

# Variance Analysis For Nonlinear Systems

by

Wei Yu

A thesis submitted to the Department of  
Chemical Engineering in conformity with the requirements for  
the degree of Doctor of Philosophy

Queen's University  
Kingston, Ontario, Canada  
October, 2007

Copyright © Wei Yu, 2007

# Abstract

In the past decades there has been considerable commercial and academic interest in methods for monitoring control system performance for linear systems. Far less has been written on control system performance for nonlinear dynamic / stochastic systems. This thesis presents research results on three control performance monitoring topics for the nonlinear systems:

1. Controller assessment of a class of nonlinear systems: The use of autoregressive moving average (ARMA) models to assess the control loop performance for linear systems is well known. Classes of nonlinear dynamic / stochastic systems for which a similar result can be obtained are established for SISO discrete systems. For these systems, the performance lower bounds can be estimated from closed-loop routine operating data using nonlinear autoregressive moving average with exogenous inputs (NARMAX) models.
2. Variance decomposition of nonlinear systems / time series: We develop a variance decomposition approach to quantify the effects of different sources of disturbances on the nonlinear dynamic / stochastic systems. A method, called ANOVA-like decomposition, is employed to achieve this variance decomposition. Modifications of ANOVA-like decomposition are proposed so that the ANOVA-like decomposition can be used to deal with the time dependency and the initial condition.
3. Parameter uncertainty effects on the variance decomposition: For the variance

decomposition in the second part, the model parameters are assumed to be exactly known. However, parameters of empirical or mechanistic models are uncertain. The uncertainties associated with parameters should be included when the model is used for variance analysis. General solutions of the parameter uncertainty effects on the variance decomposition for the general nonlinear systems are proposed. Analytical solutions of the parameter uncertainty effects on the variance decomposition are provided for models with linear parameters.

# Co-authorship

The material in Chapters 3, 4 and 5 is based on original research carried out independently by myself under the supervision of Dr. T.J. Harris. The idea for the controller assessment of a class of nonlinear stochastic systems in Chapter 3 was based on an article co-authored by Dr. T.J. Harris and myself (Harris and Yu, 2007). Preliminary results from Chapter 4 and 5 were presented at the CScE conference in Calgary, AB, Canada in October 2004.

# Acknowledge

I would like to dedicate the dissertation to my beloved wife, Jinpeng Wang. She and her feeble shoulders carried all the burdens and hardships for our family, so that I had to bear none. Her greatness is beyond words and no words of gratitude are worthy of her.

I would also like to express my deepest gratitude to my academic advisor. Dr. T.J. Harris, who provided me with continuous support and guidance throughout my doctoral career at Queen's university. His attitude of seriousness toward teaching and research and his sincerity greatly affected me as a graduate student. He is my life-long mentor.

Finally, I would like to express my sincere appreciation to Dr. J. McLellan and Dr. D.J. Thomson for serving on my advisor committee. If it were not of the help and suggestion from them, I would not have been able to attain the level of quality required of a good dissertation work.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Co-authorship</b>	<b>iii</b>
<b>Acknowledge</b>	<b>iv</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Symbols</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background Information and Literature Review</b>	<b>5</b>
2.1 Control Performance Assessment . . . . .	5
2.2 Variance Decomposition for Nonlinear Systems . . . . .	10
2.3 Parameter Uncertainty Effects on the Variance Decomposition / Sensitivity Analysis . . . . .	17
<b>3 Controller Assessment of a Class of Nonlinear Stochastic Systems</b>	<b>22</b>
3.1 Introduction . . . . .	22
3.2 Process Description . . . . .	26
3.3 Minimum Variance Performance for Linear Systems . . . . .	28
3.3.1 Minimum Variance Control / Performance Bounds . . . . .	28
3.3.2 Estimation of the Minimum Variance Performance Bounds . . . . .	31
3.4 Extension to a class of Nonlinear Systems . . . . .	32
3.4.1 Nonlinear Minimum Variance Control . . . . .	33
3.4.2 Minimum Variance Lower Bounds . . . . .	35
3.4.3 Extension to General Nonlinear State Space Models . . . . .	41
3.5 Estimation of Lower Bound from Operating Data using Polynomial Approximation . . . . .	43
3.5.1 Polynomial Approximation . . . . .	43
3.5.2 Polynomial AR Model Identification using Orthogonal Least Squares Methods . . . . .	46
3.6 Simulation Results . . . . .	49

3.6.1	Example 1: An Artificial Volterra Model . . . . .	50
3.6.2	Example 2: An Empirical NARMAX Model . . . . .	56
3.6.3	Example 3: A CSTR Model . . . . .	63
3.7	Conclusions . . . . .	71
<b>4</b>	<b>Variance Decomposition for Nonlinear Stochastic Systems / Time Series</b>	<b>73</b>
4.1	Introduction . . . . .	73
4.2	Problem Description . . . . .	76
4.3	Variance Decomposition of Nonlinear ARMA Models . . . . .	79
4.3.1	Variance Decomposition . . . . .	79
4.3.2	Estimation Methods for $Var_{A_t}[Y_t I_0]$ . . . . .	84
4.3.3	Estimation Methods for $E_{I_0}[Var_{A_t}[Y_t I_0]]$ . . . . .	93
4.3.4	Estimation of the Analysis of Variance from Input/Output Data	93
4.4	Simulations . . . . .	95
4.4.1	A Random Coefficient Autoregressive Model . . . . .	95
4.4.2	A Volterra Model . . . . .	101
4.5	Conclusions . . . . .	108
<b>5</b>	<b>Parameter Uncertainty Effects on Sensitivity Analysis / Variance Decomposition</b>	<b>110</b>
5.1	Introduction . . . . .	110
5.2	Variance-based Sensitivity Analysis . . . . .	114
5.2.1	A Special Case . . . . .	119
5.3	Parameter Uncertainty and Sensitivity Analysis . . . . .	120
5.3.1	Parameter Uncertainty and Sensitivity Indices . . . . .	120
5.3.2	Linear Models in the Parameters . . . . .	123
5.4	Examples . . . . .	130
5.4.1	A Linear Model with Uniformly Distributed Regressors . . . . .	130
5.4.2	A Linear Model with Non-uniformly Distributed Regressors (Normal Distributed Regressors) . . . . .	134
5.5	Conclusions . . . . .	137
<b>6</b>	<b>Conclusions</b>	<b>139</b>
6.1	Thesis Contributions . . . . .	140
6.1.1	Controller Assessment of a class of Nonlinear Stochastic Systems	140
6.1.2	Variance Decomposition of Nonlinear Stochastic Systems . . . . .	141
6.1.3	Parameter Uncertainty Effects on Sensitivity Analysis . . . . .	141
6.2	Future Research Topics . . . . .	142
	<b>Bibliography</b>	<b>145</b>
	<b>Appendices</b>	<b>155</b>

# List of Tables

3.1	Example 1: Initial Candidate Terms for Models . . . . .	53
3.2	Example 1: Estimates of $\sigma_{MV}^2$ using different models . . . . .	54
3.3	Nomenclature for exothermic CSTR . . . . .	57
3.4	Example 2: Estimates of $\sigma_{MV}^2$ using different models for both AR(1) and AR(2) disturbance cases . . . . .	62
3.5	Means and variances of $D_t$ , $U_t$ , and $Y_t$ from both empirical and real models . . . . .	69
3.6	Example 3: Estimates of $\sigma_{MV}^2$ using different models for the AR(2) disturbance case . . . . .	69
3.7	Example 3: Estimates of $\sigma_{MV}^2$ (converted to normalized values) using different models for the AR(2) disturbance . . . . .	70
4.1	Sets of Frequencies obtained by using the automated algorithm . . . . .	98
4.2	Example 2: Estimates of partial / total variance using MC and FAST methods with a constant initial condition . . . . .	104
4.3	Example 2: Estimates of partial / total variance using MC and FAST methods with the uncertain initial condition . . . . .	106
4.4	Example 2: Estimates of partial / total variance of the constant initial conditions and the uncertain initial condition for the different time horizons . . . . .	108
5.1	CCPV, PUPV, CCSI and PUSI for a nonlinear model with uniformly distributed regressors (parameter mean-to-variance ratio is 5, $\sigma^2 = \pi^2/3$ )	132
5.2	CCPV, PUPV, CCSI and PUSI for a nonlinear model with normally distributed regressors . . . . .	135



# List of Figures

3.1	Example 1: The stochastic realizations of $D_t$ , $U_t$ and $Y_t$ . . . . .	51
3.2	Example 1: Open loop and closed-loop responses (PI controller) . . .	52
3.3	Example 1: Closed-loop simulation for a disturbance impulse change $\pm 0.2$ with the PI controller . . . . .	52
3.4	Example 1: Comparative box plots of the quality estimates for three models . . . . .	54
3.5	Example 2: The stochastic realization of $D_t$ , $U_t$ , and $Y_t$ for AR(1) disturbance . . . . .	60
3.6	Example 2: The stochastic realization of $D_t$ , $U_t$ , and $Y_t$ for AR(2) disturbance . . . . .	61
3.7	Example 2: Open loop and closed-loop step responses and closed-loop responses for disturbance impulse changes . . . . .	61
3.8	Example 2: Comparative box plots of the quality estimates from three models . . . . .	62
3.9	Example 3: Open loop step change response . . . . .	65
3.10	Example 3: Closed-loop step change response . . . . .	65
3.11	Example 3: Closed-loop simulation for a disturbance impulse change .	66
3.12	Example 3: The stochastic realization of $D_t$ , $U_t$ , and $Y_t$ for the AR(1) model . . . . .	67
3.13	Example 3: The normalized stochastic realization of $D_t$ , $U_t$ , and $Y_t$ for the AR(1) model . . . . .	67
3.14	Example 3: The stochastic realization of $D_t$ , $U_t$ , and $Y_t$ for the AR(2) model . . . . .	68
3.15	Example 3: The normalized stochastic realization of $D_t$ , $U_t$ , and $Y_t$ for the AR(2) model . . . . .	68
3.16	Example 3: Comparative box plots of the quality estimates from three models . . . . .	70
4.1	Plot of the transformation in Eq. (4.45): (a) Sampling point (b) its empirical distribution . . . . .	92
4.2	Scatterplots of sampling points in a two-factor case: (a) transformation in Eq. (4.45) (b) transformation in Eq. (4.47) . . . . .	92
4.3	Auto-correlation, cross-correlation and cross bi-correlation for RCA(1) model . . . . .	97

4.4	Boxplots of 100 estimates of the variance decomposition for RCA(1) model; analytical values of the partial and total variances are shown by dotted lines. . . . .	99
4.5	Boxplots of 100 estimates of the variance decomposition of $Y_t$ , $t = 1, \dots, 15$ for the RCA(1) model ; analytical values of the partial and total variances $t \rightarrow \infty$ are shown by dotted lines. . . . .	100
4.6	A sample of five hundred closed-loop Volterra Model data points with the measured and unmeasured disturbances . . . . .	102
4.7	The box plots of output $Y_t$ , $t = 1, 2, \dots, 40$ . . . . .	103
4.8	The comparative box plots of the quality estimates of partial/total variance using the FAST method . . . . .	105
4.9	The comparative box plots of the quality estimates of partial/total variance with the uncertain initial condition using the FAST method . . . . .	107
5.1	Effect of mean-to-variance ratio on PUSI ( $\sigma^2 = \pi^2/3$ ) . . . . .	133
5.2	Effect of mean-to-variance ratio on PUSI ( $\sigma^2 = \pi^2/9$ ) . . . . .	136
5.3	A boxplot of the estimates of the constant conditional partial variances using FAST for normally distributed regressive variables . . . . .	137

# List of Symbols

## Roman Letters

$\hat{D}_{t+b/t}$  b-step ahead minimum-mean-square-error forecast for the disturbance  $D_{t+b}$

$\hat{Y}_{t+b/t}$  b-step ahead minimum-mean-square-error forecast for the output  $Y_{t+b}$

$a_t$  Stochastic driving force variable

$d$  Integration order

$D_t$  Univariate disturbance variable

$e_{t+b/t}$  Prediction error

$G_p$  Process transfer function

$I_0$  Initial condition

$U_t$  Input or manipulated variable

$X_i$  regressive variable or regressor

$Y_{t+b}^{MV}$  Process output under minimum variance control

$Y_t$  Univariate output variable

$Z_t$  State variables

## Greek Letters

$\delta$  Transfer function denominator coefficient , see equation (2.1)

$\eta(b)$  SISO controller performance index , see equation (2.8)

$\kappa$  Mean-to-variance ratio , see equation (5.47)

$\mu$  True mean of variable

$\omega$  Transfer function numeration coefficient , see equation (2.1)

$\phi$  Univariate model autoregressive polynomial or coefficient

$\Psi$  Impulse response function

$\sigma^2$  True variance of variable

$\sigma_{MV}^2$  True single output minimum output variance

$\Theta$  Parameter vector , see equation (2.13)

$\theta$  Univariate model moving average polynomial or coefficient

### **Operators and Symbols**

$\nabla$  Different operator,  $\nabla = (1 - q^{-1})$

$E(\cdot)$  Expected value of enclosed variable or expression

$erf(\cdot)$  Error function , see equation (4.46)

$f(\cdot)$  Function of enclosed variable , see equation (2.13)

$q^{-1}$  Backshift operator

$V(\cdot)$  Variance of enclosed variable or expression

$Var(\cdot)$  Variance of enclosed variable or expression

### **Abbreviations**

*AIC* Akaike's information criterion

*ANOVA* Analysis of variance

*AR* Autoregressive

*ARMA* Autoregressive moving average

*CCPV* Constant conditional partial variance

*CCSI* Constant conditional sensitivity indices

*CPA* Control performance assessment

*CSTR* Continuous stirred tank reactor

*DOE* Design of experiments

*err* Error reduction ratio

*FAST* Fourier Amplitude Sensitivity Test

*FOS* Fast Orthogonal Search

*GMV* General minimum variance

*iid* Identically independent distributed  
*IRF* Impulse response function  
*LAR* Linear autoregressive  
*MC* Monte Carlo  
*MCMC* Markov Chain Monte Carlo  
*MIMO* Multi-input Multi-output  
*MISO* Multi-input single-output  
*MPC* Model Predictive Control  
*MV* Minimum variance  
*MVLB* Minimum variance lower bound  
*NAR* Nonlinear autoregressive  
*NARMA* Nonlinear autoregressive moving average  
*NARMAX* Nonlinear autoregressive moving average with exogenous inputs  
*OLS* Orthogonal Least Squares  
*PAR* Polynomial autoregressive  
*PARX* Polynomial autoregressive with exogenous inputs  
*pdf* Probability distribution function  
*PID* Proportional-Integral-Derivative  
*PUPV* Parameter uncertainty partial variance  
*PUSI* Parameter uncertainty sensitivity indices  
*RCA* Random coefficient autoregressive  
*RSA* Regressive sensitivity analysis  
*SA* Sensitivity analysis  
*SISO* Single-input single-output  
*VAR* Vector autoregressive

# Chapter 1

## Introduction

Analysis of variance (ANOVA) is a useful tool which helps an analyst to identify sources of variability from one or more potential sources, sometimes referred to as "treatments" or "factors". This method is widely used in industry to help identify the source of potential problems in production processes and to identify whether variation in measured output values is due to variability between various manufacturing processes, or within them.

In traditional analysis of variance applications, the data under study are often assumed to have been sampled from a static system, i.e., a memoryless system. In the past decades, ANOVA methods for linear dynamic systems are discussed in some statistics and econometrics areas (Lütkepohl, 1991), in control performance monitoring and assessment for Single-Input Single-Output (SISO) linear systems (Harris, 1989; Stanfelj *et al.*, 1993; Seppala, 1999) and for Multi-Input Multi-Output (MIMO) linear systems (Harris *et al.*, 1996; Huang *et al.*, 1997; Seppala *et al.*, 2002), in the variance decomposition for SISO linear systems (Desborough and Harris, 1993). Far less has been written on using the ANOVA methods for nonlinear dynamic / stochastic systems.

Chemical or biochemical processes are, in general, highly nonlinear, especially

when operated over a wide range of operating conditions. The nonlinearity is generally related to reaction kinetics or the nonlinearity of physical properties (Bates and Watts, 1988; Pearson and Ogunnaike, 1997). Therefore, there is a strong motivation to develop some methodology for assessing and monitoring these processes. Although dynamic analysis of variance methods for the nonlinear systems are more difficult to use than their static and linear counterparts, the results are very valuable and useful.

The thesis is organized in chapters as follows:

Chapter 1 provides an overview of the work.

Chapter 2 provides a general literature review of the fundamental background to this work. Each chapter provides a briefly literature review specific to its own topic.

Chapter 3 addresses the problem of controller assessment for a class of nonlinear stochastic systems. The use of autoregressive moving average (ARMA) models to assess the control loop performance for processes that are adequately described by the superposition of a linear dynamic model and linear stochastic or deterministic disturbance model is well known. In this chapter, a class of nonlinear dynamic / stochastic systems for which the similar results can be obtained is established for single-input single-output discrete systems.

Nonlinear dynamic / stochastic systems are very difficult to identify and estimate directly since: i) accurate mechanistic models are often difficult to be obtained ii) the full state variables are not available and only external data (i.e., input-output data) are available . Therefore, it is often the only approach that models the nonlinear systems using the external data (Doyle III *et al.*, 2002; Diaz and Desrochers, 1988; Harber and Unbehauen, 1990; Leontaritis and Billings, 1985*a*; Leontaritis and Billings, 1985*b*; Pearson and Ogunnaike, 1997). The nonlinear autoregressive moving average with exogenous inputs (NARMAX) models proposed by Leontaritis and Billings (1985*a*) and (1985*b*) provide a unified representation for a wide class of discrete time nonlinear dynamic / stochastic systems. Therefore, the struc-

ture of the NARMAX models is used to represent the class of nonlinear dynamic / stochastic systems which is studied for the controller assessment.

For these nonlinear systems with an additive disturbance which is in the form of a linear or restricted complexity nonlinear time series, it is shown that a feedback invariant exists and can be theoretically recovered from the closed-loop routine operating data. In this chapter, variance-based bounds on performance for these nonlinear systems are approximately estimated by using nonlinear autoregressive moving average with exogenous inputs (NARMAX) models. It is necessary to know the process time delay. The fitting of these models is greatly facilitated by using efficient algorithms, such as Orthogonal Least Squares (OLS) (Chen *et al.*, 1989) and Fast Orthogonal Search (FOS) (Chon *et al.*, 1997; Korenberg, 1988) algorithms.

In Chapter 4, we develop a variance decomposition approach to quantify the effects of different types of disturbances (dynamic and measurement disturbances) on the nonlinear dynamic systems / time series. This study is motivated by the importance that decompositions of output uncertainty corresponding to each individual input have in enabling companies to achieve goals related to quality, safety and asset utilization. If the process has additional information from measurement of some components of the disturbances, the control engineer would like to use this extra information to improve the process performance such as implementation of feedforward controls and elimination or reduction of the disturbances. But before doing these, the analysis of variance and variance decomposition must be conducted.

A method, called ANOVA-like decomposition (Archer *et al.*, 1997; Saltelli, 2002), is modified to achieve this variance decomposition. Nonlinear autoregressive moving average (NARMA) models are used as the representations of a class of nonlinear dynamic systems for this study. To reduce the computation requirement, the Fourier Amplitude Sensitivity Test (FAST) method is used to estimate the results of variance decomposition. The results of this chapter can be used in investment problems,



biomathematics and control theory where the multivariate disturbances are frequently considered.

In Chapter 5, the development of the parameter uncertainty effects on the results of the variance decomposition discussed in Chapter 4 is undertaken.

The variance decomposition in Chapter 4 is based on the assumption that the model parameters are exactly known. However, parameters of empirical or mechanistic models are most often estimated from data and are thus uncertain. The uncertainties associated with parameters should be included when the model is used for the variance decomposition. In this chapter, a general solution for parameter uncertainty effects on the variance decomposition is provided. If the models are linear in their parameters, the parameter uncertainty effects on the variance decomposition can be directly/analytically deduced for the results of variance decomposition solutions proposed in Chapter 4. Extension of this analytical approach is proposed for models that are nonlinear in the parameters.

The final chapter, Chapter 6, provides a review of the conclusions of the thesis, summarizes the thesis contributions, and recommends some potential topics for further research.

# Chapter 2

## Background Information and Literature Review

The emphasis of this Ph.D. work has been on developing ANOVA methods for nonlinear dynamic / stochastic systems. In this chapter, important background information regarding the types of systems considered in this thesis is provided. Relevant results, terminology, and concepts of control performance monitoring and assessment and variance decomposition are also reviewed. More comprehensive information about control performance assessment and variance decomposition will be found in the following chapters.

### 2.1 Control Performance Assessment

Control performance assessment (CPA) is an important technology to diagnose and maintain operational efficiency of control systems. It is widely applied in the refining, petrochemicals, pulp and paper and the mineral processing industry. Several authors (Qin, 1998; Harris *et al.*, 1999; Huang and Shah, 1999; Jelali, 2006) have published excellent reviews of CPA theoretical and practical issues. CPA methods have formed a basis for many industrial control performance assessment applications (Harris, Sep-

pala, Jofriet and Surgenor, 1996; Kozub, 1996; Thornhill *et al.*, 1999; Huang and Shah, 1999; Hoo *et al.*, 2003). Whereas theoretical and practical aspects of CPA algorithms for the univariate and multivariate linear systems have been discussed quite effectively in these and many other publications, the study of CPA on nonlinear systems has never been discussed.

Although simple statistics such as the mean and variance of manipulated and controlled variables and the percentage of constraints occurrence time can be used as a simple tool for CPA, a comprehensive approach for control performance assessment usually includes the following steps (Harris *et al.*, 1999; Jelali, 2006): i) determination of a benchmark for control performance assessment, ii) detection of the poor performing loop, iii) diagnosis of the underlying causes for the poor performance, iv) suggestions for improvement.

The topic which will be discussed in Chapter 3 falls almost exclusively under step i) determination of a benchmark for control performance assessment. The background information and literature review will mainly focus on this topic.

A common benchmark for CPA is minimum variance control whose underlying principles originate from the work by Åström (1970) and Box and Jenkins (1970). This benchmark yields a performance index which is defined as a ratio of the minimum output variance (or minimum mean square error) and the actual output variance (or actual output mean square error).

**Univariate Control Performance Assessment:** Consider a process which can be represented by a linear transfer function with an additive disturbance:

$$\begin{aligned} Y_t &= G_p(q^{-1})q^{-b}U_t + D_t \\ &= \frac{\omega(q^{-1})}{\delta(q^{-1})}q^{-b}U_t + D_t \end{aligned} \tag{2.1}$$

The first term represents the plant, where  $Y_t$  is the deviation of the process variable from its setpoint (this definition is only valid in the section for the discussions of linear systems),  $U_t$  is the manipulated variable,  $b$  is the number of whole periods of process delay,  $\omega(q^{-1})$  and  $\delta(q^{-1})$  are polynomials in the backshift operator  $q^{-1}$ , defined such that  $q^{-1}Y_t = Y_{t-1}$ .

The second term, the disturbance  $D_t$ , can be take on many forms. It could be deterministic such as randomly occurring step changes or stochastic in nature. The Autoregressive-Integrated-Moving-Average (ARIMA) models can be used to represent may different types of disturbances (Box and Jenkins, 1970; Ljung, 1987). An ARIMA model of order  $(n_\phi, d, n_\theta)$  is of the form:

$$D_t = \frac{\theta(q^{-1})}{\phi(q^{-1})\nabla^d} a_t = \Psi(q^{-1})a_t \quad (2.2)$$

where  $\{a_t\}$  is a white noise sequence with mean zero and constant variance  $\sigma_a^2$ . The operator  $\nabla$  is defined as  $\nabla = (1 - q^{-1})$ . The term  $\nabla$  permits the mean of the disturbance output to change over time so that it can exhibit some nonstationary behavior. Here, we assume that all the polynomials are stable, i.e., all poles and zeros are inside the unit circle. We further assume that  $\delta$ ,  $\theta$  and  $\phi$  are monic.

The minimum variance control first derived by Åström (1970) and Box and Jenkins (1970) is feedback control which achieves minimum output variance. When the model in Eqs. (2.1) and (2.2) are known, a controller can be designed to minimize the variance of output  $Y_t$ . To derive the minimum variance controller, we need to know the  $b$ -step ahead minimum-mean-square-error forecast for the  $Y_{t+b}$ :

$$Y_{t+b} = \frac{\omega(q^{-1})}{\delta(q^{-1})} U_t + D_{t+b}$$

$$\begin{aligned}
&= \frac{\omega(q^{-1})}{\delta(q^{-1})}U_t + \hat{D}_{t+b/t} + e_{t+b/t} \\
&= \hat{Y}_{t+b/t} + e_{t+b/t}
\end{aligned} \tag{2.3}$$

$\hat{D}_{t+b/t}$  and  $\hat{Y}_{t+b/t}$  are the b-step ahead minimum-mean-square-error forecast for the disturbance and  $Y_{t+b}$  respectively and  $e_{t+b/t}$ , the prediction error, is a moving average process of order  $b - 1$  as:

$$e_{t+b/t} = (1 + \psi_1 q^{-1} + \dots + \psi_{b-1} q^{-(b-1)})a_{t+b} \tag{2.4}$$

where the  $\psi$  weights are identical with the first b-1 impulse coefficients of the disturbance transfer function in Eq. (2.2).

The first term  $\hat{Y}_{t+b/t}$  on the right-hand side in Eq. (2.3) depends on data up to time  $t$ , while the second term depends only on data after time  $t$ . Therefore, no matter what controller is used, the two terms are uncorrelated.

The control signal which results in the minimum achievable variance in the output can be obtained by solving the following relation:

$$\frac{\omega(q^{-1})}{\delta(q^{-1})}U_t + \hat{D}_{t+b/t} = 0 \tag{2.5}$$

Therefore, the process output under minimum variance control,  $Y_{t+b}^{MV}$ , will depend on only the most recent  $b$  past disturbances, i.e.,

$$Y_{t+b}^{MV} = e_{t+b/t} \tag{2.6}$$

The key observation is that no matter which feedback controller is used, as long as closed-loop stability is preserved, the prediction error,  $e_{t+k/t}$  is unaffected. Since it does not depend on the manipulated variable over the prediction interval  $k = 1..b$ , it is called feedback invariant. This point is very important in the study of the minimum

variance performance / lower bound. The minimum variance lower bound (MVLB), as measured in the mean square sense, can be written as:

$$\sigma_{MV}^2 = \text{var}\{Y_{t+b}^{MV}\} = (1 + \psi_1^2 + \dots + \psi_{b-1}^2)\sigma_a^2 \quad (2.7)$$

The most attractive feature of CPA using minimum variance control as a benchmark is that the performance lower bounds can be estimated by using only routine closed-loop operating data with *a priori* knowledge of time delay (Harris, 1989). Further work by Desborough and Harris (1992) and Desborough and Harris (1993) proposed the use of a performance index  $\eta(b)$ , which is defined as follows:

$$\eta(b) = 1 - \frac{\sigma_{MV}^2}{\sigma_y^2} \quad (2.8)$$

where  $0 < \eta(b) < 1$ ,  $b$  is the process time delay,  $\sigma_y^2$  is the variance of the controlled variable under assessment, and  $\sigma_{MV}^2$  is the minimum variance lower bound. A performance index close to 1 implies that there is a high potential for reducing the output variance by re-tuning the existing controller or implementing a new control algorithm.

The CPA methodology has been developed and evaluated for the more complex situations such as feedforward/feedback systems (Desborough and Harris, 1993), combining regulation and setpoint tracking (Desborough and Harris, 1992), combining deterministic and stochastic disturbances (Seppala, 1999; Harris *et al.*, 1999; Kozub, 1996) and time-varying systems (Huang, 2002). Due to the success of the univariate CPA work, the CPA has been extended to MIMO linear systems (Harris *et al.*, 1996; Huang *et al.*, 1997).

## 2.2 Variance Decomposition for Nonlinear Systems

Since it is a common situation in industrial plants that one or more prospective disturbances are measured, but not used in an existing control scheme, methods for decomposing and ranking the relative influences of measured disturbances on controlled outputs are valuable for control performance assessment. With the analysis of variance, it is possible to qualify and quantify the significant components of variance according to their sources. If the contribution of the measured disturbance to the variance is relatively small or insignificant then implementing a feedforward controller or reducing the variation of the disturbance is unwarranted. When, the variance contribution is significant, then implementing a feedforward controller or modifying the process may be necessary.

When and where do we need variance decomposition for closed-loop performance assessment?

- Univariate feedback linear system with minimum variance controller: Even if the univariate feedback linear system is implemented by a minimum variance controller, the output variance may still exceed process or product specifications. Desborough and Harris (1993) state that if the minimum variance performance exceeds process or product specification, then reductions in the output variance can only be achieved by i) modifying the process to reduce or eliminate the disturbances; ii) implementing controllers for different feedforward variables; iii) reducing the deadtime.
- Multivariate linear system: Univariate and multivariate performance assessment are conceptually similar, instead of focusing on controller performance assessment, variance decomposition may provide an alternative method for process performance improvement. With investigation of output variance decomposition, reducing or eliminating a certain source of disturbance may significantly

reduce the variations of all outputs.

- Fixed controller structure: Most controllers employ a fixed structure, i.e., a Proportional-Integral-Derivative (PID) controller or Model Predictive Control (MPC). It is of interest to develop performance assessment methods for these widely used systems. For some reasons, i.e., finance expense, output variance reductions are required without changing the controller structure.
- Nonlinear system:
  - Ordinary Differential Equation (ODE): ODEs are widely used to describe the system. Coefficients of ODEs also have their physical meanings which are ideally assumed to be constant. But in the industrial operating situation, their values can be observed as the random variables. Depending model structures, one or several disturbance(s) may dominate the output variance since the nonlinearity. Reducing or eliminating their variations may significantly reduce the output variance.
  - Nonlinear input-output models: many nonlinear systems are far more complex that can be described by ODE's. A general input-output nonlinear model, nonlinear autoregressive moving average with exogenous (NARMAX) (Tong, 1990; Granger and Teräsvirta, 1993; Petersen and James, 1996; Stachurski, 2003; Jazwinski, 1970), is broadly used in nonlinear system identification and control design. Similar to its linear counterpart discussed before, we may have extra information about measured disturbances, variance decomposition will provide very valuable information for nonlinear control performance assessment and process improvement.

In all above circumstances, variance decomposition plays an essential role for process performance assessment and improvement. Variance decomposition, illustrating the contribution of each individual disturbance to the output variance, will provide



the information which can guide engineers to make the decision on the control performance assessment and process improvement.

In the last decade, academic researchers and industrial practitioners interested in variance decomposition have focused on the univariate and multivariate linear systems. Sophisticated methods have only been developed for systems that can be adequately modeled by linear models with additive linear disturbances. The dynamic analysis of variance procedure was proposed by Desborough and Harris (1993) for linear multi-input single-output (MISO) control systems. The procedure allows one to estimate the output variance inflation due to the disturbances. The variance decomposition procedure is dependent on the assumption that the disturbances are statistically independent. For multivariate linear systems, using multivariate impulse response analysis, Seppala *et al.* (2002) developed the variance decomposition for vector autoregressive (VAR) models with correlated disturbances. Far less has been reported on decomposing the variance of nonlinear systems for performance assessment or analysis. This topic is of importance to control specialists and to the large number of process engineers since most of the control systems are nonlinear systems.

In the study of ANOVA of the univariate and multivariate linear systems, the variance decomposition can be archived through the Impulse Response Function (IRF). It is impossible to extend this method into the nonlinear systems since: i) in linear systems, the effect of process disturbances can always be correctly represented as an output disturbance regardless of where they actually appear in the system. This is a consequence of the principle of superposition. It does not hold for the nonlinear dynamic / stochastic systems; and ii) the complex structures of nonlinear systems.

A variance decomposition method for numerical experiments, called variance-based sensitivity analysis (SA), is employed to analyze nonlinear dynamic / stochastic systems. If the output of each model considered can be represented as an analytic

function of input variables, e.g.,  $Y = f(X_1, X_2, \dots, X_p)$ . The importance of the given independent inputs  $X_1, X_2, \dots, X_p$  can be measured via the fractional variance which is defined as the fractional contribution to the output variance due to the uncertainties in inputs. It can be calculated by using the following ANOVA-like decomposition formula for the total output variance  $V(Y)$  of the output  $Y$  (Archer *et al.*, 1997; Cox, 1982):

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12\dots p} \quad (2.9)$$

where

$$\begin{aligned} V_i &= V(E(Y|X_i = x_i^*)) \\ V_{ij} &= V(E(Y|X_i = x_i^*, X_j = x_j^*)) \\ &\quad - V(E(Y|X_i = x_i^*)) - V(E(Y|X_j = x_j^*)) \end{aligned} \quad (2.10)$$

and so on.  $E(Y|X_i = x_i^*)$  denotes the expectation of  $Y$  conditional on  $X_i$  having a fixed value  $x_i^*$ , and  $V$  stands for variance over all the possible values of  $x_i^*$ . Variance-based SA methods are well known in the engineering literature (Cukier *et al.*, 1973; McMay, 1995; Saltelli *et al.*, 1999). They have been used across various disciplines such as weather forecasting, chemical engineering and economics. Readers are referred to Frey and Patil (2002), McMay (1996) and Saltelli and Scott (2000) for recent reviews of these topics and a number of extensions and modifications.

The same decompositions are seen in the design of experiments (DOE) (Box *et al.*, 1978) which implies the decomposition of the response into terms of increasing dimensionality (mean effects, two-way interactions, higher-order interactions). DOE is an important element of the planning of a physical experiment in which the variation in the factors is often moderate and as a result the interaction terms tend to be

small. To date, variance-based SA is the extension of DOE to numerical experiment in which the factors are varied on a wider scale and the interaction effects can be very significant over the main effects. Although there are several differences between physical and numerical experiments, the ANOVA-like decomposition in Eq. (2.9) is based on the same principles of DOE. The detailed information about the links between DOE and SA can be found in the book by Saltelli and Scott (2000).

Evaluation of the decompositions in Eq. (2.9) is most often accomplished through the Monte Carlo (MC) methods. Efficient numerical methods are required high computational costs since the high dimensionality (Rabitz *et al.*, 1999). The computational cheap methods, Fourier Amplitude Sensitivity Test (FAST) (Cukier *et al.*, 1973; Saltelli *et al.*, 1999) and Sobol's (Sobol', 1993; Homma and Saltelli, 1996) (since Sobol' method is only used in Chapter 5, its information will be introduced in the next section) methods have been developed to cope with this dimensionality problem.

#### **FAST Methods:**

- The main feature of the FAST method is a pattern search method that selects points in the input variable space. The classic FAST method is only suitable for some variables with rare probability distribution function (pdf) shapes such as positively skewed pdf and symmetric U shape pdf. With a new transformation function for the uniform distribution variable developed by Saltelli *et al.* (1999), The FAST method can be applied for more general numerical experiments.
- A transformation function is used to convert values of each input to values along a search curve. As part of the transformation, a frequency is specifically assigned for each input. By using Weyl's theorem (Weyl, 1938), the contribution of input, say  $X_i$ , to the total variance can be estimated based on the Fourier coefficients associating with the frequency assigned to  $X_i$ , and higher harmonics of this frequency (Cukier *et al.*, 1973; Cukier *et al.*, 1978). The interaction

contribution between the  $X_i$  and the other inputs can also be determined at the same time.

