \quad f(u_u) \,]^T \tag{4.31}$$

which allows the following linear regression form:

$$y = \theta^T \Phi(t) + v(t) \tag{4.32}$$

The least squares estimate of the parameter vector is defined as:

$$\hat{\theta} = [\, \hat{f}_1 \quad \hat{f}_2 \quad \hat{f}_3 \quad \ldots \quad \hat{f}_u \,]^T \tag{4.33}$$

which can be defined recursively using:

$$K(t) = \frac{P(t-1)\Phi(t)}{1 + \Phi^T(t)P(t-1)\Phi(t)} \tag{4.34}$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\Big[y(t) - \hat{\theta}^T(t-1)\Phi(t) \Big] \tag{4.35}$$

$$P(t) = P(t-1) - K(t)\Phi^T(t)P(t-1) \tag{4.36}$$

where P(t) is the parameter covariance matrix and K(t) is the gain matrix.

4.1.1 ANM Algorithm

The following stages outline the operation of the ANM. They summarize the previous section and highlight the important steps and equations required to implement the ANM.

1. Determine the input range of interest and break it into discrete intervals.

2. For every sampled data, find the interval in which the input value lies.

$$[u_k, u_{k+1}] \quad \text{where} \quad u_k \le u \le u_{k+1} \tag{4.37}$$

23

3. Use linear interpolation to determine the values for the regression vector.

$$a_k = \frac{u_{k+1} - u}{u_{k+1} - u_k} \quad \text{and} \quad a_{k+1} = \frac{u - u_k}{u_{k+1} - u_k}$$

$$\Phi = \begin{bmatrix} 0 & \cdots & 0 & a_k & a_{k+1} & 0 & \cdots & 0 \end{bmatrix}^T \tag{4.38}$$

4. Determine the estimate of the table values by using RLS estimation.

$$K(t) = \frac{P(t-1)\Phi(t)}{1 + \Phi^T(t)P(t-1)\Phi(t)}$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\left[y(t) - \hat{\theta}^T(t-1)\Phi(t)\right] \tag{4.39}$$

$$P(t) = P(t-1) - K(t)\Phi^T(t)P(t-1)$$

5. Return to step 2 for the next sampled data.

The shape of the $f(\cdot)$ function can be seen by plotting the table at any time during estimation. If the number of computations required by the RLS estimation is a concern, it is possible to reduce this number by realizing the fact that for any data, all but two values in the regression vector are zero. This greatly reduces the number of multiplications required by the matrix operations in the equations of (4.39) to determine the estimates.

4.1.2 ANM Example

This example was used by Åström [3] to exhibit the operation of the ANM. It is repeated here to show the success of the ANM using RLS estimation. The input and output are related by:

$$y = u^2 \tag{4.40}$$

which is definitely a nonlinear relationship. The input and output used for estimation are shown in Figure 4.10(a&b), and the resulting parameter estimation for the table output values is shown in Figure 4.10(c&d). Figure 4.10(c) shows that the estimates arrive quickly to near their correct values, while Figure 4.10(d) shows that the table values, represented by zeroes, do a good job approximating the squared relationship. Figure 4.11 shows the ability of the ANM to reconstruct the output using the table and interpolation formula for the given signal. This confirms the successful operation of the ANM with RLS estimation.
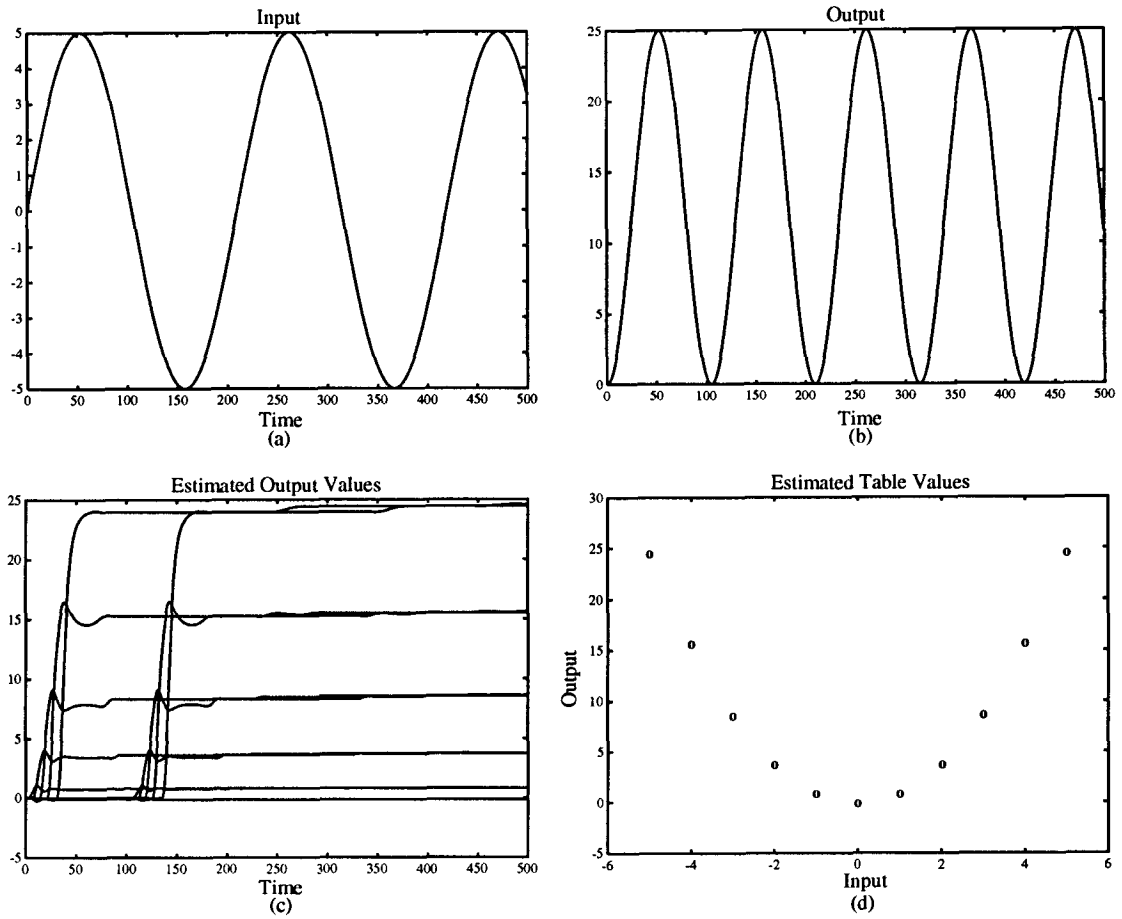
Figure 4.10: Results of the ANM estimation showing (a) the input used, (b) the resulting output, (c) the table value estimation over time and (d) a plot of the final estimated table values
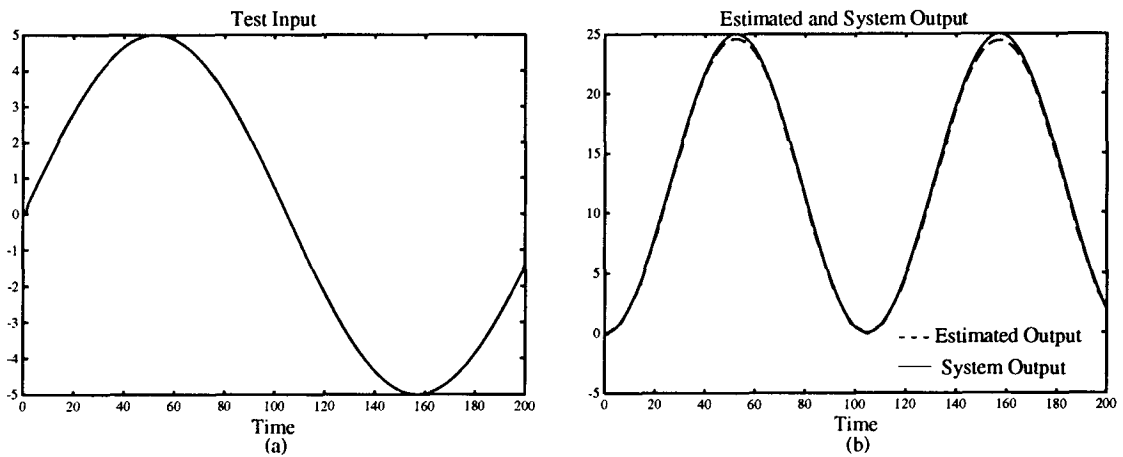
Figure 4.11: Results of using (a) the input and the ANM to (b) reconstruct the output

25

4.2 Application to Stochastic Feedback Processes

The ANM described has assumed that the input and output are obtained during steady state operation of a process. If all transients are not removed from the signals, then the static input-output relationship will not be correctly identified. These transients result from the process dynamics which are prevalent during reference changes and disturbance upsets. Prior to the installation of the ANM, some precaution must be taken to ensure that the input and output signals are in steady state. Low pass filtering of the signals will remove much of the undesired transient component of the signals as well as the effect of stochastic disturbances. Also, the ANM could be turned off during setpoint changes and turned back on once the setpoint remains constant for a period of time. These problems are very important and will be thoroughly examined by simulations in Chapter 6. Also to be examined is the problem of setting an appropriate cutoff frequency for the low pass filter.

The only remaining task is to introduce the method used to determine the minimum achievable output variance of a process. Once introduced, the details of all the estimators used for monitoring purposes will be studied using simulations.

Chapter 5

Output Variance

The variance in the process output becomes important when a constant output is desired. The controller becomes known as a regulator because it attempts to keep the output close to the reference. This chapter begins by determining the minimum variance achievable by a feedback controller. Then the method used to find this minimum level is given, for a process in closed loop regulation. Next, the details of estimating the minimum achievable output variance are presented, including an explanation of the Laguerre network used to model the process and an explanation of the Recursive Extended Least Squares method used to estimate the parameters of the Laguerre network. Lastly, the chapter ends by outlining the algorithm needed to calculate the minimum achievable output variance and demonstrating its operation by presenting a simulation result highlighting important aspects.

5.1 Minimum Achievable Output Variance

The following derivation of the minimum achievable output variance follows the method used by Åström [2] to determine minimum variance control and uses the notation introduced in equation (2.3). The goal is to find the lowest possible level of the output variance, $E\{y^2(t+k)\}$. It will then be possible to estimate this level, so that there is a comparison value for the present output variance of the process.

The derivation begins by writing the process model in terms of present and future values:

$$y(t + k) = \frac{B(q^{-1})}{A_1(q^{-1})}u(t) + \frac{C(q^{-1})}{A_2(q^{-1})\nabla^d}e(t + k) \qquad (5.41)$$

and the last term on the right can be written as:

$$\frac{C(q^{-1})}{A_2(q^{-1})\nabla^d}e(t + k) = F(q^{-1})e(t + k) + \frac{G(q^{-1})}{A_2(q^{-1})\nabla^d}e(t) \qquad (5.42)$$

which is now expressed as known and unknown quantities as of time t. The polynomial $F(q^{-1})$ is monic and of degree k-1, that is one less than the process delay. Meanwhile, the degree of $G(q^{-1})$

27

depends on q, the degree of $C(q^{-1})$, and p, the degree of $A_2(q^{-1})$, and is the larger of (p+d-1) or (q-1).

Dividing (5.42) through by e(t+k) gives a Diophantine equation:

$$\frac{C(q^{-1})}{A_2(q^{-1})\nabla^d} = F(q^{-1}) + \frac{G(q^{-1})}{A_2(q^{-1})\nabla^d}q^{-k} \tag{5.43}$$

or in a more familiar form:

$$C(q^{-1}) = F(q^{-1})A_2(q^{-1})\nabla^d + G(q^{-1})q^{-k} \tag{5.44}$$

Now substituting (5.42) into (5.41) gives:

$$y(t+k) = \frac{B(q^{-1})}{A_1(q^{-1})}u(t) + F(q^{-1})e(t+k) + \frac{G(q^{-1})}{A_2(q^{-1})\nabla^d}e(t) \tag{5.45}$$

which shows that the input can affect the future output by removing the present noise. It is not possible however, for the input to remove the future noise from the future output. Yet the value of the present noise is unknown and it is not possible to determine the input needed to remove its effect, so this noise must be replaced with something which is known, namely the present output. This is done by returning to the process model and isolating e(t):

$$e(t) = -\frac{B(q^{-1})A_2(q^{-1})\nabla^d}{A_1(q^{-1})C(q^{-1})}u(t-k) + \frac{A_2(q^{-1})\nabla^d}{C(q^{-1})}y(t) \tag{5.46}$$

This can be substituted into (5.45) and the following relation can be found:

$$y(t+k) = \frac{B(q^{-1})}{A_1(q^{-1})}\left[1 - \frac{G(q^{-1})}{C(q^{-1})}q^{-k}\right]u(t) + \frac{G(q^{-1})}{C(q^{-1})}y(t) + F(q^{-1})e(t+k) \tag{5.47}$$

Now the Diophantine equation can be manipulated to show that:

$$1 - \frac{G(q^{-1})}{C(q^{-1})}q^{-k} = \frac{F(q^{-1})A_2(q^{-1})\nabla^d}{C(q^{-1})} \tag{5.48}$$

which can be substituted into (5.47) to give:

$$y(t+k) = \frac{B(q^{-1})F(q^{-1})A_2(q^{-1})\nabla^d}{A_1(q^{-1})C(q^{-1})}u(t) + \frac{G(q^{-1})}{C(q^{-1})}y(t) + F(q^{-1})e(t+k) \tag{5.49}$$

Then the variance of y(t+k) is given by:

$$E\{y^2(t+k)\} = E\left\{[F(q^{-1})e(t+k)]^2 + \right.$$
$$\left. + \left[\frac{B(q^{-1})F(q^{-1})A_2(q^{-1})\nabla^d}{A_1(q^{-1})C(q^{-1})}u(t) + \frac{G(q^{-1})}{C(q^{-1})}y(t)\right]^2\right\} \tag{5.50}$$

There are terms missing from the above expression which contain the correlation of the future noise with past inputs and outputs. As the future noise is independent of the past inputs and outputs, the expectation of these terms is zero. There are now two terms left and it is now possible to remove the present output by the appropriate choice of u(t). The input cannot affect the future noise and so the future output will always contain these terms. The minimum achievable output variance is the level reached when the present output is removed from the future output and is:

$$E\{y^2(t+k)\} = E\left\{ \left[F(q^{-1})e(t+k) \right]^2 \right\}$$ (5.51)

As can be seen from (5.51), the minimum achievable output variance will depend on $F(q^{-1})$ and the variance of e(t). If the process delay is one, then the variance will simply be:

$$E\{y^2(t)\} = E\{e^2(t)\} = \sigma_e^2$$ (5.52)

Thus, for a single delay, the output variance will equal the disturbance variance. When the delay increases, so will the minimum achievable output variance. For a process with a delay k, the minimum achievable output variance is found from (5.51) to be:

$$\sigma_{mv}^2 = \sigma_e^2 \left(1 + f_1^2 + f_2^2 + \dots + f_{k-1}^2 \right)$$ (5.53)

The coefficients of $F(q^{-1})$ must be known along with the time delay to determine the minimum achievable output variance.

To be able to achieve the minimum achievable output variance, it is required that the input be:

$$u(t) = -\frac{A_1(q^{-1})G(q^{-1})}{B_1(q^{-1})F(q^{-1})A_2(q^{-1})\nabla^d}y(t)$$ (5.54)

The application of this control signal will result in the minimum achievable variance in the output signal and is known as the minimum variance control law. This controller is what must be used to obtain the minimum achievable output variance and any other controller will not reach this lower limit. However, this controller can require a very active control signal to obtain this minimum variance performance and the process actuators must then be able to follow this signal. Also, in order to determine the minimum variance control signal, knowledge of the disturbance filter is required. If

this is not known, then a minimum variance control algorithm cannot be used. It may be decided to use another type of controller, in which case its output variance can be compared to the minimum achievable output variance and its performance judged. Lastly, the autocorrelation function of a process under minimum variance control will exhibit an interesting and useful characteristic, which will depend on the time delay of the process. The function will have significant correlation only at lags less then the time delay. This number of lags will also correspond to the number of coefficients in the $F(q^{-1})$ polynomial.

5.2 Method of Finding Minimum Achievable Output Variance

Now that the minimum achievable output variance level has been determined, what must be done to find this level. By knowing the minimum achievable output variance, it can be compared with the present operating output variance to evaluate performance. Based on the previous discussion, the use of the autocorrelation function seems to be an obvious choice. Calculating this function would show if the minimum achievable output variance has been achieved, but the use of the autocorrelation function has several shortcomings. First, while significant correlation only at lags less than the delay indicates minimum variance performance of the controller, correlation beyond this does not give a concrete indication of how far from minimum achievable output variance the process is operating. It is then difficult to determine whether changes to the present process to improve output variance are worthwhile. It would also be difficult to determine changes in the minimum achievable output variance by just using the autocorrelation function. Though the autocorrelation would change each time it was calculated, it would be difficult to determine if the changes were significant. A better method seems to be required.

Another possible method of determining the minimum achievable output variance would consist of estimating the disturbance filter and the variance of the noise and solving the Diophantine equation (5.44) to determine the $F(q^{-1})$ coefficients. Once they are determined, the minimum achievable output variance can be calculated and compared to the present output variance. The difference between the minimum achievable output variance and operating output variance will indicate if the performance is satisfactory. The major drawback to this method exists due to the fact that the process is operating in

30

feedback control while estimation is performed. Parameter identification may not be possible during feedback control, especially if the process input is not rich enough. A way of avoiding this problem was suggested by Harris [17] and consists of estimating the closed loop noise transfer function and from this, determining the minimum achievable output variance.

Returning to the closed loop diagram of Figure 2.2, the transfer function between the noise signal and the output is found as:

$$\frac{y(t)}{e(t)} = \frac{G_d(q^{-1})}{1 + G_c(q^{-1})G_p(q^{-1})} \tag{5.55}$$

or:

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{A_1(q^{-1})C(q^{-1})}{A_2(q^{-1})\nabla^d[A_1(q^{-1}) + G_c(q^{-1})B(q^{-1})q^{-k}]} \tag{5.56}$$

Now it was shown in the previous section that the minimum achievable output variance is a function of the noise variance and coefficients of the $F(q^{-1})$ polynomial. If the above transfer function is estimated, then it is possible to find $F(q^{-1})$ and the noise sequence. It will now be shown how to find the $F(q^{-1})$ polynomial and we start by rearranging the transfer function as:

$$\frac{R(q^{-1})}{S(q^{-1})} = \frac{C(q^{-1})}{A_2(q^{-1})\nabla^d} \left(\frac{A_1(q^{-1})}{A_1(q^{-1}) + G_c(q^{-1})B(q^{-1})q^{-k}} \right) \tag{5.57}$$

and replacing $\frac{C(q^{-1})}{A_2(q^{-1})\nabla^d}$ with the Diophantine equation (5.43) (and dropping the (q^{-1})'s temporarily):

$$\begin{aligned}
\frac{R}{S} &= \left(F + \frac{Gq^{-k}}{A_2\nabla^d} \right) \left(\frac{A_1}{A_1 + G_cBq^{-k}} \right) \\
&= \frac{FA_1}{A_1 + G_cBq^{-k}} + \frac{GA_1q^{-k}}{A_2\nabla^d[A_1 + G_cBq^{-k}]} \\
&= F + \frac{FG_cBq^{-k}}{A_1 + G_cBq^{-k}} + \frac{GA_1q^{-k}}{A_2\nabla^d[A_1 + G_cBq^{-k}]} \\
&= F + \frac{[A_2\nabla^dFG_cB + GA_1]q^{-k}}{A_2\nabla^d[A_1 + G_cBq^{-k}]}
\end{aligned} \tag{5.58}$$

$$\frac{R(q^{-1})}{S(q^{-1})} = F(q^{-1}) + \frac{G'(q^{-1})q^{-k}}{S(q^{-1})}$$

The closed loop transfer function can then be expressed as a Diophantine equation where the $F(q^{-1})$ polynomial is the one required to determine the minimum achievable output variance. This is not surprising as it confirms the fact that the controller transfer function can only compensate for the

31

noise fed back to it and not compensate for the noise placed directly onto the output. After the $R(q^{-1})$ and $S(q^{-1})$ polynomials are found then the $F(q^{-1})$ polynomial can be determined by solving:

$$R(q^{-1}) = F(q^{-1})S(q^{-1}) + G'(q^{-1})q^{-k} \tag{5.59}$$

once the size of the $F(q^{-1})$ polynomial is set. This relies on knowledge of the time delay because the unknown coefficients in the $F(q^{-1})$ polynomial number one less than the delay.

It will become necessary to determine the coefficients of the $F(q^{-1})$ polynomial from a state-space representation of the process. This is because the process will be modelled by Laguerre filters which use a state-space representation. The transfer function from the noise to the input has the following state-space representation:

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}e(t)$$
$$y(t) = \mathbf{c}\mathbf{x}(t) + e(t) \tag{5.60}$$

where for a single input, single output process the following are defined. The matrix $\mathbf{A}$ will be a n by n matrix while $\mathbf{b}$ and $\mathbf{c}$ are vectors of length n. The state vector is $\mathbf{x}(t)$ which also has n elements, while the process output is $y(t)$ and the noise sequence is $e(t)$. The value of n will depend on the process represented. Now if the output is written in terms of the noise signal, the equivalent of equation (5.56), the closed loop transfer function, can be obtained. This is performed by writing the state equation as:

$$q\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}e(t)$$
$$\mathbf{x}(t) = (q - \mathbf{A})^{-1}\mathbf{b}e(t) \tag{5.61}$$

and this expression can be substituted into the output equation:

$$y(t) = \mathbf{c}(q - \mathbf{A})^{-1}\mathbf{b}e(t) + e(t)$$
$$= [1 + \mathbf{c}(q - \mathbf{A})^{-1}\mathbf{b}]e(t)$$
$$= [1 + \mathbf{c}\mathbf{b}q^{-1} + \mathbf{c}\mathbf{A}\mathbf{b}q^{-2} + \ldots \tag{5.62}$$
$$+ \mathbf{c}\mathbf{A}^{k-2}\mathbf{b}q^{-k+1} + \mathbf{c}\mathbf{A}^{k-1}(q - \mathbf{A})^{-1}\mathbf{b}q^{-k}]e(t)$$

The terms $\mathbf{c}\mathbf{b}$, $\mathbf{c}\mathbf{A}\mathbf{b}$, $\mathbf{c}\mathbf{A}^2\mathbf{b}$..., are the Markov parameters of the process. Now it has been shown that the terms in the output up to the size of the delay are not affected by any control action, so the minimum achievable output variance is defined as:

$$\sigma_{\text{mv}}^2 = \sigma_e^2\left[1 + (\mathbf{c}\mathbf{b})^2 + (\mathbf{c}\mathbf{A}\mathbf{b})^2 + \ldots + \left(\mathbf{c}\mathbf{A}^{k-2}\mathbf{b}\right)^2\right] \tag{5.63}$$

32

which can be found from the state-space representation once the matrix **A** and vectors **b** and **c** are known.

5.3 Estimation of Minimum Achievable Output Variance

The method for finding the minimum achievable output variance has now been outlined, so it is necessary to determine what must be done to estimate it. The method used by Harris [17] and Perrier [23] deals with the estimation of the time-series model previously described. The method to be used here involves modelling the closed loop noise transfer function with a state-space representation of discrete Laguerre filters. Discrete Laguerre filters are used as they avoid the necessity of choosing the degree of the closed loop polynomials, as required by time-series models. While there are parameters to be chosen so the Laguerre network can be used, estimation is more robust [25] to their choice than to the choice of polynomial degree in a time-series models. The parameters of the state-space representation are estimated using Recursive Extended Least Squares estimation as it is the least complicated method available to estimate process parameters where unknown variables exist. Unlike Harris, who assumed the time delay is known, the estimate obtained by the FMVRE will be used to determine the number of terms in the minimum achievable output variance. Finally, the minimum achievable output variance is calculated using the method outlined for a state-space representation of a process.

The following sections describe the discrete Laguerre model and how it is formulated for this situation as well as outlining Recursive Extended Least Squares estimation. The Recursive Extended Least Squares method must be applied to this problem, which requires some of the parameters of the state-space representation be estimated. The complete algorithm to find the minimum achievable output variance is outlined, including all steps and important equations necessary to find the correct value. Lastly, an example of minimum achievable output variance estimation is given.

5.3.1 Discrete Laguerre Model

In discrete time, the Laguerre set is described by the following functions [16]:

$$l_N(t) = \frac{\sqrt{1-a^2}}{(N-1)!} \frac{d^{N-1}}{dq^{N-1}} \left[q^{N-1}(1-aq)^{N-1} \right] \bigg|_{q=a}, \quad -1 < a < 1 \tag{5.64}$$

33
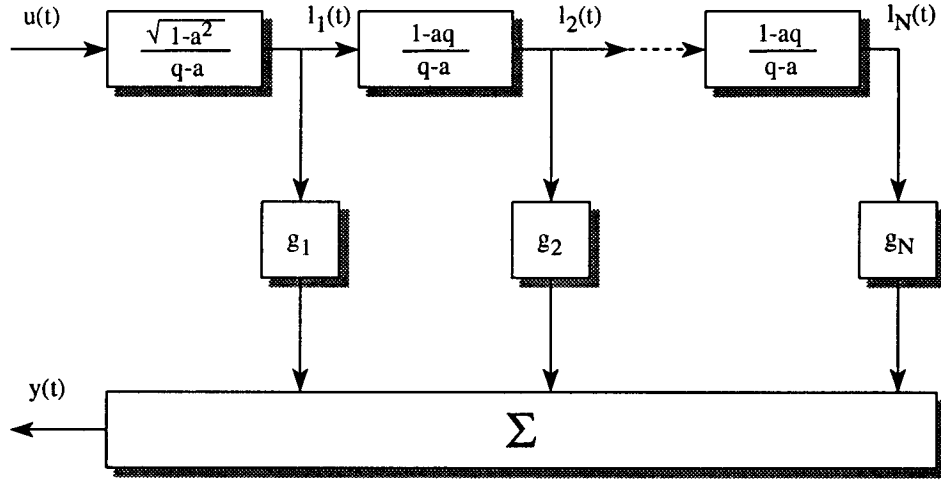
Figure 5.12: Block diagram of Laguerre filter network

where N is the order of the function while the parameter a is the function's time scale. As the Laguerre functions are orthonormal and complete in $L_2[0, \infty)$, they can represent the impulse response of any stable sampled linear process with an infinite expansion:

$$h(t) = \sum_{N=1}^{\infty} g_N \, l_N(t) \tag{5.65}$$

where g_N is the Nth Laguerre gain. The Z-transform of this set of functions results in the following Laguerre filters:

$$L_N(q) = \frac{\sqrt{1 - aq}}{q - a} \left(\frac{1 - aq}{q - a} \right)^{N-1} \tag{5.66}$$

which can then represent any sampled linear process as :

$$G(q) = \sum_{N=1}^{\infty} g_N \, L_N(q) \tag{5.67}$$

The Laguerre filters can be used to approximate a sampled linear process by using N filters which leads to the network representation in Figure 5.12.

5.3.1.1 Laguerre Network State-Space Representation

The most convenient form of a Laguerre model to use when the gains are to be estimated, is a state-space representation. The state-space representation of the discrete Laguerre filters is:

$$l(t + 1) = Al(t) + bu(t)$$

$$y(t) = c \, l(t) \tag{5.68}$$

To determine the matrix **A** and the vectors **b** and **c**, the block diagram in Figure 5.12 can be examined. A set of equations is required to define this matrix and vectors, some equations which express the future states, $l(t+1)$, in terms of present states and the input while another equation expresses the plant output in terms of the present state values and the input. From this block diagram, the following equations are found:

$$y(t) = g_1 l_1(t) + g_2 l_2(t) + \ldots + g_N l_N(t)$$

$$= [g_1 \quad g_2 \quad \cdots \quad g_N][l_1(t) \quad l_2(t) \quad \ldots \quad l_N(t)]^T \tag{5.69}$$

$$\frac{l_1(t)}{u(t)} = \frac{\sqrt{1-a^2}}{q-a} \implies l_1(t+1) = a\,l_1(t) + \sqrt{1-a^2}\,u(t) \tag{5.70}$$

$$\frac{l_2(t)}{l_1(t)} = \frac{1-aq}{q-a} \implies l_2(t+1) = a\,l_2(t) + l_1(t) - a\,l_1(t+1)$$

$$\therefore \; l_2(t+1) = a\,l_2(t) + (1-a^2)l_1(t) - a\sqrt{1-a^2}\,u(t) \tag{5.71}$$

$$\frac{l_3(t)}{l_2(t)} = \frac{1-aq}{q-a} \implies l_3(t+1) = a\,l_3(t) + l_2(t) - a\,l_2(t+1)$$

$$\therefore \; l_3(t+1) = a\,l_3(t) + (1-a^2)l_2(t) - a(1-a^2)l_1(t) + a^2\sqrt{1-a^2}\,u(t) \tag{5.72}$$

$$\frac{l_N(t)}{l_{N-1}(t)} = \frac{1-aq}{q-a} \implies l_N(t+1) = a\,l_N(t) + l_{N-1}(t) - a\,l_{N-1}(t+1)$$

$$l_N(t+1) = a\,l_N(t) + (1-a^2)l_{N-1}(t) - a(1-a^2)l_{N-2}(t) + \ldots \tag{5.73}$$

$$(-a)^{N-2}(1-a^2)l_1(t) + (-a)^{N-1}\sqrt{1-a^2}\,u(t)$$

From equations (5.70–5.73), the matrix **A** can be formed as:

$$\mathbf{A} = \begin{bmatrix} a & 0 & 0 & \ldots & 0 \\ (1-a^2) & a & 0 & \ldots & 0 \\ -a(1-a^2) & (1-a^2) & a & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (-a^{N-2})(1-a^2) & (-a^{N-3})(1-a^2) & (-a^{N-4})(1-a^2) & \ldots & a \end{bmatrix} \tag{5.74}$$

and the vector **b** is found to be:

$$\mathbf{b} = \begin{bmatrix} \sqrt{1-a^2} \\ -a\sqrt{1-a^2} \\ a^2\sqrt{1-a^2} \\ \vdots \\ (-a)^{N-2}\sqrt{1-a^2} \end{bmatrix} \tag{5.75}$$

By inspection of the **A** matrix, its elements are given by:

$$\mathbf{A} = \begin{cases} a_{ij} = a & i = j \\ a_{ij} = 0 & i < j \\ a_{ij} = (-a)^{i-j-1}(1 - a^2) & i > j \end{cases} \tag{5.76}$$

and inspection of the **b** matrix shows its elements are defined by:

$$b_i = (-a)^{i-1}\sqrt{1 - a^2}, \quad i = 1 \text{ to } N \tag{5.77}$$

From equation (5.57), the **c** vector is:

$$\mathbf{c} = \begin{bmatrix} g_1 & g_2 & \cdots & g_N \end{bmatrix} \tag{5.78}$$

and with these definitions, the state-space representation of Laguerre filter model is complete.

One slight modification is required in the above state-space representation so that it can be used for estimation of the minimum achievable output variance. This change arises from the fact that there is no delay in the closed loop noise transfer function and the noise appears directly on the output. The state-space representation becomes:

$$l(t + 1) = \mathbf{A}l(t) + \mathbf{b}e(t)$$
$$y(t) = \mathbf{c}\ l(t) + e(t) \tag{5.79}$$

where the matrix **A** and the vectors **b** and **c** are as previously defined. It should be noted that the matrix **A** and vector **b** of the Laguerre model are set once the time scale and number of filters are chosen. This allows them to be calculated prior to estimation and used directly in computations without any modifications. Now the estimation algorithm used to estimate the elements of the vector **c** as well as the e(t) values will be explained. The method chosen to do this is Recursive Extended Least Squares estimation.

5.3.2 Recursive Extended Least Squares Estimation

The Recursive Extended Least Squares (RELS) estimator will be explained using a time series model and then applied to a state-space representation. A process can be described by the following model:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t) \tag{5.80}$$

36

which is just a general case of equation (2.3) where:

$$A(q^{-1}) = 1 + a_1q^{-1} + \ldots + a_nq^{-n}$$

$$B(q^{-1}) = b_0q^{-k} + b_1q^{-k-1} + \ldots + b_mq^{-k-m} \tag{5.81}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + \ldots c_rq^{-r}$$

which can be expressed as the following linear regression model:

$$y(t) = \theta^T\Phi(t) + e(t) \tag{5.82}$$

with regressor:

$$\Phi(t) = [-y(t-1) \quad \ldots \quad -y(t-n) \quad u(t-1) \quad \ldots$$
$$u(t-m) \quad e(t-1) \quad \ldots \quad e(t-r)]^T \tag{5.83}$$

and parameter vector:

$$\theta = [\,a_1 \quad \ldots \quad a_n \quad b_1 \quad \ldots \quad b_m \quad c_1 \quad \ldots \quad c_r\,]^T \tag{5.84}$$

It would be natural to attempt to estimate the parameter vector using least squares estimation, but this method cannot be used because the e(i) terms are not known and so the regressor cannot be formed. The least squares method can be modified slightly to include estimation of unknown variables in $\Phi(t)$ from available data. Such methods, which combine the estimation of a parameter vector and unobserved components in the regressor are known as Pseudolinear Regressions (PLR). A more common PLR is the extended least squares (ELS) method.

5.3.2.1 Recursive Extended Least Squares Algorithm

The problem with attempting to use least squares estimation in the above situation is the unobserved variables in the regressor. The unobserved variables could be replaced with an estimate of them, given by rearranging (5.82):

$$e(t) = y(t) - \theta^T\Phi(t) \tag{5.85}$$

With this minor change in the regressor, the recursive extended least squares algorithm is given by:

$$\Phi(t) = [-y(t-1) \quad \ldots \quad -y(t-n) \quad u(t-1) \quad \ldots$$
$$u(t-m) \quad \varepsilon(t-1) \quad \ldots \quad \varepsilon(t-r)]^T \tag{5.86}$$

37

$$P(t) = P(t-1) - \frac{P(t-1)\Phi(t)\Phi^T(t)P(t-1)}{1 + \Phi^T(t)P(t-1)\Phi(t)} \tag{5.87}$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\Phi(t)\left[y(t) - \hat{\theta}^T(t-1)\Phi(t)\right] \tag{5.88}$$

$$\varepsilon(t) = y(t) - \hat{\theta}^T(t)\Phi(t) \tag{5.89}$$

The similarity of the RELS algorithm to that of the RLS is the main attraction of this method. Unlike other algorithms used to determine the coefficients of unobserved variables, this PLR does not require any extra filtering or checking that the filter is stable.

5.3.2.2 Convergence

This PLR is not guaranteed to converge to the true parameter vector, θ_0, and relies on the positive realness of certain transfer functions in order to converge. For a process exactly described by the following ARMAX model:

$$A_0(q^{-1})y(t) = B_0(q^{-1})u(t) + C_0(q^{-1})e(t) \tag{5.90}$$

a sufficient condition for convergence of $\theta(t)$ to θ_0 is:

$$\mathrm{Re}\left[\frac{1}{C_0(e^{i\omega})} - \frac{1}{2}\right] > 0 \quad \forall \quad \omega \tag{5.91}$$

or equally:

$$\left|C_0(e^{i\omega}) - 1\right| < 1 \quad \forall \quad \omega \tag{5.92}$$

Thus the filter $\frac{1}{C_0(e^{i\omega})} - \frac{1}{2}$ must be strictly positive real (SPR) to ensure the convergence of the parameter vector to the true values. A necessary condition to assure convergence also exists, which is:

$$\mathrm{Re}\left[C_0(e^{i\omega})\right] > 0 \quad \forall \quad \omega \tag{5.93}$$

If this condition does not hold, then there exists a A_0, B_0, and u(t) so that the probability of $\theta(t)$ approaching θ_0 is zero.

Before continuing and detailing how the Laguerre gains are estimated using RELS estimation, the convergence property will be considered. The transfer function to be modelled so that σ_{uuv}^2 can be found, differs from 5.90 in two ways. There is no $B_0(q^{-1})$ polynomial while $A_0(q^{-1})$ and $C_0(q^{-1})$ are generally complicated polynomials made up from the plant, controller and disturbance filters of the process. This means that the convergence of the RELS estimates depends on whether $R(q^{-1})$, and not $C_0(q^{-1})$ is SPR. In the simulation study the influence of this condition will be further examined.

By adopting Laguerre filters to model the closed loop noise transfer function, it is required that the Laguerre gains, which make up the **c** vector in 5.79, be estimated. The following section explains why RELS estimation is required and how it is applied to this situation, when a state-space representation is used.

5.3.2.3 Estimation of Laguerre Gains

The Laguerre model state-space representation of the closed loop noise transfer function is:

$$
\begin{aligned}
\mathbf{l}(t+1) &= \mathbf{A}\mathbf{l}(t) + \mathbf{b}e(t) \\
y(t) &= \mathbf{c}^T\mathbf{l}(t) + e(t)
\end{aligned}
\tag{5.94}
$$

the vector **c** has N elements, the Laguerre gains, which must be estimated. The parameter vector is:

$$
\theta = [c_1 \quad c_2 \quad \dots \quad c_N]^T
\tag{5.95}
$$

and the regressor is:

$$
\Phi(t) = [l_1(t) \quad l_2(t) \quad \dots \quad l_N(t)]^T
\tag{5.96}
$$

so the linear regression equation becomes:

$$
y(t) = e(t) - \theta^T\Phi(t)
\tag{5.97}
$$

Now the regressor elements are solved from the state equation:

$$
\Phi(t) = \mathbf{A}\mathbf{l}(t-1) + \mathbf{b}e(t-1)
\tag{5.98}
$$

which involves the unknown noise sequence, so RELS is required to estimate the elements of the **c** vector. The above linear regression is then rewritten to give an estimate, $\varepsilon(t)$, of each element in the noise sequence:

$$\varepsilon(t) = y(t) - \theta^T(t)\Phi(t) \tag{5.99}$$

and then the regressor is defined as:

$$\Phi(t) = l(t) = \mathbf{A}l(t-1) + \mathbf{b}\varepsilon(t-1) \tag{5.100}$$

The RELS algorithm can be carried out with these definitions and so the unknown parameter vector containing the Laguerre gains can be identified.

5.3.3 Minimum Achievable Output Variance Estimation Algorithm

Now that all the tools necessary to estimate the minimum achievable output variance have been described, the complete algorithm for its estimation will be given. Following this algorithm will obtain an estimate of the minimum achievable output variance.

1. Prior to estimation, the number of filters, N, and the time scale, a, should be chosen. The matrix **A** and vector **b** can then be determined.

$$\mathbf{A} = \begin{cases} a_{ij} = a & i = j \\ a_{ij} = 0 & i < j \\ a_{ij} = (-a)^{i-j-1}(1 - a^2) & i > j \end{cases} \tag{5.101}$$

$$b_i = (-a)^{i-1}\sqrt{1 - a^2}, \qquad i = 1 \text{ to } N$$

2. Next, an estimate of the **c** vector is obtained along with an estimate of the noise value.

$$l(t) = \mathbf{A}l(t-1) + \mathbf{b}\varepsilon(t-1)$$

$$P(t) = P(t-1) - \frac{P(t-1)l(t)l^T(t)P(t-1)}{1 + l^T(t)P(t-1)l(t)}$$

$$\hat{c}(t) = \hat{c}(t-1) + P(t)l(t)[y(t) - \hat{c}^T(t-1)l(t)] \tag{5.102}$$

$$\varepsilon(t) = y(t) - \hat{c}^T(t)l(t)$$

3. The estimate of the plant delay is found.

$$E_1(t, k_i) = \lambda E_1(t-1, k_i) + [y(t) - y(t-1)]u(t - k_i) \quad \forall \ k_i \epsilon [k_{min}, k_{max}]$$

$$E_1(t, \hat{k}(t)) = \max\{E(t, k_i)\} \quad \forall \quad k_i \ \epsilon \ [k_{min}, k_{max}] \tag{5.103}$$

4. Finally, the estimate of the minimum achievable output variance is calculated using the estimate of the plant time delay, $\hat{k}$, the estimate of the Laguerre gains, $\hat{c}$, and the estimate of the noise sequence, $\varepsilon(t)$.

$$\hat{\sigma}^2_{mv} = \hat{\sigma}^2_\varepsilon \left[1 + \sum_{i=1}^{\hat{k}-1} \left(\hat{c} A^{i-1} b \right)^2 \right] \tag{5.104}$$

5. Return to step 2 for the next sample data.

Step four does not have to be performed for every new sampled data and to save calculations, the **Ab**, **A²b**, up to **A^{kmax}b** can be computed prior to estimation.

5.3.4 Example

This example will show the ability of this algorithm to estimate the minimum achievable output variance, as well as highlight interesting aspects that result during estimation. The process used for this example is shown in Figure 5.13 and consists of a first-order plant and a ARMA(2,1) disturbance. The time delay of the plant is four samples and the plant is under Dahlin control. Along with the process, some of its properties are also shown in Figure 5.13. The transfer function to be estimated is shown along with the coefficients of the $F(q^{-1})$ polynomial. There are four coefficients in this polynomial which corresponds to the delay and are used with the variance of the noise to determine the minimum achievable output variance shown. Comparing the output variance to the minimum

$\sigma^2_e = 0.81$

$\sigma^2_y = 1.2116$

$\sigma^2_{mv} = 0.9274$

$$\frac{y(t)}{e(t)} = \frac{1 - q^{-1} + 0.31q^{-2} - 0.03q^{-3} - 0.05q^{-4} + 0.25q^{-5} - 0.03q^{-6}}{1 - 1.24q^{-1} + 0.37q^{-2}}$$

$$F(q^{-1}) = 1 + 0.24q^{-1} + 0.2376q^{-2} + 0.1758q^{-3}$$
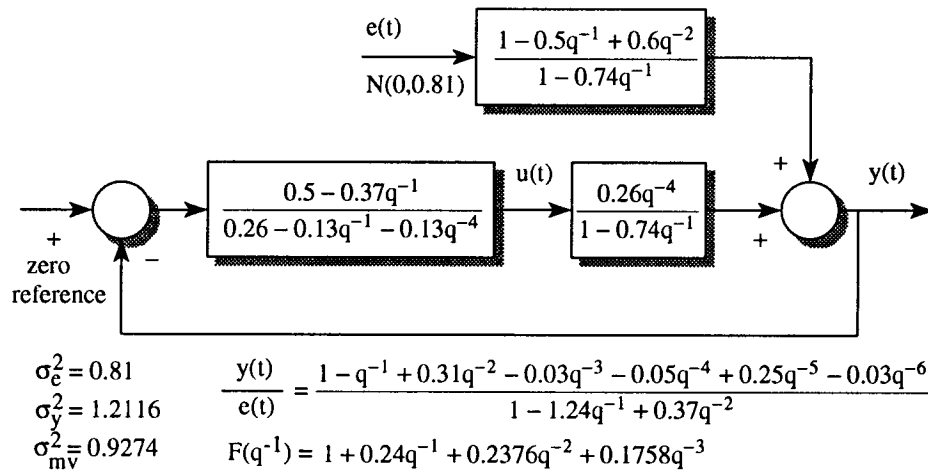
Figure 5.13: Process used in the example
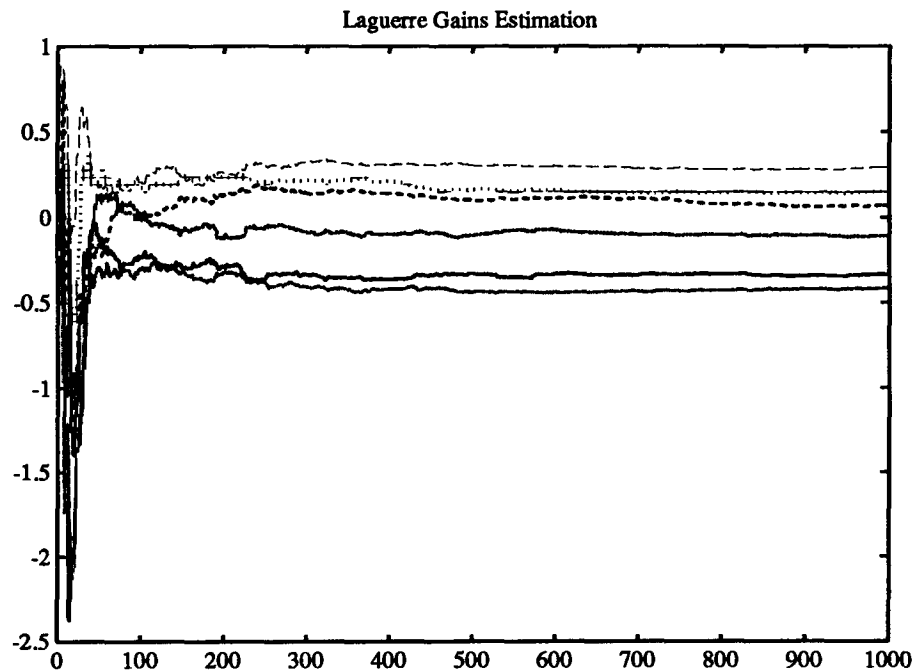
Figure 5.14: The Laguerre gains estimated during minimum achievable output variance estimation

achievable output variance, it is clear that the Dahlin controller keeps the output quite close to the minimum value.

The results of one simulation using one thousand data points and using a Laguerre network with six filters and a time scale of 0.3, are shown in Figures 5.14–5.16. This simulation used a constant reference signal, which will be the situation in a regulation problem, and in this case the reference

Number of Filters: 6

Time Scale: 0.3

$F(q^{-1}) = 1 + 0.2210\,q^{-1} + 0.2638\;q^{-2} + 0.2225\;q^{-3}$

$\sigma_\varepsilon^2 = 0.8197$
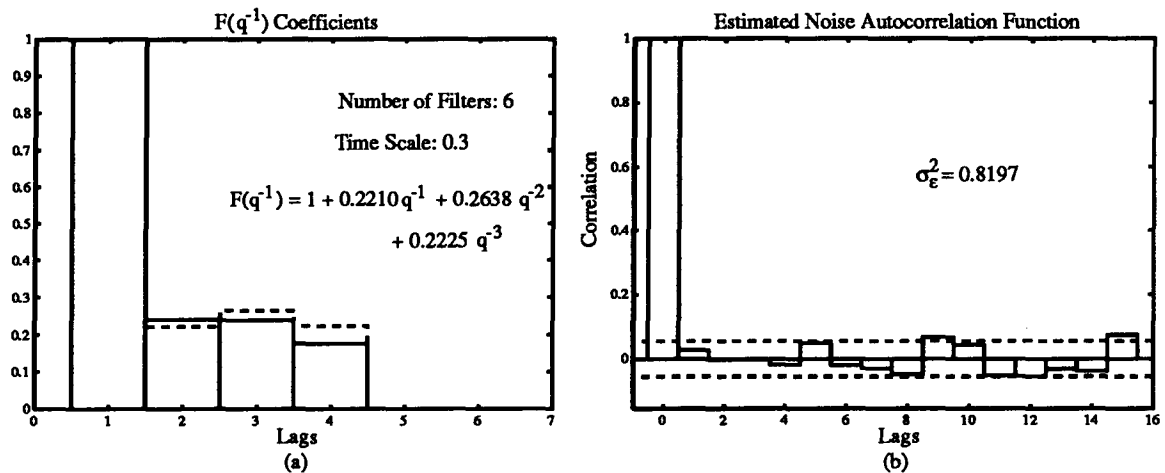
Figure 5.15: The (a) estimate of the $F(q^{-1})$ polynomial and (b) the estimated noise autocorrelation
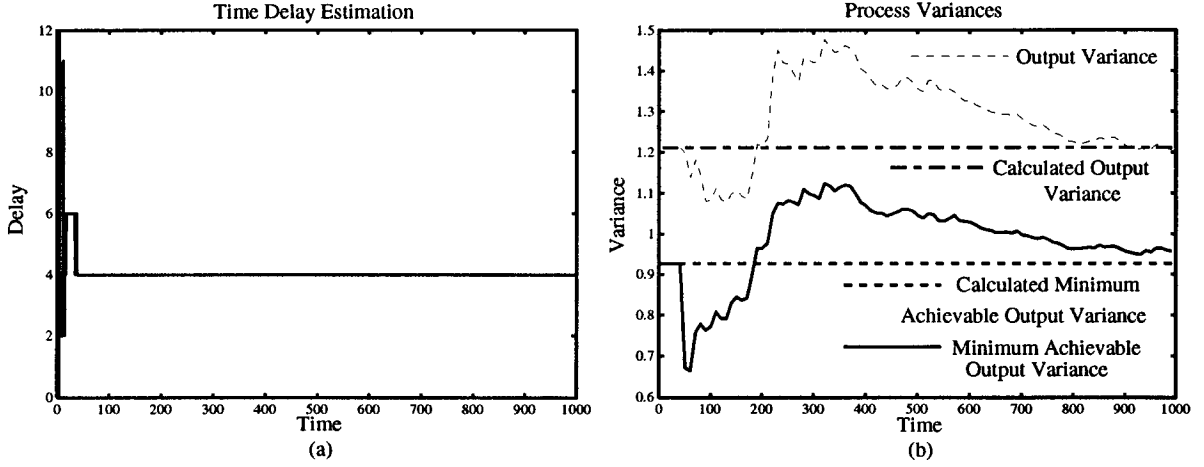
Figure 5.16: The (a) FMVRE time delay estimation and the (b) estimated minimum achievable output variance and process output variance

used was zero. The estimation of the Laguerre gains appears in Figure 5.14 and it is apparent that the gains level out after roughly four hundred points. However, these gains do not give any indication of how well the transfer function has been modelled by the Laguerre network. To see this, equation 5.63 is used to calculate the $F(q^{-1})$ coefficients, which appear in Figure 5.15(a). The close agreement suggests the Laguerre network was successful in modelling the transfer function. The autocorrelation of the estimated noise sequence appears to approximate white noise, as shown in Figure 5.15(b) as it has an impulse shape. This is what is desired as the sequence being estimated is white noise which has an autocorrelation function that has no correlation at all nonzero lags. The true delay was correctly estimated after roughly fifty samples and the estimated minimum achievable output variance is compared to the present output variance in Figure 5.16(b). As the closed loop function has an infinite impulse response, it should be realized that the calculated output variance is only an approximation of the true output variance. The estimated value, $\hat{\sigma}^2_{mv}$ approaches σ^2_{mv}, which was to be expected based on the estimation of the noise sequence and $F(q^{-1})$ coefficients. As mentioned, this simulation was performed using six Laguerre filters with a time scale of 0.3, this information is always shown on the plot of the estimated $F(q^{-1})$ coefficients, Figure 5.15(a). One simulation may not, however, give a complete sense of how well the estimation is performing and conclusions drawn based on a single simulation could be erroneous.
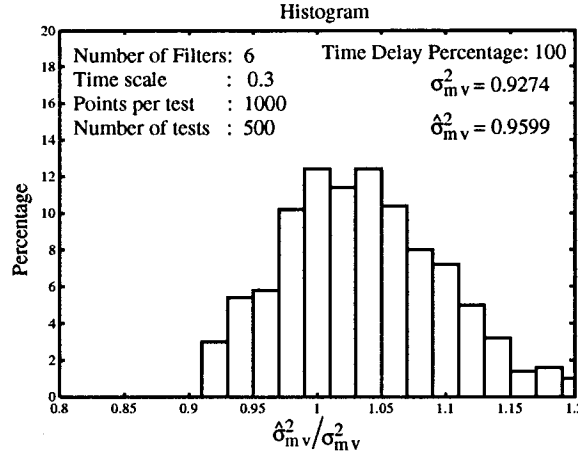
Figure 5.17: Distribution of the results of five hundred simulations

To arrive at a better sense of how well the estimation performs, this simulation was repeated five hundred times using different noise sequences to obtain a distribution of the results. These results are presented in Figure 5.17 as a histogram of the estimated minimum achievable output variance to its calculated value. From this histogram it can seen that roughly twelve percent of the 500 tests resulted in an estimated minimum achievable output variance that was between 0.99 and 1.01 of the calculated value. Also, the time delay percentage is listed in the upper right corner of the plot. The value listed corresponds to the percentage of the five hundred tests which had the delay correctly estimated. In this case, the delay was correctly estimated in all the simulations. The shape of the curve gives an indication of the distribution or variance of the values estimated. As the mean value was 0.9599, this indicates that the results obtained in Figure 5.14–5.16 can be considered typical. Also, the mean value of 0.9599 gives a measure or indicator of how well this choice of filter number and time scale performed.

Now that all of the monitoring tools have been discussed, a comprehensive simulation study will be presented which will examine the operation of the various estimators. The study is used to examine the performance of the estimators and to determine any limitations of the methods.

Chapter 6

Simulation Study

The following chapter presents various simulations to determine the performance of the monitoring tools. The simulations shown are by no means inclusive of all the simulations performed, but they do highlight the important trends encountered. Because the results of simulations depend on the process used, the results obtained are not conclusive. No recommendations should be applied universally to all processes as the conclusions are based on the specific process examined. The conclusions derived from simulations provide a good starting approach to follow when applying the monitoring to new processes.

The first study investigates the possibility of biased delay estimation occurring. Next, the effect of feedback and noise on the estimation of the static input-output relation by the ANM is examined. This is followed by estimation of two typical nonlinearities, an actuator saturation and a deadzone. The operation of the method used to determine the minimum achievable output variance is then studied. Various process properties are modified in order to see how they manifest themselves in the estimation. Finally, the choice of Laguerre network parameters is examined including how they should be modified based on the characteristics of the closed loop noise transfer function. The results of these simulations can be used when the methods are applied to data from industrial processes.

6.1 Time Delay Estimation

The FMVRE generally provides an estimate of the true delay in a process. However, there are situations where the estimate of the delay is not correct. This section presents the three different situations found in which a biased delay estimate occurred. This knowledge will enable the user of the FMVRE to determine the validity of a delay estimate. Once examples of different cases have been introduced, each will be examined individually. The biased estimation encountered usually falls into one of the following three examples, which are an incorrect delay estimate of one, an incorrect delay estimate but not at one, or an incorrect delay estimate due to insufficient excitation.

45
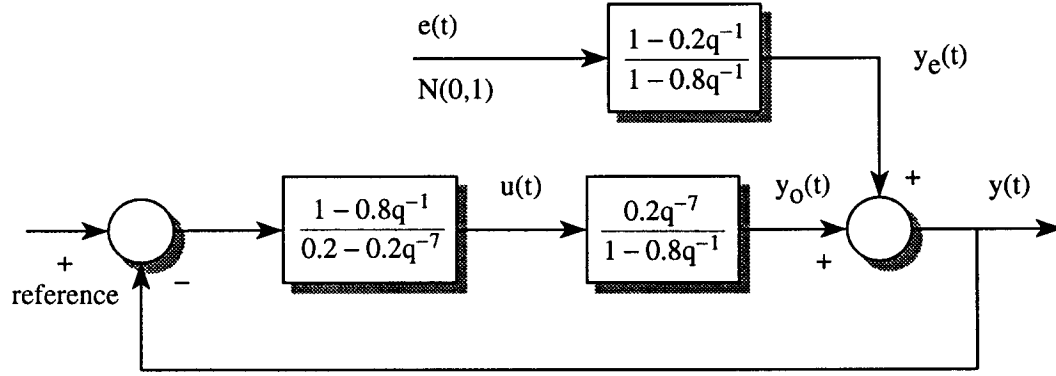
Figure 6.18: Process which results in correct delay estimate

6.1.1 Unbiased Estimation

This example of unbiased estimation is given in order to use as a comparison in the future discussion of biased estimation. By comparing aspects of both the biased and unbiased estimation, it is often possible to obtain insight into the causes of the bias. The process appears in Figure 6.18 and has a structure similar to the ones to be used which result in a biased time delay estimation. It has a time delay of seven and is controlled by a deadbeat controller. The resulting E_1 function is shown in Figure 6.19 and the true delay of seven was estimated. Now, processes will be presented that result in biased delay estimation and this process will be referred to during the examination of the causes of biased estimation.
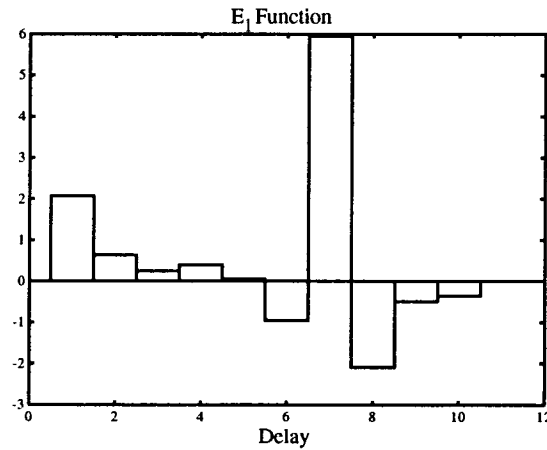
Figure 6.19: E_1 function from FMVRE estimation resulting in correct delay of seven
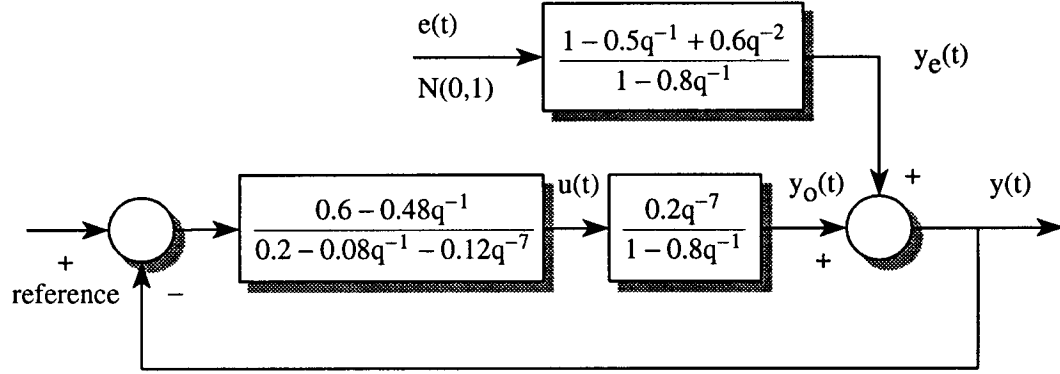
Figure 6.20: Process which results in a biased delay estimate

6.1.2 Estimated Delay of One Sampling Interval

The first situation to be examined results in a biased delay estimate of one sampling interval. The process under consideration appears in Figure 6.20 and it consists of the same first-order plant with a large time delay of seven. However, now the process uses a Dahlin controller which moves the closed loop pole to 0.4 and a second-order, ARMA(2,1) noise disturbance. The results of the FMVRE estimation appear in Figure 6.21 and show that the delay was incorrectly estimated. This example illustrates that it is possible for an incorrect time delay estimation to occur.

6.1.3 Arbitrary Bias

The next situation considered uses the process in Figure 6.22, which also has a first-order plant model, but has a time delay of two. It also uses a Dahlin controller, identical to the one used
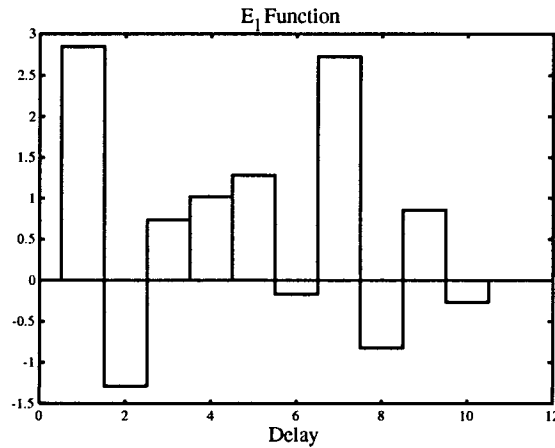
Figure 6.21: E_1 function from FMVRE estimation resulting in a bias at one

47
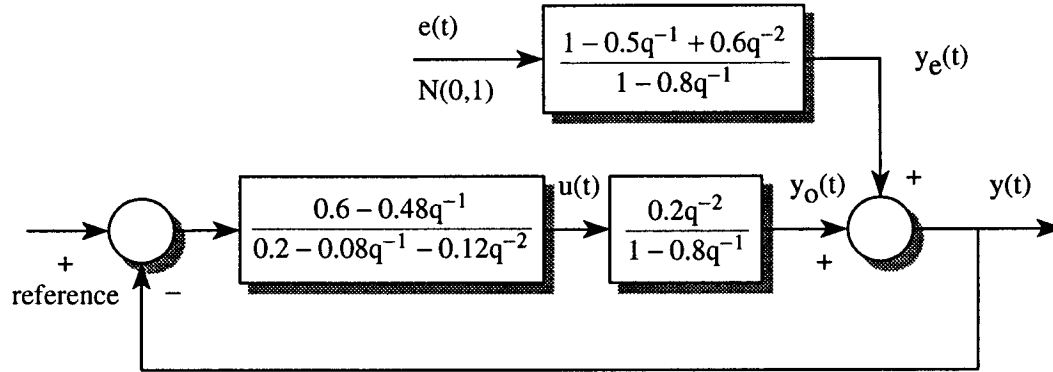
Figure 6.22: Process which results in a biased delay estimate

previously. The results of the FMVRE delay estimation appear in Figure 6.23. On this occasion, the time delay is also incorrectly estimated, but as four. This process shows that it is possible to have an arbitrary bias of the time delay.

6.1.4 Insufficient Excitation

The final process used appears in Figure 6.24 and its major difference from the previous processes considered is its use of a PID controller. The time delay of the plant is three and the noise filter is unchanged from the last example. The results of the FMVRE estimation, shown in Figure 6.25, indicate that the delay is one. This is not the correct delay and the peculiar shape of the E_1 function indicates that the cause of this bias is different than in the case of the previous examples.

Now that it has been demonstrated that it is possible to have a biased time delay estimate, the reason for its occurrence and possible solutions will be examined. It is hoped that the cause of the
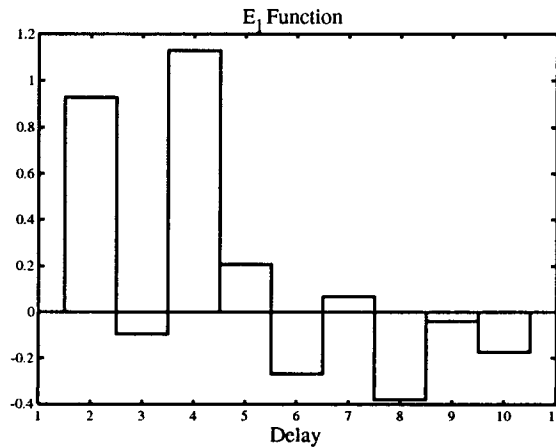
Figure 6.23: E_1 function from FMVRE estimation resulting in a bias at four
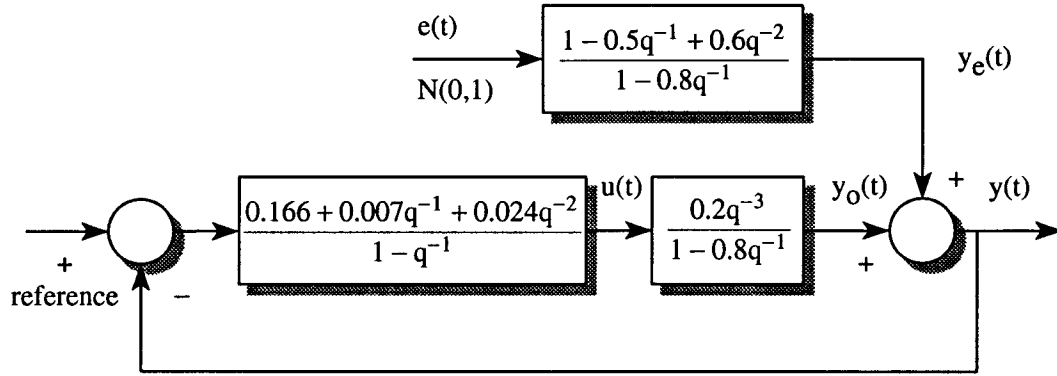
Figure 6.24: PID controlled process which results in a biased delay estimate

bias can be eliminated, but failing that, it should be possible to be able to indicate if the estimated delay is biased.

6.2 Causes and Solutions for Incorrect Delay Estimation

6.2.1 Biased Delay

The first two examples of biased delay share a similar cause, however they affect estimation in a different manner. According to Elnaggar [12], a sufficient condition exists so that delay estimation is unbiased. This condition states that in order to ensure correct delay estimation in a closed loop process, the autocorrelation function of the output of the disturbance filter, $r_{y_e}(\tau)$ must be zero for all lags over the interval $[k_{min}+k_{fb},k_{max}]$, where k_{fb} is the delay in the controller. If $r_{y_e}(\tau)$,
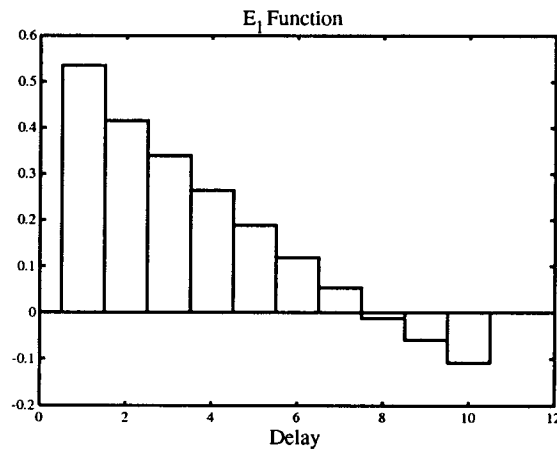
Figure 6.25: E_1 function from FMVRE estimation resulting in a bias at one
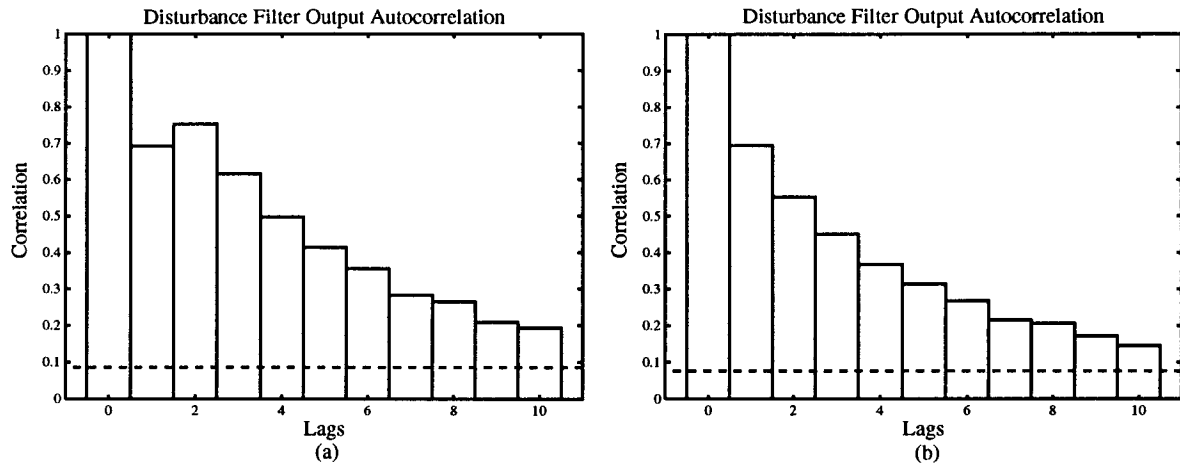
Figure 6.26: The disturbance output autocorrelation from the process (a) resulting in a delay bias, Figure 6.20, and (b) resulting in a correct delay estimate, Figure 6.18

Figure 6.26(a), is examined for the process in Figure 6.20, it is evident that the correlation at lags over the interval [0,10] is not zero. However, if $r_{y_e}(\tau)$, Figure 6.26(b), is examined for the process in Figure 6.18, it can be seen that it also has correlation at lags over the interval [0,10]. This emphasizes that the condition on $r_{y_e}(\tau)$ is only a sufficient condition and does not fully determine if a bias occurs. Therefore, something else plays a role in creating the bias. This other factor can be determined by examining the cross-correlation between the input and the disturbance filter output, $r_{uy_e}(\tau)$, which is obtained by the FMVRE when the signals u(t) and y_e(t) are used. The
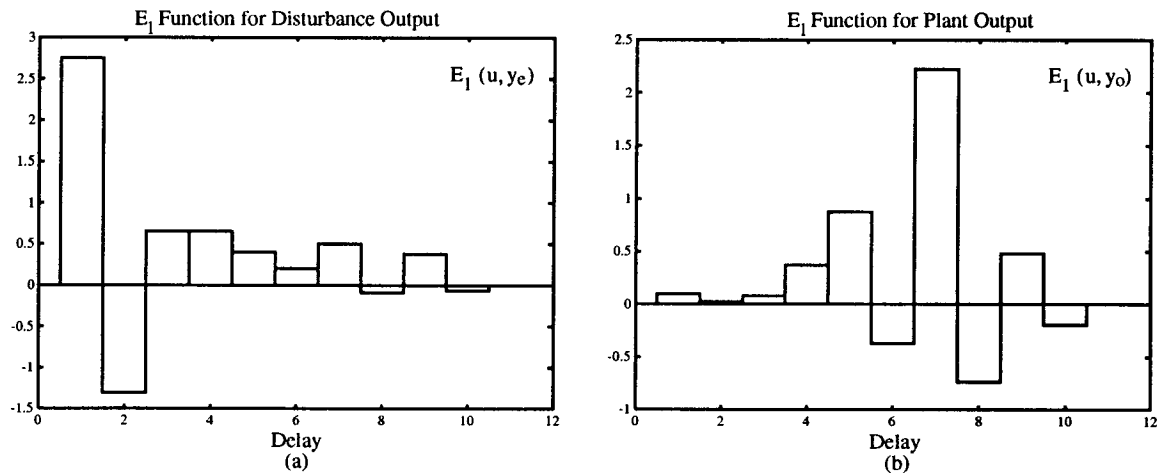
Figure 6.27: E_1 functions for (a) the input and plant output and (b) the input and disturbance output resulting in a bias at one

50
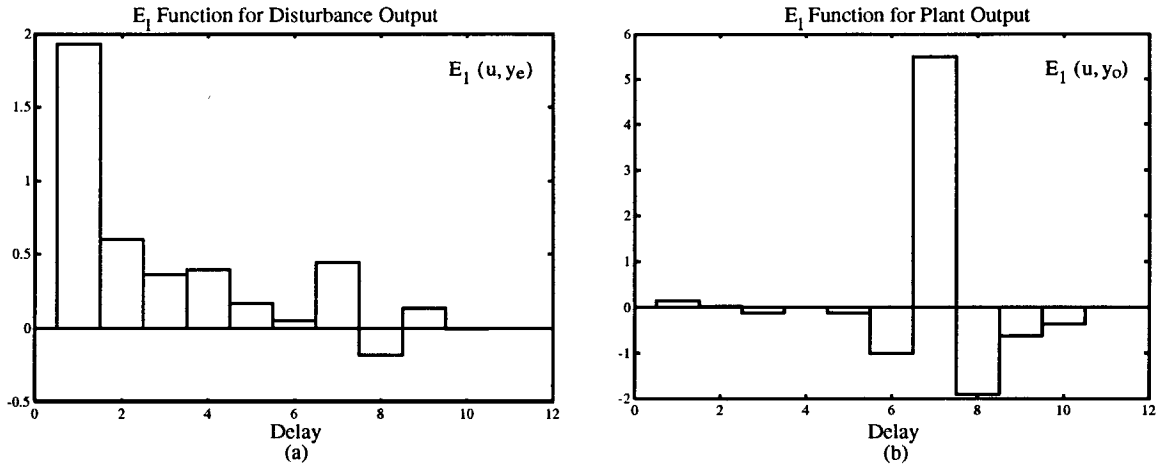
Figure 6.28: E_1 functions for (a) the input and plant output and (b) the input and disturbance output resulting in a correct estimate

calculated $E_1(u,y_e)$ function, Figure 6.27(a), shows a large bias at one, when compared to the peak in the $E_1(u,y_o)$ function corresponding to the correct delay, Figure 6.27(b), shows the value at one is larger than the correct value. When the two functions are added together, a bias occurs at one and an incorrect delay is suggested. Returning to the process in Figure 6.18, its $E_1(u,y_e)$ and $E_1(u,y_o)$ functions appear in Figure 6.28 and show that the peak at seven of the $E_1(u,y_o)$ function is larger than the peak at one of the $E_1(u,y_e)$ function and so the resulting $E_1(u,y)$ function did not have a bias. So the bias in the delay occurs when the $E_1(u,y_o)$ function, created by the plant, is corrupted by the noise $E_1(u,y_e)$ function.