- The advantages of FAST method are: i) it allows large variations in inputs, ii) the evaluation for one input can be carried out just using a single set of runs (Saltelli and Scott, 2000), iii) it is possible to develop some new transformation functions for some variables with standard distributions. The disadvantages are: i) the resolution of the variance decomposition is reduced, ii) the inputs must be statistically independent.

For nonlinear systems, a useful class of discrete-time input-output model, nonlinear autoregressive moving average with exogenous inputs (NARMAX) (Leontaritis and Billings, 1985*a*; Leontaritis and Billings, 1985*b*), is used for the study. It is shown that several well-known models such as the Volterra, Hammerstein, Wiener and bilinear models are special cases of the NARMAX model (Leontaritis and Billings, 1985*a*). This model with different sources of disturbances  $a_{i,t}$ ,  $i = 1..p$  can be represented as follows:

$$Y_t = f(Y_{t-1}, \dots, Y_{t-n_y}, U_{t-b}, \dots, U_{t-b-n_u}, a_{1,t}, \dots, a_{1,t-n_{a1}}, \dots, a_{p,t}, \dots, a_{p,t-n_{ap}}) \quad (2.11)$$

where  $f$  is a nonlinear function,  $Y_t$  and  $U_t$  are output and input with maximum lag  $n_y$  and  $n_u$  respectively,  $b$  is the process delay, and  $a_{i,t}$ ,  $i = 1..p$  is the  $i$ th disturbance with the lag  $n_{a_i}$ .

Unlike its static counterpart, variance decomposition for nonlinear stochastic / dynamic systems must include the facts of initial conditions and time dependency. ANOVA-like decomposition is modified to cope with the initial conditions and time dependence.

Time dependence adds an extra dimension to dynamic ANOVA-like decomposi-

tion. For example, the variance of a response variable might be decomposed over a time horizon. This requires significant computational capability. Instead of variance decompositions for each time interval, the variance decomposition for the infinite time is very more attractive and useful for the variance analysis since it can provide the general information of variance decomposition. It is more practical to study the output with a certain finite time interval to approximate the behavior of the output in the infinite time interval. Using the simple statistical methods such as boxplot, autocorrelation and cross-correlation, the finite time interval can be determined. In Chapter 4, the more detailed information about the determination of the finite time interval will be provided.

Initial conditions are another problem which we will encounter in variance decomposition of nonlinear systems. We cannot use the ANOVA-like decomposition in Eqs. (4.1) and (4.2) directly since the initial condition must be considered within the variance decomposition. Using the well-known variance decomposition theorem (Parzen, 1962), we can decompose the variance of output  $Y_t$ ,  $t = 1, 2, \dots, n$  as:

$$Var[Y_t] = E_{I_0}[Var_{A_t}[Y_t|I_0]] + Var_{I_0}[E_{A_t}[Y_t|I_0]] \quad (2.12)$$

where  $A_t$  denotes all of disturbances entering the system from time 1 to time  $t$ .  $E_{I_0}[\bullet]$  denotes the expectation of  $[\bullet]$  respect to  $I_0$  and  $Var_{I_0}[\bullet]$  denotes the variance of  $[\bullet]$  respect to  $I_0$ . Based on the first term on the right-hand side of Eq. (2.12), the variance decomposition of nonlinear systems can be conducted using ANOVA-like decomposition.

## 2.3 Parameter Uncertainty Effects on the Variance Decomposition / Sensitivity Analysis

The variance decomposition / variance-based sensitivity analysis, discussed in Chapter 4 is based on the assumption that the model parameters are exactly known. However, parameters of empirical or mechanistic models usually contain some uncertainties (e.g. estimated from data). The uncertainties associated with these estimated parameters should be automatically propagated when the model is used for variance decomposition / variance-based sensitivity analysis.

The variance decomposition discussed in Chapter 4 is conceptually identical with the variance-based Sensitivity Analysis (SA). The topic of parameter uncertainty effects on the variance decomposition in this thesis is not only valid for the variance decomposition in Chapter 4, but also valid for the general regressive sensitivity analysis based on the variance-decomposition. In this thesis, by *regressive sensitivity analysis* (RSA), we mean the effect of variations of the regressors (regressive variables) on the system behavior. Regressive sensitivity is useful in variance decomposition, optimization and performance assessment in various fields such as food sciences (Cliff *et al.*, 1995; Tienungoon *et al.*, 2000), social science (McCarthy *et al.*, 1995), risk analysis (Bowyer and Danson, 2004; Francos *et al.*, 2003), reliability analysis (Hopperstad *et al.*, 1999), chemical engineering performance assessment (Desborough and Harris, 1993) and econometrics (Saisana *et al.*, 2005).

A uniresponse mathematical regression model is used to demonstrate our approach. The extension of our approach to the nonlinear stochastic / dynamic models is straightforward. Consider a system described by a uniresponse mathematical regression model of the form:

$$Y = f(X, \Theta) \tag{2.13}$$

where  $Y$  is an observed value of the response variable,  $X$  is a vector including regres-

sive variables,  $\Theta^T = (\theta_1, \theta_2, \dots, \theta_p)$  is a  $1 \times p$  vector of unknown parameters,  $f$  is the expectation function which may be linear or nonlinear in the parameters. The model output is described by the state or output variables in Eq. (2.13) which we indicate in general as dependent variables changing in time and/or state space. In engineering fields, the regressors  $X$  in Eq. (2.13) include the physicochemical parameters of the model (such as those related to reaction kinetics and thermodynamic equilibria), as well as initial condition and operating conditions of the systems. The physicochemical parameters  $\Theta$  in Eq. (2.13) are measured experimentally or estimated from experiment data and therefore are always subject to uncertainties. Initial and operating conditions may also be uncertain for a variety of reasons.

The full variance-based sensitivity analysis for the model in Eq. (2.13) can be straightforwardly conducted with the consideration of both regressors  $X$  and parameters  $\Theta$  uncertainties. However, the full SA may not provide the satisfactory results for some variance analysis. In some variance analyses, the regressor uncertainty may have higher priority than the parameter uncertainty.

The problem tackled in this chapter includes two objectives: i) the main objective is to investigate the effect of variations of the regressive variables such as  $X$  described in Eq. (5.1) on the system behavior  $Y$  (For the discussion convenience, we define the term *Regressive Sensitivity Analysis* (RSA) to denote the main objective), and ii) the auxiliary objective is to evaluate the effects of the parameter uncertainty on RSA.

A new approach, conditional partial variance, is proposed for this problem. When the model parameters are treated as random variables, then the partial variance can be considered as the variance and partial variance conditioned on the certain values of parameters. Then the parameter uncertainty effects on the partial variance can be calculated using the following formula. According to the well-known variance

decomposition (Parzen, 1962), the variance of  $Y$  can be written as:

$$Var[Y] = E_{\Theta}[Var_X[Y|\Theta = \tilde{\Theta}]] + Var_{\Theta}[E_X[Y|\Theta = \tilde{\Theta}]] \quad (2.14)$$

where  $E_{\Theta}[\bullet]$  denotes the expectation of  $[\bullet]$  respect to  $\Theta$  and  $Var_{\Theta}[\bullet]$  denotes the variance of  $[\bullet]$  respect to  $\Theta$ . The second term in the right-hand side of Eq. (2.14) is the fractional contribution to the output due only to the uncertainties of the parameters. The first term in the right-hand side of Eq. (2.14) is the variance contribution to the output due to the regressors with the parameter uncertainties. The variance decomposition in Eq. (2.14) will play the key role for the investigation of the parameter uncertainty effects on the variance decomposition / sensitivity analysis.

Both variance-based sensitivity analysis methods, FAST and Sobol's methods, will be used for the regressive sensitivity analysis. The FAST method has been introduced in the above section. Here, the Sobol' method will be briefly introduced.

### Sobol' Method:

The main idea of Sobol's approach is to evaluate the decomposition of the function  $f(X)$  (assuming that the parameters are known) into summands of increasing dimensionally (Sobol', 1993), namely,

$$f(X_1, X_2, \dots, X_n) = f_0 + \sum_{i=1}^n f_i(X_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n f_{ij}(X_i, X_j) + \dots + f_{12\dots n}(X_1, X_2, \dots, X_n) \quad (2.15)$$

where the inputs to the function  $Y = f(X) = f(X_1, X_2, \dots, X_n)$  be defined on the n-dimensional unit cube:

$$K^n = \{X : 0 \leq X_i \leq 1, i = 1, 2, \dots, n\}$$

(this definition is not restrictive for any ranges of the input factors since



any bounded range can be transformed into this hypercube domain)

and where  $f_0$  is a constant and the integral of every summand over any of its own variables is zero, i.e.,:

$$\int_0^1 f_{i_1 i_2 \dots i_s}(X_{i_1}, X_{i_2}, \dots, X_{i_s}) dX_{i_k} = 0 \quad 1 \leq k \leq s \quad (2.16)$$

The consequences of Eqs. (2.15) and (2.16) are that all the functions which appear within the summands in Eq. (2.15) are orthogonal, i.e., if  $(i_1, \dots, i_s) \neq (j_1, \dots, j_k)$ , then

$$\int_{K^n} f_{i_1, \dots, i_s} f_{j_1, \dots, j_k} dX = 0 \quad (2.17)$$

It is shown that there exists a unique expansion in Eq. (2.15) for any function  $f(X_1, X_2, \dots, X_n)$  which is integrable in  $K^n$  (Sobol', 1993).

The total variance of  $f(X)$  can be written as:

$$V = \int_{K^n} f^2(X) dX - f_0^2 \quad (2.18)$$

and partial variance from each of the terms in Eq. (2.15) can be computed as:

$$V_{i_1, i_2, \dots, i_s} = \int_0^1 \dots \int_0^1 f_{i_1, i_2, \dots, i_s}^2(X_{i_1}, X_{i_2}, \dots, X_{i_s}) dX_{i_s} \dots X_{i_2} X_{i_1} \quad (2.19)$$

where  $0 \leq i_1 \leq \dots \leq i_s$  and  $s = 1, 2, \dots, n$ . The partial variance  $V_{i_1, i_2, \dots, i_s}$  is the contribution to the total variance from the term  $f_{i_1, i_2, \dots, i_s}$  which reflects the interactions among the input factors  $(X_{i_1}, X_{i_2}, \dots, X_{i_s})$ .

Combining Eqs. (2.17), (2.18) and (2.19), we have

$$V = \sum_{i=1}^n V_i + \sum_{1 \leq i < j \leq n} V_{ij} + \cdots + V_{12 \cdots n} \quad (2.20)$$

The advantages of Sobol' methods are: i) it is possible to obtain the analytical solution for the variance decomposition, ii) the integrals in Eqs. (2.18) and (2.19) can be computed with the same kind of Monte Carlo integral. Hence, the MC estimates of partial and total variances are straightforward (Sobol', 2001). The disadvantage is that it is computationally more expensive in terms of number of model evaluations.

# Chapter 3

## Controller Assessment of a Class of Nonlinear Stochastic Systems

### 3.1 Introduction

The study of minimum variance control for SISO linear systems can be traced in the work by Åström (1970) and Box and Jenkins (1970). Performance indices that provide figures of merit for the performance of a closed-loop have been developed by Harris (1989) and Stanfelj *et al.* (1993). Performance index is the ratio of the best achievable variance to the variance of the controlled variable under assessment. Furthermore, Harris (1989) was first to propose that these performance indices can be estimated directly from the routine operating data by fitting the controlled variable into a ARIMA time series model. A little work will be done in this chapter to extend of Harris' work to a class of SISO nonlinear systems.

In the last decades, researchers and industrial practitioners interested in performance assessment have focused on monitoring and assessment for the univariate linear systems (Harris, 1989; Desborough and Harris, 1992; Desborough and Harris, 1993) and multivariate linear systems (Seppala *et al.*, 2002; Huang *et al.*, 1997; Harris

*et al.*, 1996). Far less has been written on extending the methodologies for performance assessment to nonlinear systems. This topic is of importance to control specialists and the large number of process engineers since most of the control systems are nonlinear systems. However, there are several challenges:

- Complexity of Nonlinear Behavior.

Nonlinear processes can exhibit six general types of behavior (Doyle III *et al.*, 2002; Pearson, 1999):

- Harmonics arising from periodic inputs
- Subharmonics arising from periodic inputs
- Chaotic response to simple inputs
- Input-dependent stability
- Asymmetric response to symmetric inputs
- Steady-State Input and Output Multiplicities

- Non-Equivalent Representations.

It is a standard result (Ljung, 1987), that any time-invariant linear system can be completely characterized by its impulse response, or equivalently by an autoregressive model. Unfortunately, this equivalence cannot be extended to all nonlinear problems (Pearson and Ogunnaike, 1997).

- Disturbance Representation.

For any process, disturbances can enter at any point. For linear systems, these disturbances can always be represented as an additive output disturbance. Since superposition does not exist in nonlinear systems, this representation is not universal. This presents challenges both in modeling and in the determination of the minimum variance performance bound.

- Challenges in Model Determination and Parameter Estimation.

For systems that admit a linear representation, there are well-established methods for obtaining models of the closed-loop system. These methods can be automated. For nonlinear systems there are enormously rich classes of models to be entertained, and the resulting models often have many parameters to be estimated.

In the case of linear systems, the feedback invariant can be easily recovered from a time series such as autoregressive moving average (ARMA) description of the closed-loop system (Harris, 1989). The feedback invariant is then used to estimate the variance of the output that would be achieved if a minimum variance controller were to be implemented. The underlying theory relies on the development of minimum variance controllers, outlined in Åström (1970) and Box and Jenkins (1970), and the existence of a feedback invariant in Harris (1989). It was shown that the minimum variance performance bound for a linear system could be estimated from routine closed-loop data by fitting a time series such as ARIMA models to the outputs and then determining the variance of the  $b$ -step ahead predictor, where  $b$  is the process delay (Harris, 1989).

In recent two decades, the development of non-linear minimum variance (MV) controllers has been proposed by many researchers, many for typical nonlinear systems such as: bilinear systems (Svoronos *et al.*, 1981; Goodwin *et al.*, 1982; Cho and Marcus, 1987) and Hammerstein models (Keviczky *et al.*, 1979). Unfortunately, as many researchers pointed out, these kinds of controllers are not practical since the implementation of this control action will result in damage of plant equipment, excessive expense or operation of the model outside its validity bound. The generalized minimum-variance (GMV) controller (Clarke and Gawthrop, 1975; Grimbale, 2005) is used to overcome this by including weights on the control inputs within the cost function. The development of the GMV controller for the nonlinear processes that

admit a non-linear ARMAX representation can be found in Sales and Billings (1990). For the more general nonlinear plant model with an additive disturbance, the GMV controller is designed by Grimble (2005). This general nonlinear plant model might involve state-space, transfer operators, neural networks or even nonlinear function look-up tables. The disturbance is assumed to have a linear time-invariant form.

In this chapter some preliminary results are developed for determining the minimum variance performance bound for a class of nonlinear systems. It is shown that a minimum-variance-feedback invariant exists for an important class of nonlinear processes that can be described by the superposition of a nonlinear dynamic model and additive linear or partially nonlinear disturbance. In these instances, the minimum variance performance can be estimated from routine operating data. It may be necessary to fit a nonlinear time series model to the closed-loop data using both inputs and outputs. A universal approximation function, polynomial time series, is used to approximate the nonlinear time series. Since the parameters of the polynomial time series are linear, the model can be estimated using linear regression techniques such as orthogonal least squares methods.

The outline of this chapter is as follows: In Section 3.2, a general nonlinear input-output model is introduced. In Section 3.3, a review of linear minimum variance controllers and the estimation of performance lower bounds from routine operating data is provided. This is followed by the development of non-linear minimum variance controllers and a development that shows the existence of a feedback invariant for this system. In Section 3.5, a more detailed description of the polynomial ARMA representation and one general method are outlined for estimating the minimum variance performance bounds. This is followed by three simulation examples to demonstrate the essential features of the method. The chapter is concluded with a description of outstanding issues and limitations of the proposed methodology.

## 3.2 Process Description

The general SISO discrete nonlinear system will be represented as:

$$\begin{aligned} Z_t &= f(Z_{t-1}^*, U_{t-b}^*) \\ Y_t &= Z_t \end{aligned} \tag{3.1}$$

where  $Z_t$  is the deterministic output of the system in response to the inputs that are denoted by  $U_t$ .  $b$  represents the number of whole periods of delay in the system and is the number of sampling intervals that elapse between making a change in the process input and first observing its effect. The notation  $f(Z_{t-1}^*, U_{t-b}^*)$  denotes a function of previous values of  $Z_{t-i}, i = 1..n_z$  and  $U_{t-b-j}, j = 0..n_u$ .

Since in most modeling cases only external (i.e., input-output) data are available, the system described in Eq. (3.1) is usually represented in the Nonlinear AutoRegressive and Moving Average (NARMA) model (Leontaritis and Billings, 1985*a*; Leontaritis and Billings, 1985*b*), namely,

$$Y_t = f_1(Y_{t-1}^*, U_{t-b}^*) \tag{3.2}$$

In reality, the function  $f_1(\cdot)$  is generally very complex and knowledge of the form of this function is often not available. The solution is to approximate  $f_1(\cdot)$  using some known simpler functions, one of which is a polynomial approximation. The function  $f$  is to expand the finite polynomials involving summations of terms  $Z_{t-i}, Z_{t-i}Z_{t-j}, \dots, U_{t-b-i}, U_{t-b-i}U_{t-b-j}, \dots, Z_{t-i}U_{t-b-j} \dots$  (Leontaritis and Billings, 1985*a*; Leontaritis and Billings, 1985*b*; Sales and Billings, 1990). The resulting expansions produce nonlinear difference equations. Hammerstein and Wiener systems are encompassed in this framework (Pearson and Ogunnaike, 1997; Sales and Billings, 1990; Harber and

Unbehauen, 1990).

In this chapter, we are interested in systems that are also affected by disturbance. The most elementary representation is:

$$Y_t = f_2(Y_{t-1}^*, U_{t-1}^*) + a_t \quad (3.3)$$

where  $a_t$  is a white noise with zero mean and constant variance  $\sigma_{a_t}^2$ .

Equation (3.3) provides the simplest stochastic nonlinear system since only additive uncorrelated noise is considered. In reality, the disturbance might include autoregressive and moving average terms or cross-products between the disturbance and the inputs and outputs. The most general form of the nonlinear stochastic system is the NARMAX model (Leontaritis and Billings, 1985*a*; Leontaritis and Billings, 1985*b*):

$$Y_t = f_3(Y_{t-1}^*, U_{t-b}^*, a_{t-1}^*) + a_t \quad (3.4)$$

In linear systems, the effect of process disturbances can always be correctly represented as an output disturbance regardless of where they actually appear in the system. This is a consequence of the principle of superposition. In nonlinear systems, superposition does not hold. It is useful however to provide an additive disturbance, the nonlinear stochastic model studied in this chapter is of the form,

$$Y_t = f_p(Y_{t-b}^*, U_{t-b}^*) + D_t \quad (3.5)$$

where  $D_t$  is the additive disturbance which can be represented by a non-linear ARMA model as:

$$D_t = f_D(D_{t-1}^*, a_{t-1}^*) + a_t \quad (3.6)$$

The final goal of the controller assessment for the nonlinear systems is to find the solution based on the general nonlinear stochastic model described in Eq. (3.4).



Obtaining the final goal is not one day job, it needs the procedures from the easy to the hard. The study of the controller assessment for the nonlinear stochastic model in Eq. (3.5) is the first step. More detailed information about this model structure will be provided in the latter section.

### 3.3 Minimum Variance Performance for Linear Systems

#### 3.3.1 Minimum Variance Control / Performance Bounds

The discrete linear model which is used to describe the process consists of two terms, one which represents the plant dynamics and another which accounts for the disturbances. The output of the process  $Y_t$ , which combines the two terms, is given by:

$$\begin{aligned}
 Y_t &= f(Z_{t-1}^*, U_{t-b}^*) + D_t \\
 &= G_p(q^{-1})q^{-b}U_t + D_t \\
 &= \frac{\omega(q^{-1})}{\delta(q^{-1})}q^{-b}U_t + D_t
 \end{aligned} \tag{3.7}$$

The first term represents the plant, where  $Y_t$  is the deviation of the process variable from its setpoint (this definition is only valid in the section for the discussions of linear systems),  $Z_t$  is the deterministic output of the system in response to the inputs that are denoted by  $U_t$ ,  $b$  is the number of whole periods of process delay,  $\omega(q^{-1})$  and  $\delta(q^{-1})$  are polynomials in the backshift operator  $q^{-1}$ , defined such that  $q^{-1}Y_t = Y_{t-1}$ .

The second term, the disturbance  $D_t$ , can be taken on many forms. It could be deterministic such as randomly occurring step changes or stochastic in nature. The

Autoregressive-Integrated-Moving-Average (ARIMA) models can be used to represent may different types of disturbances (Box and Jenkins, 1970; Ljung, 1987). An ARIMA model of order  $(n_\phi, d, n_\theta)$  is of the form:

$$D_t = \frac{\theta(q^{-1})}{\phi(q^{-1})\nabla^d} a_t = \Psi(q^{-1})a_t \quad (3.8)$$

where  $a_t$ 's are a sequence of white noise variables with mean zero and constant variance  $\sigma_a^2$ . The operator  $\nabla$  is defined as  $\nabla = (1 - q^{-1})$ . The term  $\nabla^d$  permits the mean of the disturbance output to change over time so that it can exhibit some nonstationary behavior. In this section we assume all the polynomials are stable, i.e. all poles and zeros are inside the unit circle. We further assume that  $\delta$ ,  $\theta$  and  $\phi$  are monic.

The minimum variance control first derived by Åström (1970) and Box and Jenkins (1970) is feedback control which achieves minimum output variance. When the model in Eqs. (3.7) and (3.8) are known, a controller can be designed to minimize the variance of output  $Y_t$ . To derive the minimum variance controller, we need to know the b-step ahead minimum-mean-square-error forecast for the  $Y_{t+b}$ :

$$\begin{aligned} Y_{t+b} &= \frac{\omega(q^{-1})}{\delta(q^{-1})} U_t + D_{t+b} \\ &= \frac{\omega(q^{-1})}{\delta(q^{-1})} U_t + \hat{D}_{t+b/t} + e_{t+b/t} \\ &= \hat{Y}_{t+b/t} + e_{t+b/t} \end{aligned} \quad (3.9)$$

$\hat{D}_{t+b/t}$  and  $\hat{Y}_{t+b/t}$  are the b-step ahead minimum-mean-square-error forecast for the disturbance and  $Y_{t+b}$  respectively and  $e_{t+b/t}$ , the prediction error, is a moving average process of order  $b - 1$  as:

$$e_{t+b/t} = (1 + \psi_1 q^{-1} + \dots + \psi_{b-1} q^{-(b-1)}) a_{t+b} \quad (3.10)$$

The  $\varphi$  weights are identical with the first  $b-1$  impulse coefficients of the disturbance transfer function in Eq. (3.8).

The first term  $\hat{Y}_{t+b/t}$  on the right-hand side in Eq. (3.9) depends on data up to time  $t$ , while the second term depends only on data after time  $t$ . Therefore, no matter what controller is used, the two terms are uncorrelated.

The control signal which results in the minimum achievable variance in the output can be obtained by solving the following relation:

$$\frac{\omega(q^{-1})}{\delta(q^{-1})} U_t + \hat{D}_{t+b/t} = 0 \quad (3.11)$$

Therefore, the process output under minimum variance control,  $Y_{t+b}^{MV}$ , will depend on only the most recent  $b$  past disturbances, i.e.,

$$Y_{t+b}^{MV} = e_{t+b/t} \quad (3.12)$$

The key observation is no matter which feedback controller is used, as long as closed-loop stability is preserved, the prediction errors,  $e_{t+k/t}$  are unaffected. Since they do not depend on the manipulated variable over the prediction interval  $k = 1..b$ , it is so called feedback invariants. This point is very important in the study of the minimum variance performance bound. The minimum variance performance bound (MVPB), as measured in the mean square sense, can be written as:

$$\sigma_{MV}^2 = var\{Y_{t+b}^{MV}\} = (1 + \psi_1^2 + \dots + \psi_{b-1}^2) \sigma_a^2 \quad (3.13)$$

### 3.3.2 Estimation of the Minimum Variance Performance Bounds

If the process is controlled by a linear feedback controller, in the case of linear process with additive disturbances, Harris (1989) has shown that the lower bound on performance can be estimated from routine operating data. The theoretical minimum variance lower bound on performance can be estimated by the following methods:

- Indirect Estimation

- Using ARIMA model (Harris, 1989): A time series model of the form

$$\begin{aligned} Y_t &= \frac{\alpha(q^{-1})}{\beta(q^{-1})} a_t \\ &= (1 + \hat{\psi}_1 q^{-1} + \dots + \hat{\psi}_{b-1} q^{-(b-1)} + \dots) a_t \end{aligned} \quad (3.14)$$

is identified from a set of observations of  $Y_t$ . The data can be collected with or without feedback control. The only requirement for the data is that it contains a representative sample of process disturbances. The first  $b-1$  impulse coefficients of  $\frac{\alpha(q^{-1})}{\beta(q^{-1})}$  are estimates of the first  $b-1$  coefficients of the open-loop disturbance transfer function. The estimate of the achievable minimum variance can be calculated from the estimates of the first  $b-1$  coefficients and the residual variance,  $\hat{\sigma}_a^2$ . That is:

$$\hat{\sigma}_{MV}^2 = (1 + \hat{\psi}_1^2 + \dots + \hat{\psi}_{b-1}^2) \hat{\sigma}_a^2 \quad (3.15)$$

- Using Laguerre networks (Lynch and Dumont, 1996): in this method, a Laguerre network is used to model the output instead of the ARIMA models. The detailed algorithm can be found in Lynch and Dumont (1996).

- Direct Estimation (Desborough and Harris, 1992):

fitting an autoregressive (AR) time series model of the form:

$$Y_{t+b} = \xi(q^{-1})Y_t + \varepsilon_{t+b} \quad (3.16)$$

from routine open-loop/closed-loop with this approach the variance of the residual  $\varepsilon(t+b)$  in Eq. (3.16) is an estimate of  $\sigma_{MV}^2$ .

The extension of the performance index to the case of varying set-points has been treated in Ko and Edgar (2000), Seppala *et al.* (2002), Thornhill *et al.* (2003), and McNabb and Qin (2005). Seppala *et al.* (2002) discussed the influence of set-point changes on the performance index and demonstrated the benefits of a decomposition of the control error into the components resulting from set-point changes.

With the above methods, a number of performance indices can be estimated. They can be used at the design stage to compare the performance of different controllers. However, they are often used as part of a monitoring and diagnosis scheme for control system. The methodology has been extended to the SISO non-minimum-phase systems (Tyler and Morari, 1995), time-varying systems (Huang, 2002) and multivariate systems (Harris *et al.*, 1996; Huang *et al.*, 1997). Readers are referred to the paper by Jelali (2006) for a recent review and a number of extensions, modifications and applications.

### **3.4 Extension to a class of Nonlinear Systems**

Control performance assessment for linear systems is a well-developed field, but nonlinear systems have not been studied in this regard since the difficulties discussed in the introduction section. Before we go through the development of CPA for nonlinear systems, let's look at the recent research about nonlinear minimum variance control.

### 3.4.1 Nonlinear Minimum Variance Control

Minimum variance control of linear systems is a well-developed field, but nonlinear systems are yet to attract attention in this regard since the difficulties discussed in the introduction section. However the development of nonlinear minimum variance controllers has been considered by a number of authors. Since the nonlinear minimum variance control action will result in problems such as damage to plant equipment, excessive expense, or operation of the model outside its validity bound. The nonlinear generalized minimum-variance (GMV) controller (Clarke and Gawthrop, 1975; Grimble, 2005) is studied more intensively.

The GMV controller for the nonlinear systems that can be expressed as the NARMAX models has been concerned in Sales and Billings (1990). The general minimum variance controller is designed by a cost function including a weighting on the past control inputs and current output. To solve this cost function, the construction of the b-step ahead prediction is required. The self-tuning or adaptive form is used to deal with the control processes whose parameters are either unknown or slowly time-varying.

The GMV controller for the nonlinear systems that are the superposition of a nonlinear process model plus a linear stochastic model of the form are considered in Bittanti and Piroddi (1993) and Grimble (2005):

$$Y_t = f_{P^*}(U_{t-b}^*, W_t^*) + D_t \quad (3.17)$$

In this equation  $W_t^*$  denotes auxiliary variables that are known. The functional form of the plant model is quite general and can represent both linear and nonlinear systems. As noted by Grimble (2005),  $f_{P^*}(U_{t-b}^*, W_t^*)$  does not need to be structured. It may be represented by non-linear difference equations, fuzzy neural networks (see also (Bittanti and Piroddi, 1993)), the output of a computer code or a simple look-up

table (some restrictions will be noted in a subsequent section).

Details on the calculation of a nonlinear GMV controller are provided in Sales and Billings (1990), Grimble (2005) and Clarke and Gawthrop (1975). As noted by Grimble (2005), when the control weighting tended to zero the GMV controller reverted to the minimum variance controller. The minimum variance controller for a process described by Eq. (3.17) is straightforwardly obtained by applying the algorithm of Åström. Eq. (3.17) can be written as:

$$\begin{aligned} Y_{t+b} &= f_{P^*}(U_t^*, W_t^*) + D_{t+b} \\ &= f_{P^*}(U_t^*, W_t^*) + \hat{D}_{t+b/t} + e_{t+b/t} \end{aligned} \quad (3.18)$$

If it is possible to find the control action at time  $t$  such that  $f_{P^*}(U_t^*, W_t^*) + \hat{D}_{t+b/t} = 0$ , then the resulting controller is the minimum variance controller. The procedures of finding the minimum variance controller for the nonlinear systems are exactly the same as for the linear systems. The only difference is to replace everywhere  $\frac{\omega(q^{-1})}{\delta(q^{-1})}U_t$  by  $f_{P^*}(U_t^*, W_t^*)$ . In this derivation  $q^{-1}f_{P^*}(U_t^*, W_t^*) = f_{P^*}(U_{t-1}^*, W_{t-1}^*)$ .

It may not be possible to implement a minimum variance controller due to the fact that: i) the minimum variance controller often gives high gain, wide bandwidth and unrealistically large control signal variations, and ii) the controller is physically forbidden to take on certain values which are needed to achieve the minimum variance performance. In both of these cases, the variance of the output must exceed that for when it is possible to implement the minimum variance controller. However,  $\sigma_{MV}^2$ , in Eq. (3.13), provides a theoretical lower bound on the output variance and can be used as a useful guide for controller assessment.

### 3.4.2 Minimum Variance Lower Bounds

The utility of the minimum variance lower bound estimate is that this lower bound on performance can be estimated readily from the representative samples of closed-loop data. To archive this purpose, we have to show: i) the existence of a feedback invariant, ii) a methodology for its estimation from routine operating data. In this section, the results in Harris (1989) are extended to a class of nonlinear systems. To prove the existence of a feedback invariant we only need to show that the b-step ahead prediction error be independent of the manipulated variable action.

**Theorem 1** The b-step ahead prediction error,  $b \geq 1$ , is independent of the manipulated variable if the process plus disturbance admits a representation of the form:

$$\delta(q^{-1})Y_t = f(Y_{t-b}^*, U_{t-b}^*) + \tilde{D}_t \quad (3.19)$$

where  $\tilde{D}_t$  denotes a time series model of the form:

$$\begin{aligned} \phi(q^{-1})\nabla^d \tilde{D}_t = & a_t + \sum_{i=1}^m \theta_i a_{t-i} + \sum_{i=1}^m \sum_{j=i}^m \theta_{ij} a_{t-i} a_{t-j} \\ & + \cdots + \sum_{i_1=1}^m \cdots \sum_{i_k=i_{k-1}}^m \theta_{i_1 \dots i_k} a_{t-i_1} \cdots a_{t-i_k} \end{aligned} \quad (3.20)$$

and  $\{a_t\}$  is an white noise sequence with mean  $\mu_a$  and variance  $\sigma_a^2$ .  $\phi(q^{-1})$  and  $\delta(q^{-1})$  are monic and stable polynomials in the backshift operator. The disturbance model must be invertible (Tong, 1990). Stability and invertibility of nonlinear polynomial ARMA models are discussed in Tong (1990) and Hernández and Arkun (1996).

#### Proof - Appendix A

**Corollary 1-1** The system described in Eq. (3.19) admits a description of the form:



$$Y_{t+b} = \hat{Y}_{t+b/t} + e_{t+b/t} \quad (3.21)$$

where

$$\hat{Y}_{t+b/t} = \delta^{-1}(q^{-1})(f(Y_t^*, U_t^*)) + \hat{D}_{t+b/t} \quad (3.22)$$

and

$$\delta(q^{-1})D_t = \tilde{D}_t \quad (3.23)$$

with conditional-prediction error given by:

$$\begin{aligned} e_{t+b/t} &= Y_{t+b} - \hat{Y}_{t+b/t} \\ &= D_{t+b} - \hat{D}_{t+b/t} \end{aligned} \quad (3.24)$$

These follow immediately from Theorem 1 (Appendix).  $\tilde{D}_t$  can be interpreted as the additive disturbance and  $D_t$  as the output disturbance. (Note: It is not true that  $\delta(q^{-1})\hat{D}_{t+b/t} = \hat{\tilde{D}}_{t+b/t}$ ).

### Remarks

- By construction  $E\{e_{t+b/t}\} = 0$  (Appendix)
- In the nonlinear case  $e_{t+b/t}$  in Eq. (3.24) is a very complex function. It will include terms of  $a_{t+b}, \dots, a_{t+1}$  and may include terms  $a_{t-k}$ ,  $k \geq 0$ . The latter terms are not considered random variables when the conditional expectation is taken with respect to the information set  $I_t$ .
- In this chapter the disturbance is assumed to be an additive disturbance and to be represented by a linear time series. These assumption can be strengthened by the following comments:
  - The assumption of an additive term is valid if the interaction terms between the disturbance terms and the dynamic terms are small in comparison to

the main effect terms. As will be seen, this assumption can be tested empirically.

- The disturbance is assumed to have a linear time series representation. This is not very restrictive, since in many applications the model for the disturbance is only linear approximation.