From the point of view of monitoring, it is not reasonable to expect to make changes to the process in order to improve conditions for better identification. However, with a little consideration, it may be possible to remove the effect of this bias from estimation. Upon examining Figure 6.21, the largest value after the bias at one, is the true delay at seven. Would it be possible to change the interval $[k_{min}, k_{max}]$ to avoid the bias at one? The answer is yes if there is a time delay in the process beyond that resulting from the sampling process. In this case, though the delay may vary, it would not be possible for it to disappear, causing the true delay to correspond to one. Thus by using a lower bound on the possible delay interval above one, this biased delay problem can be avoided.
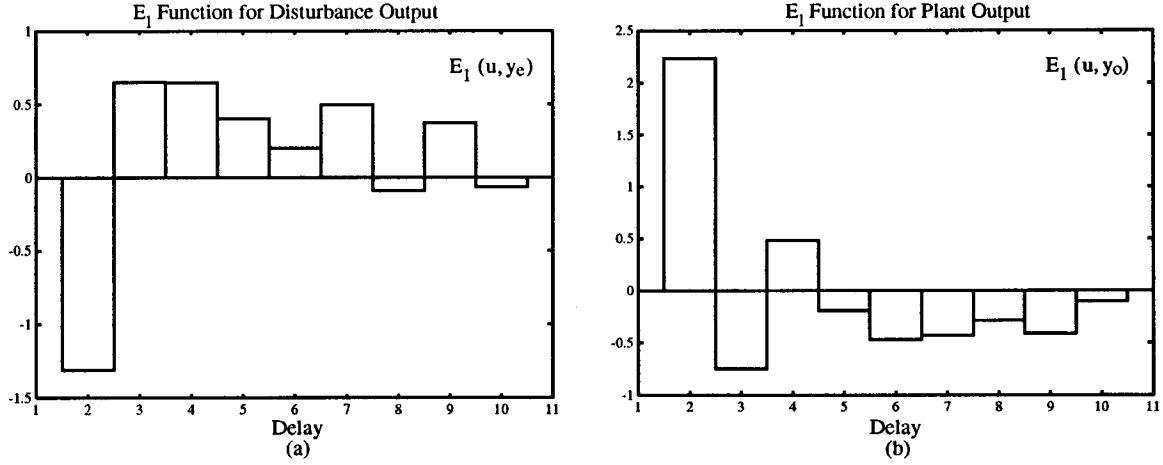
Figure 6.29: E_1 functions for (a) the input and plant output and (b) the input and disturbance output resulting in a bias at four

Next, the bias that resulted in the process shown in Figure 6.22, is considered, and it should be noted that it only differs from the process used previously in that the plant time delay was changed to two. The FMVRE estimator used a range of [2,10] as the possible delay and yet the delay was still biased. The reason for this can be understood by examining the $r_{uy_e}(\tau)$ cross-correlation function, as obtained by FMVRE estimation. Figure 6.29(a) shows the $E_1(u,y_e)$ function obtained for this process. The value of the function at two is negative, so when this function is added to $E_1(u,y_o)$, Figure 6.29(b), the correct delay value is cancelled. The maximum at four, as found in Figure 6.23, results from this cancelling effect.

It is possible to correct this bias, using an approach similar to Clarke [10], if the input and output signals are modified using a filter $G_f(q^{-1})$, so that the resulting $E_1(u_f,y_{ef})$ function does not affect the $E_1(u_f,y_{of})$ function and cause a bias. For processes considered here, the filtering is performed as:

$$G_f^{-1}(q^{-1})y(t) = G_p(q^{-1})G_f^{-1}(q^{-1})u(t) + G_f^{-1}(q^{-1})G_d(q^{-1})e(t)$$

$$y_f(t) = G_p(q^{-1})u_f(t) + y_{ef}(t) \tag{6.105}$$

which shows how the filtered signals are formed. The optimum filter to use is the disturbance filter of the process, which in this case would be:

$$G_f(q^{-1}) = \frac{1 - 0.5q^{-1} + 0.6q^{-2}}{1 - 0.8q^{-1}} \tag{6.106}$$

and shall be used here for illustration purposes. The two functions, $E_1(u_f,y_{ef})$ and $E_1(u_f,y_{of})$ were calculated and appear in Figure 6.30. They show that the negative value at two has been removed in

52
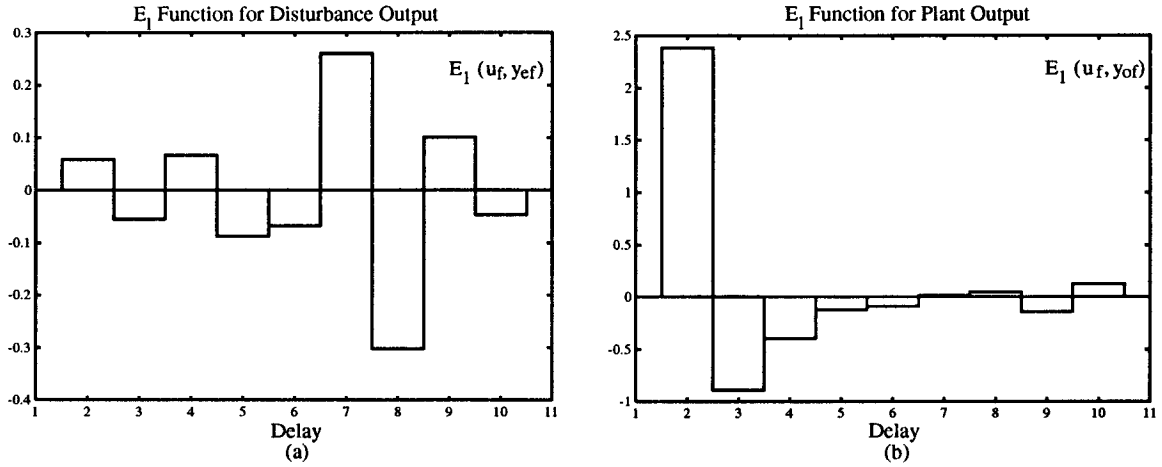
Figure 6.30: E_1 functions for (a) the filtered input and filtered plant output and (b) the filtered input and filtered disturbance output resulting in correct delay estimate

$E_1(u_f, y_{ef})$, while $E_1(u_f, y_{of})$ still has a maximum at two, the correct value. When the two functions are added together, the resulting $E_1(u_f, y_f)$ function, Figure 6.31, shows correct delay estimation. While it may not be possible to find the disturbance transfer function, Elnaggar [13] showed that the exact transfer function is not required to remove the biased delay.

6.2.2 Insufficient Excitation

The final incorrect delay estimation occurred to the process in Figure 6.24 and the resulting FMVRE estimation suggested incorrectly that the plant delay was one. This is definitely wrong and the cause of this error will be explained. It will be shown that there is nothing that can be done to
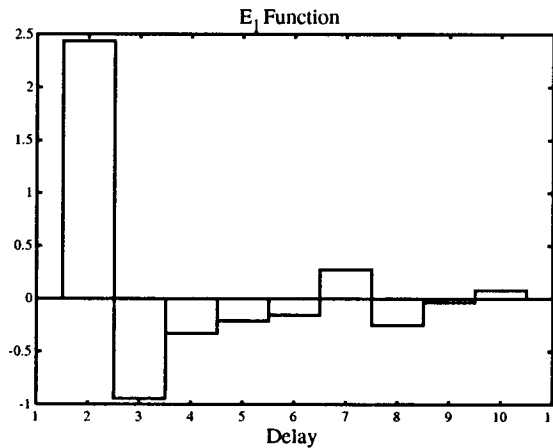
Figure 6.31: E_1 function from FMVRE estimation using filtered signals, resulting in a correct delay estimate
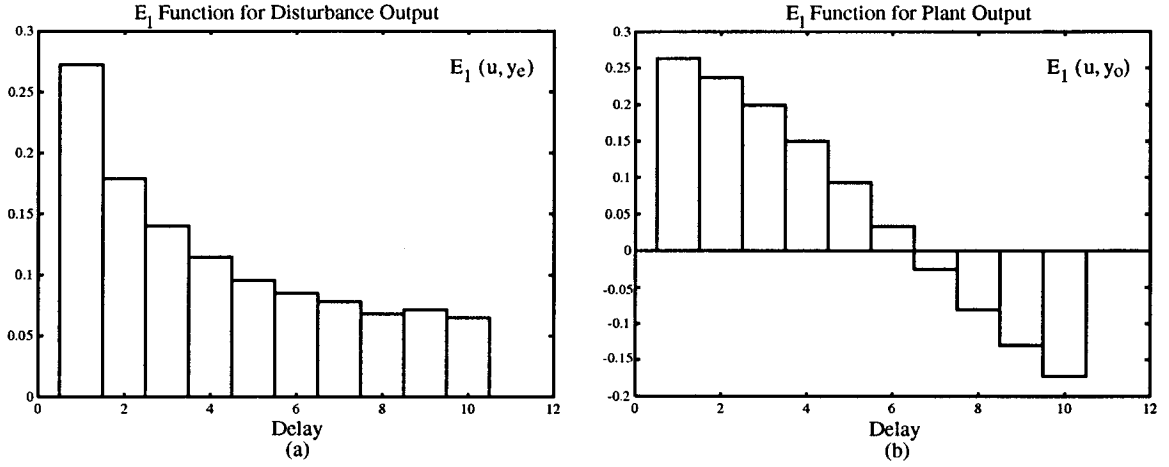
Figure 6.32: E_1 functions for (a) the input and plant output and (b) the input and disturbance output resulting in a bias at one

correct the problem, however it will be shown that it should be possible to detect that the problem is present, thus being able to warn that the biased delay estimation has occurred.

The result which suggests that the cause of this bias differs from the previous cases is shown in Figure 6.32. The FMVRE estimation using the plant output and disturbance output show that the contribution from the $E_1(u,y_e)$ function is not the cause of the bias, as the FMVRE obtain erroneous results from using the plant output, $E_1(u,y_o)$. This had not occurred in the previous cases and suggests a different cause of the bias. This led to the investigation of the excitation requirements of the input signal used by the FMVRE in order to obtain correct delay estimation. These requirements were found by Elnaggar [12] and it was shown that the input must contain frequencies in the range:

$$\pi\sqrt{2T_s\tau_{min}} > \frac{1}{f_e} > T_s(k_{max} - k_{min}) \qquad (6.107)$$

to obtain correct delay estimation, where T_s is the sampling period and τ_{min} is the fastest time constant of the process. It should be noted that the lower bound on the frequency was determined for minimum phase stable systems with a pole excess of one, which is the case for first-order plants, the type under consideration. The upper bound on the frequency ensures that there is only one maximum peak in the E_1 function over the range $[k_{min},k_{max}]$. It is the lower bound however, which is of more interest, as this is what causes the problems experienced. The lower bound is set to ensure that the phase difference between the input and output is large enough to ensure correct delay estimation.

Input Power Spectral Density Function
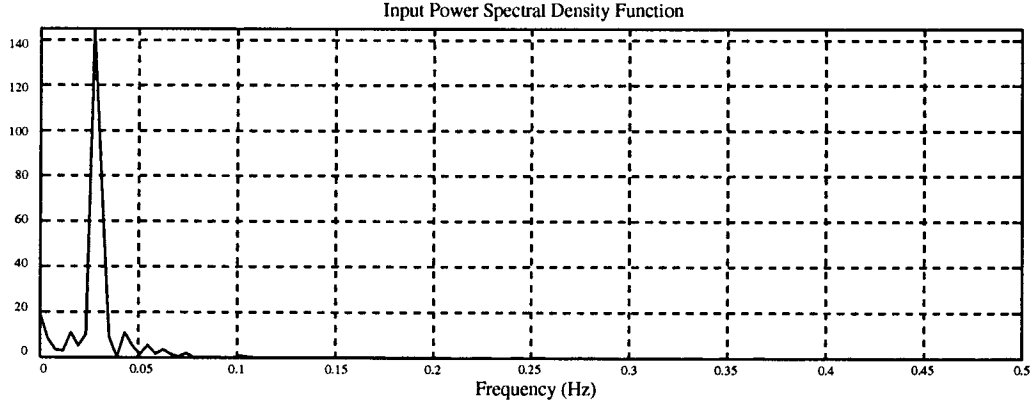
Figure 6.33: Power spectral density function for the input signal from the PID controller

Returning to the process under consideration, the lower bound on the input frequency content is (assuming $T_s=1$):

$$f_e = \left(\pi \sqrt{2 T_s \tau_{\min}} \right)^{-1} = \left(\pi \sqrt{2(1)(4.4823)} \right)^{-1} = 0.1063 \text{ Hz} \qquad (6.108)$$

Thus for unbiased delay estimation to occur, the input must contain frequency components above this level. The power spectral density function of the input in this process appears in Figure 6.33. Not surprisingly, the frequency content of the input all but disappears above 0.1 Hz. This confirms that the cause of the bias was indeed an input signal with too low a frequency content. For comparison purposes, the power spectral density function of the input for the process in Figure 6.18, where correct delay estimation occurred, appears in Figure 6.34. It shows that the input has frequency content over a wide range, which allows unbiased estimation.

Input Power Spectral Density Function
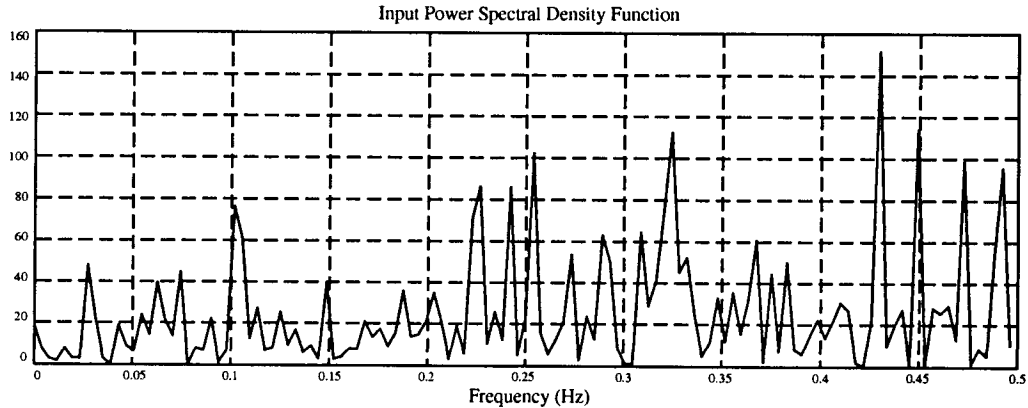
Figure 6.34: Power spectral density function for the input signal from the Deadbeat controller

55

What caused the removal of the higher frequencies from the input signal used in the process of Figure 6.33? As the process is being driven by white noise only, the input should contain components at all frequencies. The answer is found upon examination of the controller used. The magnitude response of the controller being used, Figure 6.35 (again assuming $T_s=1$), shows attenuation of the higher frequencies in the input signal. This is common when a PID or PI controller is used on a process with large delay, as stability can only be maintained by choosing controller parameters that ignore fast changes and makes only slow adjustments.

From the point of view of monitoring, little can be done to cure this problem. The controller is causing the problem and changes to it can not be justified only to improve monitoring. However, the problem itself suggests that consideration be given to the fact that the controller be changed. Though the problem cannot be fixed, its presence can be detected. It becomes apparent after examination of the shape of the E_1 function, because if it has a sinusoidal shape rather than a definite maximum, then this suggests that the input signal does not contain sufficient excitation. Also, an empirical and therefore rough limit on the lower bound can be formulated. Assuming that the sampling occurs at a reasonable rate, that being around 4 or 5 times smaller than the process time constant, and that the sampling time is known, the lower bound is found to be:

$$f_L \simeq \left(\pi \sqrt{8 T_s^2}\right)^{-1} \tag{6.109}$$

Thus for any T_s, a lower bound on the frequency content can be quickly determined to indicate the frequency requirements on the process input.
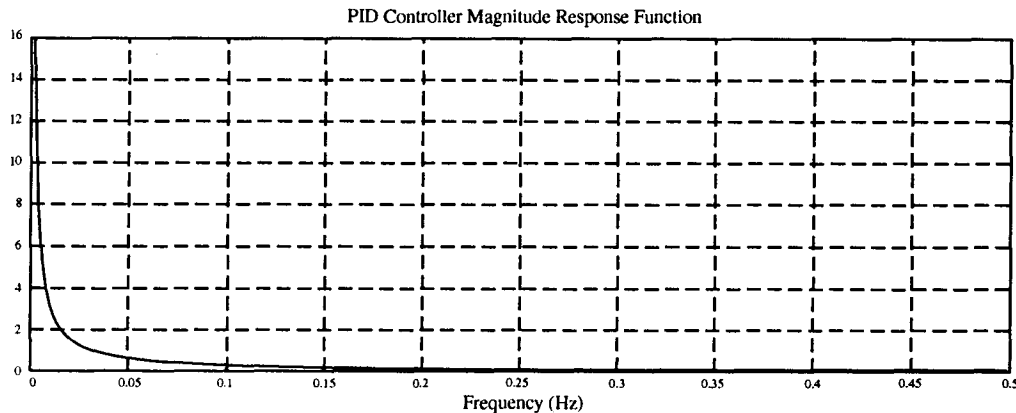
Figure 6.35: Magnitude response of the PID controller

It has been shown that the FMVRE delay estimator is not infallible. In general, when dealing with this estimator, it is not prudent to take the suggested result without examining the E_1 function. In cases where this function has a large and definite maximum value, then the delay can be treated with some confidence. However, when the E_1 function has two or more peaks close together, as is the case in the above examples, then the estimate cannot be used with complete confidence. So like all estimation, the validity of the result should be considered before it is used with abandon.

6.3 ANM Estimation of Stochastic Feedback Processes

The ability of the ANM to work when the signals are provided by a process under feedback control and/or if the signals are noisy must be investigated. It will then be possible to conclude if the ANM is suitable to be used for process loop monitoring. The purpose of these simulations is to test the ability of the ANM to correctly estimate the static relationship under these conditions and to test some procedures which attempt to lessen the adverse effect of these conditions.

The process to be simulated, Figure 6.36, consists of a first-order plant with a gain of five and a delay of two as well as a disturbance filter which colors the white noise sequence that drives it. A Dahlin controller is used for control, however it was poorly designed using a delay of three, to get a more active control signal. In order to get a yardstick with which to compare the following simulation results, an input signal, Figure 6.37, was placed through a pure gain of five only, and the output generated was used by the ANM to estimate the static gain. The results obtained from
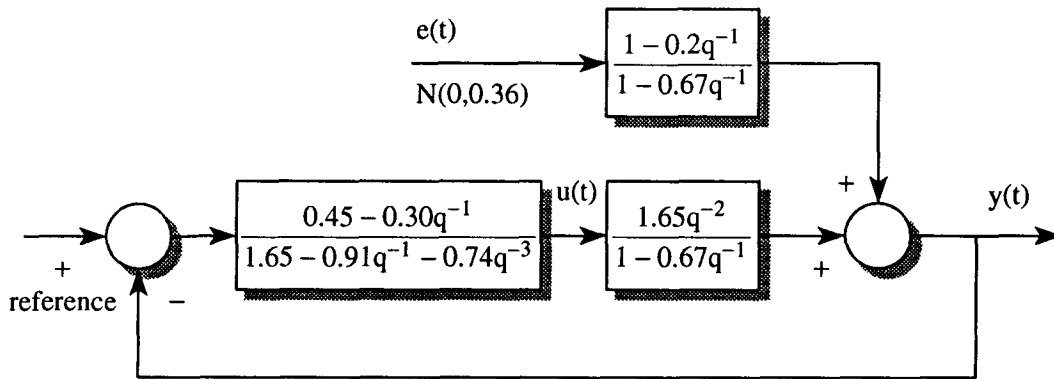
Figure 6.36: Process used to test ANM
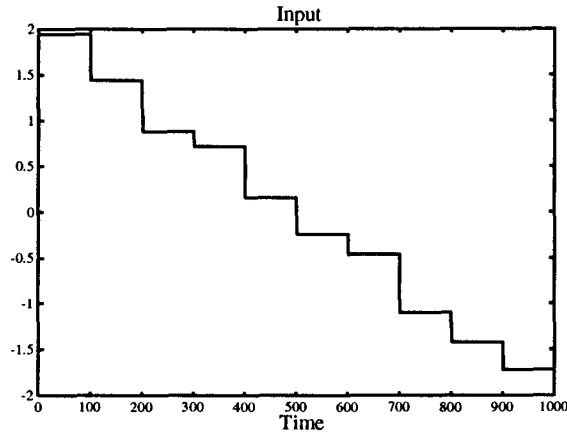
Figure 6.37: Signal used to estimate the static input-output relationship

the estimation appear in Table 6.1 beside the row marked Test A. The results are quite close to the expected value and give a sense of how close the following simulations can come to the correct value, using the given input.

The next simulation used the process in Figure 6.36, but with the disturbance filter disconnected. This was done to investigate the effect of the plant dynamics, introduced by feedback, when the reference signal changes. The reference signal used was chosen so that the steady state control signal would be the same as the signal in Figure 6.37. The input and output signals used by the ANM for modelling of the static gain appear in Figure 6.38 and the final table values estimated appear after Test B in Table 6.1. Comparing these results with Test A, indicates that the brief periods where plant dynamics were present did not greatly affect the static gain estimation. It was mentioned previously that Åström suggested turning estimation off after a reference change occurred, to let the plant dynamics die out. This was done, by turning estimation off for ten samples after the reference change, thus removing the bulk of the dynamics. The results appear in Table 6.1, Test C. As expected, the estimated values improved and practically returned to the results obtained in Test

Input	2	1.5	1	0.5	0	−0.5	−1	−1.5	−2
Output	10	7.5	5	2.5	0	−2.5	−5	−7.5	−10
Test A	9.88	7.43	4.95	2.50	−0.01	−2.47	−4.88	−7.61	−9.38
Test B	9.80	7.49	5.05	2.53	−0.13	−2.43	−4.82	−7.49	−9.36
Test C	9.85	7.43	4.95	2.50	−0.01	−2.47	−4.86	−7.61	−9.33

Table 6.1: ANM results for estimation performed under closed loop operation
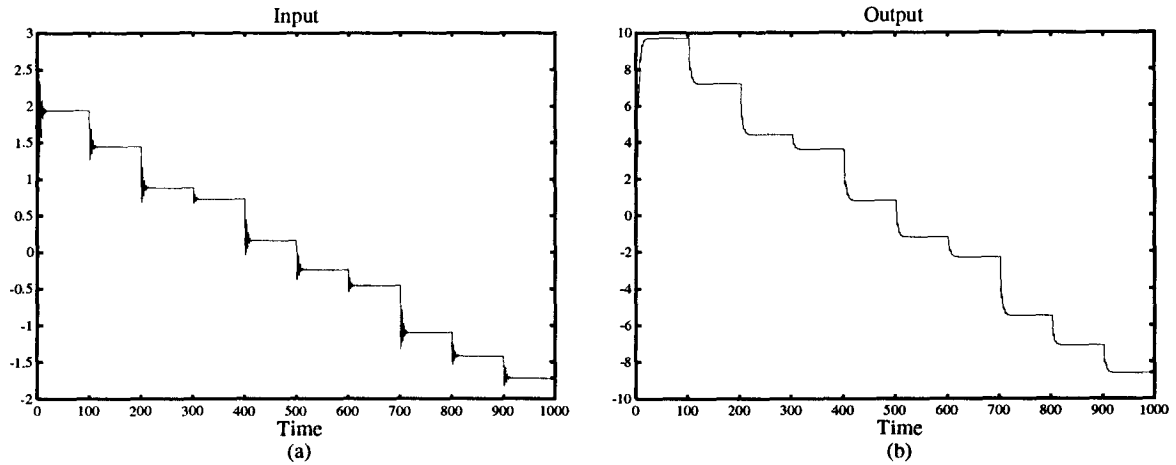
Figure 6.38: The (a) process input and (b) process output from closed loop operation

A. While turning estimation off during reference changes improves estimation, the deterioration in the first place, even for quite an active signal as in this case, was not that large. This suggests that turning the estimator off may not be necessary.

For the remaining tests, the noise in Figure 6.36 was added, and the estimator used the resulting signal. This addition will change the static input-output estimation in two ways, as it will add noise to the to the output signal, so the output is comprised of both the input and the noise. Secondly, the noise on the output will cause the output to vary from the reference, thus causing the control to be active, introducing dynamics into the output. The signals which are obtained from the process are shown in Figure 6.39 and they were used directly for estimation of the static gain. The results of
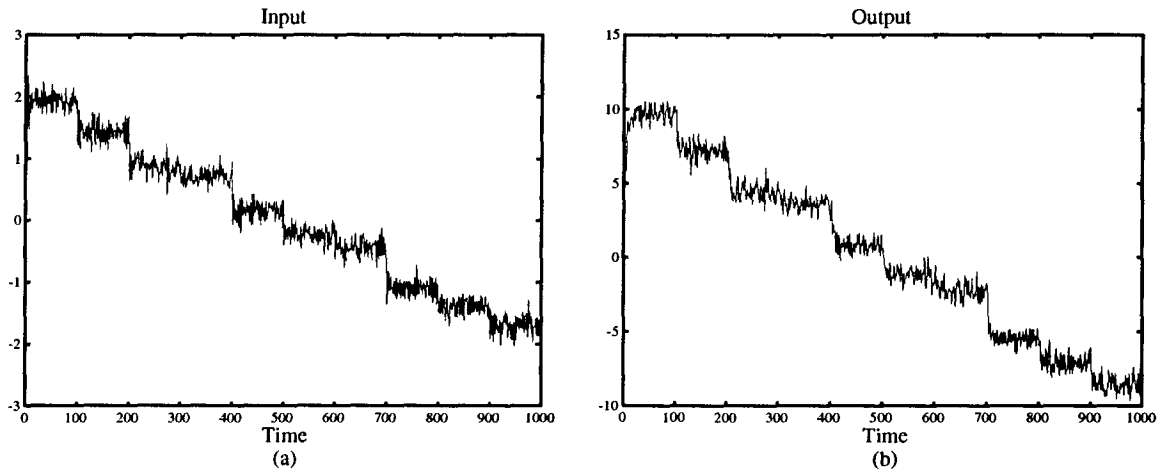
Figure 6.39: The (a) process input and (b) process output from stochastic closed loop operation

59
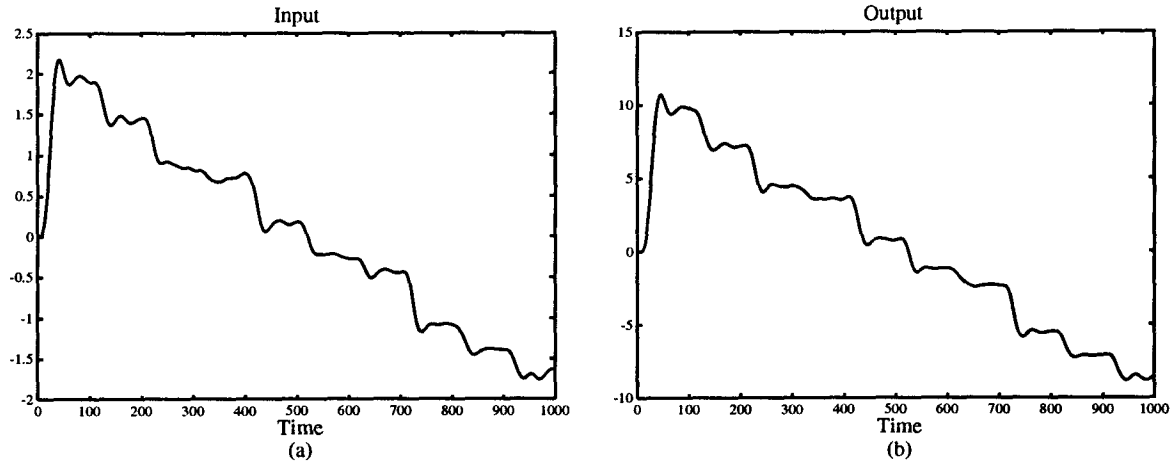
Figure 6.40: The (a) filtered process input and (b) filtered process output used by the ANM

the estimation, Test D in Table 6.2, show that the addition of the noise does indeed adversely affect estimation. This means something must be done so that the estimation can be improved.

The signals of Figure 6.39 were filtered, using a low pass Butterworth filter, in an effort to remove the effects of the noise on estimation. Three simulations were done and their results are in Table 6.1, Test E, F and G. The filter used for Test E had a cutoff frequency of 0.3 of the sampling frequency while Test F had the frequency reduced to 0.1 and Test G used a cutoff of 0.05 of the sampling frequency. The signals which resulted from the filtering are typical of those in Figure 6.40 which are filtered by a filter with a cutoff of 0.05. Below a cutoff frequency of 0.05, the results of the estimation show little improvement. This is because the noise has low frequency components that cannot be filtered without destroying the useful part of the input and output signals.

Input	2	1.5	1	0.5	0	−0.5	−1	−1.5	−2
Output	10	7.5	5	2.5	0	−2.5	−5	−7.5	−10
Test D	9.96	7.59	4.63	3.22	−0.41	−1.96	−5.26	−7.63	−8.24
Test E	9.66	7.60	4.56	3.33	−0.28	−2.28	−4.79	−7.85	−8.46
Test F	9.76	7.62	4.80	2.98	−0.03	−2.47	−4.74	−7.80	−8.84
Test G	9.81	7.66	4.90	2.84	−0.04	−2.42	−4.78	−7.72	−9.09

Table 6.2 ANM results for stochastic process using various low pass filters

These simulations show that it is possible to estimate the static input-output relationship of a process under feedback control. They show that it is helpful to turn the estimation off during reference changes and for stochastic processes, the use of a low pass filter is advised.

6.4 Estimating Nonlinearities

As it has been shown that it is possible to estimate the static input output relationship of a process under feedback control, the ability of the ANM to estimate nonlinear relationships will be investigated. By completing these simulations, it will be possible to ensure the ANM's ability to correctly estimate nonlinearities and further determine the signal conditioning required by the ANM for improved results.

6.4.1 Saturation

The first nonlinear relationship examined is a saturation curve, which is often brought about by an undersized actuator used to drive a process. A typical saturation curve appears in Figure 6.41, which shows that once the input goes above three or below minus three, the output levels off, no longer able to follow the input. The slope of the saturation region and the location of the start of the saturation both depend on individual actuators being used. For this simulation, the saturation curve shown will be used with the process in Figure 6.42 which is a first-order plant under Dahlin control. As previous simulations examined the effect of noise on the estimation of the static relationship,
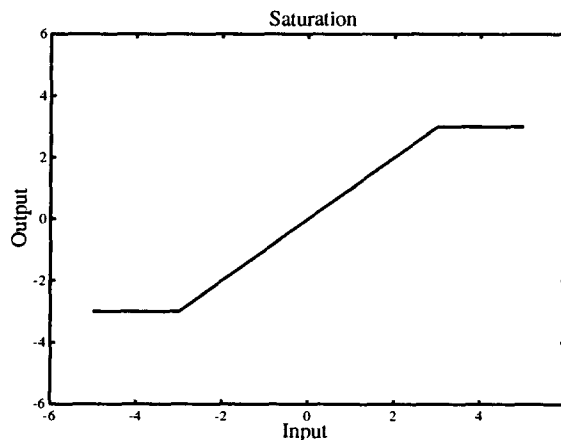
Figure 6.41: Saturation nonlinearity to be estimated

61

reference

$$\frac{1 - 0.59q^{-1}}{0.41 - 0.41q^{-1}}$$

u(t)

$$\frac{0.41q^{-1}}{1 - 0.59q^{-1}}$$

y(t)

+   −
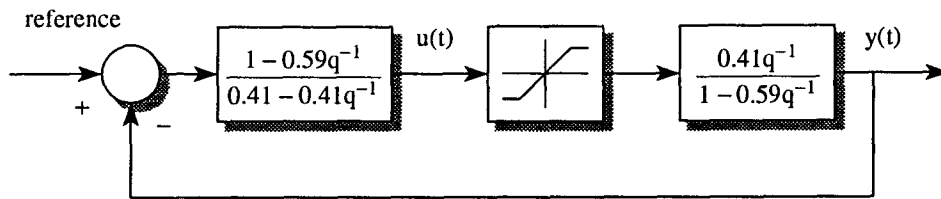
Figure 6.42: Process with saturation nonlinearity

no noise was used here, so only the ability of the ANM to estimate these nonlinear relations under feedback conditions could be judged. In order to get an idea how the ANM is able to estimate the saturation curve, the input shown in Figure 6.43(a) was placed through the saturation with the resulting output shown in Figure 6.43(b). The results of the estimation appear in Table 6.3, Test A. The estimated values vary only slightly from the expected results and confirm that the ANM is able to identify a saturation curve.

Next, the process in Figure 6.42 was simulated, using a reference signal like Figure 6.43(a), and the signals u(t) and y(t) were used to estimate the input-output static relationship of the process. These signals appear in Figure 6.44 and exhibit an interesting characteristic which results from the saturation curve and the integrating action of the controller, which attempts to drive the output to the reference value. The input signal ramps up to very large values, attempting to erase the error between the output and reference. Despite this control signal, which could be improved by introducing an anti-windup mechanism in the controller scheme, the ANM estimates the saturation curve quite well. The results, listed beside Test B in Table 6.3, show very good estimation between three and minus
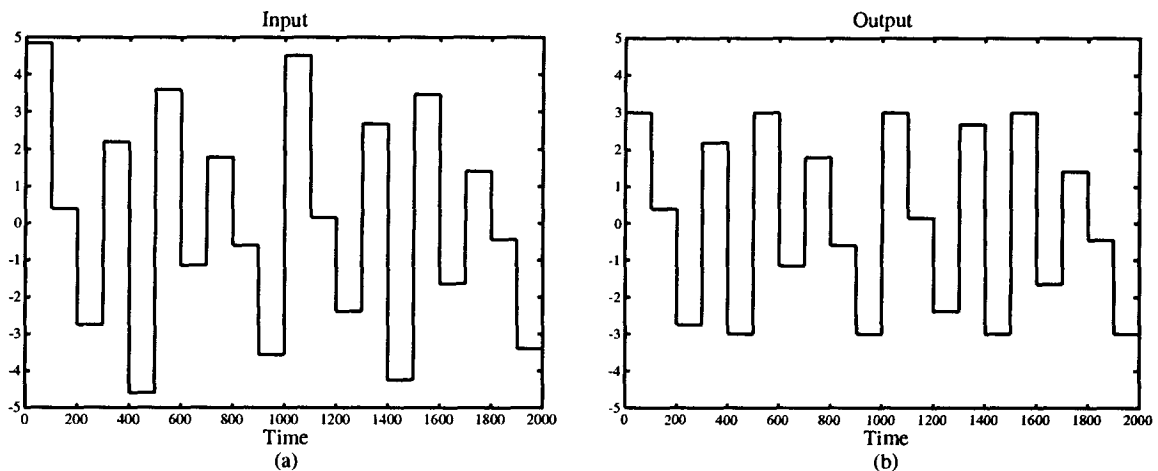
Figure 6.43: The (a) process input and (b) process output from the saturation only

Input	5	4	3	2	1	0	-1	-2	-3	-4	-5
Output	3	3	3	2	1	0	-1	-2	-3	-3	-3
Test A	2.97	2.99	2.97	2.00	0.98	0.00	-1.00	-1.99	-2.97	-3.02	-2.91
Test B	2.48	2.30	2.89	1.92	1.21	0.05	-0.96	-1.92	-2.97	-2.77	-2.42
Test C	3.51	2.14	2.90	2.00	0.99	0.01	-0.99	-2.00	-2.97	-2.77	-2.91
Test D	3.24	2.84	2.96	2.00	0.99	-0.00	-0.99	-2.00	-2.97	-2.97	-2.96

Table 6.3 ANM results for estimation of saturation

three. While the estimation is not exact, this is to be expected as the control signal spends only short periods in the range three to five and minus three to minus five.

The controller was changed to a simple gain in an effort to see how the estimation performed without the windup problem. The gain was chosen so that the process was stable and responded fairly quickly, and as a result was chosen to be three. The input and output signals used for estimation are shown in Figure 6.45. The table values estimated by the ANM are listed as Test C in Table 6.3 and show a very close estimation of the saturation curve.

The final test consisted of simply turning estimation off for ten samples after a reference change. This causes the dynamics in the signals resulting from the plant to be removed. The results of the estimation appears beside Test D in Table 6.3 and shows the improved estimation as a result of temporarily turning off the estimation.
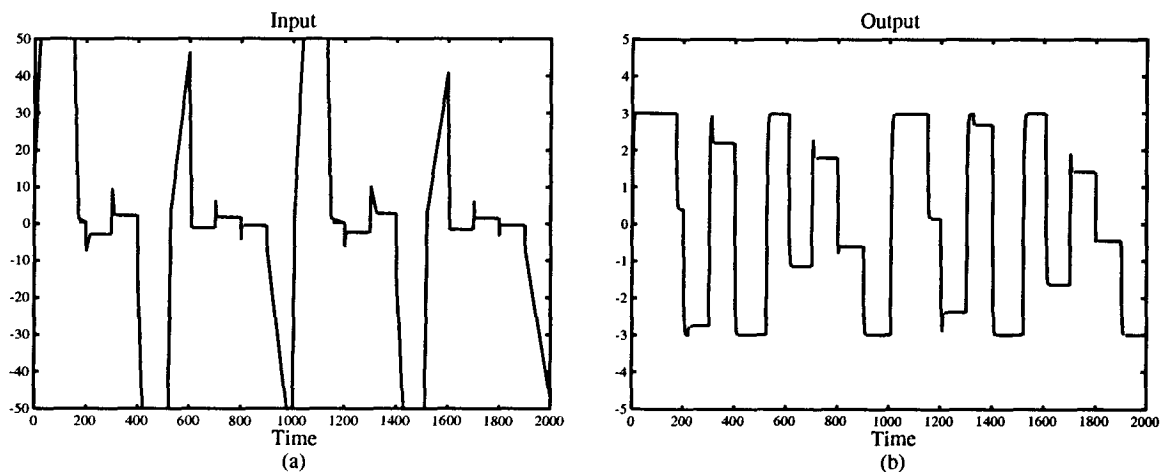
Figure 6.44: The (a) process input and (b) process output from closed loop operation with Dahlin controller
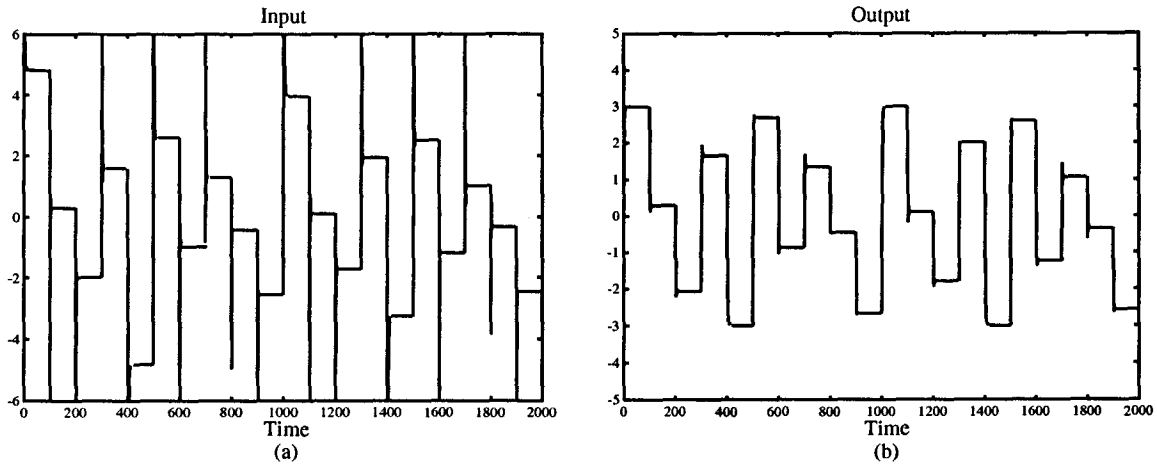
Figure 6.45: The (a) process input and (b) process output from closed loop operation with gain controller

6.4.2 Deadzone

The next nonlinear relationship to be examined is known as a deadzone. A deadzone is characterized by a range of input where no change in output occurs. Figure 6.46 displays a typical deadzone, which can be brought about by a number of reasons and often represents a malfunction in the operation of an actuator. For this particular deadzone, the output does not change over the interval [0.5,-0.5] and returns to a linear slope outside of this interval. The size of the deadzone and the slope of the curve outside the deadzone will depend on the individual situation. For these simulations, the process in Figure 6.42 was used once again, however the saturation curve was replaced by the deadzone in Figure 6.46.
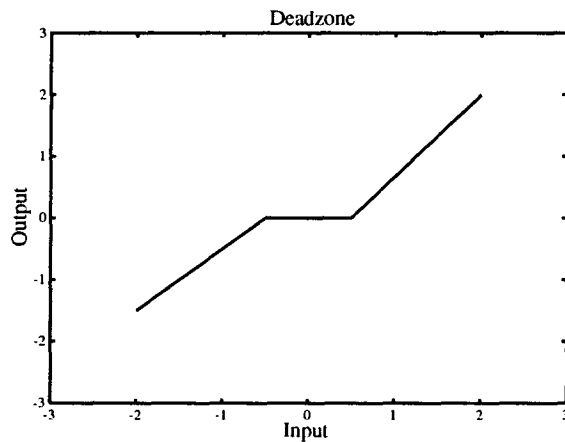
Figure 6.46: Deadzone nonlinearity to be estimated

64
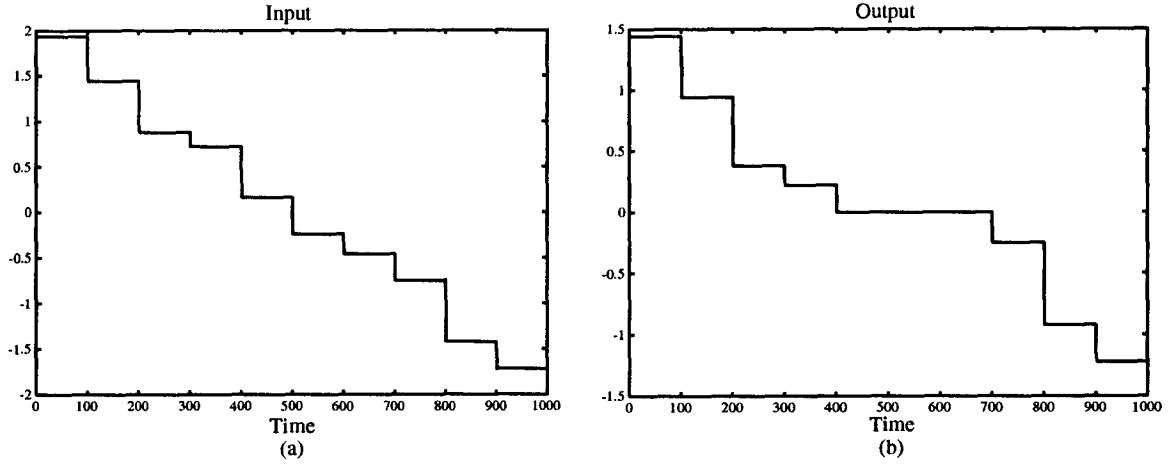
Figure 6.47: The (a) process input and (b) process output from the deadzone only

The simulations performed were similar to those done for the saturation nonlinearity and start by passing the signal in Figure 6.47(a) through the deadzone which resulted in the signal in Figure 6.47(b). The ANM estimator used the signals in Figure 6.47 and attempted to estimate the deadzone nonlinearity. The results of the estimation, which appear in Table 6.4, Test A, confirm that the ANM is able to estimate a deadzone nonlinearity.

Next, the process in Figure 6.42 was simulated using the deadzone and the signal in Figure 6.47(a) was used as the reference signal. The resulting input and output signals, u(t) and y(t) appear in Figure 6.48 and were used by the ANM to estimate the table values of the static input output relation. The results, Test B in Table 6.4, show that the ANM was able to estimate the deadzone. Except for the one value at zero, the ANM performed very well.

Finally, to see if the previous results could be improved upon, as in previous simulations, the ANM estimation was turned off for ten samples after a reference change. The signals that were used

Input	2	1.5	1	0.5	0	−0.5	−1	−1.5	−2
Output	1.5	1	0.5	0	0	0	−0.5	−1	−1.5
Test A	1.48	0.99	0.49	0.01	−0.00	−0.01	−0.47	−1.02	−1.41
Test B	1.48	0.98	0.52	0.02	0.35	0.04	−0.52	−0.94	−1.49
Test C	1.48	0.99	0.49	0.02	0.00	0.03	−0.50	−0.97	−1.48

Table 6.4 ANM results for estimation of deadzone

65
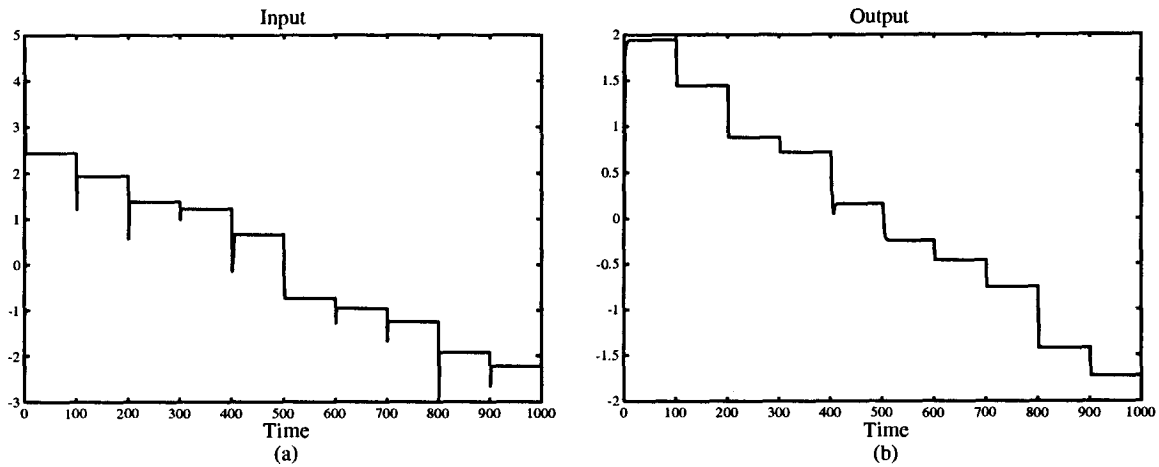
Figure 6.48: The (a) process input and (b) process output from closed loop operation with Dahlin controller

then had most of the dynamics removed. The result of the estimation, Test C in Table 6.4, show an improvement in the table values estimated.

These two simulations show the ANM ability to model typical nonlinearities. The improved results obtained by turning estimation off after reference changes suggests that this be done in all situations. There are several different nonlinearities which could appear in an industrial process. One is hysteresis, which will require some modification to the ANM so that this nonlinearity can be modelled. It will require two separate gain curves, one which is built as the input increases over time and another which is built as the input decreases over time. This will require that the direction of the input be tracked, however these small changes will enable hysteresis to be modelled.

6.5 Effect of Process Changes on Minimum Achievable Output Variance

The estimated minimum achievable output variance will change as process properties change. The goal of the following simulations is twofold. First, they will show that the operation of the minimum achievable output variance estimator is consistent with the aspects behind its calculation. Secondly, the results will be used to show how changes in estimated minimum achievable output variance can be related to changes in the process. This change in estimation can be used to determine process changes, as well as evaluate the performance of the present controller

The figure shows a block diagram with:
- $e(t)$, $N(0,0.36)$ feeding into block $\dfrac{1-0.4q^{-1}}{1-0.7q^{-1}}$
- A summing junction with "zero reference", "+", "$-$" feeding into block $\dfrac{0.7-0.47q^{-1}}{0.33-0.10q^{-1}-0.23q^{-4}}$
- Then $u(t)$ into block $\dfrac{0.33q^{-4}}{1-0.67q^{-1}}$
- Output summed to give $y(t)$

$$\sigma_e^2 = 0.36$$
$$\sigma_y^2 = 0.6114$$
$$\sigma_{mv}^2 = 0.4034$$

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{1-0.7q^{-1}+0.12q^{-2}-0.7q^{-4}+0.28q^{-5}}{1-0.97q^{-1}+0.20q^{-2}}$$

$$F(q^{-1}) = 1+0.27q^{-1}+0.18q^{-2}+0.12q^{-3}$$
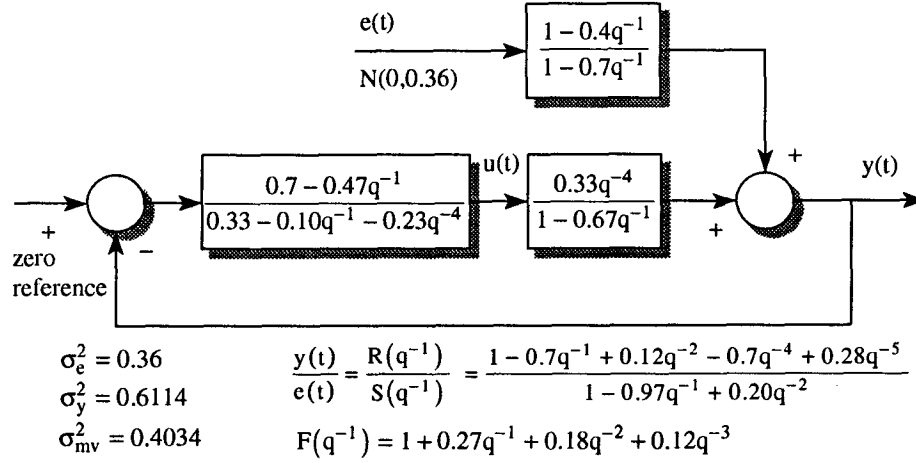
Figure 6.49: Original process and associated properties

The process used for these simulations appears in Figure 6.49, which has a first-order plant with a time delay of four. The disturbance filter is first-order, ARMA(1,1) and the noise sequence has a N(0,0.36) distribution. The process uses a Dahlin controller that compensates for the delay and moves the closed loop pole of the process to 0.3. The minimum achievable output variance and the process output variance for this process are listed in Figure 6.49, along with the closed loop noise transfer function of the process, $\frac{R(q^{-1})}{S(q^{-1})}$. This transfer function is estimated so that the minimum achievable output variance can be determined. For this process, the transfer function has a second-order denominator, while the numerator polynomial is more complicated, having order five. Lastly, the $F(q^{-1})$ polynomial whose coefficients must be found to determine the minimum achievable output variance, is given.

The results of one simulation appear in Figure 6.50. The autocorrelation function of the estimated noise sequence is white because correlation at all lags is near zero. The estimated noise has a variance of 0.3598, which is nearly identical to the process noise variance. The FMVRE delay estimator successfully estimated the correct delay, while the Markov parameters, that are determined from estimated Laguerre gains, are very close to the process values. Based on this, the estimated minimum achievable output variance is very close to the calculated value for this process. This single simulation result is confirmed by repeating it five hundred times using a different N(0,0.36) noise sequence. The histogram in Figure 6.51 shows that the mean minimum achievable output variance
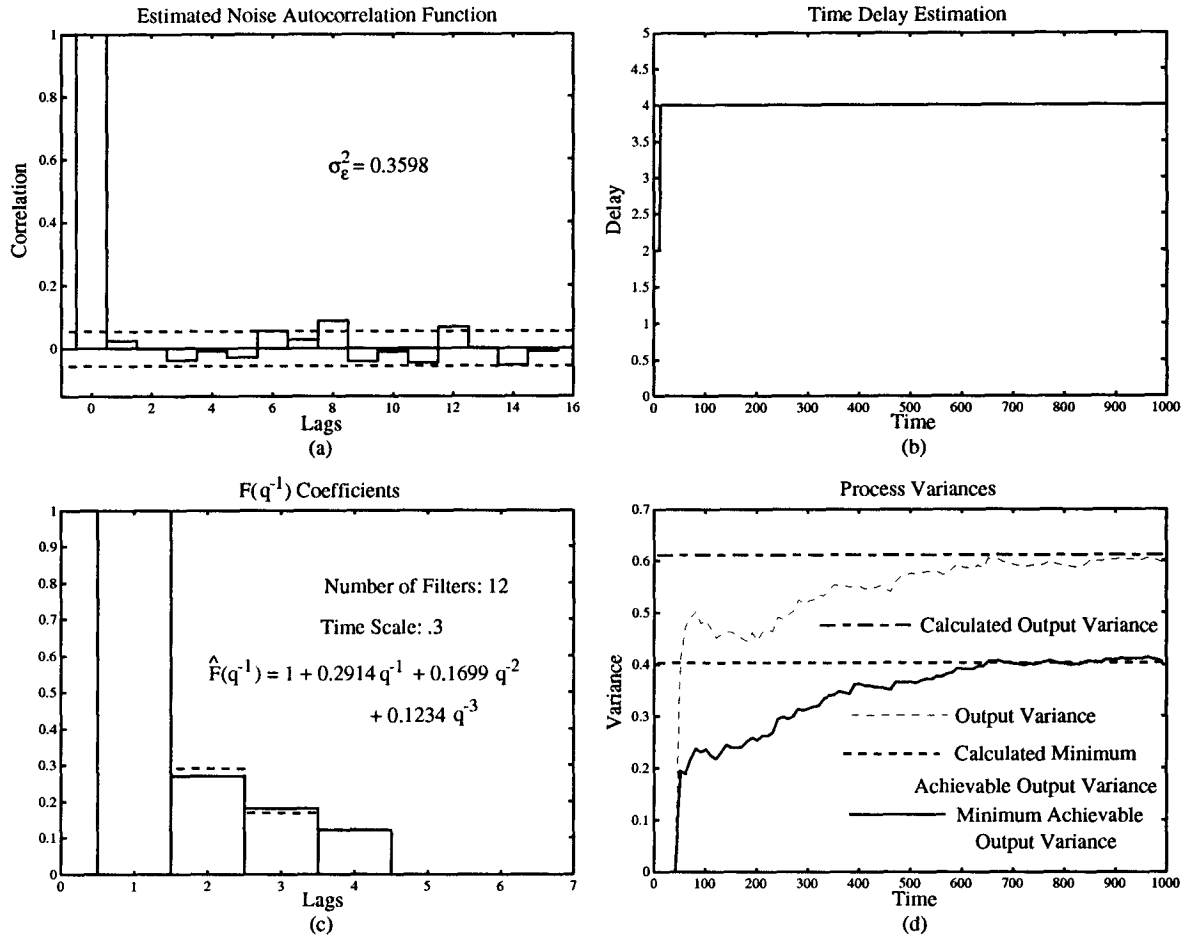
Figure 6.50: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for the original process
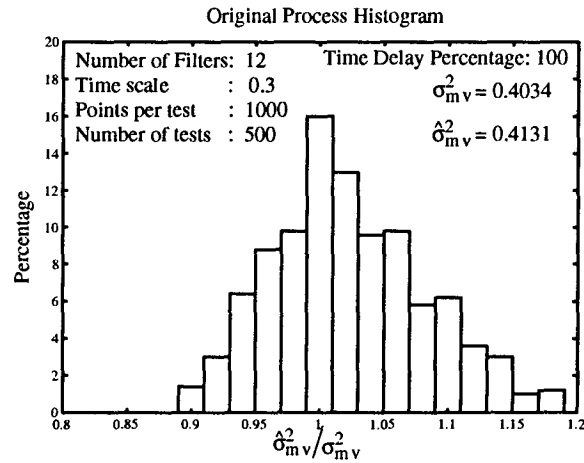
Figure 6.51: Distribution of the results of five hundred simulations for the original process

is just slightly above the calculated value, while the individual tests are closely distributed around the calculated value.

The first property changed was the delay, and it was reduced from four to two. This reduction of delay changes the $F(q^{-1})$ polynomial which will have two, rather than four coefficients:

$$F(q^{-1}) = 1 + 0.27q^{-1} \tag{6.110}$$

and then the minimum achievable output variance becomes:

$$\sigma_{mv}^2 = 0.36(1 + 0.27^2) = 0.3863 \tag{6.111}$$

The transfer function from the noise to the output becomes:

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{1 - 0.7q^{-1} - 0.58q^{-2} + 0.28q^{-3}}{1 - 0.97q^{-1} + 0.20q^{-2}} \tag{6.112}$$
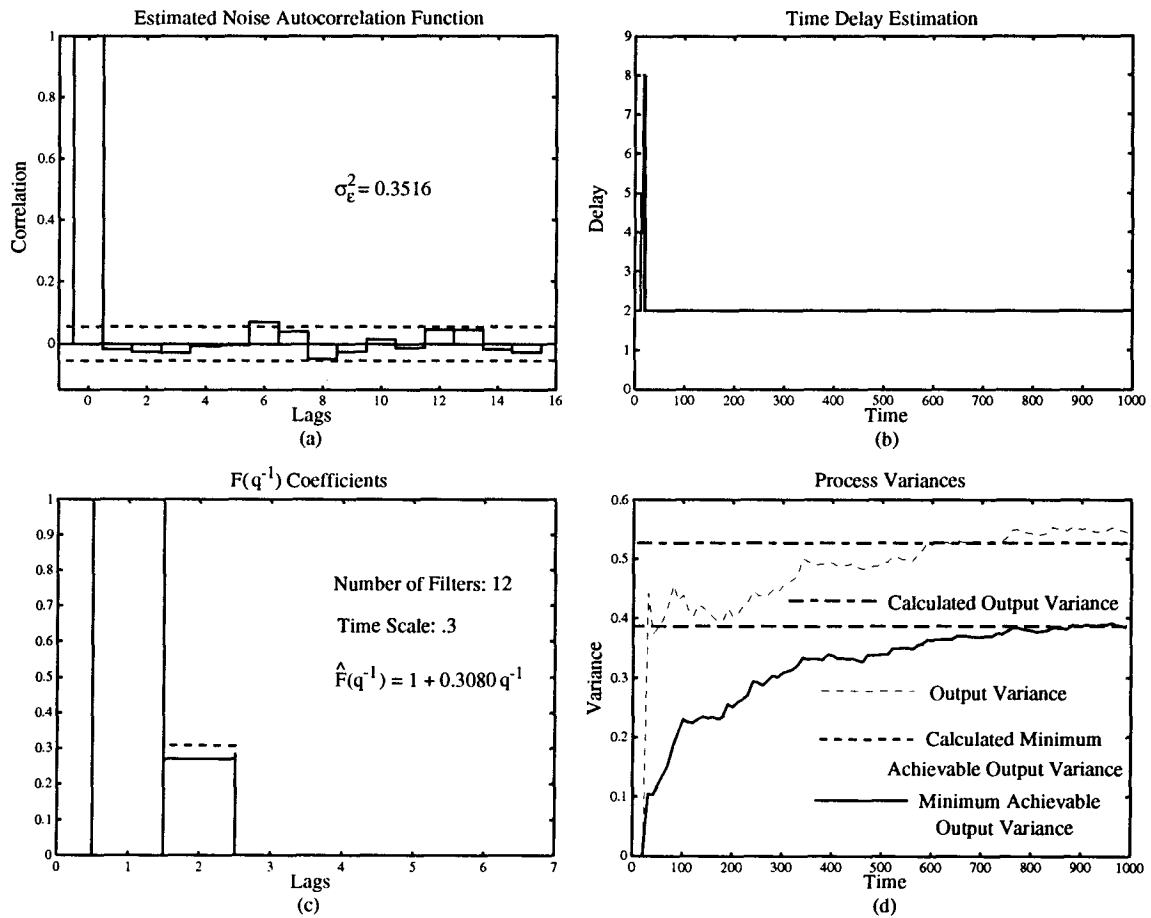
Figure 6.52: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}_{mv}^2$ and σ_y^2 for a change in delay

which changes the process output variance to:

$$\sigma_y^2 = 0.5277 \tag{6.113}$$

The results of a single simulation reflects these changes, and Figure 6.52(b) shows the FMVRE delay estimator correctly estimated the change and thus there was only two coefficients in $F(q^{-1})$. The estimation of the noise sequence variance remained closed to 0.36. Five hundred simulations were run to examine the consistency of the estimation. The mean estimated minimum achievable output variance was found to be 0.3852 and its distribution was similar to Figure 6.51, which suggests Figure 6.52 is a typical result. This simulation shows that once the FMVRE estimates a change in the process delay, there should be a change in the output variance and the minimum achievable output variance becomes smaller because $F(q^{-1})$ has fewer coefficients.

Returning the delay to four, the next change was made to the plant transfer function:

$$\frac{B(q^{-1})}{A_1(q^{-1})} = \frac{0.25q^{-4}}{1 - 0.75q^{-1}} \tag{6.114}$$

which results in a new closed loop noise transfer function of:

$$\frac{R(q^{-1})}{S(q^{-1})} = \frac{1 - 1.45q^{-1} + 0.65q^{-2} - 0.09q^{-3} - 0.70q^{-4} + 0.80q^{-5} - 0.21q^{-6}}{1 - 1.72q^{-1} + 0.93q^{-2} - 0.15q^{-3} - 0.17q^{-4} + 0.28q^{-5} - 0.11q^{-6}} \tag{6.115}$$

This change makes the process output variance:

$$\sigma_y^2 = 0.5591 \tag{6.116}$$

The minimum achievable output variance is unchanged, however, as the $F(q^{-1})$ polynomial remains the same. The results of a simulation, Figure 6.53 confirms this, as the estimated minimum achievable output variance remains near the calculated value of 0.4034, despite the change in the output variance. Figure 6.53(a&c) shows that the estimate of the $F(q^{-1})$ coefficients and noise sequence remains unchanged for the first process, while the histogram in Figure 6.54 confirms that the estimation remains virtually unchanged by the plant variation. This simulation shows that variation in only the output variance suggests that the plant transfer function has changed.
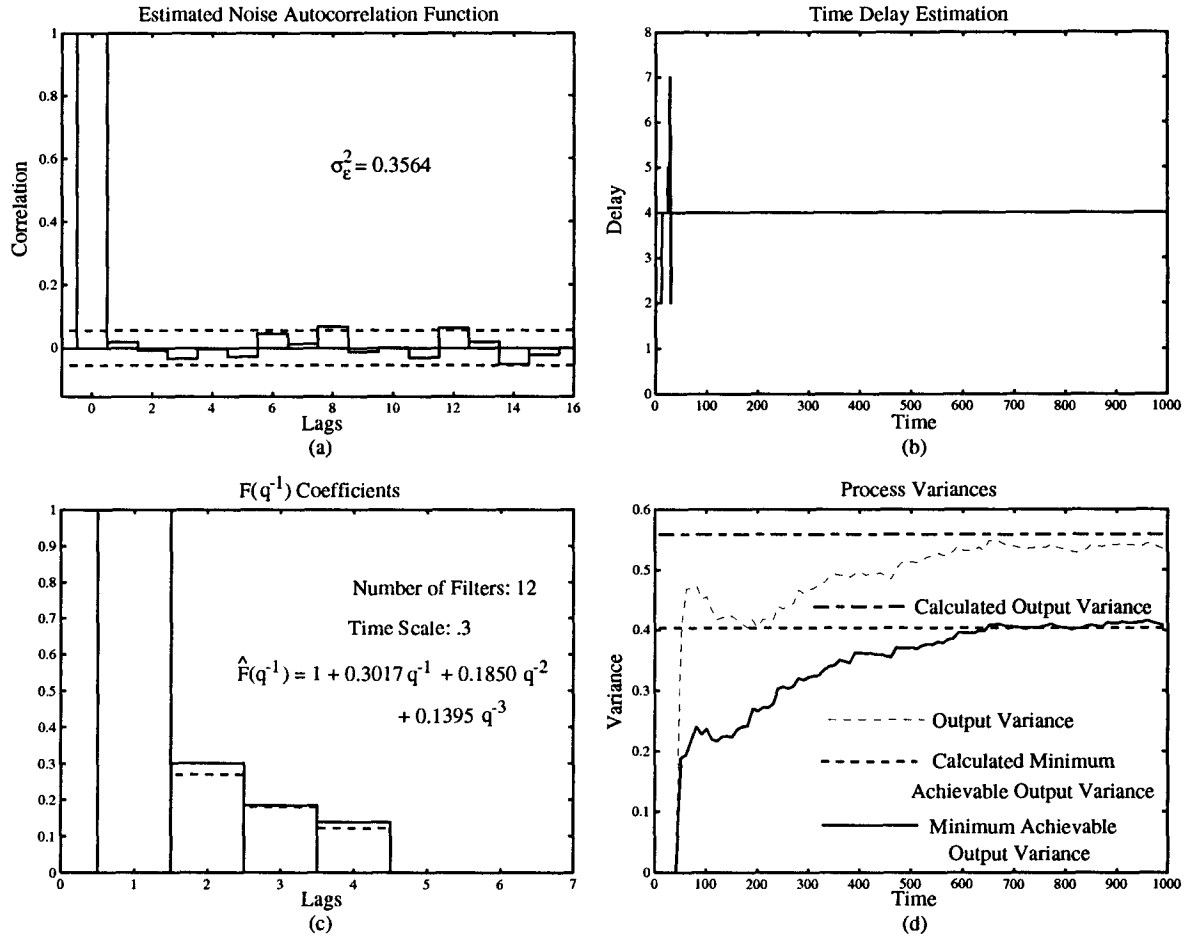
Figure 6.53: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for a change in plant
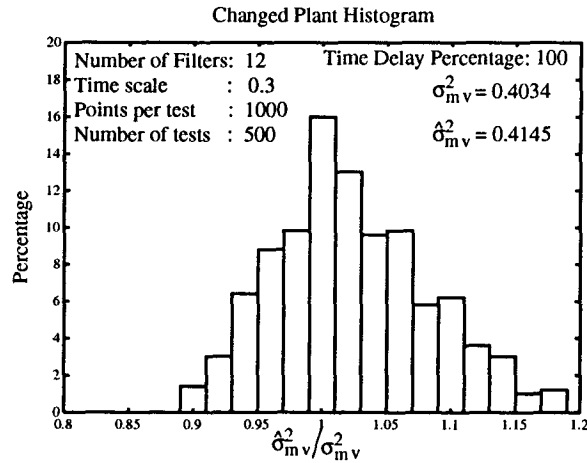
Figure 6.54: Distribution of the results of five hundred simulations for the change in the process plant

Once the plant was reverted back to its original value, the disturbance filter was changed:

$$\frac{C(q^{-1})}{A_2(q^{-1})} = \frac{1 - 0.2q^{-1}}{1 - 0.67q^{-1}} \tag{6.117}$$

which changes the $F(q^{-1})$ polynomial coefficients to:

$$F(q^{-1}) = 1 + 0.47q^{-1} + 0.32q^{-2} + 0.27q^{-3} \tag{6.118}$$

and then the minimum achievable output variance becomes:

$$\sigma_{\text{mv}}^2 = 0.36(1 + 0.47^2 + 0.32^2 + 0.27^2) = 0.4915 \tag{6.119}$$

Once again, the closed loop noise transfer function changes:

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{1 - 0.5q^{-1} + 0.06q^{-2} - 0.70q^{-4} + 0.14q^{-5}}{1 - 0.97q^{-1} + 0.20q^{-2}} \tag{6.120}$$

and then the process output variance becomes:

$$\sigma_y^2 = 0.7434 \tag{6.121}$$

From Figure 6.55, the corresponding changes in the $F(q^{-1})$ coefficients, the minimum achievable output variance and output variance are evident, while the estimated variance of the noise sequence remains the same. Figure 6.56 confirms that the mean value of the minimum achievable output variance moved towards 0.4915, the new level. Thus a change in the disturbance filter becomes noticeable, as both σ_y^2 and σ_{mv}^2 change.