A sufficient condition for Theorem 1 is that the process description not contain nonlinear terms in  $y_{t+b-k}$ ,  $k = 1 \cdots (b-1)$ , i.e., for any time periods over the delay horizon. Past dependencies on linear terms in  $Y_{t+b-k}$ ,  $k \geq 1$  are permitted. Theorem 1 may not hold if the nonlinear function  $f(\cdot)$  in Eq. (3.19) includes nonlinear terms in  $Y_{t-b+i}$ ,  $i = 1 \cdots (b-1)$ . This can be illustrated by the following artificial example:

$$Y_t = \phi Y_{t-1} U_{t-2} + a_t \quad (3.25)$$

where  $a$  is a white noise with mean  $\mu_a$  and constant variance  $\sigma_a^2$ . The outputs of  $Y_{t+1}$  and  $Y_{t+2}$  can be straightforwardly constructed as:

$$\begin{aligned} Y_{t+1} &= \phi Y_t U_{t-1} + a_{t+1} \\ Y_{t+2} &= \phi Y_{t+1} U_t + a_{t+2} \\ &= \phi(\phi Y_t U_{t-1} + a_{t+1}) U_t + a_{t+2} \end{aligned} \quad (3.26)$$

The corresponding conditional prediction and prediction error are:

$$\begin{aligned} \hat{Y}_{t+1/t} &= \phi Y_t U_{t-1} + \mu_a \\ \hat{Y}_{t+2/t} &= E\{Y_{t+2}|I_t\} \\ &= \phi(\phi Y_t U_{t-1} + \mu_a) U_t + \mu_a \end{aligned}$$

$$\begin{aligned}
e_{t+1} &= a_{t+1} - \mu_a \\
e_{t+2} &= a_{t+2} - \mu_a + \phi U_t (a_{t+1} - \mu_a)
\end{aligned} \tag{3.27}$$

The two-step ahead prediction error depends on the current control action.

**Corollary 1-2** For a linear ARIMA disturbance model  $\tilde{D}_t = \frac{\theta(q^{-1})}{\phi(q^{-1})\nabla^d} a_t$ , the prediction error and conditional mean are given by:

$$e_{t+b/t} = (1 + \varphi_1 q^{-1} + \dots + \varphi_{b-1} q^{-(b+1)}) a_{t+b} \tag{3.28}$$

where the  $\varphi$  weights are the impulse coefficients of the  $\frac{\theta(q^{-1})}{\phi(q^{-1})\nabla^d \delta(q^{-1})}$  transfer function and:

$$\begin{aligned}
\delta(q^{-1}) \hat{Y}_{t+b/t} &= f(Y_t^*, U_t^*) + \frac{P_b(q^{-1})}{\phi(q^{-1})\nabla^d} a_t \\
&= f(Y_t^*, U_t^*) + \frac{P_b(q^{-1})}{\phi(q^{-1})\nabla^d} (Y_t - \hat{Y}_{t/t-1}) \\
&= f(Y_t^*, U_t^*) + \frac{P_b(q^{-1})}{\theta(q^{-1})} \tilde{D}_t
\end{aligned} \tag{3.29}$$

$P_b(q^{-1})$  is a polynomial in the backshift operator obtained by solving the Diophantine equation:

$$\frac{\theta(q^{-1})}{\phi(q^{-1})\nabla^d \delta(q^{-1})} = 1 + \varphi_1 q^{-1} + \dots + \varphi_{b-1} q^{-b+1} + q^{-b} \frac{P_b(q^{-1})}{\phi(q^{-1})\nabla^d \delta(q^{-1})} \tag{3.30}$$

These equations follow immediately from the definition of the conditional expectation and standard results for prediction of linear time series (Åström, 1970; Box and Jenkins, 1970).

**Theorem 2** If the process in Eq. (3.19) is controlled by a linear or nonlinear feedback controller  $g(\cdot)$ , then the conditional prediction error  $e_{t+b/t}$  is feedback invariant and can theoretically be recovered from routine operating data.

**Proof** The proof follows readily by noting that with a feedback controller  $g(\cdot)$ , the process between measured output and the disturbance is of the nonlinear ARMA form:

$$\begin{aligned} Y_{t+b} &= \delta^{-1}(q^{-1})f(Y_t^*, g(Y_t^* - Y_{sp})) + \hat{D}_{t+b/t} + e_{t+b/t} \\ &= \hat{Y}_{t+b/t} + e_{t+b/t} \end{aligned} \quad (3.31)$$

**Corollary 2-1** For a linear ARIMA disturbance  $\tilde{D}_t = \frac{\theta(q^{-1})}{\phi(q^{-1})\nabla^d}a_t$ , the closed loop admits the following representations:

**Innovations or 1-step ahead representation**

$$Y_{t+1} = \hat{Y}_{t+1/t} + a_{t+1} \quad (3.32)$$

where

$$\begin{aligned} \delta(q^{-1})\hat{Y}_{t+1/t} &= f(Y_{t-b+1}^*, g(Y_{t-b+1}^* - Y_{sp})) + \frac{P_1(q^{-1})}{\phi(q^{-1})\nabla^d}a_t \\ &= f(Y_{t-b+1}^*, g(Y_{t-b+1}^* - Y_{sp})) + \frac{P_1(q^{-1})}{\phi(q^{-1})\nabla^d}(Y_t - \hat{Y}_{t/t-1}) \end{aligned} \quad (3.33)$$

This equation is of the form:

$$A(q^{-1})\hat{Y}_{t+1/t} = B(q^{-1})(f(Y_{t-b+1}^*, g(Y_{t-b+1}^* - y_{sp})) + C(q^{-1})Y_t \quad (3.34)$$

where  $A(q^{-1}) = \theta(q^{-1})$ ,  $B(q^{-1}) = \phi(q^{-1})\nabla^d$  and  $C(q^{-1}) = P_1(q^{-1})$ .

**b-step ahead representation**

$$\delta(q^{-1})\hat{Y}_{t+b/t} = f(Y_t^*, g(Y_t^* - Y_{sp})) + \frac{P_b(q^{-1})}{\phi(q^{-1})\nabla^d}(Y_t - \hat{Y}_{t/t-1}) \quad (3.35)$$

This equation is of the Nonlinear Autoregressive (NAR) form:

$$A(q^{-1})\hat{Y}_{t+b/t} = B'(q^{-1})(f(Y_t^*, g(Y_t^* - Y_{sp}))) + C'(q^{-1})Y_t \quad (3.36)$$

where  $B'(q^{-1}) = B(q^{-1})(1 + \varphi_1 q^{-1} + \dots + \varphi_{b-1} q^{-(b+1)})$  and  $C'(q^{-1}) = P_b(q^{-1})$ .

Since  $e_{t+b/t}$  is feedback invariant the minimum variance lower bounds can be estimated from routine operating data if it is possible to construct  $\hat{y}_{t+b/t}$  only with the input and output data. In linear systems, superposition applies, and it is easy to construct the predictor as the sum of disturbance and setpoint effects (Desborough and Harris, 1993). In the general nonlinear case, to model  $\hat{Y}_{t+b/t}$  may be quite difficult. The task becomes further complicated since the  $\hat{Y}_{t+b/t}$  includes both exact values of output and deviation values. However, it is possible that  $\hat{Y}_{t+b/t}$  can be approximated by a general function such as a universal polynomial or by neural networks. Then it is possible and practicable that the minimum lower bounds can be estimated from routine operating data. In next section, the detailed procedures using polynomial approximations to estimate the minimum lower bounds are discussed.

The theoretical lower bound on performance can be determined from dynamic process models described in Eq. (3.19). This form is quite general and includes NAR-MAX models, including the AR-Volterra model and superdiagonal bilinear model. A simpler case in which the nonlinear function only includes the inputs is considered, i.e.,  $f(y_{t-b}^*, u_{t-b}^*) = f(u_{t-b}^*)$ . Volterra series are a very important class of functions which satisfy this condition. They are recognized as a powerful tool in the study of nonlinear systems with memory and they have been used in a variety of situations both in applications and in the study of approximation of general nonlinear systems (Doyle III *et al.*, 2002; Pearson and Ogunnaike, 1997; Maner and III, 1997; Treichl *et al.*, 2002; Bendat, 2001; Schetzen, 1980; Pearson *et al.*, 1996).

Volterra series have theoretical justification as approximators with the following

desirable properties (Harris and Yu, 2007):

- they can be used to model nonlinear processes that have the following qualitative behavior (Doyle III *et al.*, 2002; Pearson, 1999): i) generate harmonics from periodic inputs, ii) exhibit asymmetric response to symmetric inputs, and iii) possess input multiplicities.
- many block-oriented nonlinearities, such as Hammerstein (a static nonlinearity, followed by a linear dynamic model), Wiener models (a linear dynamic model followed by a static nonlinearity), Uryson models (Hammerstein models in parallel), and projection pursuit models (Weiner models in parallel) have Volterra series representations.
- parallel and cascade Volterra models result in Volterra models
- processes described by control-affine models, i.e. admit a Volterra representation. Nonlinear control affine-models have been studied extensively in the control and chemical engineering literature, ie. (Kurtz and Henson, 1997).
- bilinear systems can be approximated with a structured Volterra representation (Tan, 2006).

### 3.4.3 Extension to General Nonlinear State Space Models

Recall a general SISO discrete nonlinear system with an additive disturbance of the form,

$$\begin{aligned} Z_t &= f(Z_{t-1}^*, U_{t-b}^*) \\ Y_t &= Z_t + D_t \end{aligned} \tag{3.37}$$

where  $Z_{t-1}^*$ ,  $U_{t-b}^*$  and  $D_t$  are defined in Section 3.2.

Using the procedures in the proof of Theorem 1, it is straightforward to show the existence of the feedback invariants. The only difference is to replace everywhere  $f(Y_t^*, U_t^*)$  by  $f(Z_{t-1}^*, U_t^*)$ .

The question is " can the prediction error  $e_{t+b/t}$  be theoretically recovered from routine operating data?". The answer is positive.

The  $i$ -step-ahead-state ( $i > 1$ ) transition function  $\Phi_i$  can be found by repeated application of the function  $f$ ,

$$\begin{aligned}
Z_{t+i} &= \Phi_i(Z_0^*, U_{t+i-b}^*) \\
&= f([Z_{t+i-1}, \dots, Z_{t+i-n_z}], Z_{t+i-b}^*) \\
&= f([f(Z_{t+i-2}^*, U_{t+i-b-1}^*), \dots, f(Z_{t+i-n_z}^*, U_{t+i-b-n_z}^*)], U_{t+i-b}^*) \\
&\quad \vdots \\
&= f([f([\dots f(Z_0^*, U_{1-b}^*)], U_{t+i-b-1}^*), \dots, f([\dots f(Z_0^*, U_{1-b}^*)], U_{t+i-b-n_z}^*)], U_{t+i-b}^*)
\end{aligned} \tag{3.38}$$

If this nonlinear system memory can be approximated to be a constant, say,  $m \geq 1$ , then the system in Eq. (3.37) can be expressed as,

$$Y_{t+m} = \Phi_m(Z_t^*, U_{t+m-b}^*) + D_{t+m} \tag{3.39}$$

### Remarks

- if we assume that the effect of the system states at time  $t$  on the system states at time  $t + m$  is small, the output of the general nonlinear system can be approximately written as,

$$Y_{t+m} \approx \Phi_m(U_{t+m-b}^*) + D_{t+m} \tag{3.40}$$

All the results from the above sections can be directly applied.

- if we assume that the states at time  $t$  are close to the zero state,  $Z_t^* \approx 0$ , the output of the general nonlinear system can also be approximated in the form of Eq. (3.40).
- if the above two assumptions are invalid, it means that the states at time  $t$  must be estimated. Usually the system memory length  $m$  is greater than the process delay  $b$  and the information about the system is only inputs and outputs. Since the  $E[Y_t] = Z_t$ , it is naturally that we use the output to approximate the state value. Then the general nonlinear system can be approximated as,

$$Y_{t+m} \approx \tilde{\Phi}_m(Y_t^*, U_{t+m-b}^*) + D_{t+m} \quad (3.41)$$

All results from above section can be directly applied in this model.

## 3.5 Estimation of Lower Bound from Operating Data using Polynomial Approximation

### 3.5.1 Polynomial Approximation

Recall the nonlinear system in Eq. (3.19) with the control action  $U_t = g_c((Y_t - Y_{sp})^*)$  as:

$$\begin{aligned} \delta(q^{-1})Y_t &= \tilde{h}(Y_{t-b}^*, U_{t-b}^*) + \tilde{D}_t \\ &= \tilde{h}(Y_{t-b}^*, g_c((Y_{t-b} - Y_{sp})^*)) + \tilde{D}_t \\ &= h(Y_{t-b}^*, (Y_{t-b} - Y_{sp})^*) + \tilde{D}_t \end{aligned} \quad (3.42)$$

For many real sampled nonlinear systems, the exact NARMAX models, described



by the function  $h$  in Eq. (3.42), are very difficult to determine. Therefore, it often necessary to approximate  $h$  by some functions. Polynomial NARMAX models have been shown to be a very good choice. The detailed procedures to estimate the MVLB using polynomial NARMAX models will be discussed in the following paragraphs. The estimation of the MVLB from operating data if the function  $h$  in Eq. (3.42) can be modeled by a Volterra series is proposed in the paper by Harris and Yu (2007). If the nonlinear function is continuous it can always be arbitrarily well approximated by polynomial models (Chen and Billings, 1989). When the system is severely non-linear then other choices of approximating functions should be considered. Being linear-in-parameters, the polynomial models can readily be estimated using linear least squares methods as can be seen in Section 3.5.2.

Let a finite polynomial series approximate the function  $h(\cdot)$ , namely,

$$\begin{aligned}
h(Y_{t-b}^*, (Y_{t-b} - Y_{sp})^*) &= h_0 + \sum_{i=b}^{m_y} h_i^y Y_{t-i} + \sum_{i=b}^{m_y} \sum_{j=i}^{m_y} h_{ij}^y Y_{t-i} Y_{t-j} + \dots \\
&+ \sum_{i=b}^{m_u} h_i^u (Y_{t-i} - Y_{sp}) + \sum_{i=b}^{m_u} \sum_{j=i}^{m_u} h_{ij}^u (Y_{t-i} - Y_{sp})(Y_{t-j} - Y_{sp}) \\
&\dots + \sum_{i=b}^{m_y} \sum_{j=i}^{m_u} h_{ij}^{yu} Y_{t-i} (Y_{t-j} - Y_{sp}) + \dots
\end{aligned} \tag{3.43}$$

Rearranging Eq. (3.43), the process can be approximated in the polynomial form:

$$\begin{aligned}
\delta(q^{-1})(Y_t) &= \tilde{h}_0 + \sum_{i=b}^m \tilde{h}_i Y_{t-i} + \sum_{i=b}^m \sum_{j=i}^m \tilde{h}_{ij} Y_{t-i} Y_{t-j} \\
&+ \dots + \sum_{i_1=b}^m \dots \sum_{i_k=i_{k-1}}^m \tilde{h}_{i_1 \dots i_k} Y_{t-i_1} \dots Y_{t-i_k} + \tilde{D}_t
\end{aligned} \tag{3.44}$$

In this equation, the nonlinear effects are confined to time intervals  $t - k$ ,  $k \geq b$  and the linear effects start at  $t - 1$ . In the case of a linear disturbance model, the

innovations representation, Eq. (3.32) and the b-step ahead representation, Eq. (3.35) have the following structure:

**Innovations or 1-step ahead representation:**

$$\begin{aligned}
A(q^{-1})\hat{Y}_{t+1/t} &= h'_0 + \sum_{i=0}^m h'_i Y_{t-i} + \sum_{i=b-1}^m \sum_{j=i}^m h'_{ij} Y_{t-i} \tilde{Y}_{t-j} \\
&+ \cdots + \sum_{i_1=b-1}^m \cdots \sum_{i_k=i_{k-1}}^m h'_{i_1 \cdots i_k} Y_{t-i_1} \cdots Y_{t-i_k}
\end{aligned} \tag{3.45}$$

**b-step ahead prediction:**

**Polynomial Autoregressive (PAR) representation:**

$$\begin{aligned}
A(q^{-1})\hat{Y}_{t+b/t} &= h'_0 + \sum_{i=0}^m h'_i Y_{t-i} + \sum_{i=0}^m \sum_{j=i}^m h'_{ij} Y_{t-i} Y_{t-j} \\
&+ \cdots + \sum_{i_1=0}^m \cdots \sum_{i_k=i_{k-1}}^m h'_{i_1 \cdots i_k} Y_{t-i_1} \cdots Y_{t-i_k}
\end{aligned} \tag{3.46}$$

**Polynomial Autoregressive with eXogenous inputs (PARX) representation:**

$$\begin{aligned}
A(q^{-1})\hat{Y}_{t+b/t} &= \tilde{h}_0 + \sum_{i=0}^m \tilde{h}_i^u U_{t-i} + \sum_{i=0}^m \tilde{h}_i^Y Y_{t-i} + \sum_{i=0}^m \sum_{j=i}^m \tilde{h}_{ij}^u U_{t-i} U_{t-j} \\
&+ \sum_{i=0}^m \sum_{j=i}^m \tilde{h}_{ij}^y Y_{t-i} Y_{t-j} + \sum_{i=0}^m \sum_{j=i}^m \tilde{h}_{ij}^{uy} U_{t-i} Y_{t-j} \cdot + C'(q^{-1})Y_t
\end{aligned} \tag{3.47}$$

Both representations can be used to estimate the variance of the b-step ahead prediction error. In this chapter, the direct b-step ahead representation will be used to estimate the minimum-variance performance which is the residual variance after fitting the model.

### 3.5.2 Polynomial AR Model Identification using Orthogonal Least Squares Methods

Direct methods to estimate the minimum lower bounds are used in this paper using the PAR model in Eq. (3.46) or PARX model in (3.47). Several methods have been proposed for this purpose: Orthogonal Least Squares (OLS) methods (Chen *et al.*, 1989) and Fast Orthogonal Search (FOS) methods (Chon *et al.*, 1997; Korenberg, 1988). Use of Artificial Neural Network (ANN) models to approximate the NARMAX models is discussed in Terasvirta *et al.* (2005) and Chen and Billings (1992). In this paper, the OLS methods are used to determine the order and estimate the parameters of the PAR/PARX models which represent the b-step ahead prediction.

When  $A(q^{-1}) = 1$ , the linear-in-parameters model in Eqs. (3.46) and (3.47) can be written matrix form as:

$$\tilde{Y} = F\Theta + \Xi \quad (3.48)$$

with

$$\tilde{Y} = \begin{bmatrix} \tilde{Y}_n \\ \vdots \\ \tilde{Y}_N \end{bmatrix}, \quad F = \begin{bmatrix} 1 & \tilde{Y}_{n-1} & \dots & \tilde{Y}_{n-m} \dots \tilde{Y}_{n-m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{Y}_{N-1} & \dots & \tilde{Y}_{N-m} \dots \tilde{Y}_{N-m} \end{bmatrix}$$

$$\Xi = \begin{bmatrix} e_n \\ \vdots \\ e_N \end{bmatrix}, \quad \Theta = \begin{bmatrix} h'_0 & h'_1 & \dots & h'_{m\dots m} \end{bmatrix} \quad (3.49)$$

$F$  is the matrix of regression variables,  $\Theta$  is the parameter vector and  $\Xi$  is the vector of b-step ahead prediction errors.  $N$ ,  $n$  and  $M$  are the data length, starting point for the regression and number of regressor variables respectively ( $n \geq M$ ). The parameter vector that minimizes  $\|\tilde{Y} - F\Theta\|^2$  ( $\|\cdot\|$  is the Euclidean norm) is given by:

$$\hat{\Theta} = (F^T F)^{-1} F^T \tilde{Y} \quad (3.50)$$

The number of possible terms in Eq. (3.49) could be very large. If the number of regressors is  $n_r$  and the maximum polynomial degree is  $n_d$ , the number of parameters is:

$$n_p = \frac{(n_r + n_d)!}{n_r! n_d!} \quad (3.51)$$

For example, if  $n_r = 8$  and  $n_d = 4$ , then  $n_p = 495$ . A substantial reduction in the number of model parameters can be achieved by an appropriate selection of the orthonormal functions. The OLS algorithm is applicable when there are a 'reasonable' number of regressors. In the case of the larger model orders and higher polynomial degrees, the Fast Orthogonal Search (FOS) (Korenberg, 1988) and Genetic Programming (GP) (Madár *et al.*, 2005) methods are recommended. Variations on least squares algorithms have also been developed for computational efficiency of large numbers of candidate regressors, e.g. Li *et al.* (2005) (such as Volterra series, a radial basis function neural network or a polynomial NARX model).

The OLS algorithm developed by Chen *et al.* (1989) involves a Q-R decomposition of the regression matrix  $F$  of the form  $F = WA$ , where  $W$  is an  $M \times M$  upper

triangular matrix and  $A$  is an  $(N - n + 1 \times M)$  matrix with orthogonal columns in the sense that  $W^T W = D$  is a diagonal matrix. ( $N - n + 1$  is the length of  $Y$  vector and  $M$  is the number of regressors.) After this decomposition one can calculate the OLS auxiliary parameter vector  $\hat{g}$  as:

$$\hat{g} = D^{-1} W^T \tilde{Y} \quad (3.52)$$

where  $\hat{g}_i$  is the corresponding element of the OLS solution vector. The sums of squares of the observed values,  $\tilde{Y}^T \tilde{Y}$  can be written as:

$$\tilde{Y}^T \tilde{Y} = \sum_{i=1}^M \hat{g}_i^2 w_i^T w_i + \hat{\Xi}^T \hat{\Xi} \quad (3.53)$$

where  $\hat{\Xi}$  is the residual error from the full model, i.e.  $\hat{\Xi} = \tilde{Y} - F\hat{\Theta}$ . An error reduction ratio,  $[err]_i$  of  $F_i$  term can be defined as:

$$[err]_i = \frac{\hat{g}_i w_i^T w_i}{\tilde{Y}^T \tilde{Y}} \quad (3.54)$$

This ratio offers a simple mean of order and select the model terms of a linear in parameters model according to their contribution to the performance of the model. The terms which have very small error reduction values, say smaller than  $\rho$ , are eliminated. The value of  $\rho$  determines how many terms will be included in the final model. This OLS algorithm can be interpreted as a forward-selection method where the reduction in sums of squares is maximized at each decision stage.

Alternatively, one can combine the OLS approach with a technique that provides a penalty for increasing model complexity. One such method is Akaike's Information

Criterion AIC(s) (Chen *et al.*, 1989):

$$AIC(\lambda) = N \ln \hat{\sigma}_{\Xi}^2 + M \cdot \lambda \quad (3.55)$$

where  $M$  is the number of the model parameters, and  $\hat{\sigma}_{\Xi}^2$  is the residual error.  $\lambda$  is a positive value chosen to provide a penalty for model complexity. Using statistical arguments, a value of  $\lambda = 4$  is recommended (Chen *et al.*, 1989; Leontaritis and Billings, 1987).

The assumption of  $A(q^{-1}) = 1$  is not restrictive. The effect of ignoring this term is to increase the number of terms in the Volterra series. When some roots of  $A(q^{-1})$  are close to the unit circle, the OLS approach may provide an inefficient form for approximation, requiring a large number of terms for the predictor in Eqs. (3.45) and (3.46). The more comprehensive model polynomial ARMA/X models can be employed to efficiently fit the closed-loop nonlinear system in Eqs. (3.42) and (3.44) efficiently. The detailed iteration procedures can be found in Chen *et al.* (1989).

For the invariant bounds to exist, it is necessary that the prediction error not depend on the manipulated variable. This assumption can be tested by regressing the prediction errors  $\hat{\Xi} = \tilde{Y} - F\hat{\Theta}Y^T$  on values of the manipulated variable, and analyzing the regression results for statistical significance.

### 3.6 Simulation Results

In this section, two examples: an artificial Volterra model and an exothermic CSTR model are provided to demonstrate the methodology outlined in this chapter.

### 3.6.1 Example 1: An Artificial Volterra Model

In the first example, a nonlinear dynamic system is assumed to be of a second order Volterra series:

$$\begin{aligned}
 Y_t = & 0.2U_{t-3} + 0.3U_{t-4} + U_{t-5} + 0.8U_{t-3}^2 \\
 & + 0.8U_{t-3}U_{t-4} - 0.7U_{t-4}^2 - 0.5U_{t-5}^2 - 0.5U_{t-3}U_{t-5} + \tilde{D}_t \quad (3.56)
 \end{aligned}$$

The disturbance is an ARIMA(2,0,0) process:

$$\tilde{D}_t = D_t = \frac{a_t}{1 - 1.6q^{-1} + 0.8q^{-2}} \quad (3.57)$$

$a_t$  is a Gaussian white noise sequence with zero mean and variance 0.1. The true value of the minimum variance lower bound is 0.6656. Since the roots of  $(1 - 1.6q^{-1} + 0.8q^{-2} = 0)$  are complex, this disturbance will display pseudo periodic behavior (Box and Jenkins, 1970). We investigate stationary linear disturbance model in simulations in this chapter, but theory is allowed that we have a nonstationary disturbance and nonlinear models. It will be an interesting topic for future research.

A proportional (P) controller and a Proportional-Integral (PI) controller are used to control the simulated process:

$$\begin{aligned}
 U_t &= -0.2(Y_t - Y_{sp}) \\
 U_t &= -\frac{0.3 - 0.2q^{-1}}{1 - q^{-1}}(Y_t - Y_{sp}) \quad (3.58)
 \end{aligned}$$

The moderate values of parameters of P and PI controllers are selected based on the performance of closed-loop response to step changes in setpoint and impulse disturbances. Special care is taken for the selection of controller parameters so that the nonlinear system is stable.

Fig. 3.1 shows a realization of  $D_t$ ,  $Y_t$  and  $U_t$  for the case when the PI controller is

used. The open loop step response to a change in the manipulated variable of  $\pm 0.5$  is shown in Fig. 3.2(a) and the closed-loop response with the PI controller to step changes of  $\pm 0.5$  in setpoint zero is shown in Fig. 3.2(b). From The open loop step response in Fig. 3.2(a), this nonlinear system has non-minimum phase behavior. The closed-loop response with the PI controller to a  $\pm 0.2$  impulse disturbance is shown in Fig. 3.3. We notice that although the open-loop process is clearly nonlinear, the nonlinearity in the closed-loop process has been reduced.

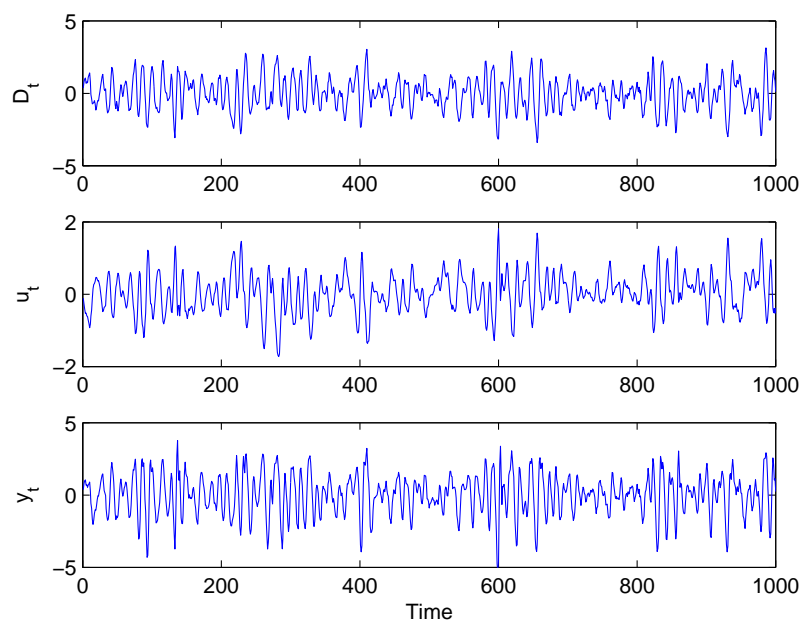


Figure 3.1: Example 1: The stochastic realizations of  $D_t$ ,  $U_t$  and  $Y_t$



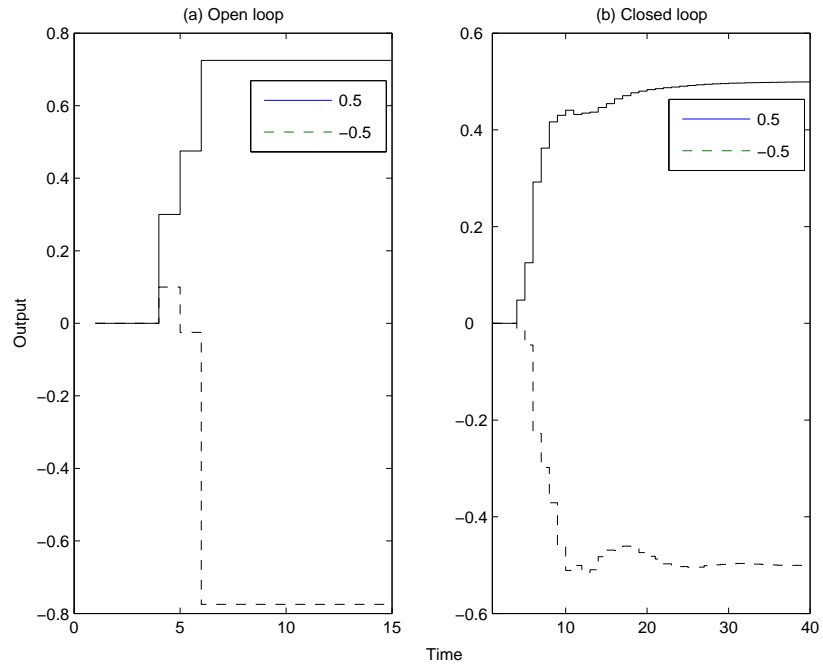


Figure 3.2: Example 1: Open loop and closed-loop responses (PI controller)

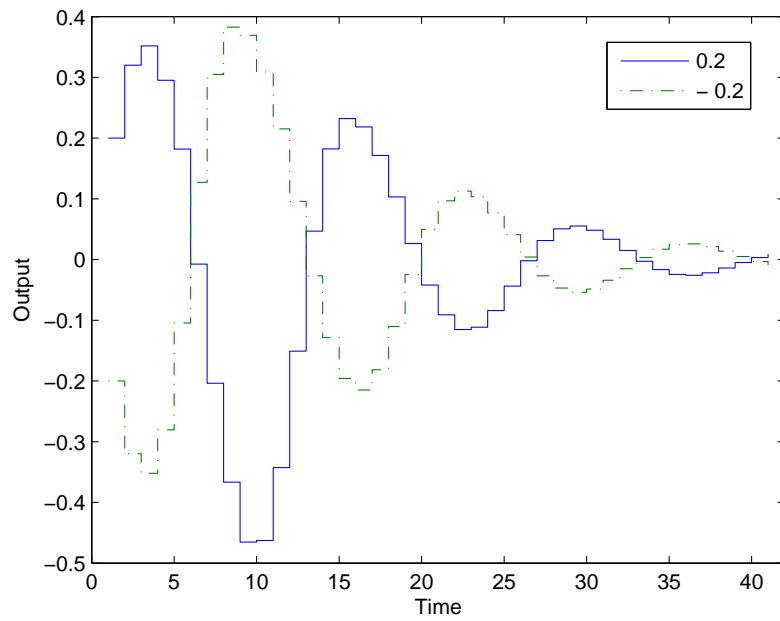


Figure 3.3: Example 1: Closed-loop simulation for a disturbance impulse change  $\pm 0.2$  with the PI controller

For estimating the minimum variance lower bounds, three direct estimation methods are used:

- Linear AutoRegressive (LAR) model:  $\hat{Y}_{t+b/t} = \sum_{i=0}^m \gamma_i \tilde{Y}_{t-i}$ . For a linear model, it is also convenient to fit the data using an ARMA representation. Calculation of the variance of the b-step ahead forecast error is straightforward once the model parameters have been identified (Harris, 1989).
- PAR model (linear and quadratic terms only) (see Eq. (3.46))
- PARX model (linear and quadratic terms only) (see Eq. (3.47))

Five hundred observations were used to fit the parameters for these models. The minimum variance bound was calculated as the residual variance from each model. This procedure was repeated five hundred times with a different white-noise sequence. For the case when the PI controller was used, the estimates of the means of inputs and outputs are  $3.15E - 5$  and  $2.93E - 4$  respectively, and the estimates of the variance of inputs and outputs are 0.4233 and 3.2061.

For the OLS method, a large number of candidate terms are selected to fit the model. The detailed information about the terms for the different models are listed at Table 3.1. For the PI controller case, the average number of the terms selected to fit the models are 14.5, 32.9 and 24.6 respective for the LAR, PAR and PARX models. For the P controller case the number of terms are 4.2, 10.8 and 5.6 for the LAR, PAR and PARX models respectively.

Table 3.1: Example 1: Initial Candidate Terms for Models

	Linear( $y$ )	Quadratic	Linear( $u$ )	Total
LAR	40	0	0	40
PAR	20	55	0	75
PARX	20	55	20	95

Table 3.2: Example 1: Estimates of  $\sigma_{MV}^2$  using different models

	PI controller		P controller	
	$\hat{\sigma}_{MV}^2$	s.d.	$\hat{\sigma}_{MV}^2$	s.d.
LAR	1.4487	0.4079	0.6826	0.0511
PAR	0.6964	0.1070	0.6697	0.0459
PARX	0.6704	0.0888	0.6683	0.0460

The estimates of the  $\sigma_{MV}^2$  using three models for the P and PI controller cases are shown in Table 3.2. The comparative box plots of the quality estimates are shown in Figure 3.4.

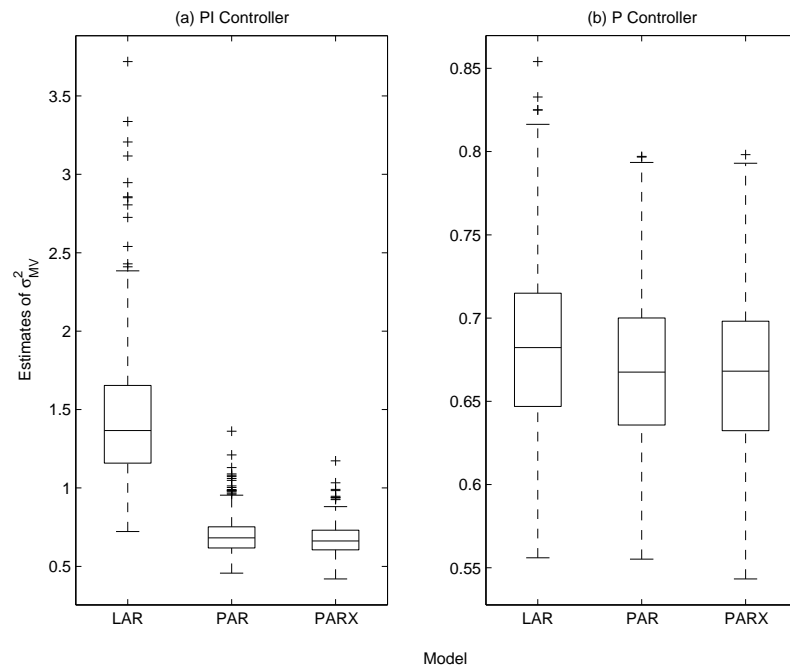


Figure 3.4: Example 1: Comparative box plots of the quality estimates for three models

Several graphic elements about box plots in this thesis are:

- The lower and upper lines of the "box" are the 25th and 75th percentiles of the sample. The distance between the top and bottom of the box is the interquartile range.