To see the effect the noise sequence had on the estimation of the minimum achievable output variance, it was changed from $N(0,0.36)$ to $N(0,1.21)$ in the original process. This change in the noise variance results in new process output variance and minimum achievable output variance:

$$\sigma_{\text{mv}}^2 = 1.356$$
$$\sigma_y^2 = 2.055 \tag{6.122}$$

One simulation was performed and it showed that the output variance and minimum achievable output variance level did change to the new values. This was a direct result of the change in noise variance, as the $F(q^{-1})$ coefficients and delay estimate remained unchanged. Five hundred simulations
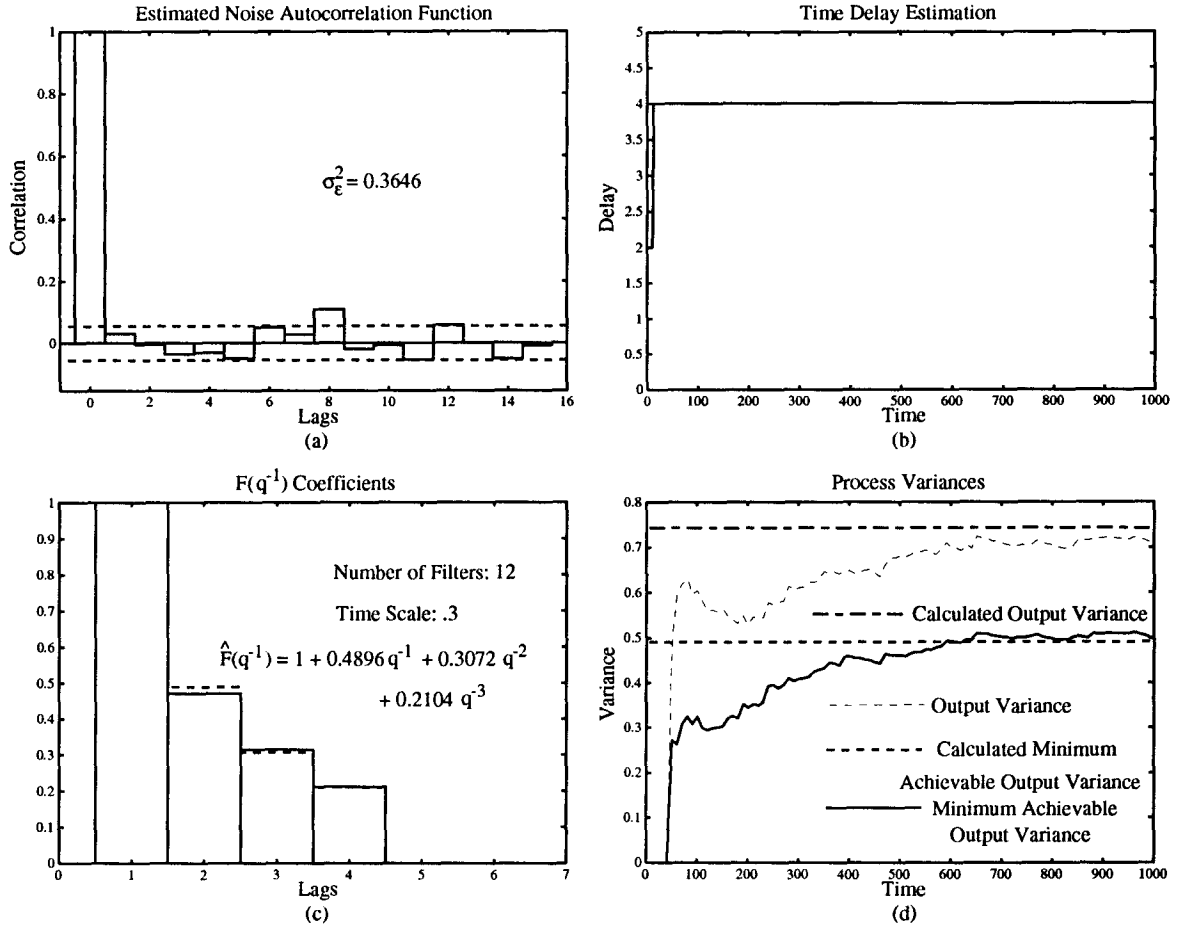
Figure 6.55: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for a change in disturbance filter
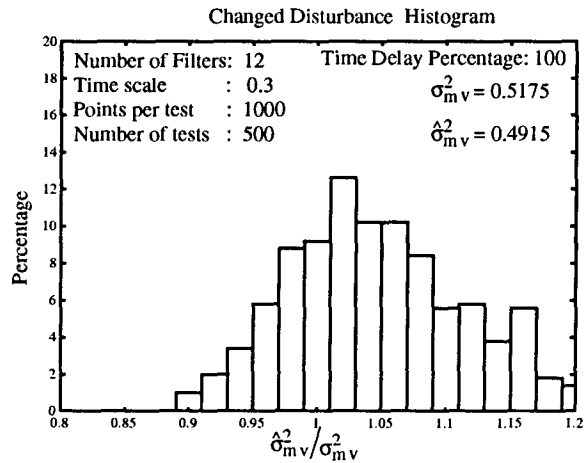
Figure 6.56: Distribution of the results of five hundred simulations for a change in disturbance filter

were performed and they confirmed this finding. As a result, if both σ_y^2 and σ_{mv}^2 change while $F(q^{-1})$ remains the same, then the noise variance must have changed.

The final part of the process to be changed is the controller. During the derivation of the method to find the minimum achievable output variance, it was determined that the minimum level could not be reduced any further by a feedback controller. The controller was changed to:

$$\frac{N(q^{-1})}{D(q^{-1})} = \frac{0.7 - 0.5040q^{-1}}{0.28 - 0.084q^{-1} - 0.196q^{-4}} \tag{6.123}$$

which is a Dahlin controller which moves the process pole to 0.3 for the plant:

$$\frac{B(q^{-1})}{A_1(q^{-1})} = \frac{0.28q^{-4}}{1 - 0.72q^{-1}} \tag{6.124}$$

This change in controller causes the closed loop noise transfer function to become:

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{1 - 0.7q^{-1} + 0.12q^{-2} - 0.70q^{-4} + 0.28q^{-5}}{1 - 0.97q^{-1} + 0.20q^{-2} + 0.12q^{-4} - 0.12q^{-5}} \tag{6.125}$$

and this changes the process output variance to:

$$\sigma_y^2 = 0.6744 \tag{6.126}$$

but it does not change the minimum achievable output variance, because the delay, noise and $F(q^{-1})$ coefficients remain unchanged. The results of one simulation, shown in Figure 6.57 proves that the controller only modifies the output variance, if compared to Figure 6.50. It is promising to note that the histogram in Figure 6.58 has a mean estimated minimum achievable output variance level close to the value found in Figure 6.51, while the shape of the histogram remains basically unchanged.

Various properties of the process have been changed, in order to study the effect on estimation of the minimum achievable output variance. This knowledge can be used to determine what has changed in the process when either σ_y^2 and σ_{mv}^2 change. For example, a change in delay will be estimated by the FMVRE, while a change in noise variance will be estimated by the RELS. Other changes in the process output variance become evident in a more subtle manner. If only the process output variance has changed, this indicates that the plant characteristics have been modified. Meanwhile, if just the value of the $F(q^{-1})$ coefficients change, then the disturbance filter has changed values. Using these direct indicators of change to signal operational changes in the process, it is possible to focus attention on a specific aspect of the process. Finally, it should be remembered that changes to the controller should only change the output variance and not affect the minimum achievable output variance.
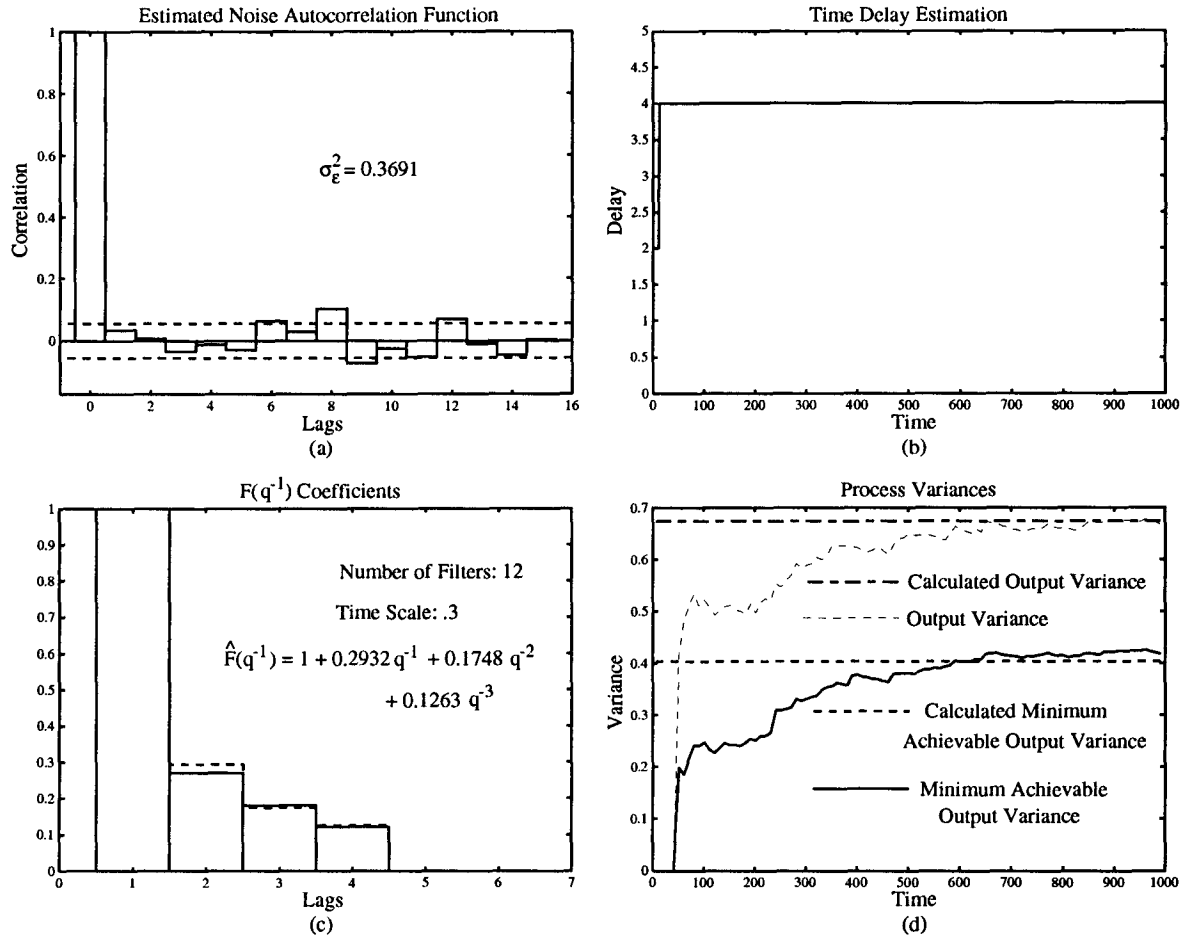
Figure 6.57: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for a change in controller
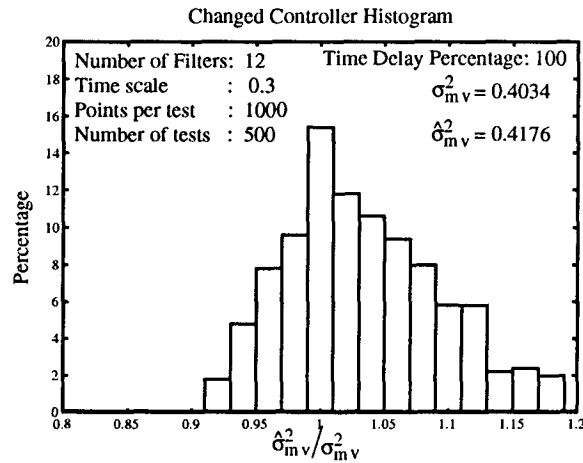
Figure 6.58: Distribution of the results of five hundred simulations for a change in the controller

6.6 Choice of Laguerre Time Scale and Number of Filters

There are two parameters of a Laguerre network that must be chosen before it is used to model a process. They are the number of filters used in the network and the time scale of these filters. The following simulations investigate the effect these parameters have on the estimation of the minimum achievable output variance. A brief discussion on what values for these parameters are usually made, is followed by the simulations. The results are discussed and following this suggestions for general choices are made.

As mentioned, the set of Laguerre functions can completely represent the impulse response of a linear process only if an infinite number of filters are used. As it is impossible to use an infinite number, the number of filters used in the Laguerre network must be chosen so that a good approximation of the process is obtained. It has been found that four to eight Laguerre filters used in the network provides a reasonable approximation [15][25] of a sampled-data plant. This work showed that good approximation also depends on the choice of the Laguerre filter time scale. It is known that the optimum choice for the time scale will depend on the process to be modelled. If a process exhibits a heavily damped behavior, which suggests that it has real poles, the best choice for the Laguerre filter time scale is a value which corresponds to the dominant process pole. However, if the Laguerre network is modelling a process which has complex poles, evident as the process will exhibit a lightly damped behavior, a different approach to choosing the time scale is required. Fu showed in [15] that the Laguerre network approximation of a sampled-data system improved as the time scale is moved toward zero.

However, the previous conclusions were drawn from situations where the input to be used for estimation was known. In this application, the transfer function to be modelled is driven by an unknown white noise sequence, that also must be estimated along with the unknown parameters. The process determines the closed loop noise transfer function, which is given by:

$$y(t) = \frac{R(q^{-1})}{S(q^{-1})} e(t) \qquad (6.127)$$

where the numerator polynomial is often unlike any polynomial usually found in a sampled-data plant. As a result of these differences, it has been found that the best choice for the number of filters

76

and time scale for the Laguerre network, when used with RELS estimation, depends on properties of $S(q^{-1})$ and $R(q^{-1})$. The choice of filter number and time scale depends on whether the closed loop noise transfer function has real or complex closed loop poles, that is the roots of $S(q^{-1})$, or if the numerator polynomial, $R(q^{-1})$, is strictly positive real (SPR). Combinations of these situations will be simulated to see if some general choices for the time scale and filter number can be made to obtain good estimation.

6.6.1 Real Closed Loop Noise Transfer Function Denominator Roots

The first two simulations examine the case where the closed loop poles are real and the $R(q^{-1})$ polynomial is strictly positive real. These examples are the least complicated and most likely the least encountered situations, however, they are important as they show the behavior of estimation in a uncomplicated situation.

The first process examined appears in Figure 6.59 and it uses a gain as the controller, so that $R(q^{-1})=C(q^{-1})$. The properties of the process are also shown in Figure 6.59, indicating the transfer function to be identified, which has real roots, and the calculated minimum achievable output variance value which must be estimated. Because the $C(q^{-1})$ polynomial is first-order, it is strictly positive real. The results of one simulation, Figure 6.60, shows that the autocorrelation function of the estimated noise sequence is white and has a variance close to the expected value, while the delay and $F(q^{-1})$ coefficients were correctly estimated. As a result, the estimated minimum achievable output variance

$$\sigma_e^2 = 0.49$$
$$\sigma_y^2 = 0.6435$$
$$\sigma_{mv}^2 = 0.6125$$

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{1 - 0.2q^{-1}}{1 - 0.7q^{-1} + 0.1225q^{-2}}$$

$$F(q^{-1}) = 1 + 0.5q^{-1}$$
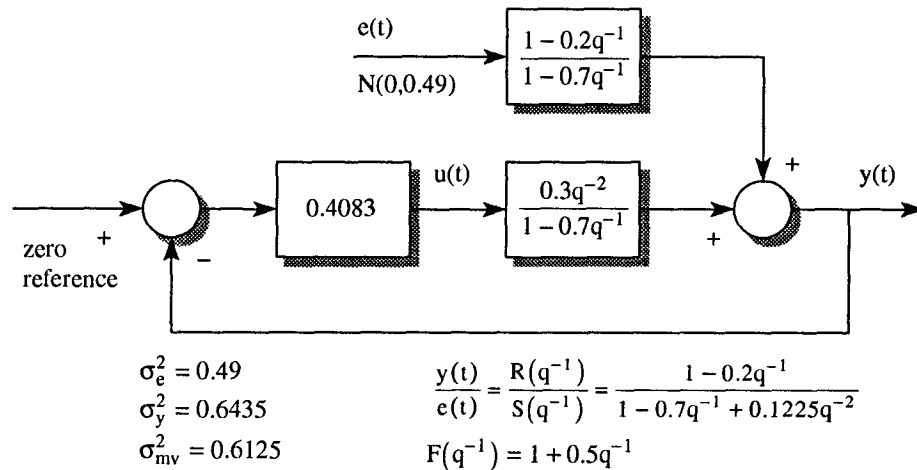
Figure 6.59: Gain controlled process

Figure 6.60: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for process with real poles
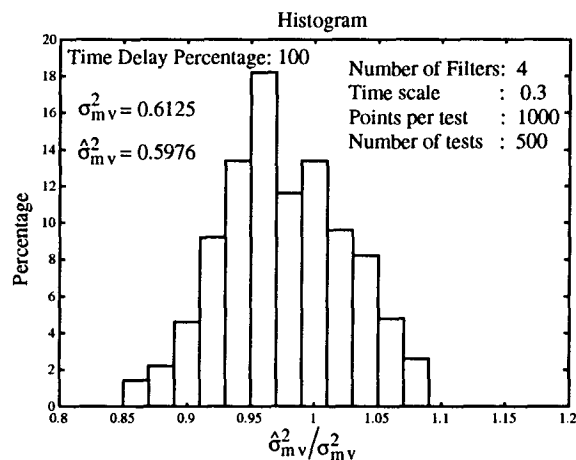
Figure 6.61: Distribution of the results of five hundred simulations for process with real poles
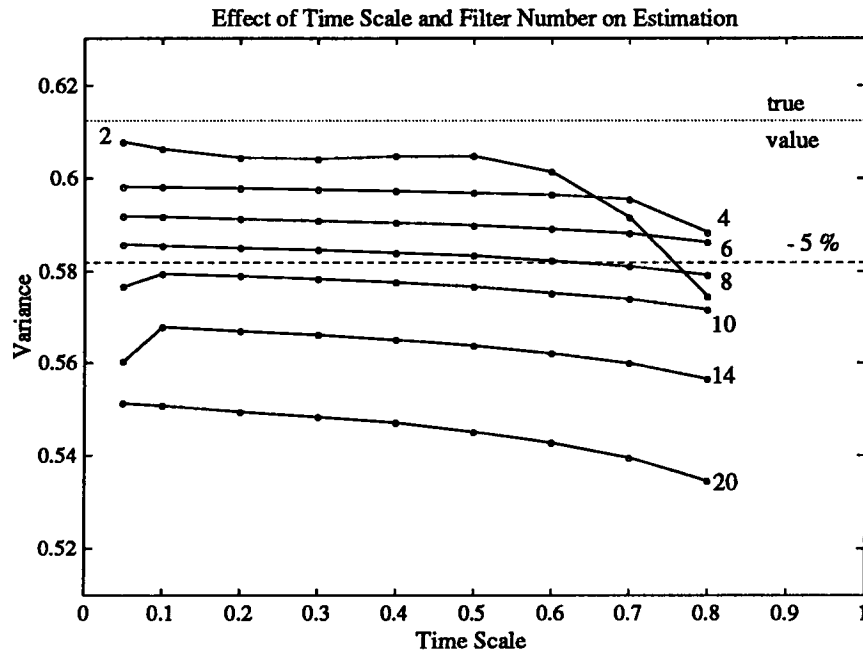
Figure 6.62: Effect of changing the time scale and filter number on $\hat{\sigma}^2_{mv}$ for process with real poles

is very close to the calculated value, Figure 6.60(d). This single test was confirmed by Figure 6.61, which shows that after five hundred simulations using different N(0,0.49) noise sequences, the mean estimated minimum achievable output variance was just under the calculated value.

Now in order to obtain an idea of how the number of filters and time scale used in the Laguerre network affects the estimated minimum achievable output variance, the test shown in Figure 6.61 was performed with different time scales and filter numbers. The mean value obtained after each test is then an indicator of the performance of that choice of time scale and filter number. The results are in the plot of Figure 6.62, which presents variance versus the time scale for a different number of filters, listed next to the corresponding curve. The plot also shows the calculated minimum achievable output variance along with minus five percent of this value. Lines are used only to join results which share the same number of filters used. When only two filters are used, the estimated minimum achievable output variance already overfits the calculated value and this continues as the number of filters is increased. Except for when two filters are used, the choice of time scale does not have much effect on the minimum achievable output variance estimated. These results suggest that the Laguerre network is not behaving as expected, perhaps because it is being used with RELS. Before further consideration of the results, another process will be examined to confirm these findings.

$\sigma_e^2 = 1.69$

$\sigma_y^2 = 2.1091$

$\sigma_{mv}^2 = 1.7576$

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{1 - q^{-1} - 0.19q^{-2} + 0.22q^{-3} - 0.03q^{-4}}{1 - 1.2q^{-1} + 0.35q^{-2}}$$

$$F(q^{-1}) = 1 + 0.2q^{-1}$$
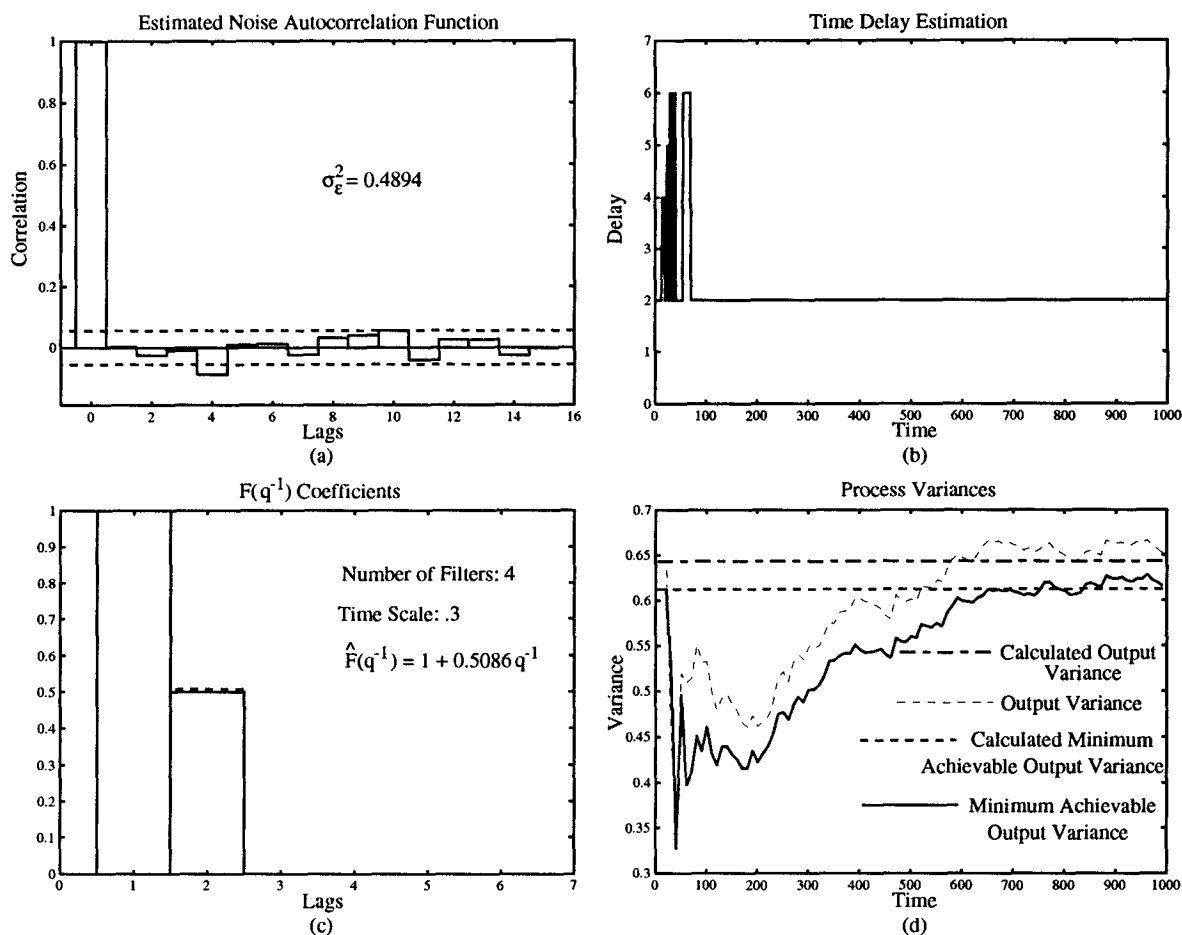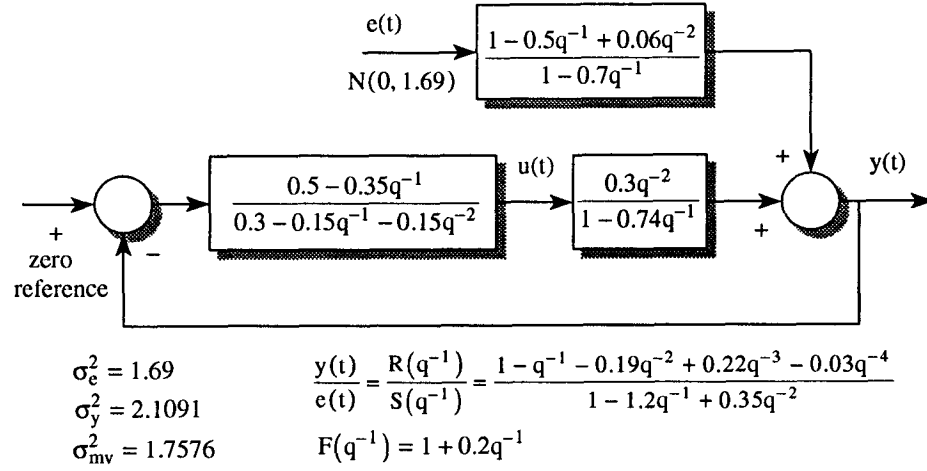
Figure 6.63: Dahlin controlled process with real poles
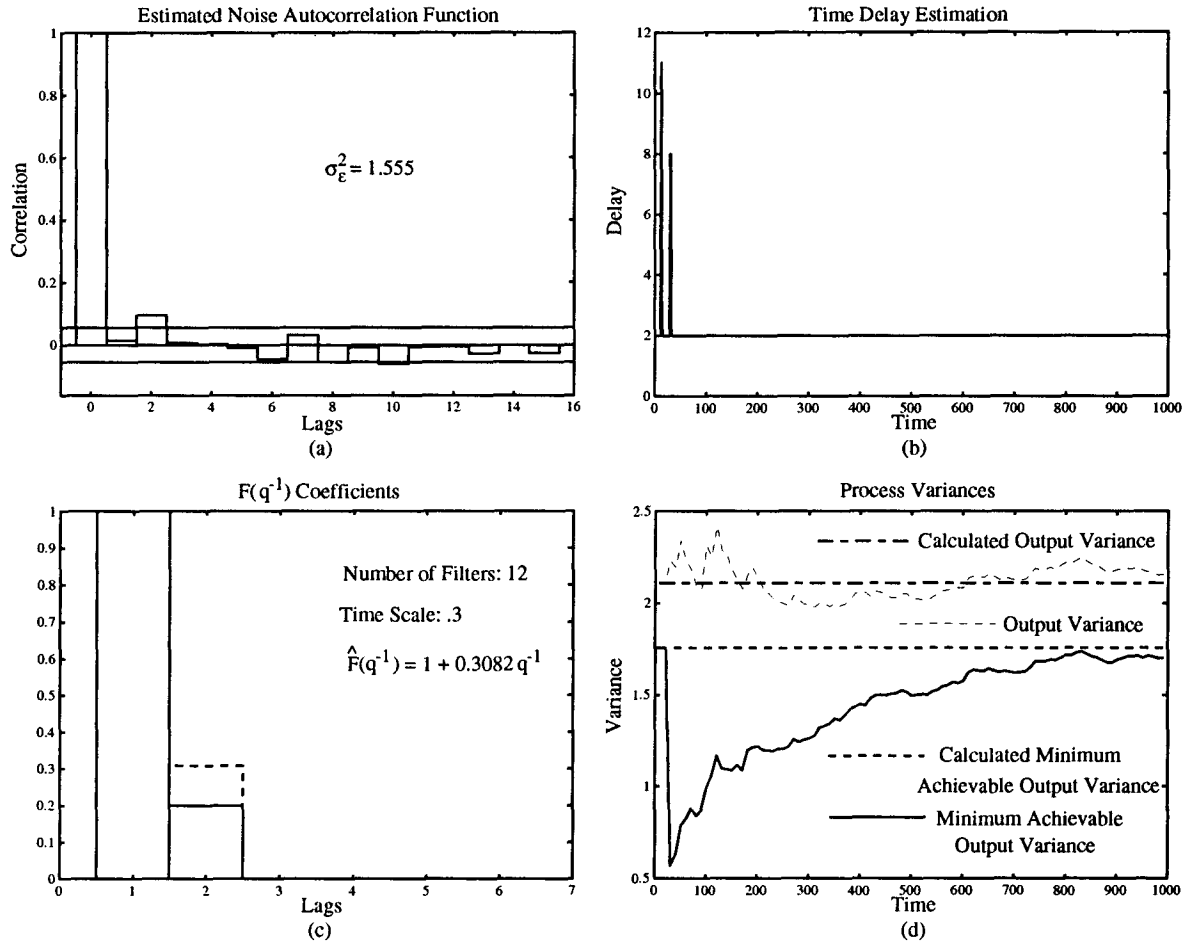
Figure 6.64: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}_{mv}^2$ and σ_y^2 for Dahlin controlled process with real poles

80

The next process examined appears in Figure 6.63 and the major difference is that this process is controlled by a Dahlin controller which makes the closed loop noise transfer function more complicated. Despite this, the numerator polynomial is still strictly positive real, so the results of this simulation can be compared to the previous results. The results of one simulation appears in Figure 6.64, which uses twelve filters and a time scale of 0.3 and the minimum achievable output variance is quite closely estimated. Examination of Figure 6.65 shows a similar trend as found in Figure 6.62, in that the minimum achievable output variance was overfitted for almost all choices of filter number. It is believed that due to the increased complication of the transfer function to be identified, the number of filters required before the overfitting begins is larger.

The results of single simulations are examined in order to determine the reason why the minimum achievable output variance is overfitted. The mean value of 1.69 found from Figure 6.65 can be explained by closely examining Figure 6.64(a&c). While the estimated noise autocorrelation function is white, the variance is less than expected, also, the $F(q^{-1})$ coefficient is larger than expected. The two effects combine and result in $\hat{\sigma}^2_{mv}$ being less than the calculated value. Similar results were found from the gain controlled process. It appears that the Laguerre network is including the noise

Figure 6.65: Effect of changing the time scale and filter number on $\hat{\sigma}^2_{mv}$ for Dahlin controlled process with real poles

sequence into its estimation of the closed loop noise transfer function, resulting in an incorrect noise variance and $F(q^{-1})$ coefficients. However, close estimation is still possible for both cases for a wide range of time scales and filter number.

6.6.2 Complex Closed Loop Noise Transfer Function Denominator Roots

In the processes considered next, the polynomial $R(q^{-1})$ has complex poles. The time scale of the Laguerre network used is real and it has been found that the number of filters used by the network must be increased [15] in order to arrive at satisfactory approximation of the process. By simulating lightly damped processes, it can be determined if the number of filters need be increased to obtain good estimation of the minimum achievable output variance.
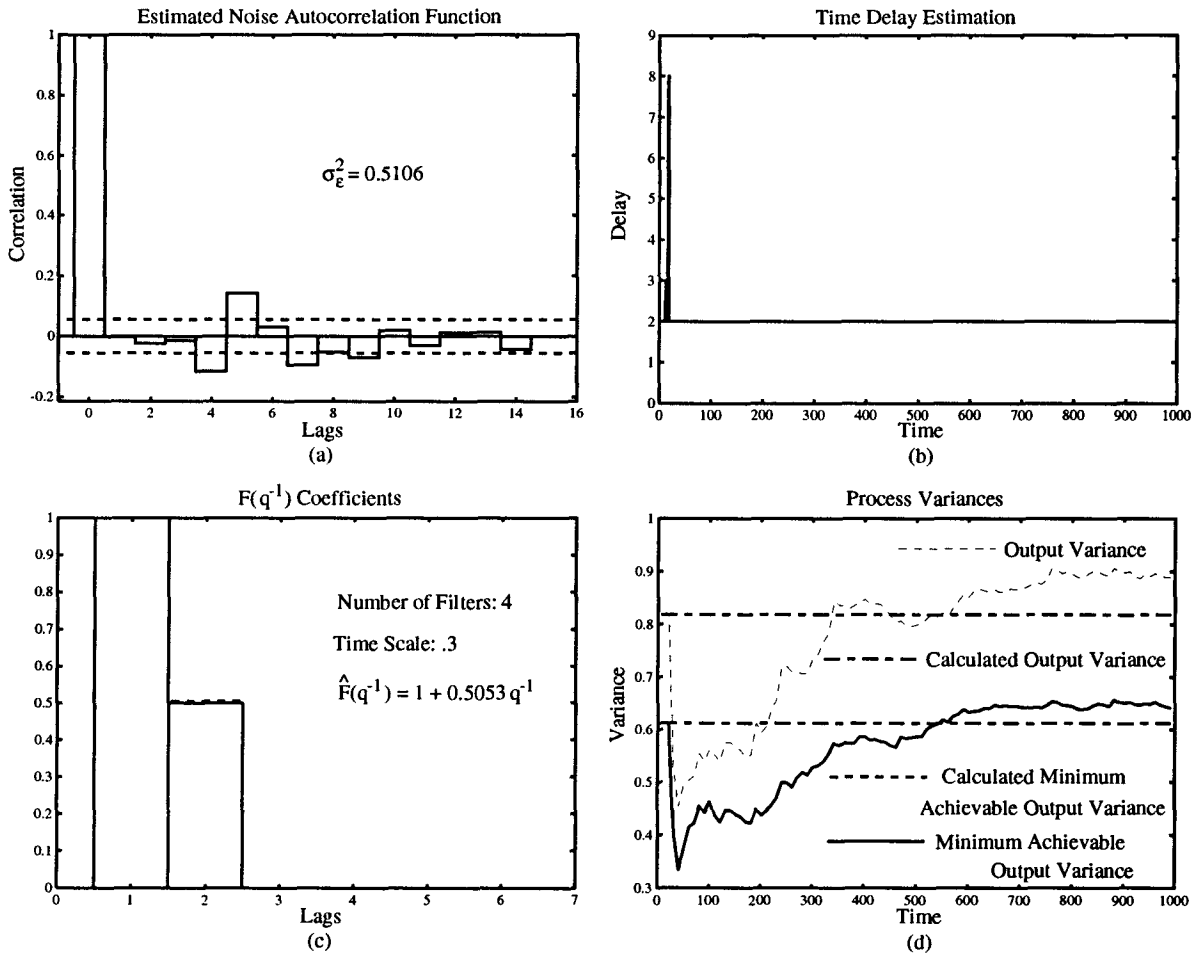

Figure 6.66: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for process with complex poles

The first process used, Figure 6.59, controlled by a pure gain, is used once again, but this time the gain is two so that the roots of $R(q^{-1})$ are complex. The effects of this change make:

$$\sigma_y^2 = 0.819 \qquad \frac{R\left(q^{-1}\right)}{S(q-1)} = \frac{1 - 0.2q^{-1}}{1 - 0.7q^{-1} + 0.6q^{-2}} \tag{6.128}$$

so the controller does not change the minimum achievable output variance. The results of one simulation, Figure 6.66, illustrates how good the estimation performs for four filters and a time scale of 0.4. The autocorrelation of the estimated noise sequence for this simulation, Figure 6.66(a), is not exactly an impulse, as was found in the previous simulations, and its variance is larger than the expected value of 0.49. The oscillatory nature of the autocorrelation function suggests that it has been influenced by the complex roots in $S(q^{-1})$. However, the estimated minimum achievable output variance comes very close to the calculated value of 0.6125 in this simulation. From Figure 6.67, it is apparent that the mean value of $\hat{\sigma}_{mv}^2$ is just slightly larger than the calculated value.