- The line in the middle of the box is the sample median. If the median is not centered in the box, that is an indication of skewness.
- The "whiskers" are lines extending above and below the box. They show the extent of the rest of the sample (unless there are outliers). Assuming no outliers, the maximum of the sample is the top of the upper whisker. The minimum of the sample is the bottom of the lower whisker. an outlier is a value that is more than 1.5 times the interquartile range away from the top or bottom of the box.
- The plus sign at the top of the plot is an indication of an outlier in the data. This point may be the result of a data entry error, a poor measurement or a change in the system that generated the data.

From the data in Table 3.2 and with reference to Figure 3.4, the following observations can be made:

- When using linear AR models, the bias in the estimate of the minimum variance lower bounds in P controller case is much smaller than PI controller case. The reason is that the quadratic contributions are less important using the P controller than the PI controller. While it is possible that the inclusion of more autoregressive terms might provide a 'better' estimate  $\sigma_{MV}^2$  when using the linear models, the AIC criterion is used to provide a practical means of avoiding overfitting the data.

An ARMA(p,p-1) model was fit to the data with increasing values of p. The model order, p, was chosen to minimize the AIC criterion. The variance of the b-step ahead forecast error was obtained from this model using standard techniques. The results were essentially indistinguishable from those obtained using the method described in the previous paragraph.

- There is a slight bias of the estimate of the performance bound for the PI controller case using PAR models. With the PI controller, the b-step prediction

should include the infinite linear and quadratic autoregressive terms for our example. However, in our implementation, only finite terms were used.

- The estimates using PARX models are better than PAR models specially for PI controller case since the PARX models are closer to the true prediction functions in finite linear and quadratic inputs form.

### 3.6.2 Example 2: An Empirical NARMAX Model

The study example is a CSTR with a first-order exothermic reaction provided by Doyle III *et al.* (1989). The dynamic behavior for this CSTR can be described using the following nondimensional normalized equations:

$$\begin{aligned}\dot{x}_1 &= -x_1 + D_a(1 - x_1)e^{\frac{x_2}{1+x_2/\gamma}} \\ \dot{x}_2 &= -x_2 + BD_a(1 - x_1)e^{\frac{x_2}{1+x_2/\gamma}} - \beta(x_2 - x_c)\end{aligned}\quad (3.59)$$

$x_1$  and  $x_2$  are the reactor dimensionless concentration and temperature respectively. The case study under consideration is a regulation of outlet reactant concentration  $x_1$ . Coolant temperatures  $x_c$  is the manipulated variable. One set of parameter values  $B = 1.0$ ,  $\beta = 0.3$ ,  $\gamma = 20.0$  and  $D_a = 0.072$  which yields an open-loop system with a single stable steady state for all fixed values of the input is selected (Knapp and Budman, 2000). The detailed nomenclature for this exothermic CSTR is listed in Table 3.3.

An empirical NARMAX model which was identified by Knapp and Budman (2000) from the input/output data generated from an exothermic CSTR model in Eq. (3.59) using three-level PRBS signals in  $x_c = 5, 14$  and  $23$  (Knapp and Budman, 2000) is used to test our methodology. The resulting input/output data set was normalized by these authors. The relationships between the true variables and normalized variables for both input and output are:

Table 3.3: Nomenclature for exothermic CSTR

$B$ dimensionless heat of reaction: $B = -\Delta H C_f \gamma / C_p T_{f_0}$
$Da$ Damkohler number: $Da = (V k_0 e^{-\gamma}) / Q_r$
$x_1$ dimensionless concentration: $x_1 = (C_f - C) / C_f$
$x_2$ dimensionless temperature: $x_2 = (T - T_{f_0}) \gamma / T_{f_0}$
$x_c$ dimensionless coolant temperature: $x_c = (T_c - T_{f_0}) \gamma / T_{f_0}$
$\beta$ dimensionless cooling rate: $\beta = UA / Q_f C_p$
$\gamma$ dimensionless activation energy: $\gamma = E_a / RT_{f_0}$
$A$ heat transfer surface area ( $dm^2$ )
$C$ reactant concentration ( $mol/L$ )
$C_f$ feed concentration of reactant ( $mol/L$ )
$C_p$ heat capacity ( $J/(g \cdot K)$ )
$E_a$ activation energy ( $J/mol$ )
$\Delta H$ heat of reaction ( $J/mol$ )
$k_0$ reaction rate constant ( $min^{-1}$ )
$Q_f$ mass feed flowrate ( $L/min$ )
$R$ ideal gas constant ( $J/((mol \cdot K))$ )
$T_c$ coolant temperature ( $K$ )
$T_f$ feed temperature ( $K$ )
$T_{f_0}$ nominal feed temperature ( $K$ )
$U$ overall heat transfer coefficient ( $J/(min \cdot K \cdot dm^2)$ )
$V$ reactor volume ( $L$ )

$$\begin{aligned}
Y &= 0.3861Y_{nor} + 0.5780 \\
U &= 9U_{nor} + 14
\end{aligned} \tag{3.60}$$

where  $Y_{nor}$  and  $U_{nor}$  denote the normalized variables.

A second-order NARMAX model of the process is given as:

$$\begin{aligned}
Z_{t+1} = F^z \begin{bmatrix} Z_t \\ Z_{t-1} \end{bmatrix} + F^u \begin{bmatrix} U_t \\ U_{t-1} \end{bmatrix} + [ Z_t \quad Z_{t-1} ] F^{zz} \begin{bmatrix} Z_t \\ Z_{t-1} \end{bmatrix} \\
+ [ Z_t \quad Z_{t-1} ] F^{zu} \begin{bmatrix} Z_t \\ U_{t-1} \end{bmatrix} + [ U_t \quad U_{t-1} ] F^{uu} \begin{bmatrix} U_t \\ U_{t-1} \end{bmatrix}
\end{aligned} \tag{3.61}$$

where

$$\begin{aligned}
F^z &= \begin{bmatrix} 0.6132 & -0.1167 \\ 0.1639 & -0.1392 \\ -0.1392 & 0.1682 \end{bmatrix} & F^u &= \begin{bmatrix} 0.4427 & 0 \\ -0.2266 & 0 \\ 0.1180 & 0 \end{bmatrix} & F^{zz} &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\
F^{zz} &= \begin{bmatrix} 0.1639 & -0.1392 \\ -0.1392 & 0.1682 \end{bmatrix} & F^{zu} &= \begin{bmatrix} -0.2266 & 0 \\ 0.1180 & 0 \end{bmatrix} & F^{uu} &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}
\end{aligned} \tag{3.62}$$

The normalized input range is  $[-1 \ 1]$  corresponding to the real value range  $[5 \ 23]$ . The output is bounded at range  $[-0.987 \ 0.777]$  corresponding to the real value range  $[0.1969 \ 0.8780]$  by the normalized input (Knapp and Budman, 2000). A linear additive disturbance  $D_t$  is assumed to add the output  $Z_t$ , the output  $Y_t$  can be written as:

$$Y_t = Z_t + D_t \tag{3.63}$$

According to the study of the stability and performance regions of linear PI controller for this example (Gao and Budman, 2005), a regular discrete PI controller is designed for our simulation study,

$$U_t = \frac{0.6 - 0.3q^{-1}}{1 - q^{-1}}(Y_{sp} - Y_t) \quad (3.64)$$

In the Matlab simulation, we use  $U_t = U_{t-1} + (0.6 - 0.3q^{-1})(Y_{sp} - Y_t)$ . If manipulated variable  $U_t$  is over the low and high limits, the value of  $U_t$  is set to equal to the limits.

For this example, two kinds of disturbance models, AR(1) and AR(2) models, are used to test our approach.

**AR(1) disturbance model:** the first disturbance model is an AR(1) model defined as:

$$D_t = \frac{a_t}{1 - 0.95q^{-1}} \quad (3.65)$$

where  $a_t$  is a Gaussian white noise with zero mean and variance 0.86E-2. Special care must be taken to select the variance of  $a_t$ , the large variance will cause the input and output to reach their limit bounds frequently. The true value of minimum variance lower bound is 2.337E-2. This model is close to a random walk model which has been widely used to describe the behavior of the disturbance. The neighboring values in this model are similar and the series exhibits marked trends (see Fig. 3.5).

**AR(2) disturbance model:** the second disturbance model AR(2) is the same as the disturbance model used in the first example.  $a_t$  is a Gaussian white noise with zero mean and variance 0.667E-2. The true value of minimum variance lower bound is 0.0444. The variance of  $a_t$  is selected so that the variance of disturbance  $D_t$  from this AR(2) disturbance model is equal to the variance of disturbance from the AR(1) disturbance model described in Eq. (3.65). Since the variances of  $D_t$  from



both AR(1) and AR(2) model are same, say, 0.08823, it give us a chance to compare the effects of the disturbance structure on the estimates of minimum variance lower bounds.

Realizations of  $D_t$ ,  $Y_t$  and  $U_t$  are shown in Figs. 3.5 and 3.6 for the AR(1) and AR(2) disturbance models respectively. The open loop response to the input step changes of  $\pm 0.5$  is shown in Fig. 3.7 (a). The closed-loop response with the PI controller to step changes of  $\pm 0.5$  in setpoint zero is shown in Fig. 3.7(b). And the closed-loop response with the PI controller to a  $\pm 0.2$  impulse disturbance is shown in Fig. 3.7(c). From the open loop response in Fig. 3.7 (a), we can observe that nonlinearity of the output of this second-order NARMA model is asymmetric. From the closed-loop response in Fig. 3.7 (b), we can also observe that the asymmetric nonlinearity of output is reduced by the effects of the feedback PI controller.

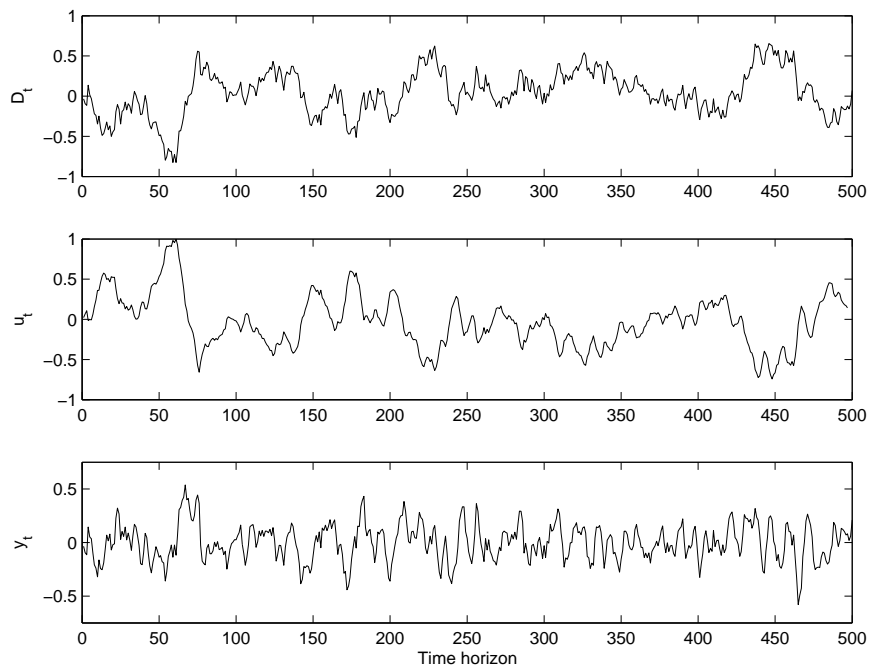


Figure 3.5: Example 2: The stochastic realization of  $D_t$ ,  $U_t$ , and  $Y_t$  for AR(1) disturbance

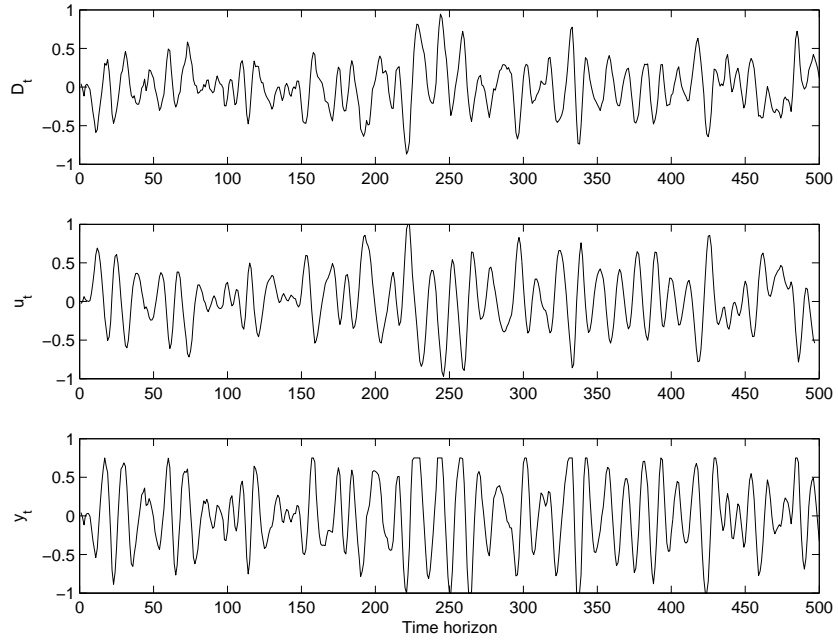


Figure 3.6: Example 2: The stochastic realization of  $D_t$ ,  $U_t$ , and  $Y_t$  for AR(2) disturbance

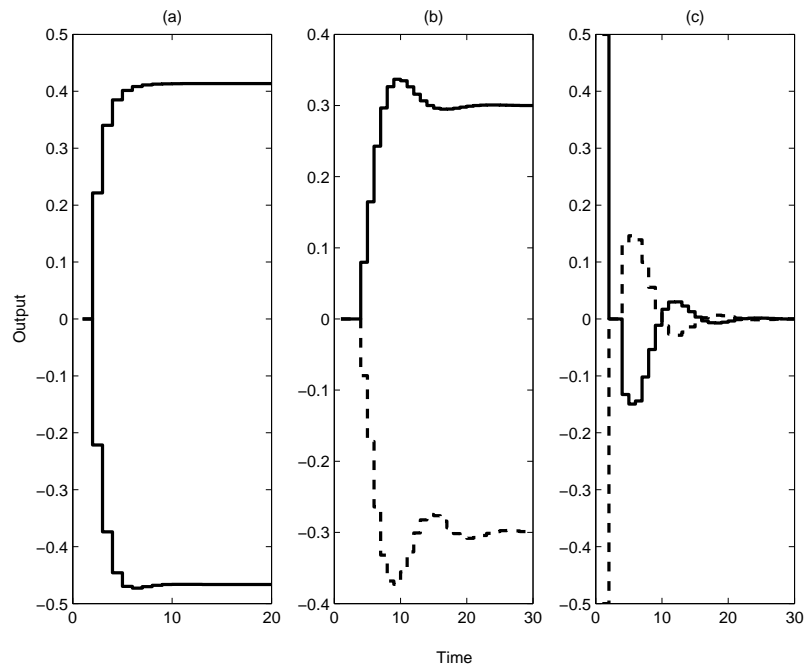


Figure 3.7: Example 2: Open loop and closed-loop step responses and closed-loop responses for disturbance impulse changes

Five hundred observations were used to estimate the minimum variance bound. This procedure was repeated five hundred times with a different white-noise sequence. The estimates of the  $\sigma_{MV}^2$  using three models for both AR(1) and AR(2) disturbance models are shown in Table 3.4.

Table 3.4: Example 2: Estimates of  $\sigma_{MV}^2$  using different models for both AR(1) and AR(2) disturbance cases

	AR(1)		AR(2)	
	$\hat{\sigma}_{MV}^2$	s.d.	$\hat{\sigma}_{MV}^2$	s.d.
LAR	2.46E-2	3.11E-3	5.73E-2	2.67E-3
PAR	2.41E-2	1.05E-3	4.65E-2	2.59E-3
PARX	2.38E-2	1.00E-3	4.54E-2	2.69E-3

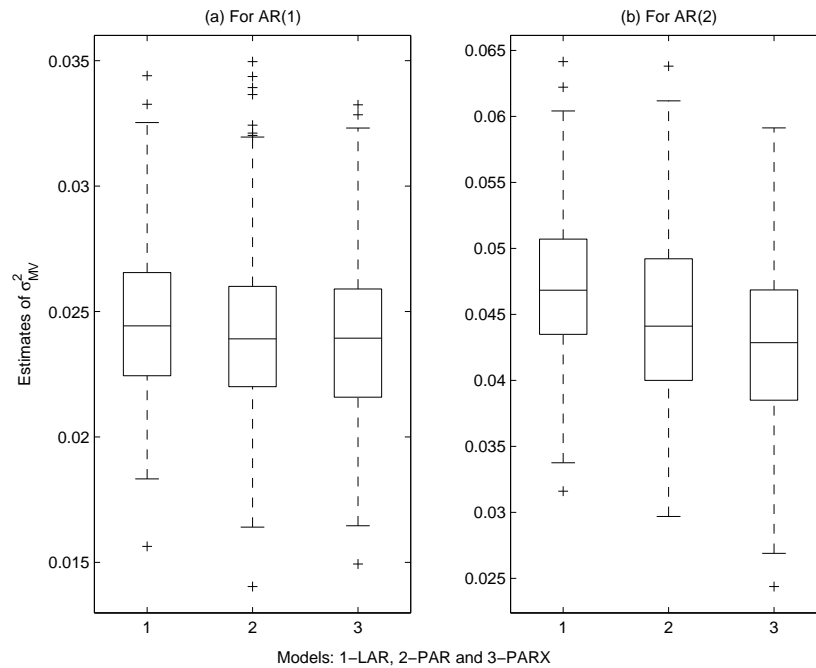


Figure 3.8: Example 2: Comparative box plots of the quality estimates from three models

The comparative box plots of the quality estimates for both AR(1) and AR(2) disturbance models are shown in Figs. 3.8. From the Tables 3.4 and Figure 3.8 (a) and (b), the following observations can be concluded:

- Similar observations discussed in the first example can also be concluded from this simulation. There are three reasons: i) the same AR (2) disturbance model is used for both simulations, ii) following the methodology in the paper by Diaz and Desrochers (1988) a Volterra model can be obtained from the NARMAX model, iii) the simulation NARMAX model is of the polynomial form which is the same as the Volterra structure.
- The disturbance model structure doesn't affect the estimates of MVLB when the PAR and PARX model are used for the estimations for this example.
- The disturbance model structure does affect the estimates of MVLB if only the LAR model is used for the estimation. For the AR(1) model which is close to a random walk model, the estimate bias of MVLB from the LAR model is much smaller than the result for the AR(2) model. The disturbance structure help the PI controller reduce the variance of the output. The minimum variance controller is a function of the b-step prediction of the disturbance  $D_t$  (see the definition of MV control in Eq. (3.11), say,  $U_t = f_{MV}(\hat{D}_{t+b/t})$ ). If the disturbance model is close to a random walk, the b-step prediction of the disturbance can be approximated as  $\hat{D}_{t+b/t} = \sum_{i=0}^{\infty} a_{t-i}$ . Then the MV controller is of the form  $u_t = f_{MV}(\sum_{i=0}^{\infty} a_{t-i})$ . The linear PI controller which is of the form  $K(\sum_{i=0}^{\infty} a_{t-i})$  ( $K$  is a constant) can be considered as a linear approximation of the MV controller. This fact can be shown from the simulation data, the estimate of output variance is 0.0312 when the disturbance variance is 0.0882. The ratio between the MVLB and the variance output is 0.74:1. The linear terms dominate the output data.

### 3.6.3 Example 3: A CSTR Model

In this simulation, we do not use the empirical model described in Eq. (3.61) in Section 3.5.2. The real CSTR system in Eq. (3.59) is used to test our methodology.

It is a typical chemical engineering process which is intensively studied at the control and system identification areas. This CSTR system is selected as a case study example due, on the one hand, to the simplicity of its mathematical representation with only two dynamical states, i.e., reactor temperature and reactor concentration, and on the other hand, to its inherent nonlinearity. The starting situation is a stable steady situation with the states  $x_{10} = 0.6219$  and  $x_{20} = 3.7092$  and input  $u_t = 14$ . In this section, all values of the input and output are the real values.

The output  $Y_t$  with an additive linear disturbance is:

$$Y_t = (x_1)_t + D_t \quad (3.66)$$

where  $D_t$  is an additive disturbance. A delay of  $b=3$  samples is assumed to present a transport delay on the output. In this simulation, the same PI controller which is used in Example 2 is also used. From the relationship between the true variables and the normalized variables in Eq. (3.60), the parameters of PI control are found as:

$$U_t = \frac{(14 - 7q^{-1})}{1 - q^{-1}}(Y_{sp} - Y_t) \quad (3.67)$$

Constraints on process input amplitude  $U_{min} = 5$ ,  $U_{max} = 23$  and output amplitude  $Y_{min} = 0$ ,  $Y_{max} = 1$  should also be considered. When the input and output exceed their limits, the limit values are assigned to them. The sampling period  $T=2$  time units. The responses of step changes for the open loop are shown in Fig. 3.9. The closed-loop response with the PI controller to step changes of  $\pm 0.2$  in setpoint 0.3 is shown in Fig. 3.10. The closed-loop response with the PI controller to a  $\pm 0.2$  impulse disturbance is shown in Fig. 3.11. In Figs. 3.9, 3.10 and 3.11, the values of output in the (a) part are the real values of outputs and in the (b) part are the normalized values of outputs. From the (b) parts we can compare this real model with the empirical model which is discussed in Section 3.6.2.

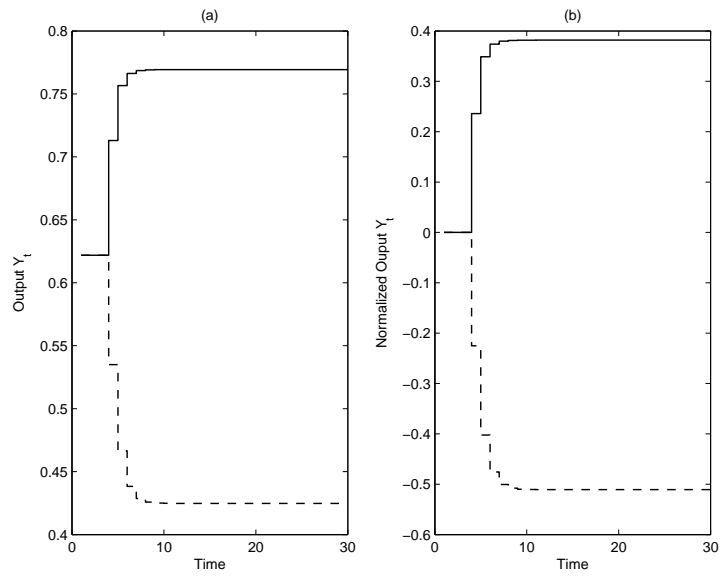


Figure 3.9: Example 3: Open loop step change response

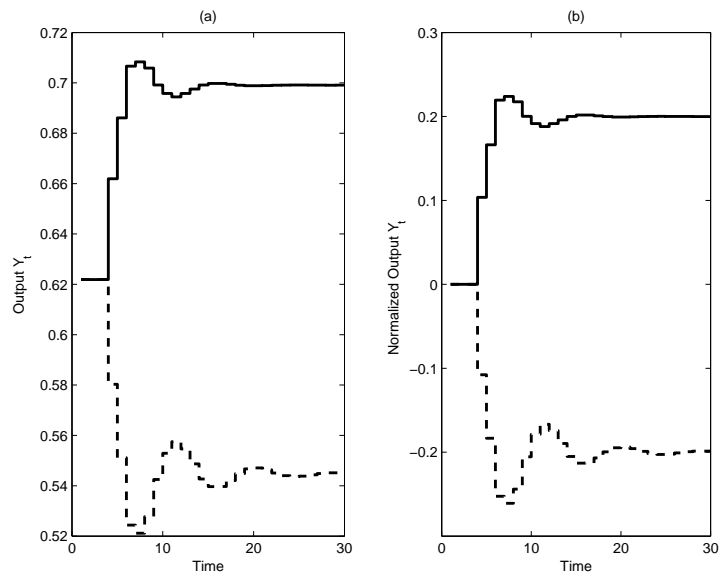


Figure 3.10: Example 3: Closed-loop step change response

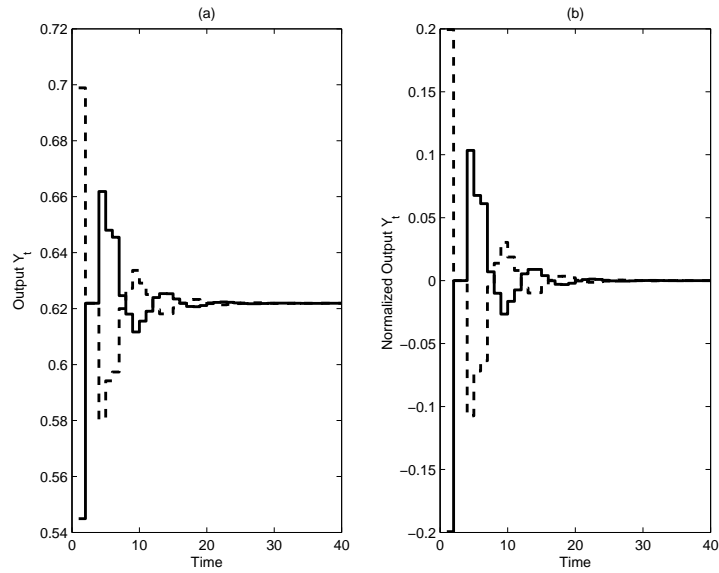


Figure 3.11: Example 3: Closed-loop simulation for a disturbance impulse change

The same AR(1) and AR(2) disturbance models are also used for this simulation. The normalized output values in the second example are approximately 0.361 time of the real values. To keep the same scale of the disturbances for this simulation, the variances of disturbance are selected to be 0.15 times of the variance of disturbance used in the second example. The variances of the Gaussian white noises are assigned to  $1E - 3$  and  $1.29E - 3$  for AR(1) and AR(2) models respectively. The true values of minimum variance lower bounds are  $3.51E-3$  and  $6.66E-3$  for the AR(1) and AR(2) disturbance models respectively.

The realizations of  $D_t$ ,  $Y_t$  and  $U_t$  are shown in Fig. 3.12 for the AR(1) disturbance model and in Fig. 3.14 for the AR(2) disturbance model. The realizations in Figs.3.12 and Fig. 3.14 are normalized using the relationship in Eq. 3.60. The normalized realizations are shown in Figs. 3.13 and 3.15.

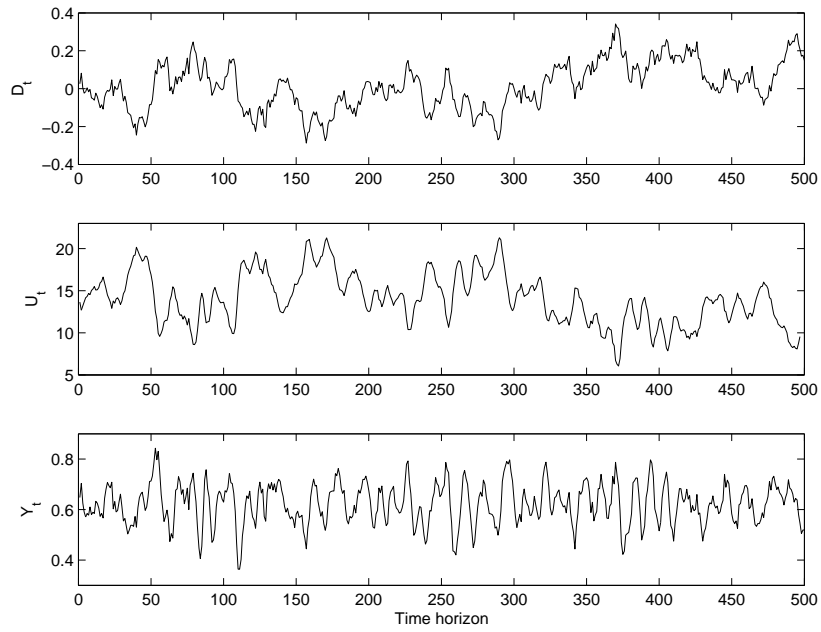


Figure 3.12: Example 3: The stochastic realization of  $D_t$ ,  $U_t$ , and  $Y_t$  for the AR(1) model

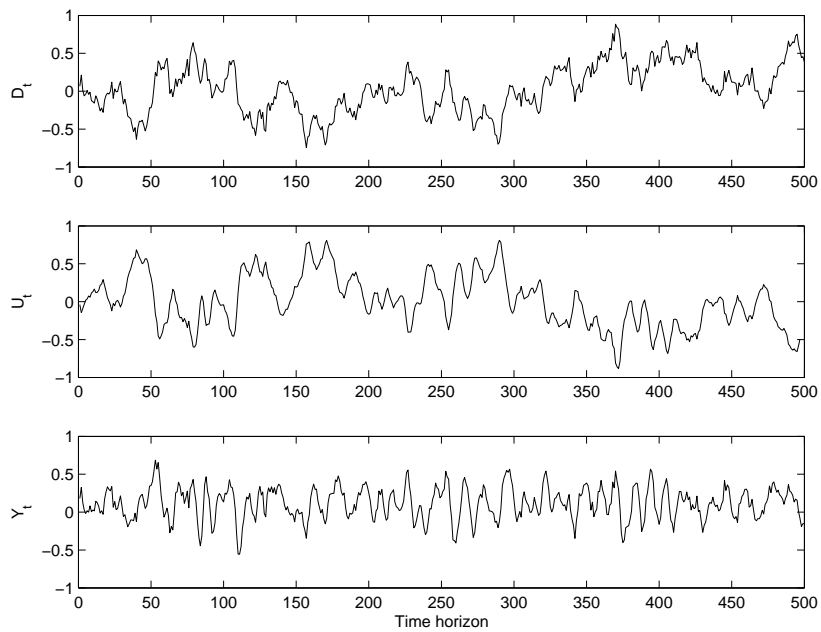


Figure 3.13: Example 3: The normalized stochastic realization of  $D_t$ ,  $U_t$ , and  $Y_t$  for the AR(1) model



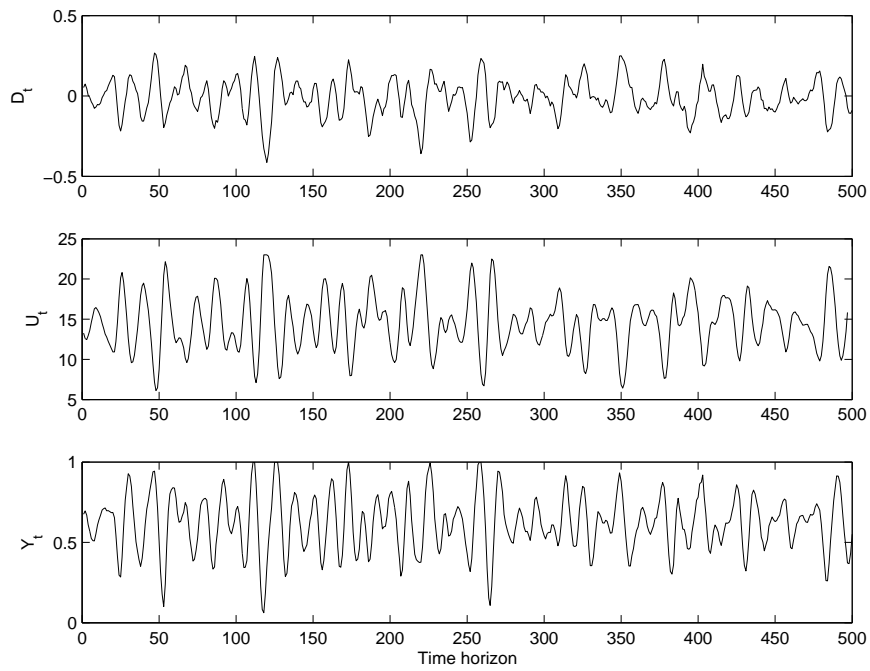


Figure 3.14: Example 3: The stochastic realization of  $D_t$ ,  $U_t$ , and  $Y_t$  for the AR(2) model

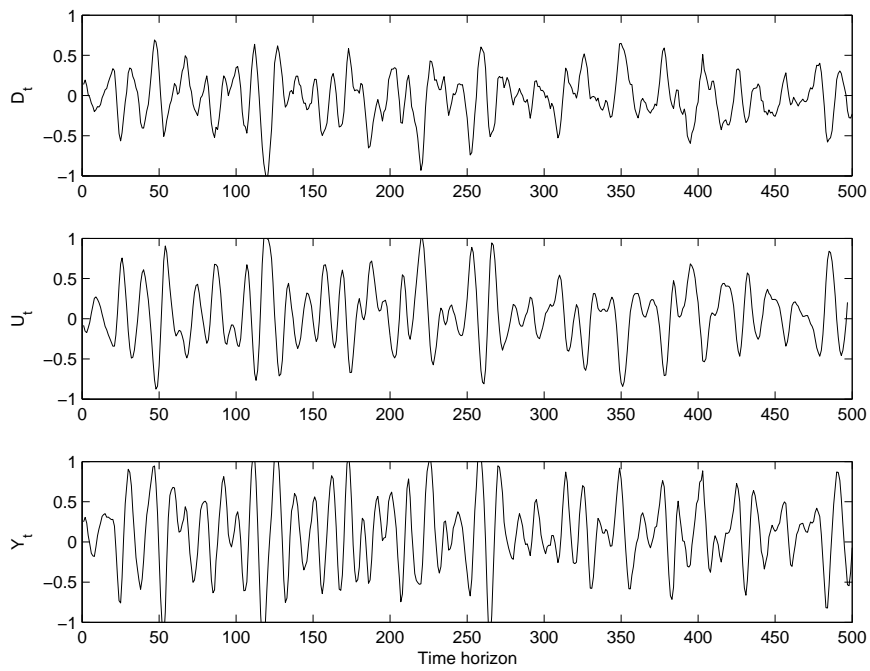


Figure 3.15: Example 3: The normalized stochastic realization of  $D_t$ ,  $U_t$ , and  $Y_t$  for the AR(2) model

The means and variances of  $D_t$ ,  $U_t$ , and  $Y_t$  from both empirical and real model with the AR(2) disturbance are listed in Table 3.5. The estimates of means and variances are based on the data shown in Figs. 3.6 and 3.15. The real values of  $D_t$ ,  $U_t$ , and  $Y_t$  from the real CSTR model shown in Fig.3.14 are converted to the normalized values. From Table 3.5, we can conclude that the empirical NARMAX model is very close to the real CSTR model in the input range [5 23].

Table 3.5: Means and variances of  $D_t$ ,  $U_t$ , and  $Y_t$  from both empirical and real models

	Estimated Mean			Estimated Variance		
	$D_t$	$U_t$	$Y_t$	$D_t$	$U_t$	$Y_t$
Empirical	-1.3E-4	2.08E-2	1.6E-3	8.98E-2	0.138	0.198
Real	-4.3E-3	3.4E-2	3.8E-4	8.6eE-2	0.143	0.206

Five hundred observations were used to estimate the minimum variance bound. This procedure was repeated five hundred times with a different Gaussian white-noise sequence. The estimates of the  $\sigma_{MV}^2$  using three models for both AR(1) and AR(2) disturbance models are shown in Table 3.6.

Table 3.6: Example 3: Estimates of  $\sigma_{MV}^2$  using different models for the AR(2) disturbance case

	AR(1)		AR(2)	
	$\hat{\sigma}_{MV}^2$	s.d.	$\hat{\sigma}_{MV}^2$	s.d.
LAR	3.70E-3	4.91E-4	8.26E-3	1.17E-3
PAR	3.61E-3	5.85E-4	6.87E-2	1.08E-3
PARX	3.56E-3	6.07E-4	6.71E-2	1.06E-3

The comparative box plots of the quality estimates for both AR(1) and AR(2) disturbance models are shown in Fig. 3.16.

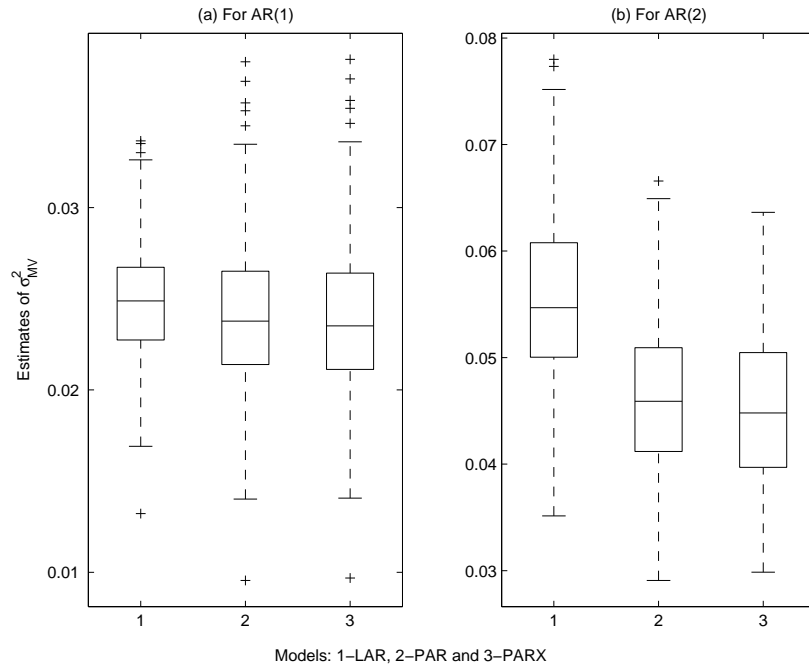


Figure 3.16: Example 3: Comparative box plots of the quality estimates from three models

Since the minimum variance lower bounds for this simulation are a linear moving average series, we can directly convert their values into the normalized values. The results are listed in Table 3.7. The similar observations in the example 2 can

Table 3.7: Example 3: Estimates of  $\sigma_{MV}^2$  (converted to normalized values) using different models for the AR(2) disturbance

	AR(1)		AR(2)	
	$\hat{\sigma}_{MV}^2$	s.d.	$\hat{\sigma}_{MV}^2$	s.d.
LAR	2.48E-2	3.29E-3	5.54E-2	6.03E-3
PAR	2.42E-2	3.92E-3	4.61E-2	5.81E-3
PARX	2.39E-2	4.61E-3	4.50E-2	5.77E-3

also be concluded in this simulation. The only different fact is that the bias of the MVLB estimates from the PAR and PARX models will include the bias caused by the polynomial approximation.

## 3.7 Conclusions

Reports of theoretical and applied work on control performance assessment and monitoring of SISO linear systems have appeared in the literature (Stanfelj *et al.*, 1993; Desborough and Harris, 1992; Harris, 1989). This chapter has provided a preliminary work to extend the methodology to a class of nonlinear dynamic / stochastic systems.

A class of nonlinear dynamic / stochastic systems for which there exist minimum variance feedback invariant performance bounds has been established. For this class of systems, the minimum variance bound can be estimated using closed-loop data using a nonlinear PAR or PARX model to estimate the b-step ahead prediction of the process. It is necessary to know the process delay. Applications of the methodology to simulation examples indicate that this approach gives very credible estimates of the minimum variance performance bound. The simulation results indicate that the Orthogonal search method is effective.

The extension of our approaches to the general nonlinear state space models is provided. It is theoretically shown that the class of nonlinear dynamic / stochastic can be broadly used to represent the general nonlinear state space models with some assumptions or modification. Applications of the methodology to the empirical NARMA model and the CSTR model indicate that this approach gives very credible estimates of the minimum variance performance bound for the general nonlinear state space models.

From the simulation results, we find that the qualities of the MVLB estimates using the linear models are dependent of the controller and the disturbance structure. On the contrary, the qualities of the MVLB estimates using the PAR or PARX model are independent of the controller and the disturbance structure. This fact can be used as the nonlinearity indicator. If the differences between the linear and nonlinear performance bounds are important, then further investigation into the source of the nonlinearities might be justified, or the use of a nonlinear controller might be

considered.

The extended horizon performance indices has been developed for linear systems (Desborough and Harris, 1992; Harris *et al.*, 1996; Kozub, 1996; Thornhill *et al.*, 1999). It strengthens the applicability of the performance indices since the process delay may be unknown or can be precisely estimated. Further development of the methods outlined in the current paper to extended-horizon predictions for the nonlinear systems would be very meaningful for the industrial applications. However, the extended horizon performance indices for the nonlinear systems will be very difficult. The reason is that the nonlinear models which provide good predictions over one time horizon, can provide poor predictions over extended horizons (Ramsey, 1996).

Our methods are restricted by the assumption of the additive linear/nonlinear disturbance. Although aforementioned this assumption is a good approximation for many nonlinear system, it is still limited for the general nonlinear systems. To extend our method to the general nonlinear systems will be a valuable project for the further research.

# Chapter 4

## Variance Decomposition for Nonlinear Stochastic Systems / Time Series

### 4.1 Introduction

Analysis of Variance (ANOVA) refers to the task of decomposing the variance of a response variable into contributions arising from the inputs, and assessing the magnitude and significance of each of their contributions. The initial techniques of the analysis of variance were developed by the statistician and geneticist R.A. Fisher in the 1920s and 1930s. In traditional analysis of variance applications, the data under study are often collected from a static system. In the last few decades, researchers and industrial practitioners interested in variance decomposition have focused on systems with memory such as the time series (Hamilton, 1994; Lütkepohl, 1991) and univariate and multivariate linear dynamic systems. Desborough and Harris (1993) gave an analytical solution of the output variance decomposition for the univariate linear system which can be expressed using infinite impulse weights and in which the distur-

bances are uncorrelated. Using multivariate impulse response analysis, the variance decomposition for Vector AutoRegressive (VAR) models with correlated disturbances was developed by Seppala *et al.* (2002).

Far less has been written on assessing the multivariate disturbance effects on the process performance for nonlinear systems. There are two main complications: i) the study of nonlinear control performance assessment is still in its infancy, and ii) complexity of nonlinear systems, since nonlinear systems have complex structure generalizations are difficult. This topic leads us to the variance decomposition which is able to decompose the variance of outputs according to each factor uncertainty. A knowledge of how the output can vary with respect to the variations of the disturbances can yield insights into the behavior of the model and can actually assist the closed-loop performance assessment or variance reduction. For example, if a process has an additional measurement of some components of the disturbance, control engineers would like to use this extra information to improve the process performance such as implementation of feed-forward controls. But before doing that, they must conduct the analysis of variance or variance decomposition. If the measurable disturbance does significantly contribute to the output variance, feed-forward control should be considered.

Two basic types of disturbance can be distinguished, namely, *dynamic disturbances*, which acts directly on the dynamics, and *measurement disturbances*, which are only added to the signal of the dynamics. Dynamic disturbances may be caused by aging of components, changes in environmental conditions, or calibration errors. One typical model including the dynamic disturbance is called the linear system with multiplicative disturbance. It has been intensively studied in modeling, prediction and control at variety of areas such as chemistry, biology, ecology, economics, physics and engineering (Kampen, 1992; Bailey *et al.*, 2004; Stachurski, 2003; Ghaoui, 1995; Jimenes and Ozaki, 2002), but there is little detailed treatment of variance decompo-

sition for these systems.

In the study of ANOVA of stable univariate and multivariate linear systems, the variance decomposition can be achieved through the Impulse Response Function (IRF) analysis. In linear systems there is a direct relationship between the impulse response and the variance. It is not true for nonlinear systems, and the IRF cannot be used in the variance decomposition for nonlinear systems in general. A variance decomposition method, called ANOVA-like decomposition, commonly for the static systems (Saltelli and Scott, 2000), is employed on the nonlinear dynamic systems. In order that the ANOVA-like decomposition is able to deal with the specific properties of the nonlinear dynamic systems such as time dependence and initial conditions, two modifications of the ANOVA-like decomposition method are proposed. Details will be addressed in the following sections. If the output of a static system considered can be represented as an analytic function of input variables, e.g.,  $Y = f(X_1, X_2, \dots, X_p)$ . The importance of the given independent inputs  $X_1, X_2, \dots, X_p$  can be measured via the fractional variance which is defined as the fractional contribution to the output variance due to the uncertainties in inputs. It can be calculated by using the following ANOVA-like decomposition formula for the total output variance  $V(Y)$  of the output  $Y$  (Archer *et al.*, 1997; Cox, 1982):

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12\dots p} \quad (4.1)$$

where

$$\begin{aligned} V_i &= V(E(Y|X_i = x_i)) \\ V_{ij} &= V(E(Y|X_i = x_i, X_j = x_j)) \\ &\quad - V(E(Y|X_i = x_i)) - V(E(Y|X_j = x_j)) \end{aligned} \quad (4.2)$$



and so on, where  $E(Y|X_i = x_i)$  denotes the expectation of  $Y$  conditional on  $X_i$  having a fixed value  $x_i$ , and  $V$  stands for variance over all the possible values of  $x_i$ . In this chapter some preliminary results are presented for variance decomposition for the nonlinear systems which can be represented as the nonlinear ARMA models.

The layout of this chapter is as follows, In Section 4.2, a general nonlinear input-output model with multivariate disturbances, nonlinear autoregressive moving average (NARMA) model which is the nonlinear model for variance decomposition study, is introduced. In Section 4.3, the development of the variance decomposition for the nonlinear ARMA models is discussed. Two modifications of the ANOVA-like decomposition method are addressed. A computationally cheap method, FAST, is introduced for quantifying the analysis of variance decomposition. Since the ANOVA-like decomposition method is model independent, using this method do not require that nonlinear system models should be known. In Section 4.4, two simulation examples are used to illustrate the essential features of the proposed methods. The chapter is concluded with a description of outstanding issues and limitations of the proposed methodology.

## 4.2 Problem Description

Let us consider the following generic nonlinear stochastic problem in a dynamic state-space form (Jazwinski, 1970):

$$\dot{\mathbf{X}}(t) = f(\mathbf{X}(t), \mathbf{U}(t), \mathbf{v}(t)) \quad (4.3)$$

$$\mathbf{Y}(t) = h(\mathbf{X}(t), \mathbf{U}(t), \mathbf{d}(t)) \quad (4.4)$$

where Eqs. (4.3) and (4.4) are called state equation and measurement equation, respectively;  $\mathbf{X}(t)$  represents the state vector,  $\mathbf{Y}(t)$  is the measurement vector,  $\mathbf{U}(t)$

denotes the systems input vector (as driving force) in a controlled environment;  $f : \mathbb{R}^{N_x} \longrightarrow \mathbb{R}^{N_x}$  and  $g : \mathbb{R}^{N_x} \longrightarrow \mathbb{R}^{N_y}$  are two vector valued functions, which are potentially time-varying;  $\mathbf{v}(t)$  and  $\mathbf{d}(t)$  represent the dynamic (process) disturbance and measurement disturbance respectively, with appropriate dimensions.

The modeling of nonlinear processes is complicated, not only by the richness of the models available, but more so by the behavior that these processes can exhibit. In practice however, we are more concerned about the class of models are those that can be modeled using an input/output representation. State models requiring the knowledge of states are excluded. Since the input/output data are most often discrete, the discrete input/output nonlinear models are broadly used in nonlinear systems research such as nonlinear identification and control design. For a discrete nonlinear model, the general single-input-single-output (SISO) form of the description is:

$$Z_t = f_1(Z_{t-1}^*, U_{t-b}^*) \quad (4.5)$$

where  $Z_t$  is the deterministic output of the system in response to the inputs that are denoted by  $U_t$ .  $b$  represents the number of whole periods of delay in the system and is the number of sampling intervals that elapse between making a change in the process input and first observing its effect. The notation is that  $f_1(Z_{t-1}^*, U_{t-b}^*)$  denotes a function of previous values of  $Z_{t-i}, i = 1..n_z$  and  $U_{t-b-j}, j = 0..n_u$ .

In this chapter we are interested in the nonlinear systems that are affected by both dynamic and measure disturbances. A general SISO nonlinear model, nonlinear autoregressive moving average with exogenous (NARMAX) (Tong, 1990; Granger and Teräsvirta, 1993; Petersen and James, 1996; Stachurski, 2003; Jazwinski, 1970), is used:

$$Y_t = f_2(Y_{t-1}^*, U_{t-b}^*, \xi_{1,t}^*, \dots, \xi_{n_\xi,t}^*, \varepsilon_{1,t}^*, \dots, \varepsilon_{n_\varepsilon,t}^*) \quad (4.6)$$

the measurement disturbance terms  $\varepsilon_{i,t}$ ,  $i = 1..n_a$  and dynamic disturbance terms  $\xi_{i,t}$ ,  $i = 1..n_\xi$  are assumed to be identically independent distributed (iid) variables with mean zero and variance  $\sigma_{a_i}^2$   $i = 1..n_a$  and  $\sigma_{\xi_i}^2$   $i = 1..n_\xi$  respectively.

A specific class of models for the system with multiplicative disturbance is defined by the time invariant discrete state equation as (Ghaoui, 1995; Jimenes and Ozaki, 2002):

$$\begin{aligned}\mathbf{X}_{t+1} &= (\mathbf{A} + \sum_{i=1}^{m_1} \mathbf{A}_{s,i} \xi_t^i) \mathbf{X}_t + (\mathbf{B} + \sum_{i=1}^{m_2} \mathbf{B}_{s,i} \varepsilon_t^i) \mathbf{U}_t + \mathbf{F} \omega_t \\ \mathbf{Y}_t &= \mathbf{C} \mathbf{X}_t + \mathbf{a}_t\end{aligned}\tag{4.7}$$

where  $\mathbf{X}_k \in \mathbb{R}^{n_x}$  is the state vector,  $\mathbf{Y}_k \in \mathbb{R}^{n_y}$  and  $\mathbf{U}_k \in \mathbb{R}^{n_u}$  are the output and input vectors of interest,  $\omega_t \in \mathbb{R}^{n_\omega}$  and  $\mathbf{a}_t \in \mathbb{R}^{n_a}$  are the process and measurement disturbance vectors,  $\xi \in \mathbb{R}^{n_\xi}$  and  $\varepsilon \in \mathbb{R}^{n_\varepsilon}$  are the multiplicative disturbances,  $\mathbf{A}$ ,  $\mathbf{A}_{s,i}$ ,  $\mathbf{B}$ ,  $\mathbf{B}_{s,i}$ ,  $\mathbf{C}$  and  $\mathbf{F}$  are constant matrices which have proper dimensions. The independent random variables  $\omega_t$ ,  $\mathbf{a}_t$ ,  $\xi_t$  and  $\varepsilon_t$  have zero-mean with variance-covariance  $\Omega^\omega$ ,  $\Omega^a$  and  $\Omega^\xi = \Omega^\varepsilon = I$  respectively. When there is no control action,  $\mathbf{U}_t = \mathbf{0}$ , a time series  $\mathbf{X}_t$  in Eq. (4.7) is called a random coefficient autoregressive model of order k, RCA(k) (Tong, 1990). The only difference between a bilinear model and the model in Eq. (4.7) is that the random coefficient  $\mathbf{A} + \sum_{i=1}^{m_1} \mathbf{A}_{s,i} \xi_t^i$  are not necessarily independent of the process or measurement disturbance.

## 4.3 Variance Decomposition of Nonlinear ARMA Models

### 4.3.1 Variance Decomposition

Although nonlinear stochastic systems in Eq. (4.6) have been studied in some depth in the control and identification theory literatures (see, e.g., (Leontaritis and Billings, 1985*a*; Leontaritis and Billings, 1985*b*; Pearson and Ogunnaike, 1997; Billings and Chen, 1989; Chen *et al.*, 1989; Chen and Billings, 1992)), the statistical analysis of these models is still in its infancy.

In linear systems, the effect of a dynamical disturbance can always be correctly represented as an additive output disturbance regardless of where it actually appears in the system. This is of the principle of superposition. Distinguishing the types of disturbance is not very important for the linear cases. In nonlinear systems, superposition does not hold. Thus, for the analysis of nonlinear dynamic systems, it is a fundamental problem to characterize the type of disturbance and to quantify the amount of disturbance and their effects. Although estimation of the level of dynamic disturbance has to be system-dependent and difficult, recently, a method was proposed whereby dynamical disturbance and measurement disturbance can be measured very precisely for the known dynamics (Heald and Stark, 2000) and unknown dynamics (Siefert *et al.*, 2003).

For discussion simplicity, no driving-force in the dynamic system (which is often referred to the stochastic control problem) is considered initially in this chapter. The extension to the more general situation is straightforward and will be illustrated by a simulation of a Volterra series model with two additive linear disturbances.

Recall the NARMA with the different sources of disturbances  $a_{i,t}$ ,  $i = 1..p$  for the single output case:

$$Y_t = f(Y_{t-1}^*, a_{1,t}^*, \dots, a_{p,t}^*) \quad (4.8)$$

where  $Y_{t-1}^* = [Y_{t-1}, \dots, Y_{t-n_y}]$  and  $a_{i,t}^* = [a_{i,t}, \dots, a_{i,t-n_{a_i}}]$  (here  $a_{i,t}$  are used to present both dynamic and measurement disturbances).

**Assumptions:**

- Eq. (4.8) can be solved numerically subject to the initial condition,  $I_0$ , to give  $Y_t$  for any choice of  $a_{1,t}^*, \dots, a_{p,t}^*$ .
- the different disturbances are uncorrelated.
- for each source of disturbance  $a_{i,t}$   $i = 1, 2, \dots, p$ ,  $a_{i,t}$   $t = 1, 2, \dots$  are independent identically distributed (iid).
- the disturbances entering the systems after time  $t=0$  and the initial condition  $I_0$  are independent.
- the initial condition  $I_0$  is a random vector with probability density function  $P(I_0)$ .

We are interested in determining the sensitivity of output  $Y_t$ , at each time interval  $t = 1, 2, \dots, n$  in Eq. (4.8) to variation of each disturbance group  $A_{i,t} = [a_{i,1}, \dots, a_{i,t}]$ ,  $i = 1..p$ . Most often, in the nonlinear stochastic systems and nonlinear time series, the future behavior of  $Y_t$  is dependent on initial conditions. For this situation, we cannot use the ANOVA-like decomposition in Eqs. (4.1) and (4.2) directly since the initial condition must be considered within the variance decomposition. Using the well-known variance decomposition theorem (Parzen, 1962), we can decompose the variance of  $Y_t$ ,  $t = 1, 2, \dots, n$  as:

$$Var[Y_t] = E_{I_0}[Var_{A_t}[Y_t|I_0]] + Var_{I_0}[E_{A_t}[Y_t|I_0]] \quad (4.9)$$

where  $A_t = [A_{1,t}, A_{2,t}, \dots, A_{p,t}]$  denotes all of disturbances entering the system from time 1 to time  $t$ .  $E_{I_0}[\bullet]$  denotes the expectation of  $[\bullet]$  with respect to  $I_0$  and  $Var_{I_0}[\bullet]$  denotes the variance of  $[\bullet]$  with respect to  $I_0$ . The second term in the right-hand side of Eq. (4.9) is the fractional contribution to the output due to only the uncertainties of the initial condition. The first term in the right-hand side of Eq. (4.9) is the variance contribution to the output due to the disturbances  $A_t = [A_{1,t}, A_{2,t}, \dots, A_{p,t}]$  with the initial condition uncertainty. From Eq. (4.9), it is straightforward to obtain  $Var[Y_t] \geq E_{I_0}[Var_{A_t}[Y_t|I_0]]$ . The special situation,  $Var[Y_t] = E_{I_0}[Var_{A_t}[Y_t|I_0]]$  will be discussed in the following paragraphs.

The conditional variance given initial condition  $I_0$ ,  $Var_{A_t}[Y_t|I_0]$ , can be decomposed directly using the ANOVA-like decomposition method as:

$$Var_{A_t}(Y_t|I_0) = \sum_i Var_{A_{i,t}}|I_0 + \sum_i \sum_{j>i} Var_{A_{i,t}A_{j,t}}|I_0 + \dots + Var_{A_{1,t}\dots A_{p,t}}|I_0 \quad (4.10)$$

where

$$\begin{aligned} Var_{A_{i,t}}|I_0 &= Var_{A_{i,t}}(E_{(A_{i,t})'}(Y_t|A_{i,t}, I_0)) \\ Var_{A_{i,t}A_{j,t}}|I_0 &= Var_{A_{i,t}A_{j,t}}(E_{(A_{i,t}A_{j,t})'}(Y_t|A_{i,t}, A_{j,t}, I_0)) - Var_{A_{i,t}}|I_0 - Var_{A_{j,t}}|I_0 \\ &\vdots \end{aligned} \quad (4.11)$$

where the index  $(A_{i,t})'$  stands for "all  $A_t = [A_{1,t}, \dots, A_{p,t}]$  but  $(A_{i,t})$ " – that is, complementary to  $(A_{i,t})$ .

The variance decomposition with consideration of the initial condition can be obtained by simply calculating the expectation of the conditional variance decomposition in Eq. (4.10) with respect to the initial condition  $I_0$ . This procedure is not necessary if the initial condition has (can be approximately assumed to have) a

linear relationship with the output  $Y_t$ . The variance decomposition can be calculated with the results of the conditional variance decomposition in Eq.(4.10) based on the mean values of initial condition. Detailed information about this topic can be found in Chapter 5.

**Special case:**

If the stochastic process in Eq.(4.8) is a geometrically ergodic Markov chain, the variance decomposition in Eqs. (4.8) can be directly obtained by applying the ANOVA-like decomposition method as shown below. The criteria guaranteeing a Markov chain to be geometrically ergodic can be found in the book (Tong, 1990). If the stochastic process in Eq. (4.8) is a geometrically ergodic Markov chain, then given the initial condition  $I_0 = [y_0^*, a_{1,0}^*, \dots, a_{p,0}^*]$  there exists a limiting probability as (Tong, 1990):

$$\lim_{t \rightarrow \infty} P(Y_t|I_0) = \pi \geq 0 \quad (4.12)$$

The limiting probability  $\pi$  is independent of the initial condition. The sufficient condition of the geometrically ergodic nonlinear time series was given by (Tong, 1990, P. 127).

Recall Eq. (4.9)  $Var[Y_t] = E_{I_0}[Var_{A_t}[Y_t|I_0]] + Var_{I_0}[E_{A_t}[Y_t|I_0]]$ . For a geometrically ergodic Markov chain,  $E_{A_t}[Y_t|I_0]$  is equal to a constant value for any different initial conditions for  $t \rightarrow \infty$ . Therefore  $Var_{I_0}[E_{A_t}[Y_t|I_0]]$  is equal to zero. The limiting variance of output  $Y_t$  can be written as:

$$Var[Y_{t \rightarrow \infty}] = E_{I_0}[Var_{A_t}[Y_{t \rightarrow \infty}|I_0]] \quad (4.13)$$

Since the output  $Y_{t \rightarrow \infty}$  is independent of the initial condition  $I_0$ , Eq. (4.13)

can be written as:

$$Var[Y_{t \rightarrow \infty}] = Var_{A_t}[Y_{t \rightarrow \infty} | I_0] \quad (4.14)$$

The results of variance decomposition using the method in Eq. (4.10) will not depend on the initial condition. The finite series, approximated to  $m$  terms which is the reasonable memory length of the nonlinear system of interest, is used for the variance decomposition. Since the polynomial NARMA models are most often used to represent nonlinear systems, auto-correlation, cross-correlation and cross bi-correlation between output and disturbances can be used to find  $m$ . Autocorrelation, cross-correlation and cross bi-correlation are mathematical tools used frequently in signal processing for analyzing functions or series of values. They are reliable with large samples (Billings and Chen, 1989; Pearson *et al.*, 1996).

The auto-correlation can be defined in Eq. (4.15) and may be estimated from data from Eq. (4.16).

$$r_{yy}(\tau) = E[(Y(t + \tau) - \bar{Y})(Y(t) - \bar{Y})] \quad (4.15)$$

$$\hat{r}_{yy}(\tau) = \frac{1}{N} \sum_{t=1}^N [(y(t + \tau) - \bar{y})(y(t) - \bar{y})] \quad (4.16)$$

The cross-correlation between the disturbance  $a_{i,t}$  and the output  $Y_t$  is defined in Eq. (4.17) and can be estimated from data from Eq.(4.18).

$$r_{ya_i}(\tau) = E[(Y(t + \tau) - \bar{Y})a_{i,t}] \quad (4.17)$$

$$\hat{r}_{ya_i}(\tau) = \frac{1}{N} \sum_{t=1}^N [(y(t + \tau) - \bar{y})a_{i,t}] \quad (4.18)$$

The auto-bi-correlation is defined in Eq. (4.19) and can be estimated from



data from Eq.(4.20).

$$t_{yyy}(\tau_1, \tau_2) = E[(Y(t) - \bar{Y})(Y(t - \tau_1) - \bar{Y})(Y(t - \tau_2) - \bar{Y})] \quad (4.19)$$

$$\hat{t}_{yyy}(\tau_1, \tau_2) = \frac{1}{N} \sum_{t=1}^N [(y(t) - \bar{y})(y(t - \tau_1) - \bar{y})(y(t - \tau_2) - \bar{y})] \quad (4.20)$$

The cross bi-correlation between the disturbance  $a_{i,t}$  and the output  $Y_t$  is defined in Eq. (4.21) and can be estimated from data from Eq.(4.22).

$$t_{ya_i a_j}(\tau_1, \tau_2) = E[(Y(t) - \bar{Y})a_{i,t-\tau_1}a_{j,t-\tau_2}] \quad (4.21)$$

$$\hat{t}_{ya_i a_j}(\tau_1, \tau_2) = \frac{1}{N} \sum_{t=1}^N [(y(t) - \bar{y})a_{i,t-\tau_1}a_{j,t-\tau_2}] \quad (4.22)$$

The auto/cross correlations  $r_{yy}$  and  $r_{ya_i}$  measure the "second-order" statistical dependency. The auto/cross correlation  $t_{yyy}$ ,  $t_{yya_i}$  and  $t_{ya_i a_j}$  measure the "third-order" statistical dependency (Billings and Chen, 1989; Pearson *et al.*, 1996; Koh and Powers, 1985). In practice we do not know  $a_{i,t}$ ,  $i = 1, 2, \dots, p$ , instead,  $\hat{a}_{i,t}$  will be used to calculate the cross bi-correlations.

### 4.3.2 Estimation Methods for $Var_{A_t}[Y_t|I_0]$

#### Monte Carlo Method

It is impossible to obtain the analytical solution for the variance decomposition for Eq. (4.10) for most nonlinear time series. However Monte Carlo (MC) methods can be used to estimate the results of the variance decomposition (Archer *et al.*, 1997; Helton and Davis, 2003; Saltelli *et al.*, 2005a; Sobol', 2001; Homma and Saltelli, 1996). The MC estimates of  $f_0$ ,  $V$  and  $V_i$  are given by the following formulae,

$$\hat{f}_0 = \frac{1}{N} \sum_{m=1}^1 f((A_t)_m) \quad (4.23)$$

$$\hat{D} = \frac{1}{N} \sum_{m=1}^N f^2((A_t)_m) - \hat{f}_0^2 \quad (4.24)$$

$$\hat{V}_{A_i} = \frac{1}{N} \sum_{m=1}^N f(((A_{i,t})'_m)^{(1)}, (A_{i,t})_m^{(1)}) f(((A_{i,t})'_m)^{(2)}, (A_{i,t})_m^{(1)}) - \hat{f}_0^2 \quad (4.25)$$

where  $N$  is the number of samples generated to obtain the MC estimates,  $(A_t)_{(m)} = [(A_{1,t})_{(m)}, \dots, (A_{p,t})_{(m)}]$  is the sampled points in  $(p \times t)$ . The superscripts <sup>(1)</sup> and <sup>(2)</sup> in Eq. (4.25) indicate that we use two sampling data matrices for  $A_t$ . Both matrices have dimension  $(N, t \times p)$ . Hence, Eq. (4.25) means that for computing  $V_{A_i}$  we multiply values of  $f$  corresponding to  $A_t$  from matrix (1) by values of  $f$  from a different matrix (2), but with the columns involving  $A_{i,t}$  taken from matrix (1). The higher-order partial variances can be computed in the same manner.

The random data can be generated using several sampling methods, ordinary random sampling (McMay *et al.*, 1979), winding stairs method (Jansen *et al.*, 1994), Latin hypercube sampling (McMay *et al.*, 1979; Stein, 1987) and quasi-random sampling (Davis and Rabinowitz, 1984; Stone, 1971).

MC procedures require very intensive computation, specially for the high dimensional case. For example, if a model  $Y = f(X)$  has  $r$  factors, the total number of terms in Eq. (4.1) is as high as  $2^r - 1$ . Rabitz *et al.* (1999) defined this problem as the ‘curse of dimensionality’. The FAST method (Cukier *et al.*, 1978; Saltelli, 2002) and Sobol’s (Sobol’, 1993; Homma and Saltelli, 1996) method have been developed to cope with this dimensionality problem. In the following section, detailed information about the FAST method will be discussed.

## FAST Method

Now let us move to the Fourier Amplitude Sensitivity Test (FAST) method. It is a procedure that has been developed for uncertainty and sensitivity analysis (Cukier *et al.*, 1973; Cukier *et al.*, 1978; Saltelli, 2002). Let us assume that  $X$  is random vector with an assumed pdf  $P(X_1, X_2, \dots, X_n)$ . The  $r^{th}$  moment of  $Y$  in Eq. (4.1) is then given by:

$$\langle Y^{(r)} \rangle = \int \dots \int f^r(X_1, X_2, \dots, X_n) P(X_1, X_2, \dots, X_n) dX_1 \dots dX_n \quad (4.26)$$

The key concept of the FAST method is to convert the  $m$ -dimensional integral in Equation (4.26) into an equivalent one-dimensional integral (Cukier *et al.*, 1973; Saltelli, 2002). This conversion is made by the introduction of a transformation function  $G_l(l = 1, 2, \dots, n)$  such that:

$$X_l = G_l(\sin(\omega_l s)) \quad (4.27)$$

where  $G_l(l = 1, 2, \dots, n)$  are a set of known functions,  $\omega_l(l = 1, 2, \dots, n)$  are a set of incommensurate frequencies (defined on the next page) with one frequency assigned to each input  $X_i$ , and  $s$  is a scalar variable. By means of this transformation variations of the  $n$  inputs are transformed into the variation of the single scalar variable  $s$ . By variation of  $s$  over the range  $-\infty < s < \infty$ , Eq. (4.27) traces out a space-filling curve (defined on the next page) in the  $m$ -dimensional input space.

The conditions for this equivalence were deduced by Weyl (1938) and in terms appropriate for the FAST method by Cukier *et al.* (1973) and Cukier *et al.* (1978).

Let  $G_l(X_l)$  be the solution of

$$\pi(1 - X_l^2)^{1/2} P_l(G_l) \frac{dG_l(X_l)}{dX_l} = 1 \quad (4.28)$$

with the boundary condition  $G_l(0) = 0$ , where  $P_l$  is the pdf of  $X_l$ . Then, it is possible to transform the  $n$ -dimensional input  $X$  with the probability  $P(X)$  into the sample space  $s$ . Weyl (Weyl, 1938) demonstrated that:

$$\bar{Y}^{(r)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^T f^r(X_1(s), X_2(s), \dots, X_m(s)) ds \quad (4.29)$$

is identically equal to  $\langle Y^{(r)} \rangle$  in Equation (4.26).

$$\langle Y^{(r)} \rangle \equiv \bar{Y}^{(r)} \quad (4.30)$$

The variance  $V$  of the output  $Y$  is:

$$V = \langle Y^{(2)} \rangle - \langle Y^{(1)} \rangle^2 = \bar{Y}^{(2)} - (\bar{Y}^{(1)})^2 \quad (4.31)$$

and can be computed by evaluating one-dimensional integral in Equation (4.29).

The set of frequencies  $\{\omega_l\}$  is incommensurate, if the linear combination of the set of frequencies satisfies (Weyl, 1938):

$$\sum_{l=1}^m r_l \omega_l = 0 \quad -\infty < r_l < \infty \quad (4.32)$$

for an integer set of  $\{r_l\}$  if and only if  $r_l = 0, l = 1, 2, \dots, m$ . If the frequencies  $\{\omega_l\}$  are, in fact, incommensurate, the search curve in  $s$  space is space-filling. An incommensurate sequence is that at most one of the frequencies can be rational, with all others being irrational. Since a computer can only represent an irrational approximately, the set of integer values  $\{\omega_l\}$  used in computation cannot be truly in-

commensurate. Schaibly and Shuler (1973) proposed a sequence of approximations of incommensurability. A set of rational number  $\{\omega_l, l = 1, 2, \dots, m\}$  is approximately incommensurate in order  $W$  if:

$$\sum_{l=1}^m r_l \omega_l \neq 0 \quad (4.33)$$

for

$$\sum_{l=1}^m |r_l| \leq W + 1 \quad (4.34)$$

It is clear from this definition that incommensurability corresponds to  $W \rightarrow \infty$ .

For convenience we shall write  $f(X_1(s), X_2(s), \dots, X_m(s))$  as  $f(s)$ . By considering  $f(s)$  in the finite interval  $(-\pi, \pi)$ , Equations (4.29) and (4.31) become:

$$\bar{Y}^{(r)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^r(s) ds \quad (4.35)$$

$$\begin{aligned} \hat{V} &= \bar{Y}^{(2)} - (\bar{Y}^{(1)})^2 \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(s) ds - \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds \right]^2 \end{aligned} \quad (4.36)$$

$f(s)$  can be expanded in a Fourier series:

$$Y = f(s) = \sum_{i=-\infty}^{\infty} [A_i \cos(is) + B_i \sin(is)] \quad (4.37)$$

where the Fourier coefficients  $A_i$  and  $B_i$  are defined as:

$$\begin{aligned} A_i &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos(is) ds \\ B_i &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin(is) ds \end{aligned} \quad (4.38)$$

From Parseval's theorem we have:

$$\langle Y^{(2)} \rangle = \sum_{i=-\infty}^{\infty} [A_i^2 + B_i^2] \quad (4.39)$$

and

$$\langle Y^{(1)} \rangle^2 = A_0^2 + B_0^2 \quad (4.40)$$

The spectrum of the Fourier series expansion is defined as  $\Lambda_i = A_i^2 + B_i^2$  with  $i \in Z = \{-\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty\}$ . Since  $f(s)$  is a real-value function,  $\Lambda_i, A_i$  and  $B_i$  have the following properties:  $\Lambda_i = \Lambda_{-i}$ ,  $A_i = A_{-i}$  and  $B_i = B_{-i}$  (Saltelli et al. 1999). Using Eqs. (4.39) and (4.40) we can express the total variance  $V$  in terms of the Fourier coefficients as:

$$V = 2 \sum_{i=1}^{\infty} [A_i^2 + B_i^2] = 2 \sum_{i=1}^{\infty} \Lambda_i \quad (4.41)$$

and  $V_i$  that is the portion variance of the output variance  $V$  arising from the uncertainty of input  $X_i$  as:

$$V_i = 2 \sum_{p=1}^{\infty} [A_{p\omega_i}^2 + B_{p\omega_i}^2] = 2 \sum_{p=1}^{\infty} \Lambda_{p\omega_i} \quad (4.42)$$

The ratio  $V_i/V$ , denoted by  $S_i$ , is the estimate of the main effect of  $X_i$  on  $Y$ . Its magnitude is independent on the choice of the set of frequencies  $\{\omega_l\}$ .