The plot of the mean estimated minimum achievable output variance values, Figure 6.67, for various filter numbers and time scales suggests that more filters are required to obtain results similar to Figure 6.62. The effect of the complex poles becomes evident when results using two filters is examined. The noise sequence and $F(q^{-1})$ coefficients appear in Figure 6.68, and while the $F(q^{-1})$
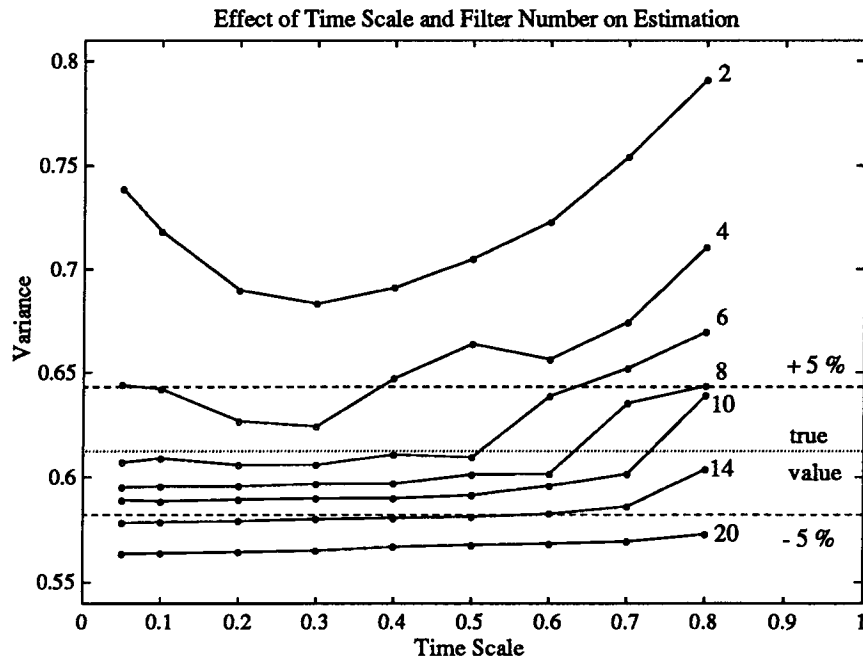
Figure 6.67: Effect of changing the time scale and filter number on $\hat{\sigma}_{mv}^2$ for process with complex poles
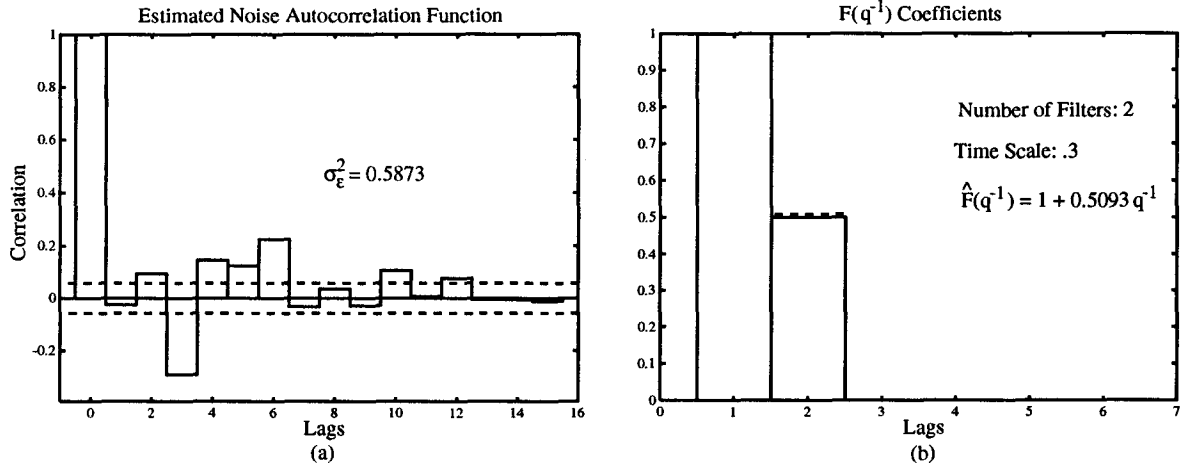
Figure 6.68: Effect of reducing filter number on (a) noise autocorrelation function and (b) the estimated $F(q^{-1})$ coefficients

coefficient is correctly estimated, Figure 6.68(a), shows that the problem lies with the estimation of the noise sequence. The autocorrelation function has a more oscillatory shape than previously found and the variance is much higher than expected. The cause of the error in the estimation $\hat{\sigma}^2_{mv}$ results from the error in the noise variance, which is caused by the complex pole pair.

The next process considered, Figure 6.69, is controlled by a Dahlin controller. This controller is poorly designed however, using a plant transfer function slightly different from the true one:

$$\frac{B(q^{-1})}{A(q^{-1})} = \frac{0.2q^{-2}}{1 - 0.8q^{-1}} \tag{6.129}$$

$$\sigma^2_e = 1.69$$
$$\sigma^2_y = 2.4513$$
$$\sigma^2_{mv} = 1.7576$$

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{1 - q^{-1} - 0.19q^{-2} + 0.22q^{-3} - 0.03q^{-4}}{1 - 1.2q^{-1} + 0.6q^{-2} - 0.25q^{-3}}$$

$$F(q^{-1}) = 1 + 0.2q^{-1}$$

Figure 6.69: Process controlled by Dahlin controller with complex poles

Figure 6.70: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for Dahlin controlled process with complex poles

This mismatch causes the introduction of a pair of complex poles at $0.18 \pm 0.515j$. Processes of this form, with a complex pole pair resulting from a mismatch between the true plant and the model used for control, are common and therefore important to study. The results of one simulation appear in Figure 6.70 and show that the autocorrelation function is once again oscillatory, though in this case $\hat{\sigma}^2_{mv}$ is quite close to the calculated value. The result of five hundred simulations, 1.741, found from Figure 6.71, confirms that the mean estimated $\hat{\sigma}^2_{mv}$ for this choice of time scale and filter number is nearly perfect. Figure 6.71 plots the remaining mean estimated minimum achievable output variance values and comparing this plot to Figure 6.65, the increase in the number of filters required to obtain a similar result is evident. When four filters are used rather than eight, Figure 6.72, $\hat{\sigma}^2_{mv}$ is
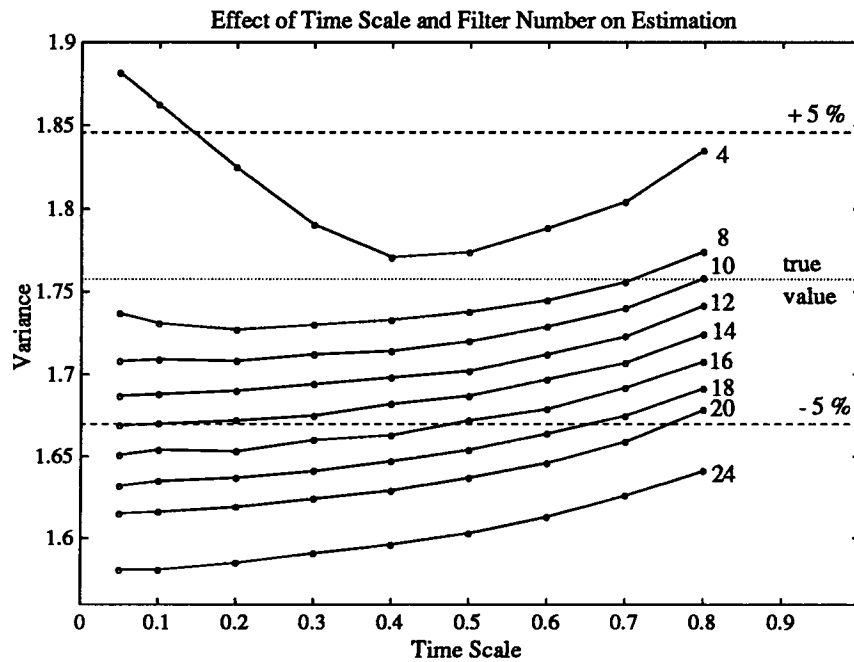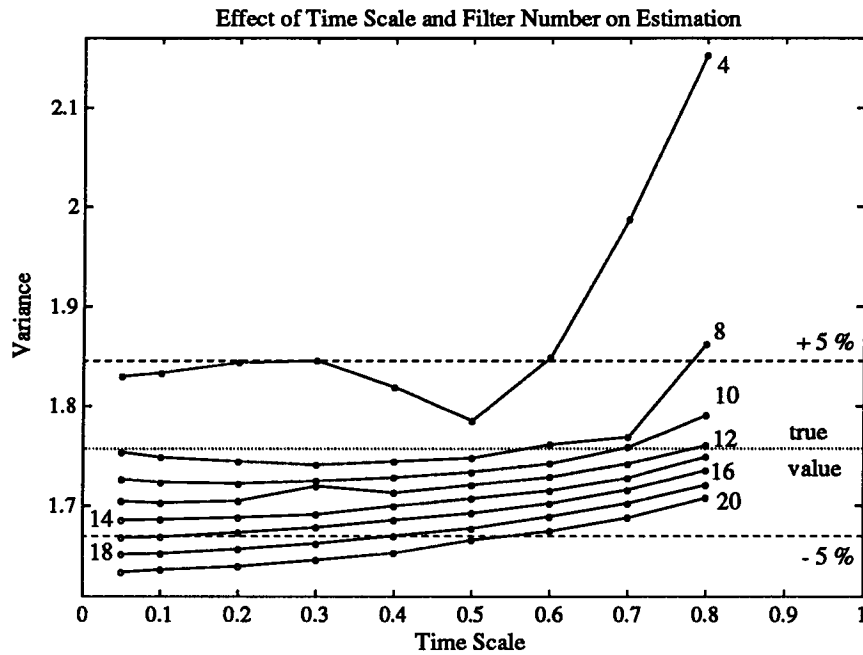
Figure 6.71: Effect of changing the time scale and filter number on $\hat{\sigma}_{mv}^2$ for Dahlin controlled process with complex poles

overestimated because the autocorrelation function of the estimated noise sequence is more oscillatory than in the previous case, resulting in an inflated noise variance.

These process simulations suggest that the number of filters must be increased in order to get the same results as for a similar process which has real poles. However, the plots of Figure 6.71 and Figure 6.65 suggests that it is possible to get estimation within plus or minus five percent
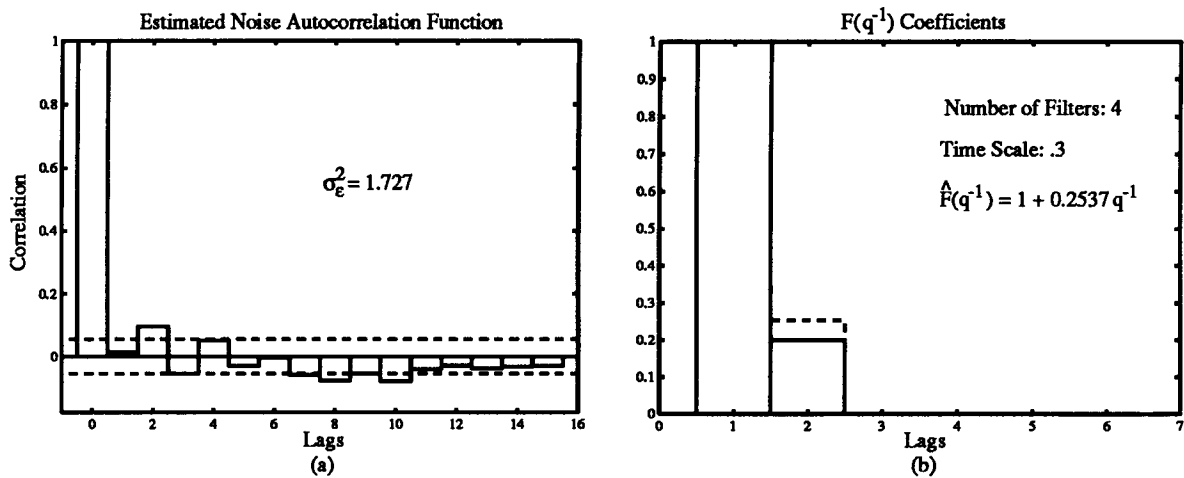
Figure 6.72: Effect of reducing filter number on (a) noise autocorrelation function and (b) the estimated $F(q^{-1})$ coefficients

of the calculated minimum achievable output variance, with the same choice of Laguerre network parameters, for both situations.

6.6.3 Closed Loop Noise Transfer Function Numerator Not SPR

The final situation, when the numerator polynomial of the closed loop noise transfer function is not strictly positive real, is considered. The numerator polynomial is given by:

$$R(q^{-1}) = C(q^{-1})A_1(q^{-1})H(q^{-1}) \qquad (6.130)$$

so it is made up of the process disturbance numerator, the plant denominator and the denominator of the controller. Combining these polynomials can easily result in $R(q^{-1})$ having a high order, which will then most likely mean it is not SPR. The SPR property can be explained in the following manner. A transfer function will have a SPR numerator if a white noise input results in an output which has an autocorrelation function close to an impulse. The higher the numerator order, the more noise coloring that o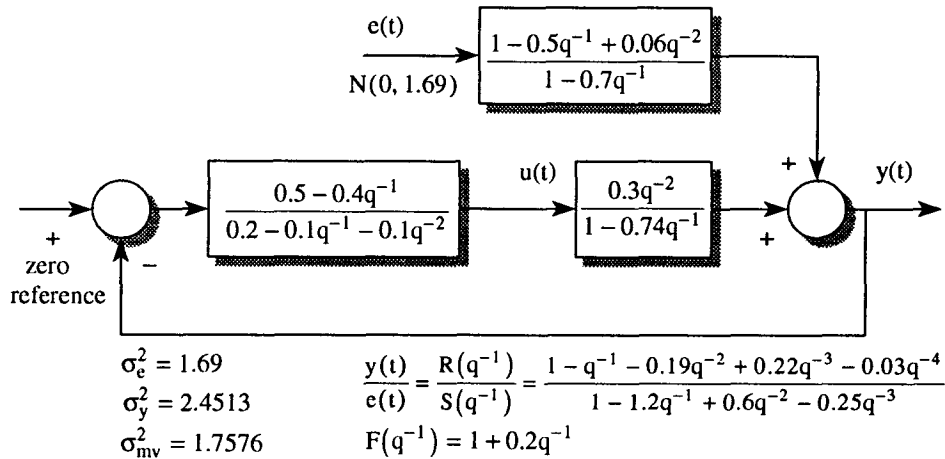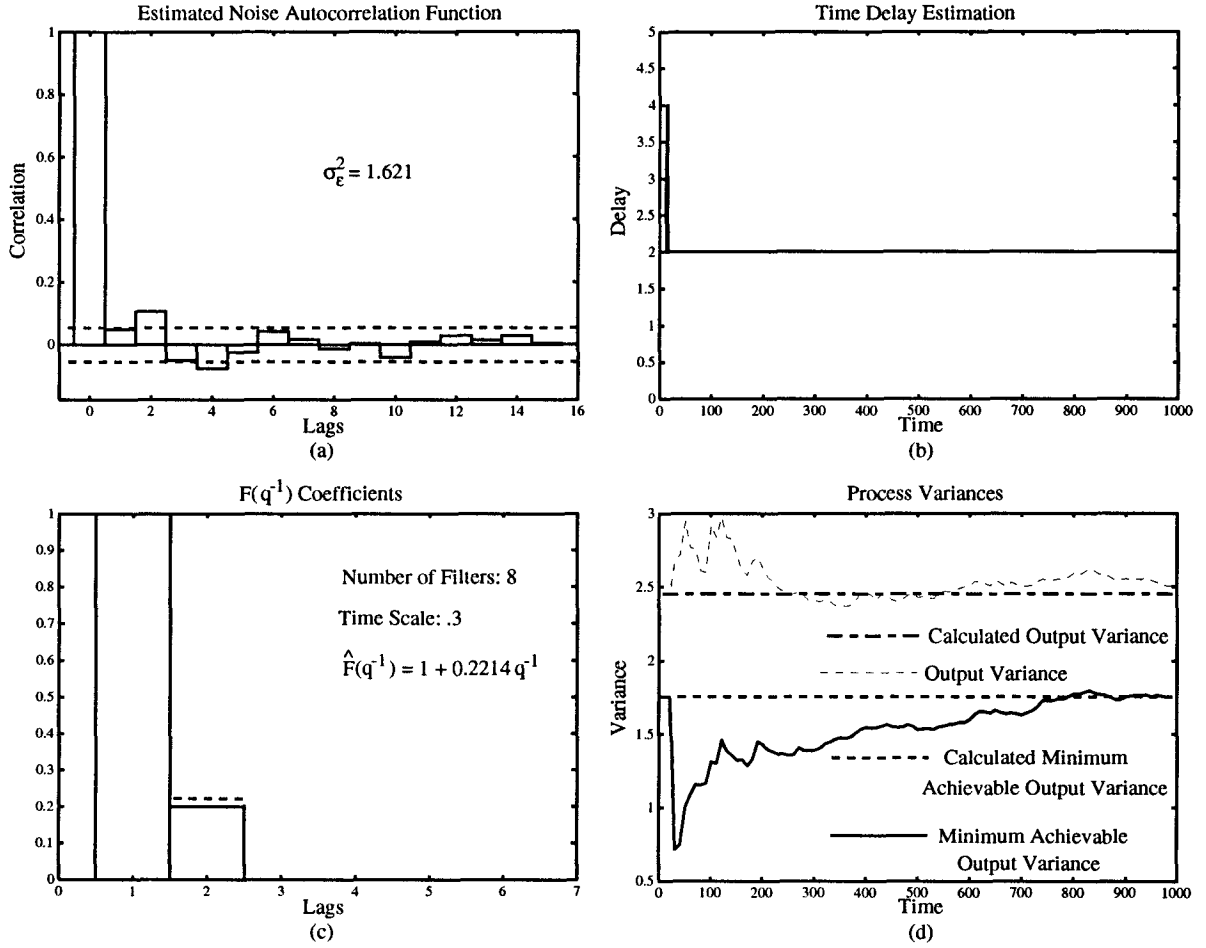ccurs, and so the autocorrelation function will no longer be an impulse. Then there will be a large chance that the numerator is not SPR.

It is important that the $R(q^{-1})$ polynomial is positive real for the estimation of the minimum achievable output variance. According to a condition of RELS estimation, it is necessary for $R(q^{-1})$ to be SPR for convergence of the estimates to be assured. However, the nature of the condition means that the estimates may converge even if $R(q^{-1})$ is not SPR. The first example illustrates this

$$\sigma_e^2 = 0.36$$
$$\sigma_y^2 = 0.9236$$
$$\sigma_{mv}^2 = 0.6256$$

$$\frac{y(t)}{e(t)} = \frac{R(q^{-1})}{S(q^{-1})} = \frac{1 - 0.2q^{-1} - q^{-4} + 0.2q^{-5}}{1 - 0.8q^{-1}}$$

$$F(q^{-1}) = 1 + 0.6q^{-1} + 0.48q^{-2} + 0.384q^{-3}$$
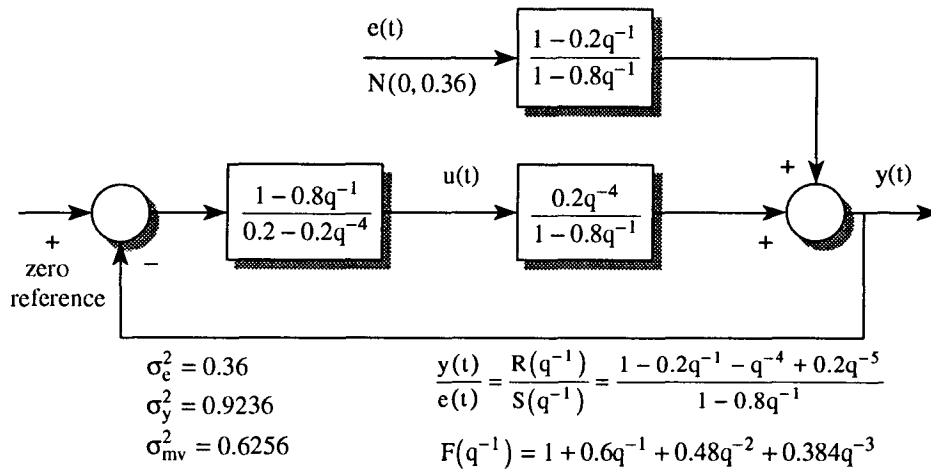
Figure 6.73: Process controlled by Dahlin controller without SPR $R(q^{-1})$
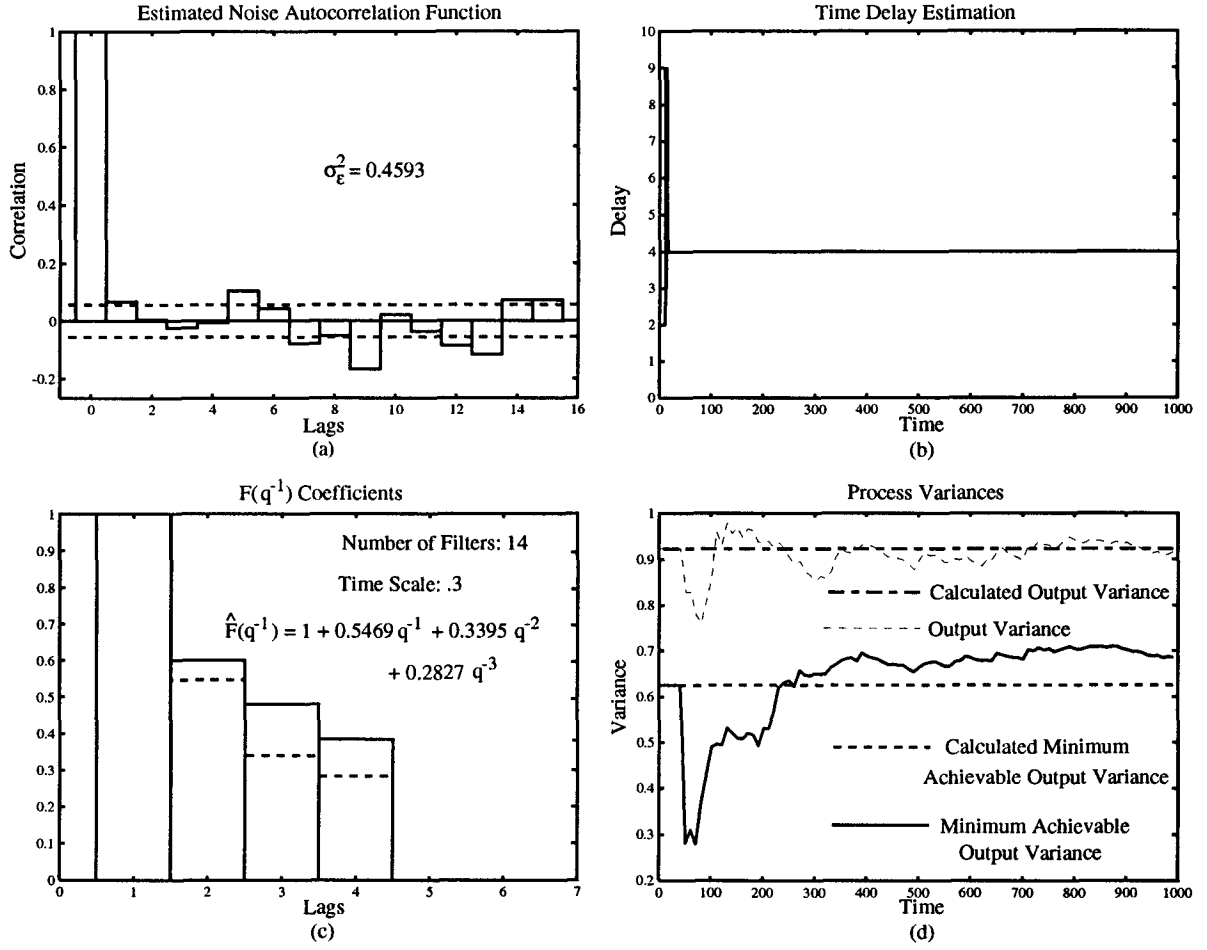
Figure 6.74: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for Dahlin controlled process without SPR $R(q^{-1})$

fact and shows how the choice of filter number and time scale are affected by the realness condition. The second process illustrates limitations of the estimator caused by this condition.

The process in Figure 6.73, similar to those used previously, but it uses a Deadbeat controller which results in a $R(q^{-1})$ polynomial that is not SPR. Results of a single simulation, Figure 6.74, show that the error between $\hat{\sigma}^2_{mv}$ and σ^2_{mv} appears to be caused by an inability to obtain a correct estimate of the noise sequence and incorrect $F(q^{-1})$ coefficients. It has been observed for various processes with a numerator that is not SPR, that a good $\hat{\sigma}^2_{mv}$ value can be obtained despite poor estimation of the noise sequence and $F(q^{-1})$ coefficients, . The histogram constructed from the $\hat{\sigma}^2_{mv}$ values obtained from five hundred simulations, Figure 6.75, confirms that the mean is indeed close to the calculated value. However, the distribution of the histogram differs, in that it has a larger variance, which is

Histogram
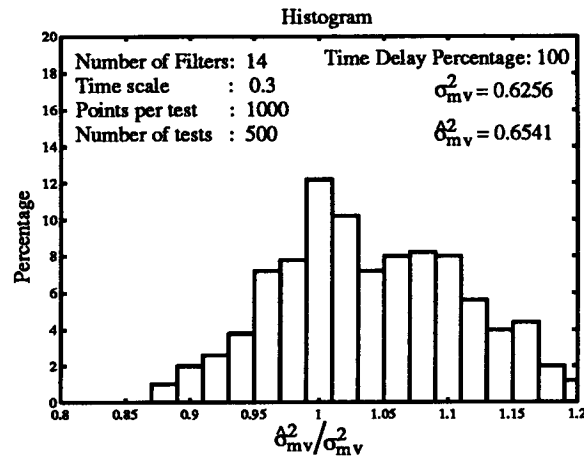
Figure 6.75: Distribution of the results of five hundred simulations for Dahlin controlled process without SPR $R(q^{-1})$

also a common occurrence for processes with a numerator that is not SPR. The unrealness of the numerator affects the choice of the Laguerre parameters used to obtain good estimates. A plot of the mean $\hat{\sigma}^2_{mv}$ determined for different choices, indicates the number of filters required to obtain good estimation has increased a great deal over previous situations. Figure 6.76 shows that even if
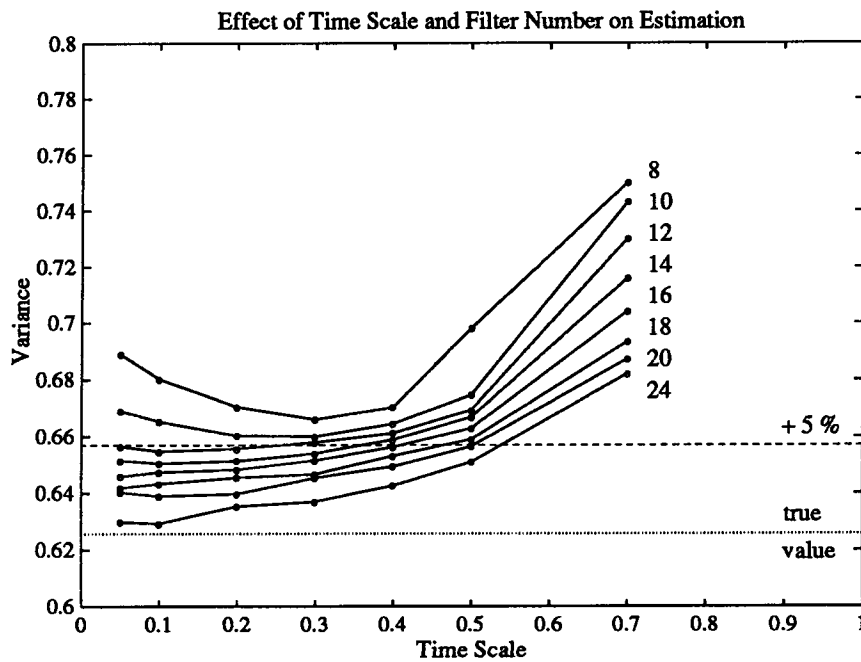
Figure 6.76: Effect of changing the time scale and filter number on $\hat{\sigma}^2_{mv}$ for Dahlin controlled process without SPR $R(q^{-1})$

twenty four filters are used, the mean value does not drop below the calculated value. Also, the choice of the time scale becomes more important then previously, with a value in the range [0.2,0.4] a suitable choice. It appears that if $R(q^{-1})$ is not SPR, the estimation of the minimum achievable output variance is noticeably affected, but it can be compensated for by the prudent choice of a time scale and filter number.

To illustrate the limitations of the RELS method, when a numerator is not SPR, the process in Figure 6.59 is used, except that the noise filter numerator has been changed. This new choice of the $C(q^{-1})$ polynomial is:

$$C(q^{-1}) = 1 + 1.5q^{-1} + 0.75q^{-2} \qquad (6.131)$$
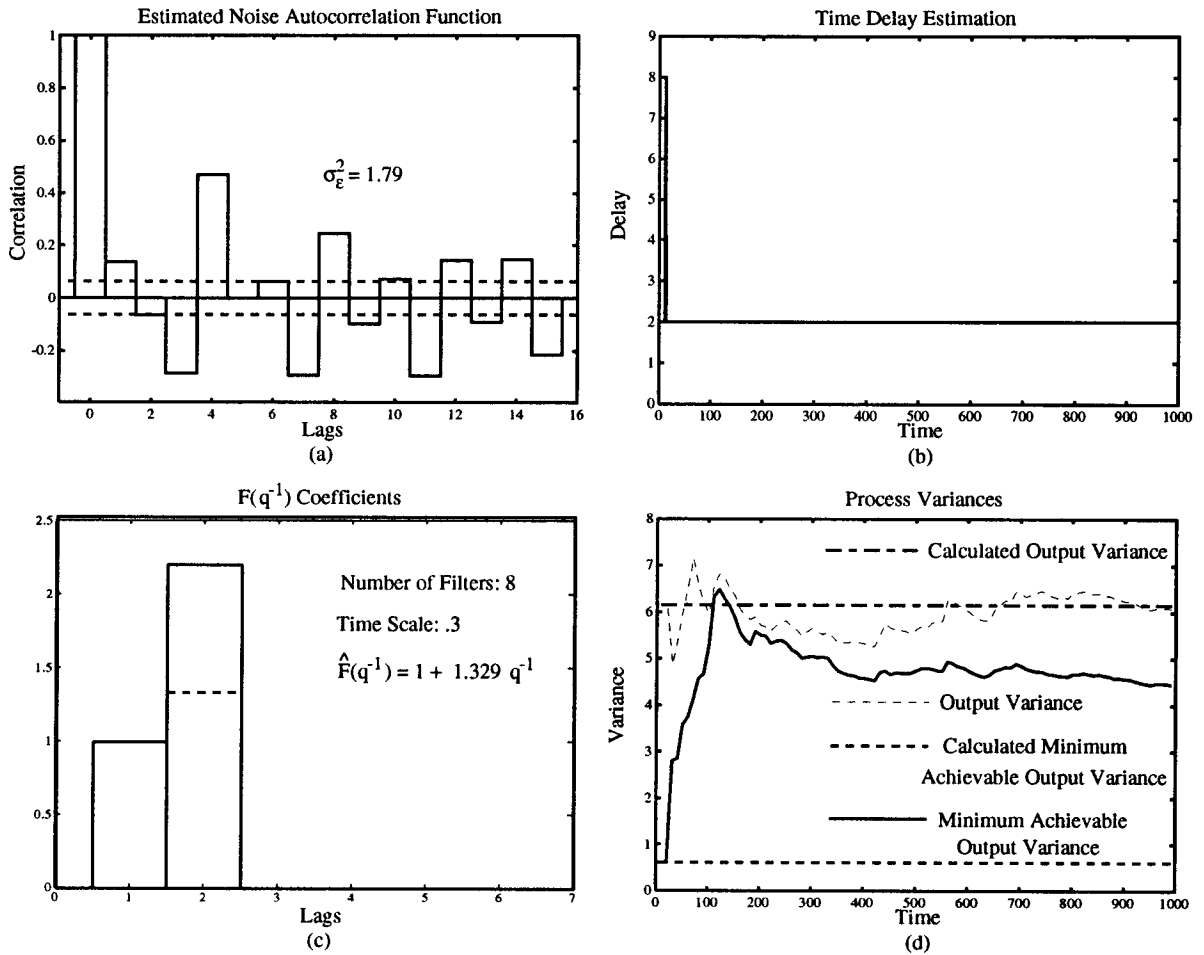
Figure 6.77: The (a) estimated noise autocorrelation function, (b) the FMVRE delay estimation, (c) the estimated $F(q^{-1})$ coefficients and (d) $\hat{\sigma}^2_{mv}$ and σ^2_y for process without SPR $R(q^{-1})$

which because the controller is a pure gain equals the $R(q^{-1})$ polynomial. The changes to the other process properties is:

$$\sigma_y^2 = 6.1521 \quad \frac{R(q^{-1})}{S(q^{-1})} = \frac{1+1.5q^{-1}+0.75q^{-2}}{1-0.7q^{-1}}$$

$$\sigma_{mv}^2 = 2.2616 \quad F(q^{-1}) = 1 + 2.2q^{-1}$$

This polynomial is used extensively in Pseudolinear Regression literature [21] as an example for which RELS estimation will not converge. Attempts to estimate the minimum achievable output variance, as in Figure 6.77, illustrate its failure to come close to σ_{mv}^2. Both the estimated noise sequence and $F(q^{-1})$ coefficients are nowhere near their calculated values. Figure 6.78 confirms the inability of the method to converge, as the mean $\hat{\sigma}_{mv}^2$ values are not close to the calculated value. This process illustrates that estimation may not be possible for all processes, however, the problem will be evident after the autocorrelation function of the estimated noise sequence is examined. The autocorrelation function will have large correlation at numerous lags, as shown in Figure 6.77(a), when the nonrealness of the $R(q^{-1})$ polynomial is affecting estimation. Fortunately, this problem does not seem to occur for the majority of cases when $R(q^{-1})$ is not SPR.

The results of these simulations suggest that some trial and error may be necessary to obtain suitable values for the time scale and number of filters used in the Laguerre network. There is often
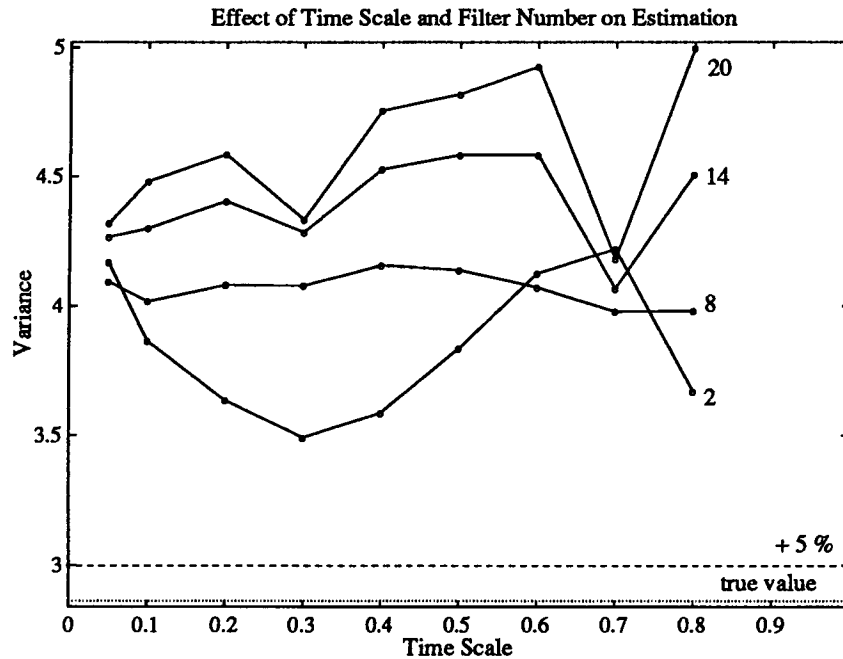

Figure 6.78: Effect of changing the time scale and filter number on $\hat{\sigma}_{mv}^2$ for process without SPR $R(q^{-1})$

a wide range of Laguerre parameters which result in good estimation, that being plus or minus five percent of the calculated value. Based on the previous simulations, some general recommendation will be made.

The estimated noise sequence autocorrelation function is an important tool that should be used when choosing a filter number. If the autocorrelation function has correlation at lags above zero, this indicates the number of filters should be increased. If increasing the filter number does not improve the autocorrelation function shape, then estimation may not converge, due to the fact the $R(q^{-1})$ polynomial is not strictly positive real. In this situation, an alternative estimator, such as Maximum Likelihood should be employed. If the noise autocorrelation function has an impulse shape, then decreasing the number of filters used should be attempted, as long as the autocorrelation function of the estimated noise remains white. Using the lowest number of filters which result in an impulse autocorrelation, will provide the best estimate of σ_{inv}^2. Generally, the simulations obtained good results for a range of Laguerre network filters, regardless of whether $R(q^{-1})$ was SPR or $S(q^{-1})$ has real or complex poles. The choice of time scale does not appear as crucial as the choice of filter number. Except when $R(q^{-1})$ is not SPR, the $\hat{\sigma}_{inv}^2$ value found varied only slightly for different time scale values for a constant filter number. Processes where $R(q^{-1})$ positive realness is a concern, the time scale should be used in the range of [0.2,0.5]. Based on various simulations, this time scale range can be used with a moderate degree of confidence, in all situations.

6.7 Conclusions

There were several different simulations presented in this chapter, all attempting to examine relevant details encountered when using the estimation methods presented for monitoring. The results of each section will now be summarized in order that the important details can be emphasized.

The first simulations performed were concerned with determining the possible causes of biased delay estimation and suggesting solutions able to remove the bias. If the autocorrelation function of the disturbance output is not zero for lags which correspond to the possible interval of delay, a biased delay estimate may occur. The bias may occur at one sample, but this also corresponds to the delay caused by the sampling procedure and does not need to be included in the possible range

of delay, if the process is known to have delay. It is also possible for the bias to occur at a value other than one. This bias can be removed by filtering the signals as explained previously. The other source of bias arises when the input signal does not contain sufficient excitation to estimate the delay. Unfortunately, it is difficult to justify removing this bias based solely on the fact that the delay is required only for monitoring. An equation was determined that only requires the sampling interval and the lower frequency required to estimate the delay can be determined. Thus by examining the power spectral density function of the input it is possible to determine if the excitation requirements are met. In general, the safest course to follow when estimating the delay is to examine the E_l function to check if the maximum value definitely outweighs the other values. If this occurs, then the delay can be used with confidence.

The next study was concerned with determining the performance of the ANM on closed loop processes with noise. The first simulation examined a linear stochastic process and examined the use of low pass filters and the benefit of turning off the ANM for a short period after a reference change. The results showed the benefits of these techniques and it is suggested that they are implemented if it is not too difficult. Any nonlinearity in the process will degrade its performance, so it is important that they can be estimated by the ANM. Two typical nonlinearities were used, a deadzone and saturation, and both were correctly estimated by the ANM. By including filtering to reduce the noise and turning off the ANM after reference changes, the ANM is capable of estimating the static input-output relation.

Focus then shifted to the estimation of the minimum achievable output variance and interest was in determining a suitable choice for the Laguerre network parameters and determining the feasibility of using the estimated properties, required to calculate $\hat{\sigma}^2_{\text{mv}}$, as indicators of process change. The FMVRE will estimate changes in delay and RELS will estimate any changes to the noise variance. Only σ^2_y will change if the process plant changes, while if only the Markov coefficients change, modifying $\hat{\sigma}^2_{\text{mv}}$, then the process disturbance filter is different. Using this knowledge it is possible to determine changes in the process. It was found that the choice of Laguerre parameters depends on the properties of the closed loop noise transfer function. The best way to determine if the Laguerre network is suitably modelling the closed loop noise transfer function is by examining the estimated

noise autocorrelation function. As few filters as possible should be used so that it has a white noise autocorrelation function. In general more filters will be required when the denominator of the closed loop noise transfer function has complex roots or if the numerator in not SPR. There is also the possibility that the estimates will not converge to their true values when the numerator in not SPR. This is a limitation of RELS estimation, which can be overcome by using a more complicated estimation algorithm. Fortunately, this problem is evident as the noise autocorrelation function will not be near an impulse and the problem itself does not seem to occur that frequently. While some trial and error may required to determine suitable Laguerre values, a good starting point is around ten filters with a time scale in the range of [0.2,0.5].

While investigating the method used to determine the minimum achievable output variance, the advantages of using a Laguerre network over a time-series model were displayed. Despite changing every part of a process, which resulted in a different closed loop noise transfer function, one choice of Laguerre parameters was able to provide a good estimate of σ^2_{mv}. Meanwhile, an ARMA model would have to be chosen with some care if a time-series representation was used and it would likely have to be modified as the closed loop noise transfer function changed. There also is comfort in the knowledge that increasing the size of the Laguerre network will result in improved estimation because the computational load will increase. With an ARMA model there is no guarantee that the estimation will improve though the computational load also increases. While investigating a suitable choice of Laguerre parameters, it was found that a wide range will result in good estimation, a robustness not possessed by an ARMA model. While the Laguerre gains do not offer the same insight into a process as do the roots of the time-series polynomials, in this application it is not important as it is the Markov coefficients that are of interest and must be calculated regardless of which model is used. In conclusion, the use of a Laguerre network provides good estimation of a changing closed loop noise transfer function while avoiding the problem of choosing the ARMA model order.

Chapter 7

Industrial Data

The most crucial test faced by the monitoring tools is now presented, the application to industrial data. The data were gathered from three different processes, a pressure screen, a reject refiner and a Kamyr digester. Monitoring was performed in an effort to test the operation of the tools on industrial processes as well as demonstrate the usefulness of monitoring. For each process presented, a brief overview of its operation is followed by the goal of monitoring in this specific case. The output and input signals are presented and then analyzed. The monitoring is performed, outlining what was required in order that the presented results could be achieved. Finally, attempts are made to verify the results obtained and once they are considered reasonable, the results are interpreted to evaluate the process performance.

7.1 Pressure Screen Reject Flow

A pressure screen is an integral component of the pulping process. They are often used after wood chip refining to remove unwanted components from the pulp, such as oversized fibres, dirt and bark. A screen will typically have two outlets, one for the accepted pulp and the other for rejected pulp. The reject flow is usually held constant, removing a sufficient percentage of the unwanted components in the pulp. The pressure screen reject flow loop is a basic flow loop, where the output is the flow of pulp and the valve position is the input. In this application, the flow is controlled by a PI controller and the process is sampled every two seconds.

This flow loop is typical of the type of processes which can be monitored using the methods presented. The data will be examined to determine the performance of the PI controller. Once the minimum achievable output variance is determined, it can be compared to the process output variance to determine any possible improvement.

The process data appears in Figure 7.79, and shows approximately 1200 samples or about forty minutes of the screen operation. The flow reference was six hundred litres per minute, and the output stays about this value, and has a mean of six hundred litres per minute. The input is also
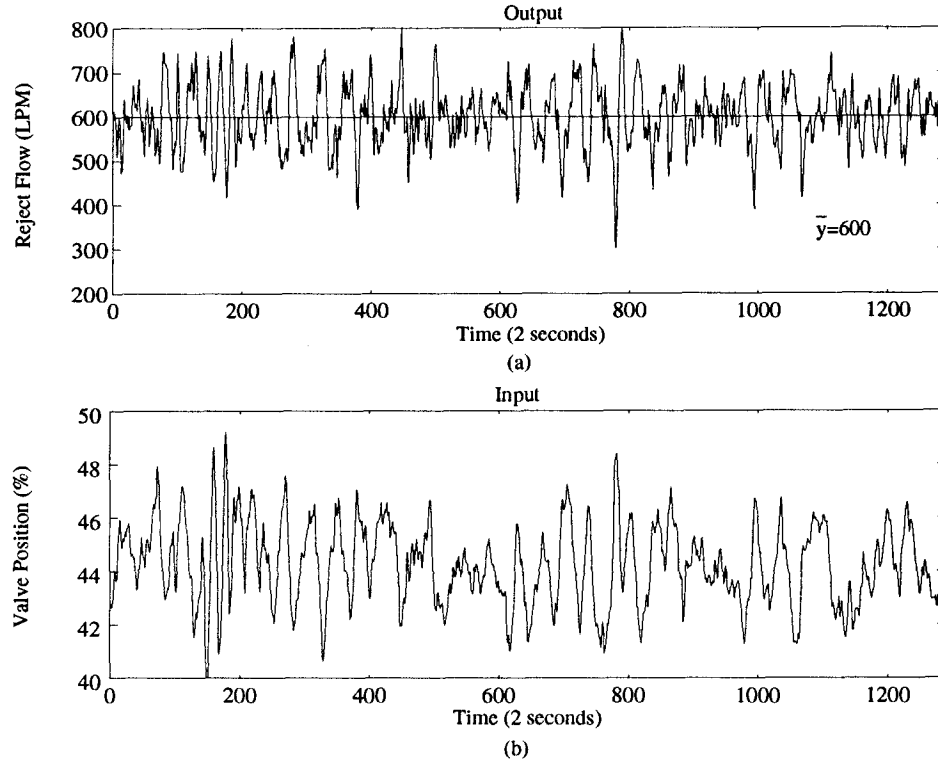
Figure 7.79: The pressure screen reject flow input and output

shown, the position of the valve, and it is given as a percentage of the fully open position. Before applying the monitoring tools, the input and output signals are examined. Figure 7.80(a) shows the autocorrelation function of the process output. The process variance after 1200 samples was 5128 and there is significant correlation at most lags, even up to fifteen. This indicates that unless the delay is fifteen samples or thirty seconds, which is unlikely for a flow loop, the process is definitely not under minimum variance control. There is an indication at this point then, that the process variance is not at the minimum level. The variation of the output signal seems high as a standard deviation of seventy is quite large when compared to the reference. Also shown on Figure 7.80(b) is the power spectral density of the input signal, which will be examined to determine if it will be possible to estimate the delay with the FMVRE. Using equation 6.109, the lower limit on the frequency is determined:

$$f_L \simeq \left(\pi \sqrt{8 T_s^2} \right)^{-1} = \left(\pi \sqrt{8 (2^2)} \right)^{-1} = 0.056 \text{ Hz} \qquad (7.133)$$

Comparing this number with the power spectral density, it can be seen that the power in the signal dies out around 0.05 Hz which indicates that there may be difficulty in estimating the delay.
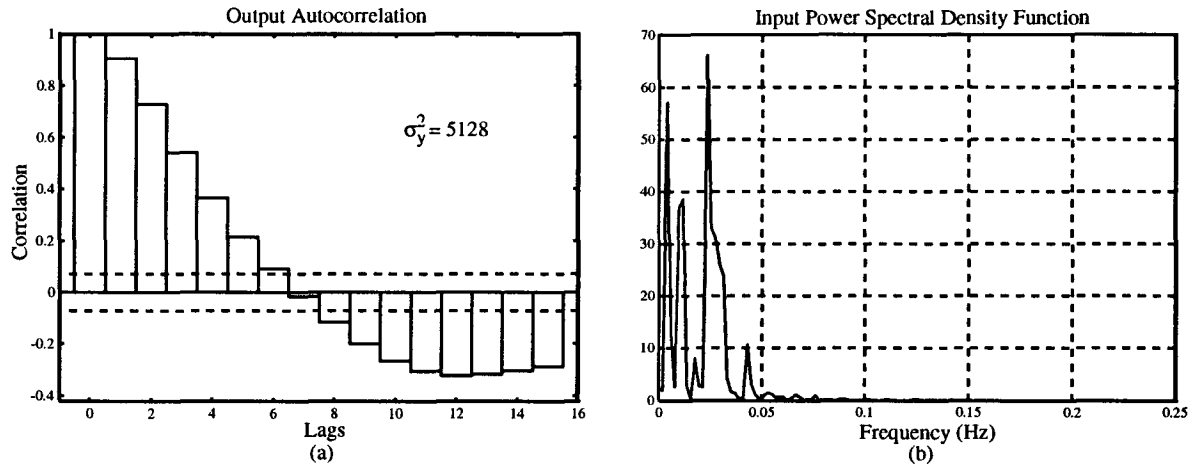
Figure 7.80: The output autocorrelation function and input power spectral density function
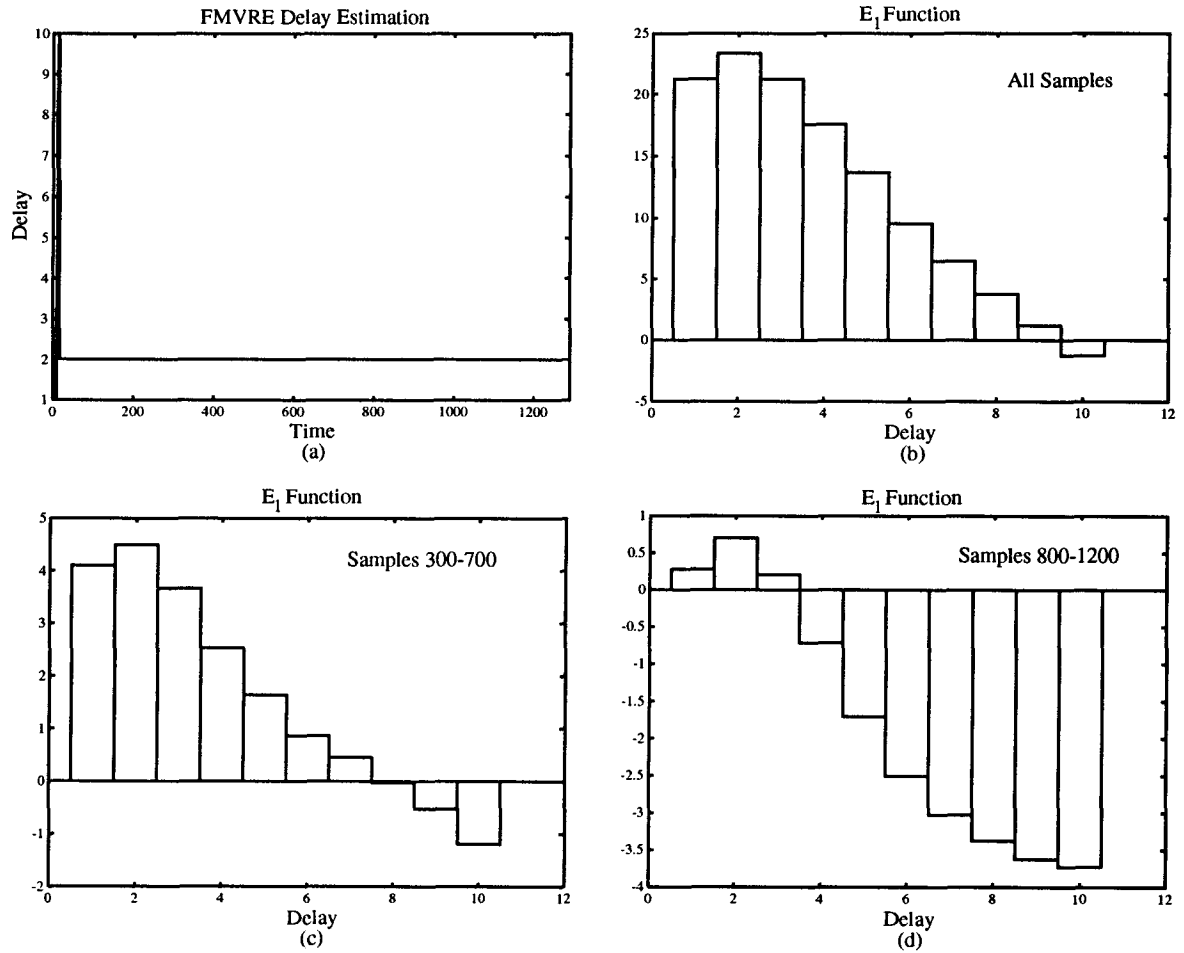
Figure 7.81: The FMVRE (a) estimation of the delay and E_1 function for (b) all samples, (c) samples 300 to 700 and (d) samples 800 to 1200

The estimation of σ_{mv}^2 begins by first determining the time delay of the process, using the FMVRE. The results of the estimation appear in Figure 7.81 and Figure 7.81(a) indicates that the delay of the process is two samples. Examining the E_1 function, Figure 7.81(b), the peak is at two, though the values at one and three are not that far off. It was shown that the power in the input signal dies out near the lower frequency limit required by the FMVRE to estimate the delay and this accounts for the shape of the E_1 function. The consistency of the estimation is checked in Figure 7.81(c) and Figure 7.81(d), by using two subsets of the data. They also suggest a delay of two. A delay of two shall be used throughout the remaining section for determining the minimum achievable output variance.

Next, in order to find the closed loop noise transfer function and noise sequence, the number of filters and their time scale must be decided. The filter number was chosen so that the noise autocorrelation function had an impulse shape. The results from two different choices are shown in Figure 7.82. It can be seen that the autocorrelation function has some oscillation, but is closer to an impulse when ten filters are used. It was found that increasing above ten filters did not improve the shape of the autocorrelation function.

The minimum achievable output variance estimation is shown in Figure 7.83, for the Laguerre network parameters discussed above. The results are quite similar for both choices of filter number as the variance ratios are within five percent. It was found that despite using a wide range of filter number (4 to 20) and time scale (0.1 to 0.7), the majority of the variance ratio lay between 2.1 and 2.6. Once again the Laguerre network shows its ability to provide good estimation without too much concern with the parameters chosen.

It is time to step back a moment and consider the results obtained, to see if they are reasonable for this process. Some experience has been gained by controlling this process, which indicates that the delay is around four seconds. Also, the process is known to be quite oscillatory, which is confirmed by the estimated noise autocorrelation function. Based on the a-priori knowledge available, the results obtained seem reasonable. As a final check of the estimates, the closed loop noise transfer function is modelled using an ARMA(5,5) model whose parameters are estimated using Maximum Likelihood Estimation. This is the method that was used by Harris [17] and as the Maximum
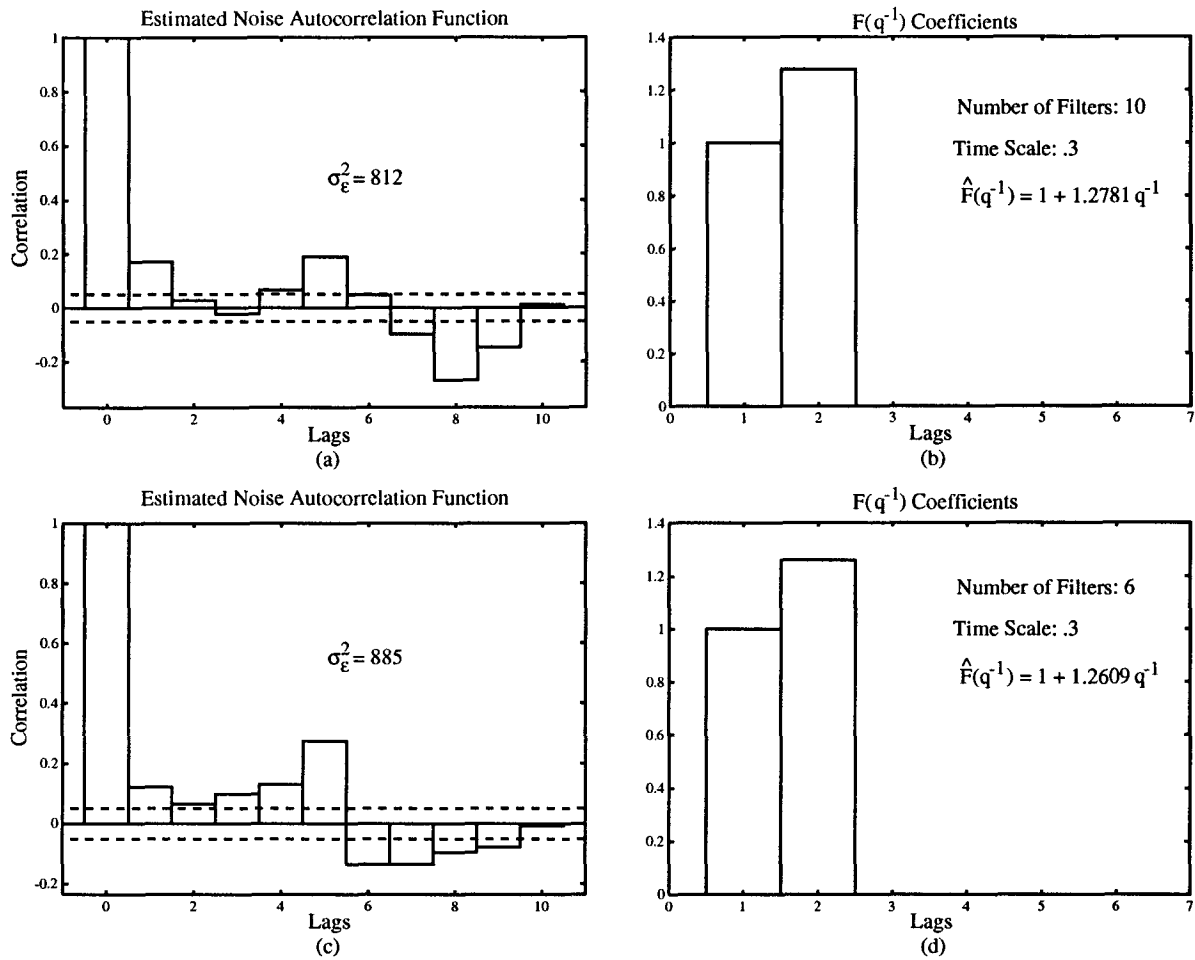
Figure 7.82: The estimated noise sequence and $F(q^{-1})$ coefficient for (a&b) ten filters and (c&d) six filters
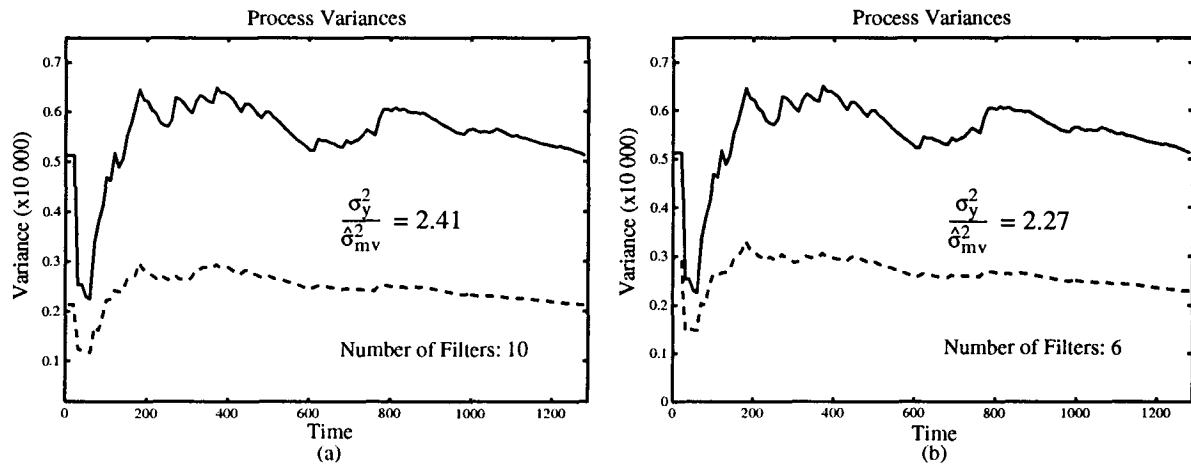
Figure 7.83: The estimated minimum achievable output variance using (a) ten filters and (b) six filters

99

Likelihood estimator used is an off-line batch estimator, rather than an on-line recursive estimator, the estimation will not be biased and can be compared to the values estimated using RELS. The result of this estimation found:

$$\frac{\hat{R}(q^{-1})}{\hat{S}(q^{-1})} = \frac{1 - 1.01q^{-1} + 0.55q^{-3} - 0.28q^{-4} - 0.23q^{-5}}{1 - 2.33q^{-1} + 1.9q^{-2} - 0.94q^{-4} + 0.5q^{-5}} \qquad \hat{\sigma}_{\varepsilon}^{2} = 738$$

$$\hat{F}(q^{-1}) = 1 + 1.316q^{-1} \qquad \frac{\sigma_{y}^{2}}{\hat{\sigma}_{uv}^{2}} = 2.5$$

The variance ratio, estimated noise variance and estimated $F(q^{-1})$ coefficient all compare very closely with the values obtained previously.

Finally, what do the results mean? They suggest that the output variance could be reduced by over two times if a minimum variance controller was used. This is a fair amount, but if the PI controller is desired rather than a minimum variance controller, it should be retuned to attempt to lower the ratio. If this is not possible, then perhaps a different control algorithm, such as a Dahlin controller should be employed. Also, it should be considered if the minimum standard deviation of forty-five is acceptable. If it is not, then changes to the process will be required to reduce it further, or feedforward control must be employed. No time should be wasted retuning the controller if the minimum standard deviation is not satisfactory.

7.2 Reject Refiner Motor Load

The next process examined is the motor load loop of a reject refiner. A reject refiner is a smaller version of a main refiner, one which appears in Figure 7.84. The reject refiner is typically found downstream of a pressure screen, as it refines the rejected pulp, reducing the pulp size further so that it is suitable to be used. The mechanical operation of a double-disc refiner, which has two plates, will now be explained. The plates are rotated and pushed together by large motors and a hydraulic cylinder [11]. Fibres and dilution water are fed into the center of the plate gap and a fibre pad forms between the plates. The pulp works its way out the periphery of the plates, in smaller pieces than when it entered. The energy per unit mass of wood fibres is a major factor in the resulting pulp quality and depends on the chip feedrate into the refiner and its motor load. The motor load is changed by manipulating the gap between the plates which increases as the gap decreases. An

100
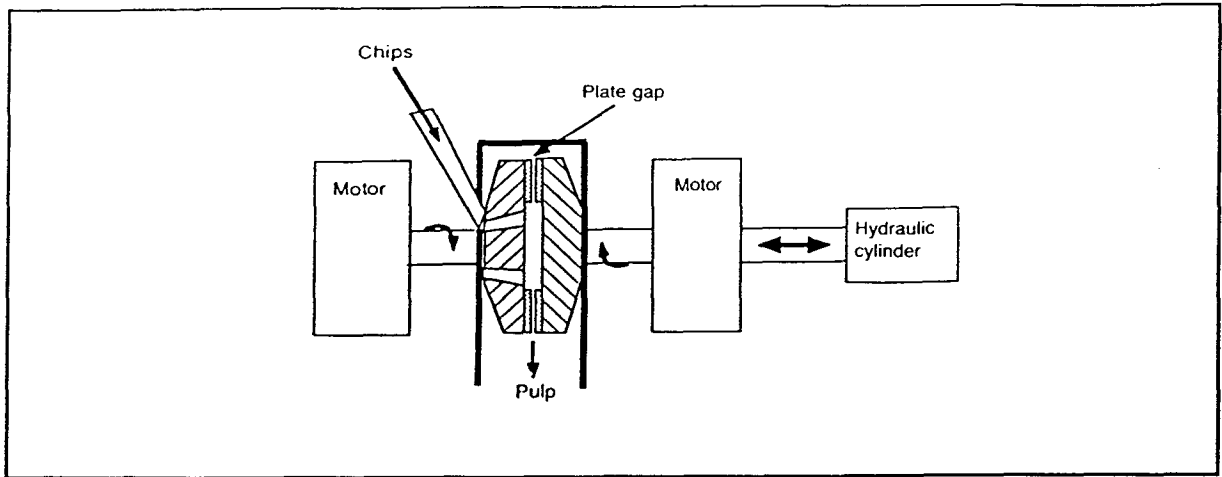
Figure 7.84: A double-disc refiner

interesting property of this relation is the fact that the gain will change sign, which occurs if the plates are moved too close, as the pad of fibres breaks down, allowing the plates to move together easier, and ultimately they clash together. This means that there is a maximum motor load setpoint which is obtainable. The goal of controlling the motor load is to keep it as high possible, all the while avoiding a pad collapse. The reject refiner motor load is controlled by an Adaptive Constrained Minimum Variance (CMV) controller and it is sampled every second.

Though this loop will typically be in regulation mode for large periods of time, this data was gathered during a period when the reference motor load was changed often. This is because the control action was being tested by moving the motor load so the refiner operated in near the pad collapse region. The reference signal changed often over the period it was collected, so it is not possible to examine the minimum achievable output variance. However, the ANM will be used on this data, in order that the static input-output relation can be estimated.

The input and output signals appear in Figure 7.85, and the 2700 samples corresponds to about forty-five minutes of operation of the refiner, including seventeen reference changes made in this period. It is obvious from the output that this process is quite noisy. The motor load reference and output is measured in megawatts (MW) and the plate gap is given as a percentage, where 100% corresponds to the completely closed position. The motor load reference is set to values which are chosen to induce a pad collapse, but the control algorithm successfully moves the refiner operation

101
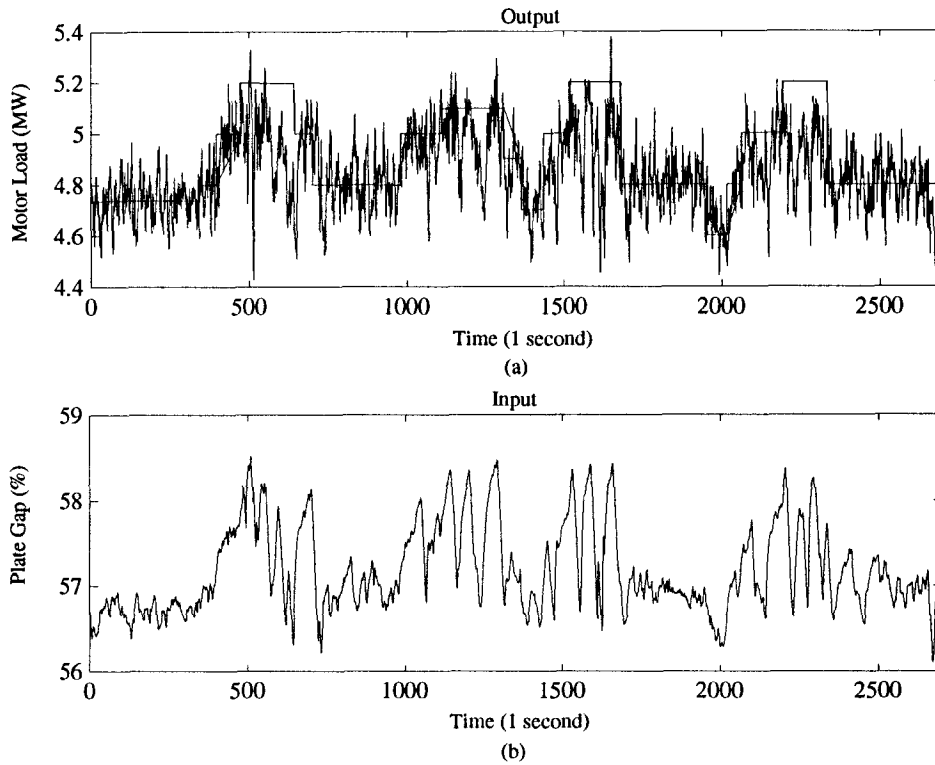
Figure 7.85: The input and output signals of the reject refiner

away from a possible collapse. This action provides several cases where the refiner operation may have entered the region of slope change, so that the ANM may be able to estimate the slope change.