The minimum sample size to be used in classic FAST (Cukier *et al.*, 1973) is

$$N_s = 2M\omega_{max} + 1 \quad (4.43)$$

where  $M$  is an integer number indicating how many higher harmonics are considered that is the maximum value for the  $p$  in Equation (4.42). Generally  $M$  is taken to be 4 or higher, and  $\omega_{max}$  is the largest frequency. More detailed information about FAST

method such as aliasing, discrete sampling, frequency selection and sample size can be found in Saltelli (2002).

**Choice of frequency** Usually, a high value is assigned to the frequency  $\omega_{A_{i,t}}$  for the group  $A_{i,t}$  and a different set of low values is assigned to all the other frequencies,  $\omega_{(A_{i,t})'}$ , for the remaining groups. Thus, by evaluating the spectrum at the frequencies  $\omega_{(A_{i,t})'}$  and their higher harmonics  $M\omega_{(A_{i,t})'}$  we can estimate the partial variance  $V_{(A_{i,t})'}$ . Then, it is easy to get the Total Sensitivity Index for the group  $A_{i,t}$  as  $ST_{A_{i,t}} = 1 - V_{(A_{i,t})'}/V$ . An automated algorithm has been proposed by (Saltelli, 2002) to select the frequencies. It is structured as follows:

- the maximum allowable frequencies for the complementary groups  $(A_{i,t})'$  must be  $max(\omega_{(A_{i,t})'}) = \frac{\omega_{A_{i,t}}}{2M}$ , where  $M$  was defined previously.
- the frequencies for the complementary set are chosen from the range  $[1, max(\omega_{(A_{i,t})'})]$  as: i) the step between frequencies must be as spread as possible; ii) the number of the factors assigned to the same frequency must be as low as possible.

The Total Sensitivity Index (TSI)  $ST_i$  of one factor  $X_i$  is defined as the sum of all the sensitivity indices involving that factor. For example, in a three-factor case, the TSI can be written as:

$$\begin{aligned}
 ST_1 &= S_1 + S_{12} + S_{13} + S_{123} \\
 ST_2 &= S_2 + S_{12} + S_{23} + S_{123} \\
 ST_3 &= S_3 + S_{23} + S_{13} + S_{123}
 \end{aligned}
 \tag{4.44}$$

**A transformation function for normal variables:**

Several transformation functions are proposed such as long-tailed and positively skewed pdf by (Cukier *et al.*, 1973) and symmetric U shape pdf by (Koda *et al.*, 1979).

The general transformation function for uniform variables was developed by (Saltelli, 2002). Based on this transformation, the transformation function for any standard distributed variable can be easily found. Since the normal variable is commonly used to describe the uncertainty of the disturbance and will be used in our simulation, the transformation for a normal variable  $x_l$  with mean zero and variance  $\sigma_{x_l}^2$  can be written as:

$$\begin{aligned} x_l &= G_l(\sin(\omega_l s)) \\ &= \sqrt{2}\sigma_{x_l} \operatorname{erf}^{-1}\left(\frac{2}{\pi} \arcsin(\sin(\omega_l s))\right) \quad l = 1, 2, \dots, p \end{aligned} \quad (4.45)$$

where  $\operatorname{erf}$  is the error function which can be written as:

$$\operatorname{erf}(x_l) = \frac{1}{\sqrt{2\pi}} \int_0^{x_l} \exp(-t^2/2) dt \quad (4.46)$$

The transformation (4.45) always returns the same points in  $K^p$  as  $s$  varies in  $(-\pi, \pi)$ . To make more efficient use of the model evaluation, Saltelli et al. (Saltelli, 2002) introduced the random phase-shift for their transformation for uniform distribution variables. Here we apply this idea for our transformation, the modified transformation (4.45) is:

$$x_l = \sqrt{2}\sigma_{x_l} \operatorname{erf}^{-1}\left(\frac{2}{\pi} \arcsin(\sin(\omega_l s + \varphi_l))\right) \quad l = 1, 2, \dots, p \quad (4.47)$$

where  $\varphi_l$  is a random phase-shift chosen uniformly in  $[0, 2\pi)$ . The advantage of (4.47) is that the starting point of the curve can be anywhere within  $K^p$ .

Space filling curves from Eq. (4.45) with  $\omega_l = 11$  and variance  $\sigma_{x_l}^2 = 1$  are plotted in Fig. 4.1 (a). A histogram of the empirical distribution of  $x_l$  corresponding to its transformation is given in Fig. 4.1 (b). This is based on a sample 629 points chosen



at a regular interval  $[-\pi, \pi]$  along the search-curve  $s$ . the histogram can be regarded as normal distribution.

To illustrate the space filling for the transformations in Equation (4.45) and (4.47), we have plotted in Fig. 4.2, (a) and (b), the sample points generated in a case with two factors in which  $\omega_1, \omega_2 = 11, 21$ .

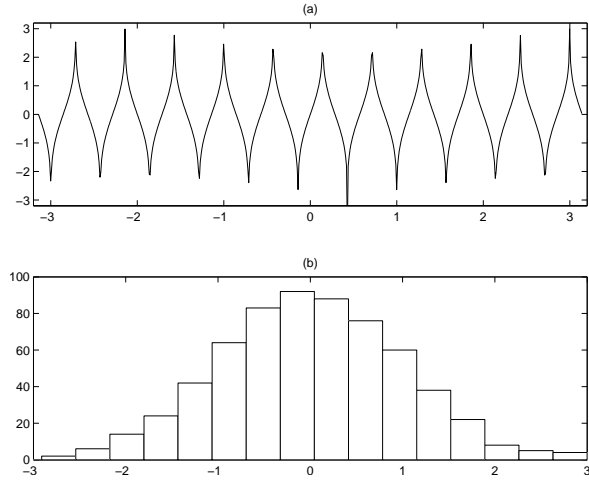


Figure 4.1: Plot of the transformation in Eq. (4.45): (a) Sampling point (b) its empirical distribution

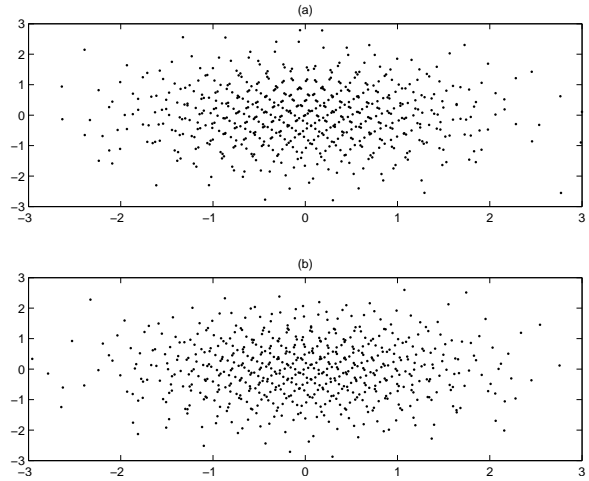


Figure 4.2: Scatterplots of sampling points in a two-factor case: (a) transformation in Eq. (4.45) (b) transformation in Eq. (4.47)

### 4.3.3 Estimation Methods for $E_{I_0}[Var_{A_t}[Y_t|I_0]]$

Due to the complexity of nonlinear systems, it is impossible to obtain analytical solution for the term  $E_{I_0}[Var_{A_t}[Y_t|I_0]]$ , Monte Carlo methods can be used to estimate these effects. An ordinary random sampling method (McMay *et al.*, 1979) is employed to generate the random initial condition. Since the initial sample may be unrepresentative, it is usual to ignore the first few iterations in Markov Chain Monte Carlo (MCMC) methods (Gilks *et al.*, 1996). Iterates within an initial transient phase or *burn-in* period are discarded. This idea will be applied to estimate  $E_{I_0}[Var_{A_t}[Y_t|I_0]]$ , the detailed procedures are described in the following steps: i) A burn-in period, say  $t_{burnin}$ , is used in order that the sequence is allowed to move from the starting value to a more general situation, ii) if the simulation starts at time 0, the initial condition  $I_0$  is collected at the end of burn-in period  $t_{burnin}$ . Then the conditional variance decomposition  $Var_{A_t}[Y_t|I_0]$  will be estimated using MC and FAST methods, iii) to repeat step i) and ii), iv) the average values of the conditional variance decomposition from step iii) are the estimate of  $E_{I_0}[Var_{A_t}[Y_t|I_0]]$ .

### 4.3.4 Estimation of the Analysis of Variance from Input/Output Data

Most often, the nonlinear model structure and the location of the measured disturbances are not available. However, these need not be known in order to perform the analysis of variance such as variance decomposition. Because the function  $f$  in Eq. (4.8) can be assumed as a variety of forms, the identification of nonlinear systems becomes a much more difficult task than the linear counter-part. However, it is often possible to approximate function  $f$  by some functions such as polynomial models and exponential time-series models (Billings and Chen, 1989). If the nonlinear function is continuous it can always be arbitrarily well approximated by polynomial models

(Chen and Billings, 1989). When the system is severely non-linear other choices of approximate functions could be considered. Being linear-in-parameters, the polynomial models can readily be estimated using linear least squares methods as can be seen in Section 3.5.2. Many real-world systems exhibit complex nonlinear characteristics, the neural network method have been developed for modeling complex nonlinear systems(Chen *et al.*, 1990; Chen and Billings, 1992). It is shown that a neural network which has the capability to model sophisticated nonlinear relationships can provide a ideal means of modeling complicated nonlinear systems.

Input/ouput data can be used to estimate nonlinear models using polynomial models as the following procedures:

- step 1: The first step is to fit individual linear/nonlinear autoregressive models to the measured disturbance:

$$D_{i,t} = f_{Di}(D_{i,t-1}^*, \hat{a}_{i,t-1}^*) + \hat{a}_{i,t} \quad (4.48)$$

$\hat{a}_{i,t}$ ,  $i = 1..p$  is an estimate of the independent driving force for the  $i^{th}$  measured disturbance.

- step 2: The next step is to fit a lagged regression to the output into the polynomial NARMAX model using the orthogonal least square algorithm, with the past values of  $Y_t, U_t$  and the measured disturbance  $(\hat{a})_{i,t}$ ,  $i = 1..p$ :

$$Y_t = f_P(Y_{t-1}^*, U_{t-b}^*, \hat{a}_{1,t}^*, \dots, \hat{a}_{q,t}^*, \hat{a}_{0,t-1}^*) + \hat{a}_{0,t} \quad (4.49)$$

The model order can be determined using an Error reduction ratio (err) procedure (Chen and Billings, 1992). Akaike's information criterion (AIC) can be used to compromise between the performance of the model (variance of residuals) and complexity of the model. Once the model structuring is completed,

the model parameter are estimated by using the least-square algorithm. The detailed information about model order selection and validation can be found in (Chen *et al.*, 1989; Billings and Chen, 1989).  $\hat{a}_{0,t}$  is an estimate of the independent driving force for the unmeasured disturbance.

- step 3: the cross-correlation between two various disturbances can be estimated by calculating the sample cross-correlation of the residuals  $\hat{a}_{i,t}$ ,  $i = 1, 2$  as:

$$\phi_{\hat{a}_i \hat{a}_j}(\tau) = \frac{1}{N} \sum_{t=1}^N (\hat{a}_{i,t+\tau} \hat{a}_{j,t}) \quad \forall \tau \text{ and } i \neq j \quad (4.50)$$

If  $\phi_{\hat{a}_i \hat{a}_j}(\tau)$  is statistically zero for  $i \neq j$ , then the uncorrelated assumption of the measured and unmeasured disturbances is confirmed.

With the approximate nonlinear model built in the above 3 steps, the variance decomposition can be conducted using the aforementioned methods. Instead of using the polynomial approximation for the nonlinear model identification in the above 3 steps, neural network techniques also can be used.

## 4.4 Simulations

### 4.4.1 A Random Coefficient Autoregressive Model

Consider the first order random coefficient autoregressive RCA(1) model:

$$Y_t = (\alpha + a_{1,t})Y_{t-1} + a_{2,t} \quad (4.51)$$

where  $a_{i,t}$ ,  $i = 1, 2$  are i.i.d. normally distributed with mean zero and variance  $\sigma_i^2$  and independent of initial condition  $Y_0$ .  $\{a_{1,t}\}$  and  $\{a_{2,t}\}$  are independent of each other and  $\alpha$  is a real constant. Using the criteria given in (Nicholls and Quinn, 1982; Tong,

1990; Pourahmadi, 1986),  $\alpha^2 + \sigma_1^2 < 1$  is a sufficient condition for model (4.51) to be ergodic.

The infinite solution of  $Y_t$  is given by:

$$Y_t = \sum_{j=0}^{\infty} \pi_j a_{2,t-j} \quad (4.52)$$

where  $\pi_0 = 1$  and

$$\pi_j = \prod_{i=0}^{j-1} (\alpha + a_{1,t-i}), \quad j = 1, 2, \dots$$

Letting  $\tau = \alpha^2 + \sigma_1^2 < 1$ , it is straightforward to show (Hwang and Basawa, 2005):

$$E[Y_{t \rightarrow \infty}] = 0 \quad \text{and} \quad \text{var}(Y_{t \rightarrow \infty}) = \frac{\sigma_2^2}{1 - \tau^2} \quad (4.53)$$

Using the method in Eqs. (4.31) and (4.32), the analytical solution for the variance decomposition at time  $t \rightarrow \infty$  is:

$$\begin{aligned} V_1 &= 0 \\ V_2 &= \frac{\sigma_2^2}{1 - \alpha^2} \\ V_{12} &= \frac{\sigma_1^2 \sigma_2^2}{(1 - \alpha^2)(1 - \tau^2)} \end{aligned} \quad (4.54)$$

Firstly, the variance decomposition for an ergodic nonlinear time series is illustrated using the model in (4.51). The parameters are assigned as 0.6, 0.3 and 0.4 respective to  $\alpha^2$ ,  $\sigma_1^2$  and  $\sigma_2^2$ .

Since this example is initial condition independent according to the criteria (Tong, 1990; Pourahmadi, 1986),  $Y_0$  is set to zero. Since the initial condition includes terms  $Y_{t-1}$ ,  $a_{1,t} Y_{t-1}$ , and  $a_{2,t}$ ,  $t \leq 1$ , the auto-correlation  $r_{yy}$ , cross-correlation  $r_{ya_2}$  and cross bi-correlation  $t_{yya_1}$  defined in Section 4.3.1 are estimated using 500 realizations.

The results are shown in Fig. 4.3. It shows that this RCA(1) model's memory can be approximated to be 10. It means that effect of the initial condition  $Y_{t-1}, a_{1,t}Y_{t-1}$ , and  $a_{2,t}, t \leq 1$  on the present output value  $Y_{10}$  would be insignificant.

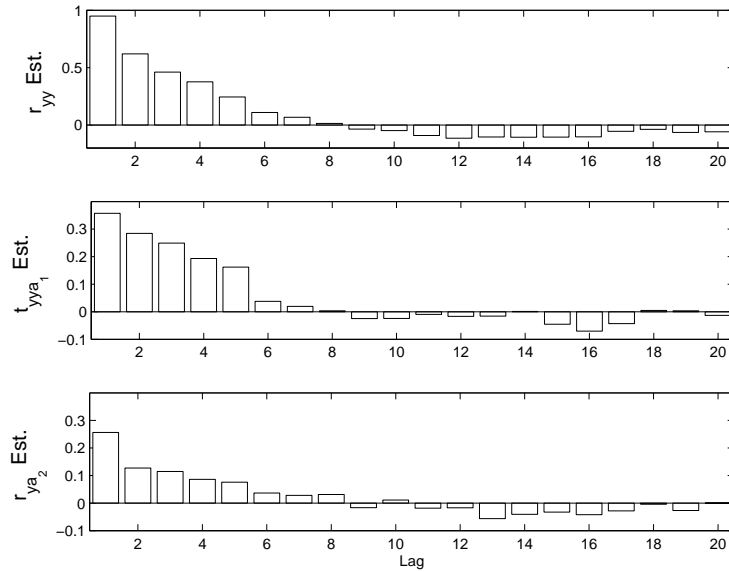


Figure 4.3: Auto-correlation, cross-correlation and cross bi-correlation for RCA(1) model

Now the infinite series described in Eq. (4.52) can be approximately written as:

$$Y_{t \rightarrow \infty} \simeq Y_{10} = \sum_{j=0}^{m=10} \pi_j a_{2,t-j} \quad (4.55)$$

where  $\pi_0 = 1$  and

$$\pi_j = \prod_{i=0}^{j-1} (\alpha + a_{1,t-i}), \quad j = 1, 2, \dots$$

The variance of  $Y_{10}$  is:

$$Var(Y_{10}) = \frac{(1 - \tau^{20})\sigma_2^2}{1 - \tau^2} \quad (4.56)$$

For our simulation,  $Var(Y_{10}) = 1.148$  is accounting for 97.6% of the limiting variance  $Var(Y_{t \rightarrow \infty}) = 1.176$ .

To decompose the variance for  $Y_{10}$ , we use the different choices of frequencies

shown in Table 4.1.  $N_s$  denotes the sample size used in FAST. Fig. 4.4 shows an example of the extend FAST method with the transformation in Eq. (4.45) applied to the RCA(1) model with finite memory 10. One hundred estimates, obtained using different starting points, of the partial and total variances are computed. The boxplots of their summary statistics, for each disturbance group, are plotted against the sample size  $N_s$ . The estimates converge to the analytical values and the precision of the estimates increases, as the number of samples (or the spread of frequencies) increases.

The results of variance decompositions for  $Y_t$ ,  $t = 1, \dots, 15$  are plotted in Fig. 4.5. The high frequency and its complementary frequencies are chosen to be 152 and  $\{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$  (with step 2) respectively. It shows that the estimates of the variance decomposition results appears to converge to the true values when the time horizon increases.

Table 4.1: Sets of Frequencies obtained by using the automated algorithm

Simulation No.	$N_s$	High Freq.	Complementary Freq.		Step
			Max	Low Freq.	
1	65	8	1	$\{1,1,1,1,1,1,1,1\}$	0
2	641	80	10	$\{1,2,3,4,5,6,7,8,9,10\}$	1
3	1217	152	19	$\{1,3,5,7,9,11,13,15,17,19\}$	2
4	2369	296	37	$\{1,5,9,13,17,21,25,29,33,37\}$	4

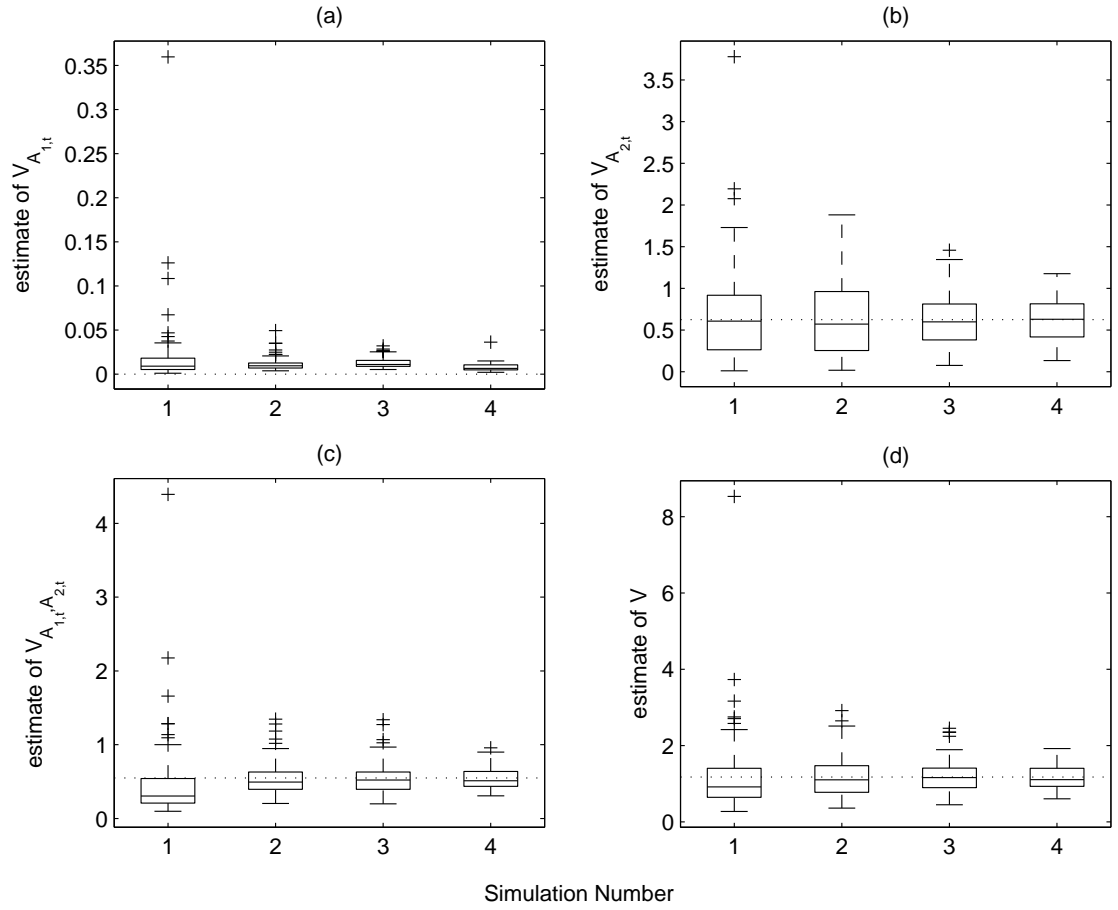


Figure 4.4: Boxplots of 100 estimates of the variance decomposition for RCA(1) model; analytical values of the partial and total variances are shown by dotted lines.



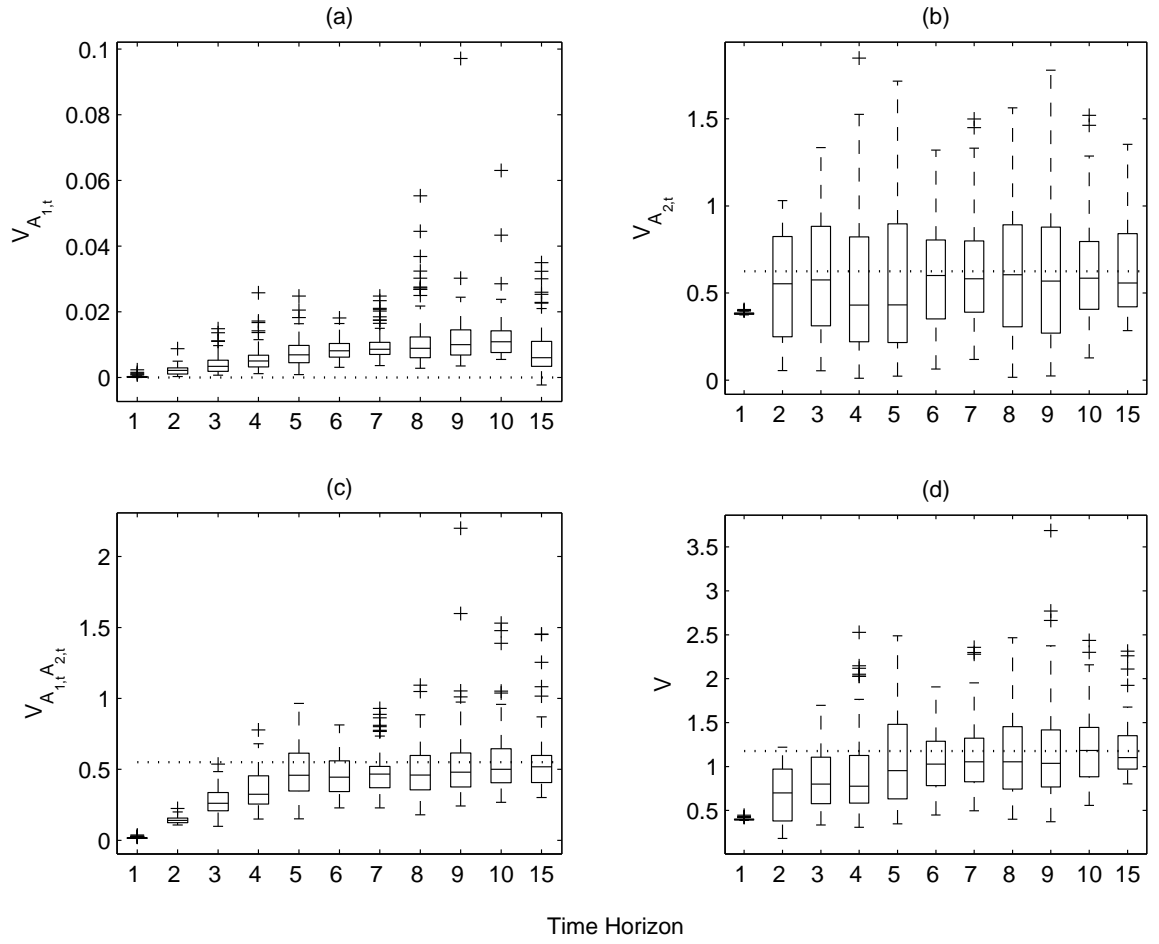


Figure 4.5: Boxplots of 100 estimates of the variance decomposition of  $Y_t$ ,  $t = 1, \dots, 15$  for the RCA(1) model ; analytical values of the partial and total variances  $t \rightarrow \infty$  are shown by dotted lines.

#### 4.4.2 A Volterra Model

A simple nonlinear system, example 1 studied in Chapter 3, is used for our first simulation example. Two additive linear disturbances are assumed to affect the output  $Y_t$  directly. The model can be expressed as:

$$\begin{aligned}
 Y_t = & 0.2U_{t-3} + 0.3U_{t-4} + U_{t-5} + 0.8U_{t-3}^2 \\
 & + 0.8U_{t-3}U_{t-4} - 0.7U_{t-4}^2 - 0.5U_{t-5}^2 - 0.5U_{t-3}U_{t-5} + \sum_{i=1}^2 D_{i,t} \quad (4.57)
 \end{aligned}$$

where the disturbance  $D_{1,t}$  is the measured disturbance which is in the form of an ARIMA(2,0,0) process:

$$D_{1,t} = \frac{a_{1,t}}{1 - 1.6q^{-1} + 0.8q^{-2}} \quad (4.58)$$

and the second disturbance  $D_{2,t}$  is the unmeasured disturbance which can be represented as an AR(1) process:

$$D_{2,t} = \frac{a_{2,t}}{1 - 0.9q^{-1}} \quad (4.59)$$

$a_{1,t}$  and  $a_{2,t}$  are the i.i.d. normal variables with zero mean and variance 0.03 and 0.05 respectively. The variances of the disturbances  $D_{1,t}$  and  $D_{2,t}$  are equal to 0.1997 and 0.2632 respectively. The disturbances  $D_{1,t}$  and  $D_{2,t}$  are uncorrelated. The same PI controller  $U_t = -\frac{0.3-0.2q^{-1}}{1-q^{-1}}(Y_t - Y_{sp})$  which is used in the simulation at Section 3.6.1 is also used here. A closed-loop data set consisting of five hundred samples for  $D_{1,t}$ ,  $D_{2,t}$ ,  $U_t$  and  $Y_t$  is shown in Fig. 4.6.

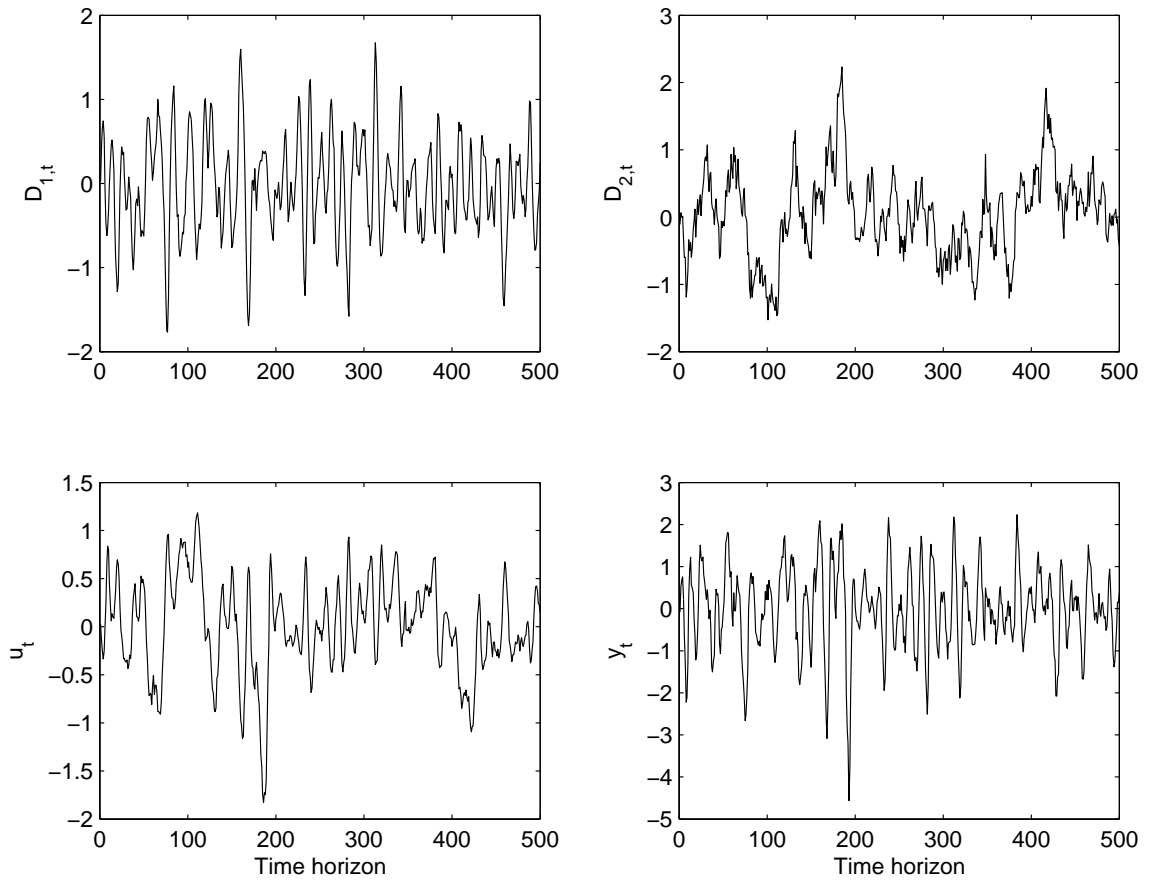


Figure 4.6: A sample of five hundred closed-loop Volterra Model data points with the measured and unmeasured disturbances

## Variance Decomposition with a Initial Condition $I_0 = 0$

Since it is impossible to obtain an analytical solution for the variance decomposition, a numerical solution, Monte Carlo (MC) method, is used to estimate the partial / total variances. Random sampling is used to generate the samples for the MC calculation (McMay *et al.*, 1979). Based on the initial condition  $I_0 = 0$ , the variations of output  $Y_t, t = 1, 2, \dots, 40$  is shown Fig. 4.7. From Fig. 4.7, we can observe that the variations of output  $Y_t$  between time  $t = 20$  and time  $t = 40$  are not significantly different. So finite memory length, say 20, is picked, the variance decomposition for this memory length can be considered as a approximation for the variance decomposition for the infinite memory length.

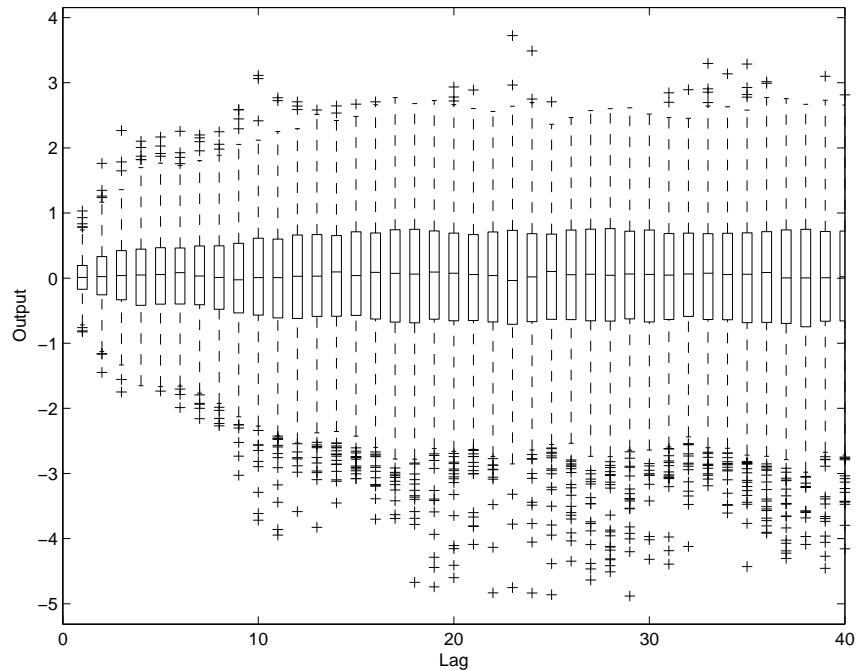


Figure 4.7: The box plots of output  $Y_t, t = 1, 2, \dots, 40$

The procedures for estimating the partial variance  $V_1$  is shown at the following steps: i) one sample set of  $A_{1,t}^1 = [a_{1,1}, \dots, a_{1,20}]$  are generated, ii) the other sample set of  $A_{2,t}^1 = [a_{2,1}, \dots, a_{2,20}]$  are generated, the output  $y_{20}^1$  is calculated, iii) the step ii) is repeated for two hundred times to collect the output  $y_{20}^i, i = 1, \dots, 200$ . iv) to estimate

the condition mean  $E(y_t|A_{1,t}^1)$  using the output data from step iii), v) to repeat steps i)-iv) two hundred times and to estimate the partial variance  $V_1$  using two hundred means. The same procedures must be repeated for estimating the partial variance  $V_2$ . The estimates of the partial and total variances are listed in column 2 at Table 4.2.