The first step that must be performed is to filter out the noise in the signals. This was done using a Butterworth low pass filter with a cutoff frequency at 0.05 Hz. The signals which result from this filtering appear in Figure 7.86. They have a similar trend as the original signals but have no high frequency components. Next, the interval for the ANM must be decided on, so that the static input-output relationship can be estimated. First, the interval boundaries were chosen, set so that most of the input falls in that range. The interval of [56.6,58.4] was decided upon and the ANM estimation for this range appears in Figure 7.87(a). The slope of this line indicates the gain of the process over this interval and it highlights a limitation associated with using a single estimate for the gain of this process. This result suggests that the gain is constant over the interval, however, the ANM will show this is not the case. The second curve in Figure 7.87(a), using the intervals [56.6,57.2,57.8,58.4] indicates that the slope is relatively constant below 57.8, but seems to change above this value.
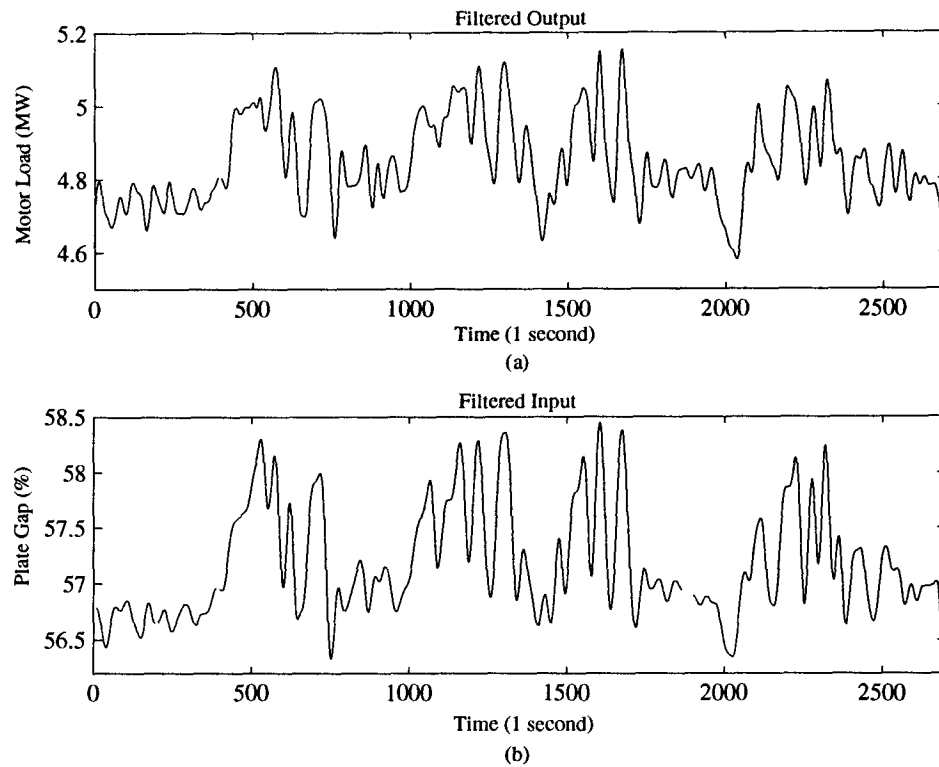
Figure 7.86: The filtered input and output signals of the reject refiner

Breaking the interval down into six segments, Figure 7.87(b), gives a better indication of the area where the slope change occurs. In Figure 7.87(c), further breaking down of the region above 57.8 occurs, where the slope changes. The ANM was also run using the unfiltered data, to see the effect of the noise on the estimation. Figure 7.87(d) compares the results with the single interval result and though the bias created by the noise is evident, the same shape is there.

As this is data from a real process, it is not possible to know if these results are correct, based only on the data supplied. However, the results do make sense when they are compared with what is known about the process. For this process, a change in slope is expected, so the results of ANM estimation are consistent with this knowledge. There is an advantage in knowing where the slope change occurs, as it can be used to operate the motor load loop. The motor load reference can be chosen based on the results of the estimation, to avoid pad collapse and plate clash. Also, the static gain curve can be used by the a controller to signal that a reference change is required to avoid a plate clash.
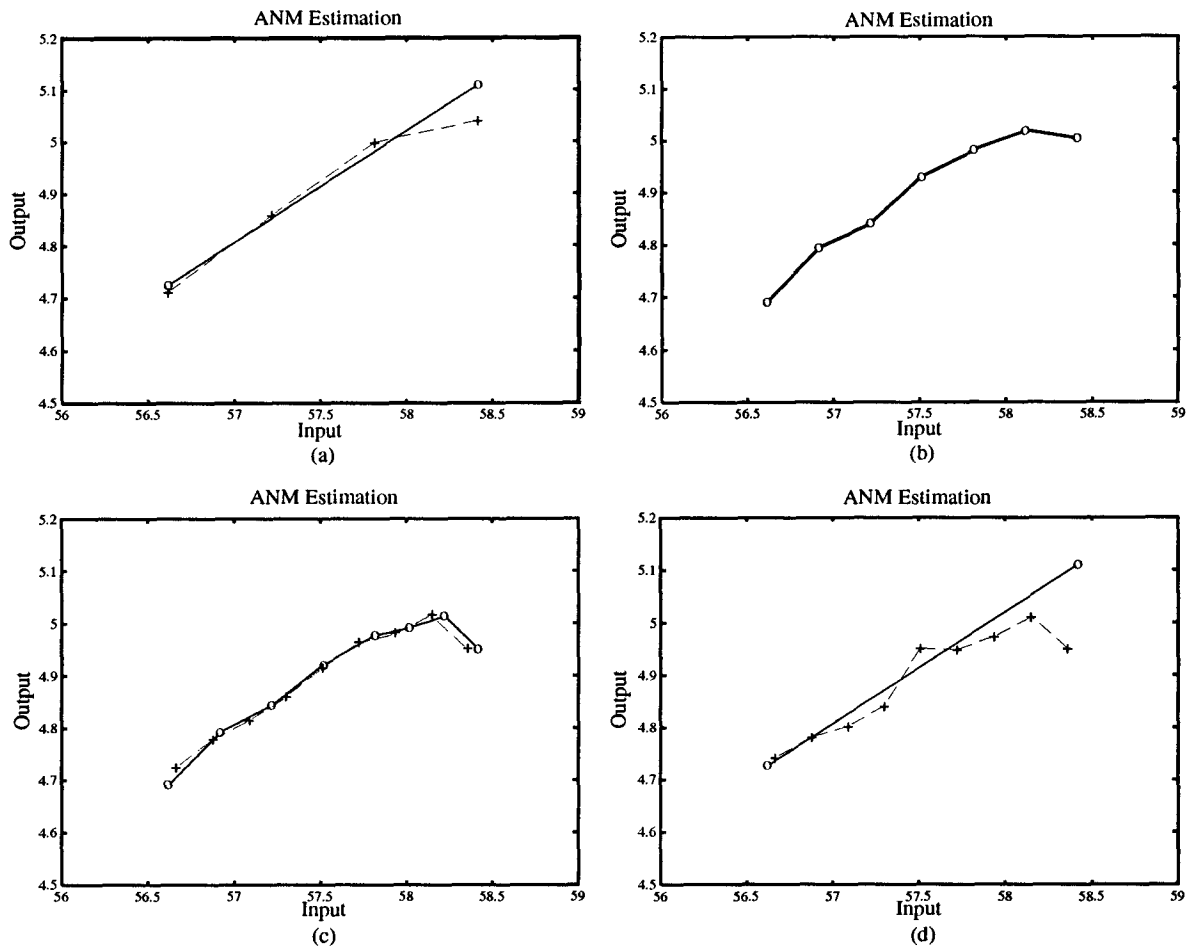
Figure 7.87: Various results from the ANM estimator

7.3 Kamyr Digester Chip Level

A digester continuously cooks wood chips, the first step on their road to becoming used in paper products. The properties of the cooked chips depend on the extent of the reaction in the digester. The most direct effect on this reaction is the residence time of the chips in the digester and the most common method of controlling this time is by manipulating the chip level. This control problem was examined by Allison [1] and some of the data collected during this trial is examined. The chip level was controlled by manipulating the blow flow using an Adaptive Generalized Predictive Control (AGPC) method. The blow flow is the rate at which cooked pulp is released from the digester. Chip level control is very difficult because it is a time-varying and operating point sensitive process. This is because the chip level is difficult to measure and typically the strain gauges used result in
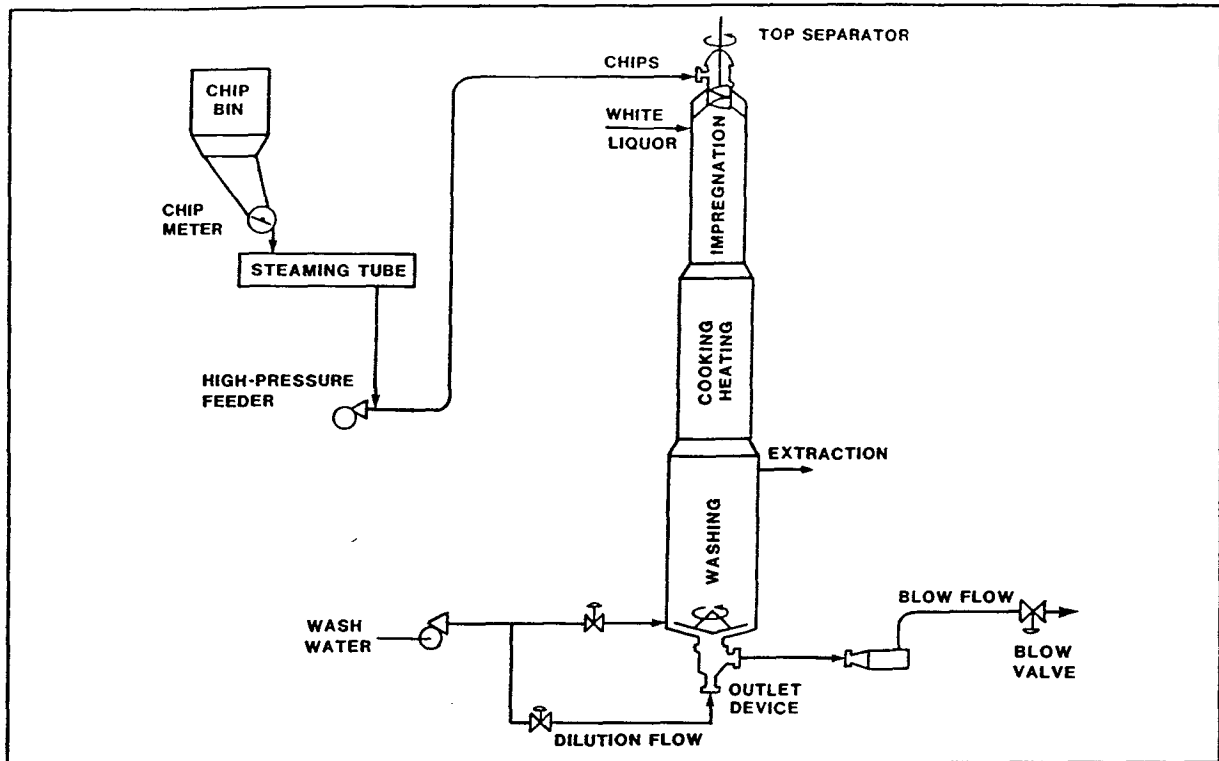
Figure 7.88: A typical Kamyr digester

measurements high in noise and of poor resolution. Also, the chip packing in the digester depends on the size, density, species and moisture content of the chips which will change on a regular basis. A typical Kamyr digester appears in Figure 7.88, showing the flow through the digester. Also available is data collected when the digester was under P control and Constrained Minimum Variance control. All the data was collected every five minutes from the process.

As mentioned, the data was collected during periods when different control schemes were being used. The various schemes will be compared using the ratio of output variance to $\hat{\sigma}^2_{muv}$. Also, one data set contains approximately three days of digester operation under AGPC. These data will be split into four sections to compare the operation of the process over the different intervals.

The chip level for the various schemes appears in Figure 7.89, with only the output shown, so that it can be compared to the reference and the variance of the input signal. Thus it is possible to see how close the output is to the reference and how active the input was to achieve the results. Figure 7.89(a) presents thirty-three hours of operation under P control where the large dip at sample 140, corresponds
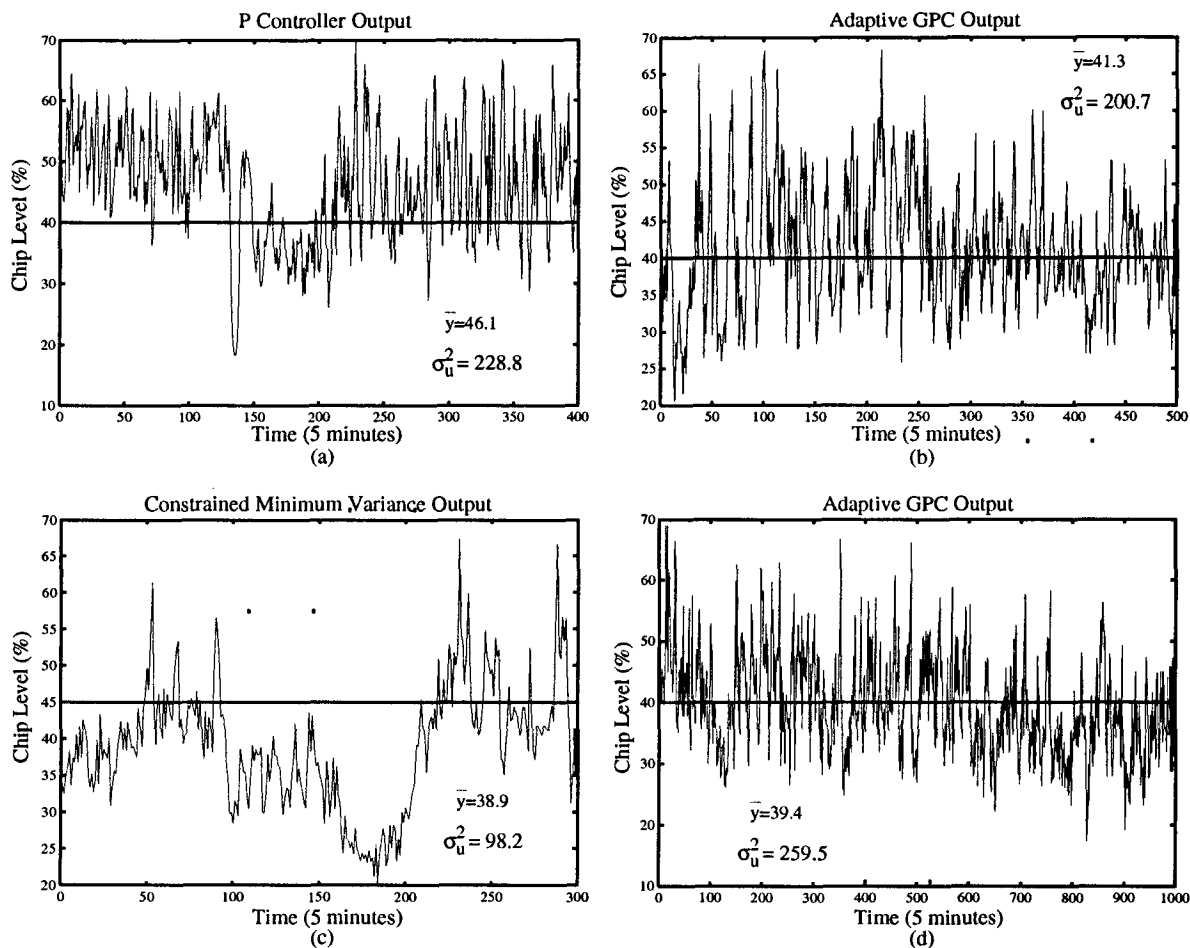
Figure 7.89: The output signals of the digester for (a)
P control, (b) AGPC, (c) CMV control and (d) more AGPC

to a short manual period in an attempt to bring the output closer to the reference. This was typical of the control used on the digester, as a PI controller did not provide satisfactory performance and the operator was required to provide the integral action. The operation was then switched to AGPC control shown in Figure 7.89(b), which lasted for approximately forty hours. The mean of the output is much closer to the reference, indicating better control. Twenty-five hours of control under CMV control is shown in Figure 7.89(c), which exhibits the worst control. Finally, Figure 7.89(d) presents some more AGPC control at a later date, about eighty-three hours of operation. The autocorrelation function of these signals appears in Figure 7.90, confirming the above observations, that the AGPC scheme appears to provide the best control and CMV controller seems to provide the worst control.

106

P Controlled Output Autocorrelation Function

$\sigma_y^2 = 88.9$

$\sigma_s^2 = 125.4$

(a)

Adaptive GPC Output Autocorrelation Function

$\sigma_y^2 = 75.6$

$\sigma_s^2 = 76.2$

(b)

Constrained MV Output Autocorrelation Function

$\sigma_y^2 = 70.2$

$\sigma_s^2 = 107.0$

(c)

Adaptive GPC Output Autocorrelation Function

$\sigma_y^2 = 66.0$

$\sigma_s^2 = 66.2$

(d)
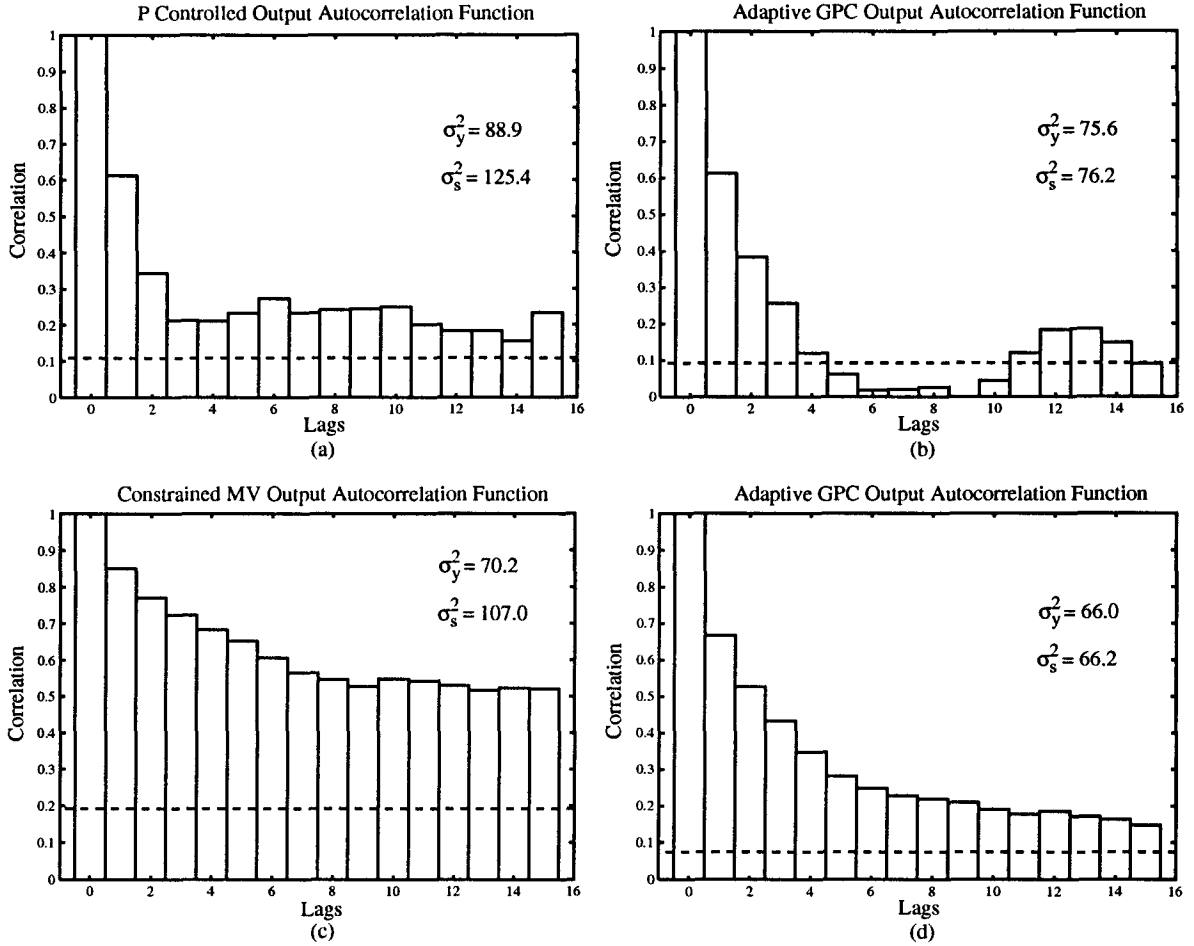
Figure 7.90: The autocorrelation functions of the output signals of the Kamyr digester for (a) P control, (b) AGPC, (c) CMV control and (d) more AGPC

It should also be noticed that two variances are listed in Figure 7.90, σ_y^2, the variance of the output about its mean, while σ_s^2 is the variance of the output about the reference, that being the mean square value of the control error. The reason σ_s^2 must be considered can be explained by examining Figure 7.90(c). While σ_y^2 for the CMV control is only around seventy, which is close to seventy-six, the variance of the AGPC in Figure 7.90(b), the output response shows that the AGPC is superior to the CMV. Comparison of the σ_s^2 values for these processes, one-hundred and six to seventy-six, gives a better indication of the superior control strategy. This approach penalizes the performance of a control which may keep the process variance down, but does not stay near the output. The σ_s^2 value is quickly calculated using:

$$\sigma_s^2 = \sigma_y^2 + (\bar{y} - y_s)^2 \qquad (7.135)$$

where $\bar{y}$ is the mean of the output and y_s is the reference value.

Also, an apparent contradiction exists if Figure 7.90(b) and Figure7.90(d) are compared, as Figure 7.90(d) has a lower variance, yet the autocorrelation function dies out quicker in Figure 7.90(b). Based on this information, an inconsistency seems to exist, but it will be explained after the minimum achievable output variance of the process is calculated.

Unfortunately, it was not possible to estimate the delay of this process. The noise created a bias at one, but even if this was ignored, the data gave an inconsistent delay estimate, ranging between two and four. While this was in the range of likely values [1], there is no guarantee that the estimated change was actually a change in the delay and not caused by the disturbance. As a result, the delay suggested in [1] as occurring most often will be used to determine $\hat{\sigma}^2_{mv}$, that being two samples.

The data were manipulated using various filter numbers and time scales and a choice of ten filters and a time scale of 0.2 was settled on, as it produced an estimated noise autocorrelation function that was closest to an impulse. The results are listed in Table 7.5 and confirm that the AGPC provides the best control, while the CMV control does not perform near as well. The ratio for the two instants of AGPC control are very similar, which suggests that the σ^2_s is lower for the second case because the process has changed slightly, not that the controller is doing a better job. The slightly lower ratio of 1.32 for the first instance of AGPC control could also explain the decrease in correlation in lags five through ten apparent in its autocorrelation function.

A closer look at the data in Table 7.5 confirms the nonstationary behavior of the digester. Even though the AGPC and gain control methods share the same reference, variation in the wood chips entering the digester are evident by examining $\hat{\sigma}^2_{\varepsilon}$ and the $F(q^{-1})$ coefficients. The change in operating point, which occurs during the CMV control, appears to have a large effect on the properties of the

Controller	$\sigma^2_s/\hat{\sigma}^2_{mv}$	σ^2_s	$\hat{\sigma}^2_{mv}$	$\hat{\sigma}^2_{\varepsilon}$	$\hat{F}(q^{-1})$coefficient	$\sigma^2_y/\hat{\sigma}^2_{mv}$
P	1.78	125.5	70.8	49.7	0.652	1.26
GPC(a)	1.32	76.2	58.0	43.8	0.568	1.29
CMV	4.15	107	25.74	16.7	0.737	2.73
GPC(b)	1.4	66.2	47.4	37.2	0.524	1.39

Table 7.5 The results of minimum achievable output variance estimation

digester as the estimated noise standard deviation drops to four from seven. Comparing these figures to knowledge available during operation of the digester, such as the type of chips used, may enable improvements in the process performance to be made.

Next, the output of Figure 7.89(d) was broken into four equal segments, each representing about twenty-one hours of digester operation. The purpose of this is to show how monitoring σ_{mv}^2 can be used to indicate changes in the process. Once these changes are associated with a cause, steps can be taken to minimize their effect on performance.

The control performance, $\sigma_s^2/\hat{\sigma}_{mv}^2$ was calculated using ten Laguerre filters and a time scale of 0.2. The results appear in Table 7.6. The control performance ratio improves between period one and two, stays approximately the same for period three, but worsens in period four. Comparison of these ratios to the variance performance ratio, $\sigma_y^2/\hat{\sigma}_{mv}^2$ shows that the this ratio stays approximately the same, which is encouraging as it suggests that the AGPC enjoys the same success in minimizing the variance over the three days. However, the control fails to keep the output close to the reference all the time, and this drift results in an inflated control ratio in period four.

Some system identification was performed in order to obtain a time-series model of the digester [1]. One day of data was used to fit the model with the digester operating at a reference of forty-five percent. The following model was decided upon:

$$y(t) = \frac{-0.076}{1-0.91q^{-1}}u(t-2) + \frac{1-0.5q^{-1}}{\nabla}e(t)$$

$$\sigma_e^2 = 31 \qquad F(q^{-1}) = 1 + 0.5q^{-1}$$

(7.136)

These numbers are similar to those estimated in Table 7.5 and Table 7.6, while the time-varying and operating point dependency of the digester found are to be expected, based on the operation of the digester.

Period	$\sigma_s^2/\hat{\sigma}_{mv}^2$	σ_s^2	$\hat{\sigma}_{mv}^2$	$\hat{\sigma}_e^2$	$\hat{F}(q^{-1})$coefficient	$\sigma_y^2/\hat{\sigma}_{mv}^2$
1-250	1.57	77.7	49.4	36.6	0.591	1.43
251-500	1.37	61.3	44.8	34.3	0.552	1.35
501-750	1.42	54.7	38.2	28.3	0.589	1.41
751-100	2.06	72.1	35.0	28.2	0.4912	1.43

Table 7.6 The results of minimum achievable output variance estimation on the AGPC control

Using the methods presented in earlier chapters and applying the conclusions reached in the simulation study, it was possible to successfully monitor some industrial processes. Results of monitoring the power spectral density function indicated that the performance of the PI controller is two and a half times worse than is possible, though retuning it may result in better performance. The application of the ANM to the motor load control of a reject refiner resulted in the estimation of its static gain. The knowledge of the slope change is very useful and could be used either in control or in choosing the motor load reference. The performance of various control algorithms, used to regulate the chip level in a Kamyr digester, was determined using the monitoring tools. They indicated which performed the best and this was consistent with what was suggested by the autocorrelation function of the outputs. The time-varying characteristic of this process was also noted, as the process properties changed quite often. These examples point out the benefits of monitoring and the extra insight gained about the process.

Chapter 8

Conclusion

The important role that control loop monitoring plays in maintaining high process performance can not be understated. The monitoring methods presented in this thesis are concerned with the regulation control problem so that the output variance can be kept as small as possible. The ideas and results will be summarized in the order that they were presented in this thesis.

It is possible to estimate the delay of a stochastic process in regulation by using the Fixed Model Variable Regressor Estimator. Successful delay identification was obtained for a wide range of processes and knowledge of the time delay can be used to guarantee that the control method correctly compensates for it. Also, changes in delay over time will often signal the start of deterioration of the process operation. The delay estimate can sometimes be biased, which will be caused by the disturbance or if the input signal lacks sufficient excitation. It is possible to correct the bias caused by the disturbance, but the bias due to insufficient excitation can not be fixed. However an approximation of the lower limit of the input signal frequency content can be easily found with only the knowledge of the sampling interval. The insufficient excitation problem is most often encountered when PI controllers are used for control.

The Adaptive Nonlinear Modeller was successfully applied to closed loop stochastic processes as it was able to estimate their static input-output relationship. The use of Recursive Least Squares estimation to find the table values was not computationally intensive, even if a large number of intervals are used. This is because all but two of the entries in the regressor are zero, which reduces the total number of calculations required. The Adaptive Nonlinear Modeller was able to identify various nonlinearities including a deadzone and actuator saturation, both which will reduce the process performance. Two steps can be employed to improve the estimation of the static input-output relation. The estimator can be turned off for a short period after setpoint changes, which reduces the plant dynamics present in the signals. Also, for stochastic processes, the use of a low pass filter is important to reduce the effect of noise on estimation.

111

The final process property determined was the minimum achievable output variance. The estimation of the minimum achievable output variance involved the successful application of a Laguerre network. The Laguerre gains were correctly estimated for the majority of processes considered, using Recursive Extended Least Squares estimation. This justifies the use of this estimator over other, more complicated ones, even though in some situations estimation may not converge. This may occur if the closed loop noise transfer function numerator is not strictly positive real. Simulations showed a wide range of Laguerre parameters can be used to obtain high estimation accuracy. If the autocorrelation function of the estimated noise sequence is close to an impulse, then the corresponding choice of Laguerre parameters can be used with a high degree of confidence. It was also found that examination of the estimated noise sequence and Markov coefficients, required to determine the minimum achievable output variance, can be used to indicate if the plant or disturbance of the process change. Besides being able to use a large range of Laguerre network parameters, a single choice of parameters were able to model numerous changes in the closed loop noise filter. It should always be remembered that changing control will not change the minimum achievable output variance. The minimum achievable output variance will only change with the noise properties of the process.

The last stage in this thesis involved applying the monitoring tools to data from industrial processes. This stage moved the study from simulation to the real world and the results were successful. Much useful information about the processes considered was provided by monitoring. The results show how monitoring can be used to compare various control algorithms, by determining which provides the best reduction in the output variance. Also, once the minimum variance level is known, decisions can be made as to the best approach to try in order that the process output variance can be reduced. If the minimum achievable output variance is not low enough, further reduction can be made by either retuning or replacing the control algorithm or by modifying the process. Also, because changes in the process plant and disturbance are detected during estimation, this knowledge can be used to improve process performance. Lastly, this industrial data shows that PI controllers and high noise in processes can cause problems in the estimation of the time delay. The application of monitoring to industrial processes will result in improved process performance.

There are several possible avenues envisioned that would continue this work. A very useful

exercise would involve the transportation of the monitoring tools to a distributed control system to enable the monitoring of an industrial process for a long period of time, such as days or tens of thousands of samples. This would confirm the benefits derived from monitoring. Also, the minimum achievable output variance could be determined for other control structures, including the case where feedforward control is used. It was found that the number of filters in the Laguerre network had to be increased to maintain good estimation for closed loop noise transfer functions with complex poles. It is possible to use a Laguerre network which employs a complex time scale [24], which should be done if the extra computation required by the network size is a concern. It would also be possible to use the estimated process properties with a knowledge based system. Decisions could be made based on the results obtained from monitoring the process. Finally, attention could be focused on the possibility of monitoring more process properties. It would be useful to have an idea of the servo performance of the process and to use this in conjunction with the variance performance to obtain a controller that satisfies both criteria. This would be very helpful when applied to cascade loops. The inner loop in this configuration generally performs a servo role while the outer loop acts as a regulator. The availability of an indicator of the overall performance of the cascade loop would be very beneficial.

References

[1] B. J. Allison, G. A. Dumont, L. H. Novak, and W. J. Cheetham. Adaptive-Predictive Control of Kamyr Digester Chip Level. *AIChE Journal*, 36(7):1075–1086, July, 1990.

[2] K. J. Åström. *Introduction to Stochastic Control Theory*. Academic Press, Inc., New York, New York, 1970.

[3] K. J. Åström. The Adaptive Nonlinear Modeller. Technical Report TFRT-3178, Department of Automatic Control Lund Institute of Technology, 1985.

[4] K. J. Åström. Assessment of Achievable Performance of Simple Feedback Loops. *International Journal of Adaptive Control and Signal Processing*, 5(1):3–19, 1991.

[5] K. J. Åström and B. Wittenmark. *Computer Controlled Systems*. Prentice–Hall, New Jersey, 1984.

[6] K. J. Åström and B. Wittenmark. *Adaptive Control*. Addison-Wesley Publishing Company, Don Mills, Ontario, 1989.

[7] W. L. Bialkowski. Bleach Plant Process Control. In *Tappi Bleach Plant Operations Short Course*, Hilton Head, S.C., June, 1990.

[8] W. L. Bialkowski. Dreams vs. Reality: A View from Both Sides of the Gap. In *Control Systems '92: Dream vs. Reality*, Whistler, B.C., September, 1992.

[9] G. Box and G. Jenkins. *Time Series Analysis*. Holden-Day, San Francisco, California, 1976.

[10] D. W. Clarke, C. Mohtadi, and P. S. Tuffs. Generalized Predictive Control Parts I and II. *Automatica*, 23(2):137–160, 1987.

[11] G. A. Dumont and K. J. Åström. Wood Chip Refiner Control. *IEEE Control Systems Magazine*, pages 38–43, April, 1988.

[12] A. Elnaggar. *Variable Regression Estimation of Unknown System Delay*. PhD thesis, University of British Columbia, 1990.

[13] A. Elnaggar. Direct Delay Estimation in Stochastic Closed Loop Systems. Submitted to 12th World Congress IFAC for publication, 1992.

[14] G. Ferretti, C. Maffezzoni, and R. Scattolini. Recursive Estimation of Time Delay in Sampled Systems. *Automatica*, 27(4):653–661, 1991.

[15] Y. Fu and G. A. Dumont. Optimum Laguerre Time Scale and Its On-line Estimation. *IEEE Transactions on Automatic Control*, publication February 1993.

[16] S. Gunnarsson and B. Wahlberg. Some Asymptotic Results in Recursive Identification Using Laguerre Models. *International Journal of Adaptive Control and Signal Processing*, 5:313–333, 1991.

[17] T. J. Harris. Assessment of Control Loop Performance. *The Canadian Journal of Chemical Engineering*, 67:857–862, October, 1988.

[18] T. J. Harris and J. F. MacGregor. An Overview of Discrete Stochastic Controllers:Generalized PID Algorithms with Dead-Time Compensation. *The Canadian Journal of Chemical Engineering*, 60:425–432, June, 1982.

[19] R. Isermann. *Digital Control Systems*. Springer–Verlag, Berlin, Germany, 1981.

[20] L. Ljung. *System Identification*. Prentice–Hall, New Jersey, 1987.

[21] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, Massachusetts, 1983.

[22] J. P. Norton. *An Introduction to Identification*. Academic Press, Inc., New York, New York, 1986.

[23] M. Perrier and A. A. Roche. Towards Mill-Wide Evaluation of Control Loop Performance. In *Control Systems '92: Dream vs. Reality*, Whistler, B.C., September, 1992.

[24] B. Wahlberg. Identification of Resonant Systems Using Kautz Filters. In *Proceedings of the 30th IEEE Conference on Decision and Control*, Brighton, England, December, 1991.

[25] C. C. Zervos. *Adaptive Control Based on Orthonormal Series Representation*. PhD thesis, University of British Columbia, 1988.

[26] C. C. Zervos and G. A. Dumont. Deterministic Adaptive Control based on Laguerre Series Representation. *International Journal of Control*, 48(6):2333–2359, 1988.