Table 4.2: Example 2: Estimates of partial / total variance using MC and FAST methods with a constant initial condition

	MC	FAST
$V_1$	0.8540	0.8405(0.7071)
$V_2$	0.2061	0.1904(0.1661)
$V_{12}$	0.0246	0.0276(0.0978)
$V$	1.0847	1.0585(0.7551)

The partial and total variances are also estimated using the FAST method with a transformation function for a normal variable. The high frequency 16 is assigned to the unmeasured disturbance  $D_{2,t}$  and a low one 2 to the measured disturbance  $D_{1,t}$ . The sample size is 129 for each estimation. The estimation is repeated two hundred times. Unlike the MC methods, the estimations must be done individually, the partial and total variances can be estimated simultaneously from the FAST method for the two factor case. The estimates of the partial and total variances using the FAST method are also listed in column 3 at Table 4.2. The values in parentheses in column 3 at Table 4.2 is the standard deviation of the estimates of partial and total variance using FAST method. The simulation was taken by a Compaq 5340CA 5000 series with *Intel*<sup>®</sup> *Pentium*<sup>®</sup> 4 processor 1.7 GHZ and 512 MB SyncDRAM. The calculation times are 1.41 and 0.056 min respective to MC and FAST methods. The calculation time of FAST methods will significantly save time in the following simulation which includes the consideration of initial condition uncertainty. The comparative box plots of the quality estimates are shown in Fig. 4.8.

The sample size  $200 \times 200 \times 20$  is used to estimate the  $V_1$  using the MC method.

The first 200 samples for  $A_{2,20}$  are used to estimate  $E_{A_{2,20}}[Y_{20}|A_{1,20}]$  and the second 200 samples are used to estimate  $V_{A_{1,20}}(E_{A_{2,20}}[Y_{20}|A_{1,20}])$ . The third sample size 20 is from memory length 20. The estimates of the partial and total variances using the MC method are also listed in column 2 at Table 4.2. The estimate of  $V_1$  using the FAST method with only  $200 \times 129 \times 20$  sample size is close to the MC estimate. The sample size for FAST method means: 129 samples for FAST method, 20 is the memory length, and 200 is the repeat number for the FAST method. Furthermore, with the same sample set, the FAST method can compute all values of  $V_1$ ,  $V_2$ ,  $V_{12}$  and  $V$  at the same time. The new sample set is required by the MC method for each new estimate of  $V_2$ ,  $V_{12}$  and  $V$ . The MC method is computationally more expensive in terms of model evaluations.

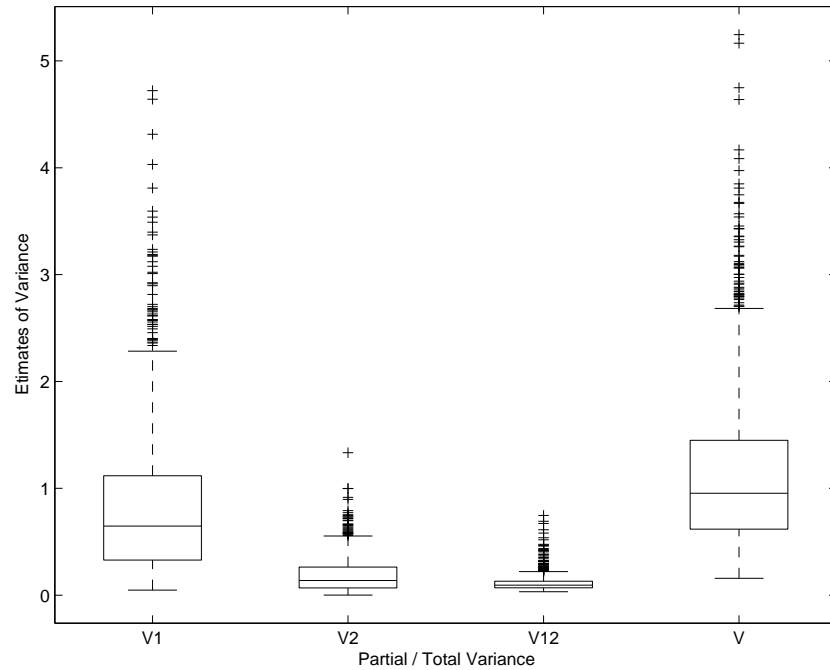


Figure 4.8: The comparative box plots of the quality estimates of partial/total variance using the FAST method

## Variance Decomposition with Uncertain Initial Condition $I_0$

For our simulation, the burn-in period is 30. The sample size  $200 \times 200 \times 200 \times 20$  is used to estimate the  $V_1$  using the MC method. The first sample size 200 is for the estimation of the partial variance  $V_1$  with consideration of initial condition uncertainty  $E_{I_0}[V_{A_{1,20}}(E_{A_{2,20}}[Y_{20}|A_{1,20}])]$ . The rest sample sizes have the same definition used in the above simulation. Similarly, the sample size  $200 \times 200 \times 129 \times 20$  is used for the FAST method. The effects of the uncertainties of the initial conditions on the output variance decomposition are listed in Table 4.3. The calculation times are 246.4 and 15.2 min respective to MC and FAST methods. The FAST method will save a lot of computation time for this simulation. The comparative box plots of the quality estimates using the FAST method are shown in Fig. 4.9.

Table 4.3: Example 2: Estimates of partial / total variance using MC and FAST methods with the uncertain initial condition

	MC	FAST
$V_1$	0.7302	0.7267(0.6436)
$V_2$	0.2103	0.2055(0.1769)
$V_{12}$	0.0731	0.0775(0.0798)
$V$	1.0136	1.0097(0.7143)

Since the expected values of the initial conditions are equal to zero, the results of variance decomposition with a constant initial condition listed in Table 4.2 can be considered as the variance decomposition of the expected values of the initial conditions for example  $V_{A_{1,20}}(E_{A_{2,20}}[Y_{20}|(A_{1,20}, E[I_0])])$ . The results in Table 4.3 are the expected values of the initial conditions of the variance decomposition  $E_{I_0}[V_{A_{1,20}}(E_{A_{2,20}}[Y_{20}|A_{1,20}])]$ . In general these two calculations are not equal. The results in Table 4.2 and Table 4.3 illustrate these inequalities. The differences between these two terms may or may not be significant, it will depend on the model structure and disturbance statistics. The differences between these two terms with

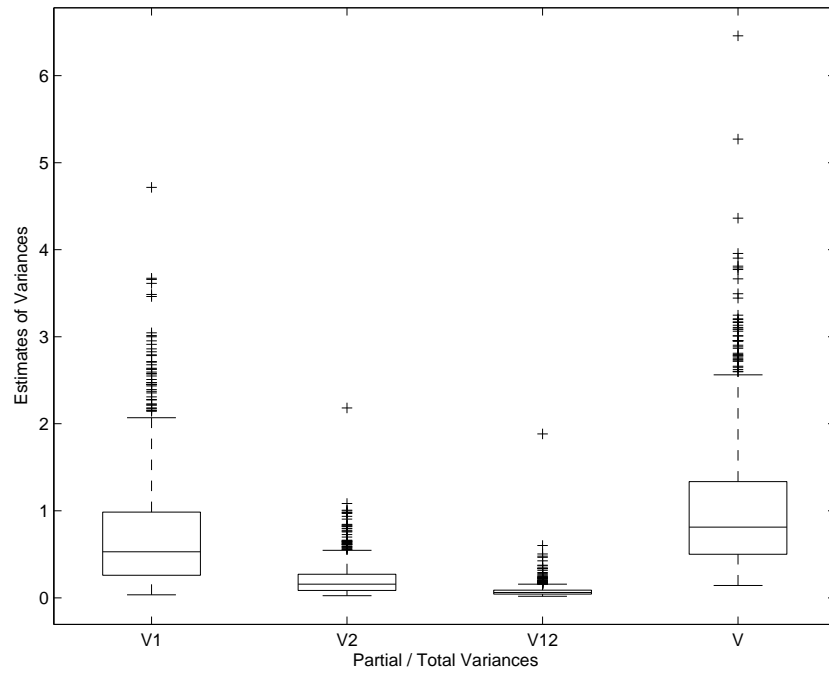


Figure 4.9: The comparative box plots of the quality estimates of partial/total variance with the uncertain initial condition using the FAST method

different time horizons are listed in Table 4.4. The data in column 3 in Table 4.4 are obtained using the MC method and the data in column 4 are calculated using the FAST method with the same parameter values such as frequency used in the previous simulation. From Table 4.4, we can see that the differences between these two terms will become smaller as the time horizon increase for this simulation example. An interesting phenomena in Table 4.4 is that the conditional output variance given a constant initial condition is much less than the output variance for the output  $Y_{10}$  case. It means that the results of the variance decomposition based on uncertain initial conditions may be significantly different from the results based on a constant initial conditions.



Table 4.4: Example 2: Estimates of partial / total variance of the constant initial conditions and the uncertain initial condition for the different time horizons

$Y_t$	Variance	Constant IC	Uncertain IC
t=10	$V$	0.8168	0.4705
	$V_1$	0.6133	0.2923
	$V_2$	0.1764	0.1572
	$V_{12}$	0.0270	0.0208
t=20	$V$	1.0847	1.0097
	$V_1$	0.8540	0.7267
	$V_2$	0.2061	0.2055
	$V_{12}$	0.0246	0.0775
t=30	$V$	1.1471	1.1653
	$V_1$	0.8526	0.8561
	$V_2$	0.1897	0.1859
	$V_{12}$	0.1048	0.1233

## 4.5 Conclusions

Reports of theoretical and applied work on performance assessment and variance decomposition of MISO processes for linear systems have appeared in the literature for the case in which the different disturbances are independent (Stanfelj *et al.*, 1993; Desborough and Harris, 1993) and the different disturbances are correlated (Seppala, 1999). This chapter has provided a preliminary analysis of variance results for MISO nonlinear processes.

We have addressed the case where there is no cross-correlation among the disturbances within the nonlinear systems which can be represented in the NARMAX models. We have shown that the variance decomposition of the nonlinear time series and nonlinear stochastic systems can be estimated using the ANOVA-like decomposition in Eq. (4.1). Since most often the variance decomposition of the nonlinear stochastic / dynamic systems is dependent of the initial conditions, the modified ANOVA-like decomposition method is proposed to cope with the initial condition

uncertainty. The cheaper computational method, extended FAST method, is used to reduce the calculation dimensionality. Applications of the methodology to the examples indicate that this approach gives very credible estimates of the variance decomposition.

The nonlinear stochastic / dynamic systems do not necessarily need to be known. The nonlinear systems can be approximately modeled by a polynomial NARMAX model. The methods for structure selection, parameter estimation, and diagnostic checking for polynomial NARMAX model have been developed (Chen *et al.*, 1989; Billings and Chen, 1989). The variance decomposition can be done based on the approximate polynomial NARMAX model.

The variance decomposition for the nonlinear dynamic / stochastic systems and time series discussed in this chapter is based on the assumption that the different disturbances are uncorrelated. This assumption is not always applicable in practice. The investigation into the effects of cross-correlated disturbances on analysis of variance for nonlinear MISO systems may be necessary for the extension of nonlinear variance decomposition problems.

# Chapter 5

## Parameter Uncertainty Effects on Sensitivity Analysis / Variance Decomposition

The variance decomposition, variance-based sensitivity analysis, discussed in Chapter 4 is based on the constant model parameters. However, parameters of empirical or mechanistic models usually contain some uncertainties (e.g. estimated from data). The uncertainties associated with these estimated parameters should be automatically propagated when the model is used for variance analysis. The topic of parameter uncertainty effects on the variance decomposition in this thesis is not only valid for the variance decomposition in Chapter 4, but also valid for the more general area, sensitivity analysis. It is the reason that in the title the term "sensitivity analysis" is used instead of "variance decomposition".

### 5.1 Introduction

Sensitivity Analysis (SA) methods are well known in the engineering literature (Cukier *et al.*, 1973; McMay, 1995; Saltelli *et al.*, 1999). They have been used across various

disciplines such as weather forecasting, chemical engineering and economics. Readers are referred to (Helton *et al.*, 2006; Helton and Davis, 2002; Ionescu-Bujor and Cacuci, 2004; Cacuci and Ionescu-Bujor, 2004; Saltelli *et al.*, 2006; Saltelli *et al.*, 2005*b*; Kleijnen and Helton, 1999; Saltelli *et al.*, 2004; Frey and Patil, 2002; McMay, 1996; Saltelli and Scott, 2000) for recent reviews of these topics and a number extension and modifications.

The objective of SA is to ascertain how the model output depends on its input factors. This behavior is described by the state or output variables which we indicate in general as dependent variables changing in time and/or state space. In engineering fields, the inputs include the physicochemical parameters of the model (such as those related to reaction kinetics and thermodynamic equilibria), as well as initial condition and operating conditions of the systems. The physicochemical parameters are measured experimentally or estimated from experiment data and therefore are always subject to uncertainties. Initial and operating conditions may also be uncertain for a variety of reasons.

Consider a system described by a uniresponse mathematical regression model of the form:

$$Y = f(X, \Theta) \quad (5.1)$$

where  $Y$  is an observed value of the response variable,  $X$  is a vector including regressive variables,  $\Theta^T = (\theta_1, \theta_2, \dots, \theta_p)$  is a  $1 \times p$  vector of unknown parameters,  $f$  is a function which may be linear or nonlinear in the parameters. In the literature, by *Parametric Sensitivity*, refer to the effect of variation of the input parameters, such as  $\Theta$  in Eq. (5.1), on the system behavior  $Y$ .

Parametric sensitivity may be useful in uncertainty analysis, model validation and reduction, examples being provided in the fields of chemical kinetics (Cukier *et al.*, 1973; Atherton *et al.*, 1975; Cukier *et al.*, 1978; Varma *et al.*, 1999), biophysical systems (Gunawan *et al.*, 1995), risk analysis (Saltelli, 2002; Draper *et al.*, 1999;

Saltelli and Tarantola, 2002) and economics (Planas and Depoutot, 2002). In this paper, by *Regressive Sensitivity*, we mean the effect of variations of the regressive variables such as  $X$  described in Eq. (5.1) on the system behavior. Regressive sensitivity is useful in variance decomposition, optimization and performance assessment in various fields such as food sciences (Cliff *et al.*, 1995; Tienungoon *et al.*, 2000), social science (McCarthy *et al.*, 1995), risk analysis (Bowyer and Danson, 2004; Francos *et al.*, 2003), reliability analysis (Hopperstad *et al.*, 1999), chemical engineering performance assessment (Desborough and Harris, 1993) and econometrics (Saisana *et al.*, 2005).

Three general classes of sensitivity analysis techniques can be defined (Campolongo *et al.*, 2000) as:

- Factor screening methods (e.g., one-at-a-time (OAT) methods, factorial designs);
- Local method (e.g., differential or nominal value analysis);
- Global method (e.g., Monte Carlo analysis, FAST and Sobol's method).

Screening methods are typically qualitative tools providing only a ranking of the importance of regressors / parameters. Screening method can usually be further characterized as retaining properties of either local or global methods. One advantage of screening methods is that they are computationally efficient. Local methods tend to address only a specific point, or local region in the regressor / parameter space. Mostly based on partial derivatives, local methods usually embody a univariate assessment framework among regressors / parameters. As such they are valid only in small range taken about the nominal value of regressor / parameter. Global methods apportion the total variance of the output to the each input factor and the interactions among the input factors. All input factors are allowed to vary simultaneously over their ranges that take into account the shapes of their probability density functions. Global

methods are far more computationally demanding and involve various methods of sampling the input factor space (e.g., random sampling, quasi-random sampling, Latin hypercube sampling(LHS), etc) (Helton and Davis, 2003). A review of global SA methods, including Monte Carlo based regression-correlation measures and variance-based method can be found in Saltelli and Scott (2000) and Frey and Patil (2002).

A class of global sensitivity analysis techniques that are known as variance-based SA methods mainly includes three methods, namely, the correlation ratio-based methods (McMay, 1995; McMay, 1996), the FAST (Cukier *et al.*, 1973; Cukier *et al.*, 1978; Saltelli *et al.*, 1999), and the Sobol' methods (Sobol', 1993). The SA in this paper is based on the FAST and Sobol' methods. The full variance-based sensitivity analysis for the model in Eq. (5.1) can be straightforwardly conducted with the consideration of both regressors  $X$  and parameters  $\Theta$  uncertainties. However, the full SA may not provide satisfactory results for some variance analyses. In some variance analyses, the regressor uncertainty may have higher priority than the parameter uncertainty. For example, in the control performance assessment study field, the SA / variance decomposition which is able to decompose the variance of output according to each disturbance will be considered by an control engineer at first. The results can be directly used to help the output variance reduction. Then the parameter uncertainty effects on these results will be considered. This problem motivates us to propose a new approach.

The problem tackled in this chapter includes two objectives: i)the main objective is to investigate the effect of variations of the regressive variables such as  $X$  described in Eq. (5.1) on the system behavior  $Y$  (For convenience, we define the term *Regressive Sensitivity Analysis* (RSA) to denote the main objective), and ii) the auxiliary objective is to evaluate the effects of the parameter uncertainty on RSA. The difference between our approach and the full SA will be demonstrated in the following section. The general solutions for general nonlinear models are proposed. It is impossible

to deduce analytical solutions for general nonlinear models, Computational methods such as Monte Carlo method must be applied.

$f(X, \Theta)$  is a nonlinear mapping when models are developed from fundamental principles that are based on the laws of conservation of mass, energy and momentum. However, in most cases, the information necessary to build a fundamental model is lacking, and the empirical models are a viable alternative. Linear-in-parameter models (e.g., polynomial models) are the most widely used for the empirical models. Polynomial models are quite simple, typically involving a bilinear product of the form  $f(X, \Theta) = H^T(X)\Theta$ , where  $H(X)$  is a polynomial in the regressive variables. It is shown that if the model is linear in its parameters, only the results of RSA based on the mean values of parameters are needed to deduce the effects of parameter uncertainties on RSA.

The layout of this chapter is as follows. In Section 5.2, the global variance based SA is introduced. In Section 5.3, an approach for incorporating the parameter uncertainty effects on RSA for general nonlinear models is proposed. This makes uses of a fundamental identity cited in (Parzen, 1962). In most cases, numerically intensive methods are required to account for these uncertainties. However, for models that are linear in the parameters, analytical solutions are readily obtained. Some simple methods to cope with the more general case are proposed. In Section 5.4, two examples are used to illustrate the parameter uncertainty effects on the regressive sensitivity analysis. This is followed by the conclusions and a discussion of areas for future work.

## 5.2 Variance-based Sensitivity Analysis

Variance-based sensitivity analysis (SA) aims to ascertain how much a model output depends on each or some of its input factors. A review of variance-based SA can

be found in (Chan *et al.*, 2000). The importance of the given independent inputs  $X_1, X_2, \dots, X_p$  can be measured via the so-called sensitivity index, which is defined as the fractional contribution to the output variance due to the uncertainties in inputs. It can be calculated by using the following decomposition formula for the total output variance  $V(Y)$  of the output  $Y$  (Archer *et al.*, 1997; Cox, 1982):

$$Var[Y] = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12\dots p} \quad (5.2)$$

where

$$\begin{aligned} V_i &= Var[E(Y|X_i = x_i^*)] \\ V_{ij} &= Var[E(Y|X_i = x_i^*, X_j = x_j^*)] \\ &\quad - Var[E(Y|X_i = x_i^*)] - Var[E(Y|X_j = x_j^*)] \end{aligned} \quad (5.3)$$

and so on.  $E[Y|X_i = x_i^*]$  denotes the expectation of  $Y$  conditional on  $X_i$  having a fixed value  $x_i^*$ , and  $Var$  stands for variance over all the possible values of  $x_i^*$ . Decomposition (5.2) is used in statistics as the analysis of variance (ANOVA) of the output  $Y$  (Archer *et al.*, 1997; Rabitz *et al.*, 1999). If a model has  $p$  factors, the total number of terms in Eq. (5.2) is as high as  $2^p - 1$ .

The sensitivity indices  $S_{i_1, i_2, \dots, i_s}$  (Sobol', 1993; Cukier *et al.*, 1973; Saltelli *et al.*, 1999) can be computed as:

$$S_{i_1, i_2, \dots, i_s} = \frac{V_{i_1, i_2, \dots, i_s}}{V} \quad (5.4)$$

All the  $S_{i_1, i_2, \dots, i_s}$  are nonnegative and their sum is:

$$S = \sum_{i=1}^n S_i + \sum_{1 \leq i < j \leq n} S_{ij} + \dots + S_{12\dots n} = 1 \quad (5.5)$$

While the initial problem formulation appears restrictive, it is possible to incor-



porate additive or multiplicative measurement uncertainties with the inclusion of an additional variable. These measurement uncertainties can also be made functions of the regressive parameters, allowing for flexible descriptions of measurement uncertainty.

Evaluation of the sensitivity indices is most often accomplished through analytical means or through variations of Monte-Carlo methods. Efficient numerical methods are required for large scale problems (Rabitz *et al.*, 1999). The FAST (Cukier *et al.*, 1973; Saltelli *et al.*, 1999) and Sobol's (Sobol', 1993; Homma and Saltelli, 1996) methods have been developed to cope with this dimensionality problem. The Sobol's indices can be computed with the Monte Carlo integral (Homma and Saltelli, 1996; Sobol', 2001). Many applications of these methods can be found in various chapters in the book (Saltelli and Scott, 2000). The details on the FAST methods have been discussed in Chapter 4. The information about Sobol's method will be introduced in the following section.

#### **Sobol's Method:**

The sensitivity method described here was developed by Sobol' (Sobol', 1993), based on his earlier work on the Fourier Haar series (Sobol', 1969). Consider a system that can be expressed by Equation (5.1). We are interested in determining the sensitivity of the output  $Y$  to variation in each input  $X_i, i = 1, 2, \dots, p$  where the input vector  $X$  is a random vector with joint probability density function  $P(X)$ . Through estimating the sensitivity of a function  $f(X, \Theta)$  with respect to each individual input  $X_i$ , the sensitivity indices are developed to reflect the contribution of each input on the output variance.

Let the inputs to the function  $Y = f(X, \Theta) = f(X_1, X_2, \dots, X_n; \Theta)$  be defined on the n-dimensional unit cube:

$$K^n = \{X : 0 \leq X_i \leq 1, i = 1, 2, \dots, n\}$$

and

$$\Theta \in R^p \quad (5.6)$$

This definition is not restrictive for any ranges of the input factors since any bounded ranges can be transformed into this hypercube domain.

It is possible to decompose  $f(X, \Theta)$  into summands of increasing dimensionality (Sobol', 1993), namely:

$$f(X_1, X_2, \dots, X_n; \Theta) = f_0 + \sum_{i=1}^n f_i(X_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n f_{ij}(X_i, X_j; \Theta) + \dots + f_{12\dots n}(X_1, X_2, \dots, X_n; \Theta) \quad (5.7)$$

where  $f_0$  is a constant and the integral of every summand over any of its own variables is zero, i.e.:

$$\int_0^1 f_{i_1 i_2 \dots i_s}(X_{i_1}, X_{i_2}, \dots, X_{i_s}) dX_{i_k} = 0 \quad 1 \leq k \leq s \quad (5.8)$$

The consequences of Eqs. (5.7) and (5.8) are that all the functions which appear within the summands in Eq. (5.7) are orthogonal, i.e. if  $(i_1, \dots, i_s) \neq (j_1, \dots, j_k)$ , then

$$\int_{K^n} f_{i_1, \dots, i_s} f_{j_1, \dots, j_k} dX = 0 \quad (5.9)$$

It is showed (Sobol', 1993) that there exists a unique expansion in Eq. (5.7) for any function  $f(X_1, X_2, \dots, X_n; \Theta)$  which is integrable in  $K^n$ . All the terms which appear within the summands in Equation (5.7) can be evaluated as follows:

$$\begin{aligned} f_0 &= \int_{K^n} f(X; \Theta) dX \\ f_i(X_i; \Theta) &= \int_0^1 \dots \int_0^1 f(X; \Theta) dX_{-i} - f_0 \\ f_{ij}(X_i, X_j; \Theta) &= \int_0^1 \dots \int_0^1 f(X; \Theta) dX_{-ij} \\ &\quad - f_i(X_i; \Theta) - f_j(X_j; \Theta) - f_0 \end{aligned} \quad (5.10)$$

and so on, where  $dX_{-(ij\dots m)}$  indicates the integration over all variables except the variables within the subscript parenthesis.

The total variance of  $f(X; \Theta)$  can be written as:

$$V = \int_{K^n} f^2(X; \Theta) dX - f_0^2 \quad (5.11)$$

and partial variance from each of the terms in Eq. (5.7) can be computed

$$V_{i_1, i_2, \dots, i_s} = \int_0^1 \cdots \int_0^1 f_{i_1, i_2, \dots, i_s}^2(X_{i_1}, X_{i_2}, \dots, X_{i_s}; \Theta) dX_{i_s} \cdots X_{i_2} X_{i_1} \quad (5.12)$$

where  $0 \leq i_1 \leq \cdots \leq i_s$  and  $s = 1, 2, \dots, n$ . The partial variance  $V_{i_1, i_2, \dots, i_s}$  is the contribution to the total variance from the term  $f_{i_1, i_2, \dots, i_s}$  which reflects the interactions among the input factors  $(X_{i_1}, X_{i_2}, \dots, X_{i_s})$ .

Combining Eqs. (5.9), (5.11) and (5.12), we have

$$V = \sum_{i=1}^n V_i + \sum_{1 \leq i < j \leq n} V_{ij} + \cdots + V_{12\dots n} \quad (5.13)$$

The sensitivity indices  $S_{i_1, i_2, \dots, i_s}$  (Sobol', 1993; Cukier *et al.*, 1973; Saltelli *et al.*, 1999) can be computed as:

$$S_{i_1, i_2, \dots, i_s} = \frac{V_{i_1, i_2, \dots, i_s}}{V} \quad (5.14)$$

All the  $S_{i_1, i_2, \dots, i_s}$  are nonnegative and their sum is:

$$S = \sum_{i=1}^n S_i + \sum_{1 \leq i < j \leq n} S_{ij} + \cdots + S_{12\dots n} \quad (5.15)$$

The Sobol's indices in Eqs. (5.11) and (5.12) can be computed with the Monte Carlo integral (Homma and Saltelli, 1996; Sobol', 2001).

### 5.2.1 A Special Case

There are however a number of important special cases such as bilinear models for which analytical results can be obtained.

Bilinear models are often used in empirical statistical models under the framework of factorial models. Frequently more complex models are approximated by bilinear models to facilitate analysis. A bilinear model is often represented as a polynomial of the form:

$$Y = \beta_0 + \sum_{i=1}^m \beta_i X_i + \sum_{i=1}^m \beta_{ii} X_i^2 + \sum_{i=1}^m \sum_{j>i}^m \beta_{ij} X_i X_j \quad (5.16)$$

The regressive variables  $X_i, i = 1..p$  are independently distributed and follow some distributions with finite moments up to order four. We will assume that the mean of the regressive variables is 0 and variance  $\sigma_i^2$ . This is not restrictive as the bilinear model can always be cast in this form. Using the ANOVA-like decomposition method, the variance decomposition of the model in Eq. (5.16) can be obtained as:

$$V_i = \begin{bmatrix} \beta_i & \beta_{ii} \end{bmatrix} \underbrace{\begin{bmatrix} \sigma_i^2 & E[X_i^3] \\ E[X_i^3] & E[X_i^4] - \sigma_i^4 \end{bmatrix}}_{M_i} \begin{bmatrix} \beta_i \\ \beta_{ii} \end{bmatrix} \quad (5.17)$$

and

$$V_{ij} = \beta_{ij}^2 \sigma_i^2 \sigma_j^2 \quad (5.18)$$

If the distribution of  $X_i$  is symmetric, then  $E[X_i^3] = 0 \forall i$ .

**Uniformly distributed variables:** If the regressive inputs are uniformly distributed:  $X_i \sim U(-a_i, a_i)$ , then  $V_i$  is:

$$V_i = \begin{bmatrix} \beta_i & \beta_{ii} \end{bmatrix} \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \frac{4}{5} \sigma_i^4 \end{bmatrix} \begin{bmatrix} \beta_i \\ \beta_{ii} \end{bmatrix} \quad (5.19)$$

**Normally distributed variables:** If the regressive inputs are normally distributed:  $X_i \sim N(0, \sigma_i^2)$ , the  $V_i$  can be written as:

$$V_{X_i} = \begin{bmatrix} \beta_i & \beta_{ii} \end{bmatrix} \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & 2\sigma_i^4 \end{bmatrix} \begin{bmatrix} \beta_i \\ \beta_{ii} \end{bmatrix} \quad (5.20)$$

## 5.3 Parameter Uncertainty and Sensitivity Analysis

### 5.3.1 Parameter Uncertainty and Sensitivity Indices

The situation to be considered now is to develop the regressive sensitivity methods recognizing that the model parameters may be random variables. The characterization of parameter uncertainty may come from a statistical analysis if the parameters were estimated from data. Alternatively, one may employ a Bayesian framework, and assign an uncertainty description to the parameters. The Bayesian interpretation relies extensively on the good judgement and skill of the analyst or modeler.

An initial approach might be to consider the augmented variables  $[X|\Theta]$  as the variables of interest and use standard sensitivity methods. If a model in Eq. (5.1) has  $m$  regressive factors and  $p$  parameters, the total number of variance decomposition terms is  $2^{m+p} - 1$ . Since these would lead to analysis where the number of factors is increased, the analysis cannot provide a suitable resolution. We are instead interested in performing the analysis with respect to the original set of input (regressor) variables, wherein we account for the variation in the model parameters. The total number of terms included in this approach is  $2^m - 1$ .

A new approach, conditional partial variance, is proposed for this problem. When the model parameters are treated as random variables, then the partial variance described in Eq. (5.2) can be considered as the variance and partial variance conditioned

on the certain values of parameters usually equal to the mean value of the distribution for the parameters. We shall denote the mean value for the parameters by the vector  $\tilde{\Theta}$ . Assuming for the purposes of analysis that parameters and regressor variables are statistically independent. To clearly note the dependence of the analysis on the values of  $\tilde{\Theta}$ , Eq. (5.3), should be written as:

$$\begin{aligned} Var_X[Y|\Theta = \tilde{\Theta}] &= \sum_i V_i|(\Theta = \tilde{\Theta}) + \sum_i \sum_{j>i} V_{ij}|(\Theta = \tilde{\Theta}) \\ &+ \cdots + V_{12\dots p}|(\Theta = \tilde{\Theta}) \end{aligned} \quad (5.21)$$

where

$$\begin{aligned} V_i|(\Theta = \tilde{\Theta}) &= Var_{(X_i)}[E_{\sim(X_i)}(|X_i = x_i^*, \Theta = \tilde{\Theta})] \\ V_{ij}|(\Theta = \tilde{\Theta}) &= Var_{(X_i, X_j)}[E_{\sim(X_i, X_j)}[Y|X_i = x_i^*, X_j = x_j^*, \Theta = \tilde{\Theta}] \\ &\quad - V_i|[\Theta = \tilde{\Theta}] - V_j|[\Theta = \tilde{\Theta}] \\ &\quad \vdots \end{aligned} \quad (5.22)$$

we use  $\sim(\bullet)$  to represent 'all but  $(\bullet)$ '.

Using the well-known variance decomposition (Parzen, 1962), the variance of Y can be written as:

$$Var[Y] = E_{\Theta}[Var_X[Y|\Theta = \tilde{\Theta}]] + Var_{\Theta}[E_X[Y|\Theta = \tilde{\Theta}]] \quad (5.23)$$

where  $E_{\Theta}[\bullet]$  denotes the expectation of  $[\bullet]$  respect to  $\Theta$  and  $Var_{\Theta}[\bullet]$  denotes the variance of  $[\bullet]$  respect to  $\Theta$ . The second term in the right-hand side of Eq. (5.23) is the fractional contribution to the output due to only the uncertainties of the parameters. The first term in Eq. (5.23) right-hand side is the variance contribution to the output

due to the regressors with the parameter uncertainties.

It is possible then to affect a variance decomposition in a manner similar to that described earlier. The only issue will be how to apportion the contributions from second term in the right-hand side of Eq. (5.23). Individual problems will suggest an appropriate apportioning scheme.

Let  $P_{i_1, i_2, \dots, i_s}$  denote the partial variance with consideration of parameter uncertainty,  $P_{i_1, i_2, \dots, i_s} = E_{\Theta}[V_{i_1, i_2, \dots, i_s} | (\Theta = \Theta^*)]$ . This will be called in this paper as the parameter uncertainty partial variance (PUPV). The sensitivity indices which are calculated based on PUPV are called parameter uncertainty sensitivity indices (PUSI). To distinguish these values from the situation where the RSA is based on the mean values of parameters  $\tilde{\Theta}$  in Eq.(5.21), the results of RSA in this fashion are called as the constant conditional partial variance (CCPV). The sensitivity indices which are calculated based on CCPV are called constant conditional sensitivity indices (CCSI).

CCSI can be considered as another alternative solution for the RSA. The advantage of this solution is that the computational requirement of variance-based SA is reduced, and the disadvantage is that the parameter uncertainty is completely ignored. In the example demonstrated in the later section, it shows that the difference between PUSI and CCSI may be significant. In our approach, an analytical solution of PUPV can be directly deduced through the CCPV if the model is linear in its parameters. If there is no significant difference between CCSI and PUSI, the CCSI can still be kept as the results of SA, but in this case the neglect of the parameter uncertainties is justified.

The difference between PUPV and CCPV depends on the uncertainties of the parameter estimators, regressor uncertainties and the model structure. It can be seen from the linear examples on the next section that the difference may be important. When the estimated parameters are unbiased and the variances of estimated

parameters are very small, the values of PUSI are close to the values of CCSI.

If the model is nonlinear in parameters, it is impossible to deduce an analytical solution of PUPV directly from the CCPV. Computational methods such as Monte Carlo method must be applied to estimate the PUPV, although the cost is the intensive computation. It can be done by the following procedures: i) generating the value for vector parameter  $\Theta^*$  using some sampling methods, ii) calculating the partial variances defined in Eq. (5.22) given  $\Theta^*$  using the variance-based SA methods, iii) repeating the steps i) and ii), iv) the average values from step iii) are the estimates of PUPVs.

### 5.3.2 Linear Models in the Parameters

Most often  $f(X, \Theta)$  is a nonlinear mapping when models are developed from fundamental principles that are based on the laws of conservation of mass, energy and momentum. Generally it is nonlinear in the parameters, why are we still interested in studying the variance-based SA for the linear-in-parameter models? The answer is that the linear-in-parameter models are widely used in science and engineering fields for nonlinear models/systems. Detailed information about the applications of the linear-in-parameter models is presented in the following paragraphs.

- In some instances, the experimenters/ modelers know the exact form of  $f(X, \Theta)$  which is nonlinear in the parameters. However linear-in-parameter models still are valuable for the cases: i) the linearization of  $f(X, \Theta)$  respect to the parameters  $\Theta$  is a suitable approximation for the original nonlinear model, and ii) the nonlinear model, so called intrinsically linear models, can be linearized with simple transformations, and the transformation form can be used for the sensitivity analysis.
- In many cases, the information necessary to build a fundamental model of the



processes is lacking and determination of the exact function  $f(X, \Theta)$  in Eq. (5.1) is impossible. In this instance, Therefore, it is necessary to approximate  $f(X, \Theta)$ . It can be done by a set of elementary functions called a basis such that all functions under study can be written as linear combinations of the elementary functions in the basis. Polynomial and trigonometric functions are widely used in this fashion. When the system is severely non-linear other choices of approximate functions could be considered. When the linear-in-parameter model is used, the model form is often quite simple, i.e. bilinear model typically involving terms of a bilinear product of the form  $f(X, \Theta) = H^T(X)\Theta$ , where  $H(X)$  is a polynomial in the regressive variables. The parameters of the models can readily be estimated using linear least squares methods.

Consider the case where the model is linear in its parameters. With this qualification, the system described in Eq. (5.1) can be written as:

$$Y = f(X; \Theta) = H^T(X)\Theta \quad (5.24)$$

where

$$\Theta^T = [\theta_1, \theta_2, \dots, \theta_p] \quad (5.25)$$

and

$$\begin{aligned} H^T(X) &= [h_1(X), h_2(X), \dots, h_p(X)] \\ h_i(X) &= \frac{\partial f(X; \Theta)}{\partial \theta_i} \quad i = 1, 2, \dots, p \end{aligned} \quad (5.26)$$

When the regressor vector  $X$  and parameter vector  $\Theta$  are independent, it follows that:

$$Var_X[Y|\Theta = \tilde{\Theta}] = \tilde{\Theta}^T M \tilde{\Theta}$$

where

$$\begin{aligned}
M &= E_X[(H(X) - \overline{H(X)})(H(X) - \overline{H(X)})^T] \\
&= E_X \begin{pmatrix} (h_1 - \bar{h}_1)^2 & \cdots & (h_1 - \bar{h}_1)(h_p - \bar{h}_p) \\ \vdots & (h_2 - \bar{h}_2)^2 & \vdots \\ \vdots & \ddots & \vdots \\ (h_p - \bar{h}_p)(h_1 - \bar{h}_1) & \cdots & (h_p - \bar{h}_p)^2 \end{pmatrix} \\
&= \left( \sum_{i=1}^n M_i + \sum_{1 \leq i < j \leq n} M_{ij} + \cdots + M_{12 \dots n} \right) \tag{5.27}
\end{aligned}$$

For notational convenience,  $h_i$  is used to represent  $h_i(X)$  and  $\overline{(\bullet)}$  denotes the mean value of  $(\bullet)$ . In Eq. (5.27):

$$\begin{aligned}
M_i &= E_{X_i}[(H(X) - \overline{H(X)})(H(X) - \overline{H(X)})^T | X_i = x_i^*] \\
M_{ij} &= E_{X_i X_j}[(H(X) - \overline{H(X)})(H(X) - \overline{H(X)})^T | X_i = x_i^*, X_j = x_j^*] \\
&\quad - M_i - M_j \quad (1 \leq i < j \leq n) \tag{5.28}
\end{aligned}$$

The nomenclature for the other entries in Eq.(5.27) follows immediately. All of the matrices  $M_i, M_{ij}, \dots, M_{12 \dots n}$  and  $M$  are symmetric by definition. Since the conditional total variance in Eq. (5.27) is always greater than zero, the matrix  $M$  must be positive definite.

As well

$$Var_{\Theta}[E_X[Y | \Theta = \tilde{\Theta}]] = Var_{\Theta}[\tilde{H}\tilde{\Theta}] \tag{5.29}$$

where

$$\tilde{H} = E_X[H(X)] \tag{5.30}$$

If the mean of output  $Y$  has been removed respect to  $H(X)$ ,  $Y' = (H(X)^T - \overline{H(X)}^T)\Theta$ . The term  $Var_{\Theta}[E_X[Y'|\Theta = \tilde{\Theta}]] = 0$ , the contribution of the output variance from the parameter uncertainties is only via interactions.

The partial variances given a nominal parameter vector  $\tilde{\Theta}$  can be written as:

$$P_{i_1, i_2, \dots, i_s} = E_{\Theta}(\tilde{\Theta}^T M_{i_1, i_2, \dots, i_s} \tilde{\Theta}) \quad (5.31)$$

The expectation of a quadratic form only depends on the mean and variance-covariance matrix of parameter  $\Theta$  (Graybill, 1983). Since the matrix  $M_{i_1, i_2, \dots, i_s}$  is symmetric, the analytical solution for the parameter uncertainty partial variance (PUPV) can be written as:

$$P_{i_1, i_2, \dots, i_s} = \tilde{\Theta}^T M_{i_1, i_2, \dots, i_s} \tilde{\Theta} + tr(M_{i_1, i_2, \dots, i_s} \Sigma_{\Theta}) \quad (5.32)$$

Recalling the constant conditional partial variance (CCPV), we have:

$$\tilde{V}_{i_1, i_2, \dots, i_s} = V_{i_1, i_2, \dots, i_s} | \bar{\Theta} = \bar{\Theta}^T M_{i_1, i_2, \dots, i_s} \bar{\Theta} \quad (5.33)$$

Comparing Equation (5.32) with Eq. (5.33), the difference between PUPV and CCPV is the term  $tr(M_{i_1, i_2, \dots, i_s} \Sigma_{\Theta})$  which depends on the model structure and parameter uncertainties.

### Sobol' and FAST method

- **Sobol' Method** In this case, all of the matrices  $M_i, M_{ij}, \dots, M_{12\dots n}$  and  $M$  can be readily expressed in terms of a functional expansion of  $H(X)$  and are shown below.

Decomposing Eq. (5.24) using Sobol's method, we have:

$$\begin{aligned}
f_0 &= E[Y] = [E[h_1], E[h_2], \dots, E[h_p]] \Theta \\
f_i(X_i) &= E[Y|X_i] - f_0 \\
&= [E[h_1|X_i] - E[h_1], \dots, E[h_p|X_i] - E[h_p]] \Theta \\
&= K_i \Theta \quad i = 1, 2, \dots, n \\
f_{ij}(X_i, X_j) &= E[Y|X_i, X_j] - f_i - f_j - f_0 \\
&= [E[h_1|X_i, X_j] - E[h_1|X_i] - E[h_1|X_j] - E[h_1], \\
&\quad \dots, E[h_p|X_i, X_j] - E[h_p|X_i] - E[h_p|X_j] - E[h_p]] \Theta \\
&= K_{ij} \Theta \quad i, j = 1, 2, \dots, n \text{ and } i < j \\
&\vdots \\
\end{aligned} \tag{5.34}$$

If the regressor vector  $X$  and parameter vector  $\Theta$  are independent, the partial variances given a parameter vector  $\tilde{\Theta}$  can be written as:

$$\begin{aligned}
V_i|\tilde{\Theta} &= \tilde{\Theta}^T \left( \int_0^1 K_i^T K_i dX_i \right) \tilde{\Theta} \\
&= \tilde{\Theta}^T M_i \tilde{\Theta} \quad i = 1, 2, \dots, n \\
V_{ij}|\tilde{\Theta} &= \tilde{\Theta}^T \left( \int_0^1 \int_0^1 K_{ij}^T K_{ij} dX_i dX_j \right) \tilde{\Theta} \\
&= \tilde{\Theta}^T M_{ij} \tilde{\Theta} \quad i, j = 1, 2, \dots, n \text{ and } i < j \\
&\vdots \\
V_{12\dots n}|\tilde{\Theta} &= \tilde{\Theta}^T \left( \int_{K^n} K_{12\dots n}^T K_{12\dots n} dX \right) \tilde{\Theta} \\
&= \tilde{\Theta}^T M_{12\dots n} \tilde{\Theta} \\
\end{aligned} \tag{5.35}$$

- **FAST Method** The FAST method uses an efficient method to sample from the requisite distribution. The desired quantities are numerically approximated with a Monte-Carlo type simulation. The efficiency of the FAST method arises from the special method in which the standard distributions are sampled using a space-filling approach.

Using the FAST method, we can determine the Fourier coefficients for the linear model as:

$$\begin{aligned}
A_i &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos(is) ds = \frac{1}{2\pi} \int_{-\pi}^{\pi} (H(s)\Theta) \cos(is) ds = H_{A_i} \Theta \\
B_i &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin(is) ds = \frac{1}{2\pi} \int_{-\pi}^{\pi} (H(s)\Theta) \sin(is) ds = H_{B_i} \Theta
\end{aligned}
\tag{5.36}$$

$V_i$ , the portion variance of the output variance  $V$  arising from the uncertainty of input  $X_i$ , is:

$$\begin{aligned}
V_i &= 2 \sum_{p=1}^{\infty} [A_{p\omega_i}^2 + B_{p\omega_i}^2] \\
&= 2\Theta^T \sum_{p=1}^{\infty} [H_{A_{p\omega_i}}^T H_{A_{p\omega_i}} + H_{B_{p\omega_i}}^T H_{B_{p\omega_i}}] \Theta \\
&= 2\Theta^T M_i \Theta
\end{aligned}
\tag{5.37}$$

## Special Cases and Extensions

- **Linear Models:** The full variance-based SA for linear models which are linear in both parameters and regressors has been discussed in (Saltelli *et al.*, 2006). It is trivial to show that the results from our results are identical to the results in (Saltelli *et al.*, 2006).
- **Bilinear Models:** A class of bilinear models was introduced in Section 5.2.

The expression for  $M_i$  is given in Eq. (5.17). The expression for  $M_{ij}$  is obtained by inspection from Eq. (5.18).

- **Extension to Nonlinear Models:** Our approach is a simple and powerful method for analyzing the effects of the parameter uncertainties on SA for the models which are linear in the parameters. Often, however, a research has these models which are nonlinear in the parameters. In such cases, linear techniques must be extended.

Considering a general nonlinear uniresponse function in Eq. (5.1). The simple extension of our approach to this nonlinear models is to use the Taylor series approximation.  $f(X; \Theta)$  can be approximately represented by taking the first term in the Taylor series at one nominal parameter vector  $\tilde{\Theta}$ .

$$\begin{aligned} f(X, \Theta) &\simeq f(X, \tilde{\Theta}) + \frac{\partial f}{\partial \Theta} \Big|_{\tilde{\Theta}} (\Theta - \tilde{\Theta}) \\ &= f(X, \tilde{\Theta}) + H(\Theta - \tilde{\Theta}) \end{aligned} \tag{5.38}$$

where

$$\begin{aligned} H &= [h_1, h_1, \dots, h_p] \\ h_i &= \frac{\partial f(X; \Theta)}{\partial \theta_i} \Big|_{\tilde{\Theta}} \quad i = 1, 2, \dots, p \end{aligned} \tag{5.39}$$

Rearranging Eq. (5.38), we have:

$$f(X, \Theta) \simeq \tilde{H}\Theta^* \tag{5.40}$$

where

$$\tilde{H} = [f(X, \tilde{\Theta}) - H\tilde{\Theta} \quad H] \quad (5.41)$$

$$\Theta^{*T} = [1 \quad \Theta] \quad (5.42)$$

Using the Taylor approximation in Eq. (5.40), the results in Section are straightforward applied for the nonlinear models. However, special care must be taken for this linear approximation approach since the nonlinearity of the model and the range of the parameter uncertainties will affect reliability of the linear approximation. Measures of nonlinearity could be used to help decide when linearized results provide acceptable approximations (Bates and Watts, 1988).

## 5.4 Examples

To demonstrate the parameter uncertainty effects on the sensitivity indices for nonlinear models which are linear in their parameters, an artificial analytical model with 3 regressive variables discussed in (Ishigami and Homma, 1990) is considered. In the first example, the input variables are assumed to be independent uniformly distributed. In the second example, regressive variables are assumed to have a multivariate normal distribution.

### 5.4.1 A Linear Model with Uniformly Distributed Regressors

In the first example, the model is described by the following equation:

$$f(X, \Theta) = \theta_1 \sin X_1 + \theta_2 \sin^2 X_2 + \theta_3 X_3^4 \sin X_1 \quad (5.43)$$

The input probability density functions are uniform:

$$P_i(X_i) = \begin{cases} \frac{1}{2\pi} & -\pi \leq X_i \leq \pi \\ 0 & \text{otherwise} \end{cases}, i = 1, 2, 3 \quad (5.44)$$

The nominal values of the parameters are given by (Ishigami and Homma, 1990):

$$\tilde{\Theta}^T = [1 \quad 7 \quad 0.1] \quad (5.45)$$

This model is linear in the parameters  $\Theta$  but nonlinear in the regressive variables  $X$ . An interaction term is expressed in the term of  $X_3^4 \sin X_1$ . Using Sobol's method discussed in Eqs. (5.34) and (5.35), the matrix relating the total variance and partial variances can be obtained analytically as:

$$\begin{aligned} M_1 &= \begin{bmatrix} \frac{1}{2} & 0 & \frac{9}{10}\sigma^4 \\ 0 & 0 & 0 \\ \frac{9}{10}\sigma^4 & 0 & \frac{81}{50}\sigma^8 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 & 9.74 \\ 0 & 0 & 0 \\ 9.74 & 0 & 189.77 \end{bmatrix}, M_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ M_{13} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{648}{225}\sigma^8 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 377.37 \end{bmatrix} \\ M &= \begin{bmatrix} \frac{1}{2} & 0 & \frac{9}{10}\sigma^4 \\ 0 & \frac{1}{8} & 0 \\ \frac{9}{10}\sigma^4 & 0 & \frac{9}{2}\sigma^8 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 & 9.74 \\ 0 & 0.125 & 0 \\ 9.74 & 0 & 527.14 \end{bmatrix} \end{aligned} \quad (5.46)$$



The substitution  $\sigma^2 = \pi^2/3$  has been made in these expressions.  $M_3$  and  $M_{23}$  are null matrices.

The analytical values for CCPV are shown in the second column of Table 5.1. Numerical values for CCPV and CCSI are shown in columns 3 and 5.

To investigate the effects of the parameter uncertainties, we will assume that the parameters,  $\Sigma_\Theta = (\theta_1, \theta_2, \theta_3)$ , have a multivariate normal distribution with mean  $\tilde{\Theta}$  and variance-covariance matrix  $\Sigma_\Theta$ . The variance-covariance matrix will be parameterized as  $P\Gamma P$  where  $P$  is a diagonal matrix whose elements are the standard deviation of each parameter and  $\Gamma$  is the correlation matrix of the parameters:

$$\Sigma_\Theta = \text{diag}\left[\sqrt{\frac{\tilde{\theta}_1}{\kappa}} \sqrt{\frac{\tilde{\theta}_2}{\kappa}} \sqrt{\frac{\tilde{\theta}_3}{\kappa}}\right] \begin{bmatrix} 1 & c_{12} & c_{13} \\ c_{21} & 1 & c_{23} \\ c_{31} & c_{32} & 1 \end{bmatrix} \text{diag}\left[\sqrt{\frac{\tilde{\theta}_1}{\kappa}} \sqrt{\frac{\tilde{\theta}_2}{\kappa}} \sqrt{\frac{\tilde{\theta}_3}{\kappa}}\right] \quad (5.47)$$

where  $\kappa$  is the mean-to-variance ratio. For our simulation, the values in matrix  $\Gamma$

Table 5.1: CCPV, PUPV, CCSI and PUSI for a nonlinear model with uniformly distributed regressors (parameter mean-to-variance ratio is 5,  $\sigma^2 = \pi^2/3$ )

Variables	CCPV Analytical	CCPV	PUPV	CCSI	PUSI
$X_1$	$\frac{1}{2}(\theta_1 + \frac{9}{5}\theta_3\sigma^4)^2$	4.346	8.241	0.314	0.327
$X_2$	$\frac{1}{8}\theta_2^2$	6.125	6.300	0.442	0.256
$X_3$	0	0	0	0	0
$X_1X_2$	0	0	0	0	0
$X_1X_3$	$\frac{72}{25}\theta_3^2\sigma^8$	3.374	11.321	0.244	0.416
$X_2X_3$	0	0	0	0	0
$X_1X_2X_3$	0	0	0	0	0

are assigned as  $c_{12} = c_{21} = c_{23} = c_{32} = 0.5$  and  $c_{13} = c_{31} = 0$ . The PUPV and PUSI are tabulated in columns 4 and 6 in Table 5.1 for a mean-to-variance ratio of 5.

As can be seen, the impact on regressive variable  $X_2$  is quite small compared to the changes associated with variables  $X_1$  and  $X_3$ . For this value of the mean-to-variance ratio, the relative importance of regressive variable  $X_1$  remains unchanged, while regressive variable  $X_2$  goes from the most important to least important variable, and regressive-variable effect  $X_1X_3$  goes from the least important to the most important effect.

The effect of the variance of the parameter estimates on the PUSI values is shown in Fig. 5.1. As the mean-to-variance ratio varies from 1 to 10, the relative importance of the variance-contribution terms changes. For large values of  $\kappa$ , the PUSI values approach the CCSI values.

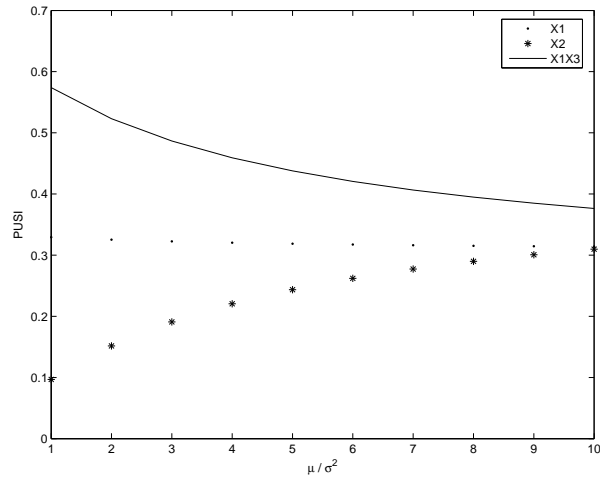


Figure 5.1: Effect of mean-to-variance ratio on PUSI ( $\sigma^2 = \pi^2/3$ )

The analysis thus far has focussed on the impact of the parameter uncertainties on the sensitivity indices. This corresponds to an analysis of variance analysis for the term  $E_{\Theta}[Var_X[Y|\Theta = \tilde{\Theta}]]$  in Eq. (5.23). To analyze the impact on the variance of  $Y$ , it is necessary to include the term  $Var_{\Theta}[E_X[Y|\Theta = \tilde{\Theta}]]$ . As indicated in an earlier section, the apportioning of the various contributions of this term cannot be made generally, and individual problems may suggest the most appropriate decomposition. For the problem at hand,  $Var_{\Theta}[E_X[Y|\Theta]] = \sigma^2/4 = 0.82$  which is only occupies 3.1%

of the total output variance. This does not affect the interpretation of the results for the case considered in this table.

### 5.4.2 A Linear Model with Non-uniformly Distributed Regressors (Normal Distributed Regressors)

In the second example, we use the same model structure discussed in Section 5.4.1, but we now assume that input  $X = (X_1, X_2, X_3)$  are independent normally distributed,  $X_i \sim N(\mu_i, \sigma_i^2)$ . The means and variances are given the values  $\mu_1 = \mu_2 = \mu_3 = 0$  and  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \pi^2/3$  which are equal to the means and variances of inputs in the previous example respectively.

The analytical and numerical values calculated from the extended FAST method for CCPV are shown in the second column of Table 5.2. Numerical values for CCPV and CCSI are shown in columns 3 and 5. These results are considerably different than the situation where the regressive variables had a uniform distribution even though in both cases the means and variances of the regressive variables were identical. The effect of the distribution assumption is most pronounced on the variance contribution from  $X_1X_3$ . The disadvantage of the FAST method is in the resolution of the results, only main factors and the combination of all interaction terms can be identified (Saltelli *et al.*, 1999).

To determine the impact of the parameter uncertainties, the matrices  $M_1$ ,  $M_2$ ,  $M_3$  and the matrix  $M_{X_1}$  for calculating the total partial variance from the interaction term involving the factor  $X_1$  defined in Eq. (5.27) are calculated for a mean-to-variance ratio of 5 in Eq. (5.48) (the bracketed numerical terms are the theoretical values):

$$\begin{aligned}
M &= \begin{bmatrix} 0.5_{(0.5)} & 0.0_{(0)} & 16.2_{(16.1)} \\ 0.0_{(0)} & 0.13_{(0.12)} & 0.01_{(0)} \\ 16.2_{(16.1)} & 0.01_{(0)} & e^{8.71(8.72)} \end{bmatrix} & M_2 &= \begin{bmatrix} 0.0_{(0)} & 0.0_{(0)} & 0.0_{(0)} \\ 0.0_{(0)} & 0.1_{(0.12)} & 0.04_{(0)} \\ 0.0_{(0)} & 0.04_{(0)} & 53.9_{(0)} \end{bmatrix} \\
M_1 &= \begin{bmatrix} 0.4_{(0.5)} & 0.0_{(0)} & 13.0_{(16.2)} \\ 0.0_{(0)} & 0.0_{(0)} & 0.01_{(0)} \\ 13.0_{(16.2)} & 0.01_{(0)} & 481_{(528)} \end{bmatrix} & M_{X_1} &= \begin{bmatrix} 0.1_{(0)} & 0.0_{(0)} & 3.2_{(0)} \\ 0.0_{(0)} & 0.0_{(0)} & 0.04_{(0)} \\ 3.21_{(0)} & 0.04_{(0)} & e^{8.62(8.63)} \end{bmatrix}
\end{aligned} \tag{5.48}$$

Table 5.2: CCPV, PUPV, CCSI and PUSI for a nonlinear model with normally distributed regressors

Variables	Theoretical CCPV	CCPV	PUPV	CCSI	PUSI
$X_1$	$\frac{(\theta_1 + 3\theta_3\sigma^4)^2}{2}(1 - e^{-2\sigma^2})$	9.74 (9.01)	18.44 (19.62)	0.123 (0.126)	0.095 (0.101)
$X_2$	$\frac{\theta_2^2}{8}(1 - e^{-4\sigma^2})^2$	6.37 (6.13)	7.60 (6.64)	0.090 (0.086)	0.039 (0.034)
$X_3$	0	0.24(0)	0.35(0)	0.0(0)	0.0(0)
interaction including $X_1$	$48\theta_3^2\sigma^8(1 - e^{-2\sigma^2})$	55.71 (56.22)	166.7 (168.50)	0.744 (0.798)	0.856 (0.871)

Note: the bracketed numerical terms are the theoretical values

Again, the theoretical values are enclosed in parenthesis. Theoretical and numerical values for PUPV and PUSI are shown in columns 4 and 6 of Table 5.2. Since total partial variance from the interaction term involving the factor  $X_1$  is only contributed by the interaction between  $X_1$  and  $X_3$ , for simplifying the explanation, in the following discussion, we only use the  $V_{13}$  to represent the summation value. Unlike the case where the regressive variables are uniformly distributed, there is little impact on the

relative ranking of the variables with the inclusion of parameter-effects uncertainty as the interaction of  $X_1$  and  $X_3$  is the main contributor to the analysis.

To investigate the impact of parameter uncertainty on the sensitivity indices, the same variance-covariance matrix in Eq. (5.47) is also used here. However, the regressors' variance is reduced to  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \pi^2/9$ . The effect of the mean-to-variance ratio is shown in Fig. 5.2. As the mean-to-variance ratio increases, the values of PUSI converge to the CCSI. The speed of convergence of this example is slower than the example shown in Fig. 5.1.

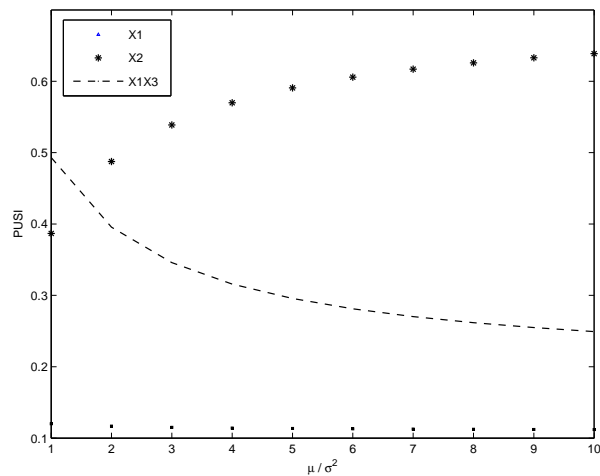


Figure 5.2: Effect of mean-to-variance ratio on PUSI ( $\sigma^2 = \pi^2/9$ )

The FAST method requires the selection of an number of 'tuning parameters' (the detailed information can be found in (Saltelli *et al.*, 1999; Cukier *et al.*, 1973; Cukier *et al.*, 1978)). In particular it is necessary to choose the number of simulations, the number of samples in a given simulation and several frequency parameters. For the examples in this chapter, the quantities of interest are obtained as the average of 500 simulations in which the sample size is  $N = 321$ . The frequency selections are 5, 40 and 3 respective to  $X_1$ ,  $X_2$  and  $X_3$ . A boxplot of the estimates of the constant conditional partial variances  $V_1$ ,  $V_2$  and  $V_{13}$  is shown in Fig. 5.3. As can be seen, while there is considerable variation in the quantities, the average value gives very

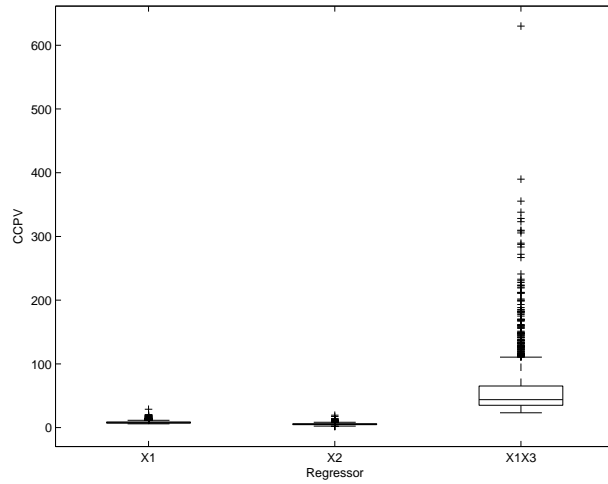


Figure 5.3: A boxplot of the estimates of the constant conditional partial variances using FAST for normally distributed regressive variables

credible estimates of the theoretical values.

## 5.5 Conclusions

Methods to included parameter uncertainty effects on regressive sensitivity analysis have been considered in this chapter. Our approach is fundamentally different to the current variance-based SA. We are interested in performing the analysis with respect to the original set of input (regressor) variables, wherein we account for the variation in the model parameters.

For general nonlinear models which are nonlinear in both regressors and parameters, Monte-Carlo type methods are generally required to undertake the analysis. When models are nonlinear in the regressors terms and linear in the parameters, explicit expressions for the parameter uncertainty effects have been derived. Modifications to both Sobol’s and FAST methods have been developed to find conditional partial variance matrices  $M_{i_1, i_2, \dots, i_p}$  described in Eq. (5.27). Parameter uncertainties effects on the sensitivity indices are presented by PUPV which is the combination of CCPV given parameter mean and variance-covariance matrix of the parameters.

Our approach can be used in abroad SA on the empirical models since most empirical models are in the form of linear in parameters. The difference between CCPV and PUSI can be analytically expressed as the term  $tr(M_{i_1, i_2, \dots, i_p} \Sigma_{\Theta})$ . A number of special cases have been considered, and extension of this method to approximate the effect of parameter uncertainties in models that are nonlinear in the parameters have been proposed.

Several related issues require further investigation: i) in the most applications, there are some correlations among the input factors, requiring the use of more flexible sampling methods and ii) further development of methods to account for parameter nonlinear. A simple modification has been proposed in this paper to enable the results to be used on nonlinear models. However, it is known in the statistics literature (Bates and Watts, 1988), that the impact of nonlinearities can significantly affect the statistical analysis of parameter uncertainty. Sophisticated mathematical methods to determine when this might be an issue have been developed and it would be interesting to modify their development for our approach.

# Chapter 6

## Conclusions

Much progress has been made in the area of control performance monitoring since the early 1990's as researchers have further developed and applied the concepts introduced by Åström (1970), and Harris (1989). The overall goal of this thesis was to contribute to this body of research by identifying and solving some theoretically interesting and /or industrially relevant control performance monitoring problems for the nonlinear dynamic/ stochastic systems. The contributions of this thesis were not solely based on extending previously reported results, but also on developing the new variance analysis methods for nonlinear dynamic / stochastic systems and time series. The material on controller performance assessment for a class of SISO nonlinear dynamical / stochastic systems in Chapter 3 is the first study dedicated to the task of nonlinear control performance assessment. Chapter 4 was a case study demonstrating how a ANOVA-like decomposition method can be used to analyzed the contributions to the output variance from different sources of disturbances after some modifications. Chapter 5 served to motivate the parameter uncertainty effects the variance decomposition for the nonlinear stochastic / dynamic systems and nonlinear times series.

In the following final sections of this thesis, the research contributions will be



reviewed on a chapter-by-chapter basis, and some topics for further research will be considered.

## **6.1 Thesis Contributions**

### **6.1.1 Controller Assessment of a class of Nonlinear Stochastic Systems**

This is the first study for nonlinear control performance assessment. A class of nonlinear dynamic / stochastic systems for which there exist minimum variance feedback invariant performance bounds has been established. This class of nonlinear systems can be nested with many block oriented nonlinear models such as Volterra series, NARMAX and Winner models. For this class of systems, the minimum variance bound can be estimated using closed-loop data using a nonlinear PAR or PARX model to estimate the b-step ahead prediction of the process. It is necessary to know the process delay. Application of the methodology to a simulation example indicates that this approach gives very credible estimates of the minimum variance performance bound. The simulation results indicate that the Orthogonal search method is effective.

Our approaches are not restricted for this class of nonlinear dynamic / stochastic systems alone. It is theoretically shown that the class of nonlinear dynamic / stochastic can be broadly used to represent the general nonlinear state space models with some assumptions or modification. Our approaches are valid for the general nonlinear state space models. Applications of the methodology to the empirical NARMA model and the CSTR model indicates that this approach gives very credible estimates of the minimum variance performance bound for the general nonlinear state space models.

### **6.1.2 Variance Decomposition of Nonlinear Stochastic Systems**

Chapter 4 was the first investigation into the effects of independent disturbances on output variance decompositions for multi-input, single-output (MISO) nonlinear systems / nonlinear time series. ANOVA-like decomposition methods are used to extend existing analysis of variance methods from static systems to the dynamic / stochastic nonlinear systems.

The case where there is no cross-correlation among the disturbances within the nonlinear systems which can be represented in the NARMAX models is studied in this section. Since the variance decomposition for the nonlinear systems will always include time dependency and initial condition, we have proposed the modifications for the ANOVA-like decomposition method in Eq. (4.1). The results from the simulations show that the modified ANOVA-like decomposition method does excellent job for the variance decomposition of the nonlinear systems. Since it is required intensive computational capacity for the MC methods, the cheaper computational method, extend FAST method, is used to reduce the calculation dimensionality and computational intension.

### **6.1.3 Parameter Uncertainty Effects on Sensitivity Analysis**

In Chapter 5 the effects of parameter uncertainties on the regressive sensitivity analysis for the nonlinear systems are investigated. we propose a new method to perform the analysis with respect to the original set of input (regressor) variables, wherein we account for the variation in the model parameters. The general solutions for the nonlinear models which are both nonlinear in the parameters and the regressors are proposed. If the nonlinear models are linear in the parameters, the parameter uncertainty effects on the variance-based sensitivity analysis can be directly obtained

from the variance-based sensitivity analysis results based on the mean values of the parameters. Both Sobol's and FAST methods are employed to find conditional partial variance matrices  $M_{i_1, i_2, \dots, i_p}$  described in Eq. (4.47). The extension of our approach to nonlinear models is proposed.

For the models which are linear in the parameters, parameter uncertainties effects on the sensitivity indices can be presented by PUPV which is the combination of CCPV given parameter mean (or Bayesian estimator) and variance-covariance matrix. It is straightforward to implement this approach on the regressive SA applications in which parameter uncertainties are ignored. Since the expectation of quadratic depends only on the mean and variance-covariance matrix, the distributions of parameter uncertainties are not required for this approach. The difference between CCSI and PUSI can be analytically expressed as the term  $tr(M_{i_1, i_2, \dots, i_p} \Sigma_\theta)$ .

## 6.2 Future Research Topics

Some suggestions for future research based on the results presented in Chapter 3 through 5 will be discussed below.

- The development thus far has been based on a class of the nonlinear process description given by Eq. (3.19) in Chapter 3. The performance bounds described in Chapter 3 have been presented under the additive disturbance assumption. The more comprehensive method which can cope with the dynamic disturbances will be developed for expanding the range of nonlinear control performance assessment.
- If the nonlinear function  $f(\cdot)$  in Eq.(3.19) in Chapter 3 is continuous it can always be arbitrarily well approximated by polynomial models. When the system is severely non-linear it would be other choices of approximate functions could be considered. Using neural networks models to estimate the lower performance

bounds will be an good alternative method for the nonlinear performance assessment.

- The CPA has been extended to linear MIMO systems by, e.g., Harris *et al.* (1996), Huang *et al.* (1997) and Huang and Shah (1999). Study of CPA for nonlinear MIMO systems will be another interesting and valuable research topic. However, since in the CPA of MIMO feedback systems, the so-called interactor matrix (or equivalences) (which plays an important role, as shown by Harris *et al.* (1996) and Huang *et al.* (1997)), in general, cannot be constructed from the knowledge of time delays only. The study of CPA for nonlinear MIMO systems will face the same problem. Since difficulties of the nonlinear system identification of MIMO nonlinear systems, this study will encounter a great challenge.
- The nonlinear control performance assessment discussed in Chapter 3 is also studied with the assumption that the setpoint conditions are constant. For the linear SISO systems, the control performance assessment under variable setpoint conditions has been successfully studied by Harris *et al.* (1999) and Seppala (1999). It is natural to investigate the nonlinear control performance assessment under variable setpoint conditions. The study of this interesting challenge can be based on the methods developed in Chapter 3 and 4.
- The variance decomposition for the nonlinear dynamic / stochastic systems and time series discussed in Chapter 4 is based on the assumption that the different disturbances are uncorrelated. This assumption is not always applicable in practice. The investigation into the effects of cross-correlated disturbances on analysis of variance nonlinear MISO systems may be necessary for the extension of nonlinear variance decomposition problems.
- The analysis of variance methods discussed in Chapter 4 were shown to be

useful for quantifying variance decompositions for MISO nonlinear systems with independent disturbances. It would be interesting to investigate more formally how such diagnostics could be used in the enhancement of the control design process.

- The parameter uncertainty effects on the variance decomposition discussed in Chapter 5 are only studied for the static nonlinear systems. It is an interesting topic to connect the results in Chapter 5 with the variance decomposition for nonlinear dynamic / stochastic systems discussed in Chapter 4.

# Bibliography

- Archer, G. E. B., Saltelli, A. and Sobol', I. M. (1997), 'Sensitivity measures, anova-like techniques and the use of bootstrap', *Journal of Statistical Computations and Simulations* **58**, 99–120.
- Åström, K. J. (1970), *Introduction to Stochastic Control Theory*, Academic Press, New York.
- Atherton, R. W., Schaihker, R. B. and Ducot, E. R. (1975), 'On the statistical sensitivity analysis of models for chemical kinetics', *AIChE Journal* **21**, 441–448.
- Bailey, B. A., Doney, S. C. and Lima, I. D. (2004), 'Quantifying the effects of dynamical noise on the predictability of a simple ecosystem model', *Environmetrics* **15**, 337–355.
- Bates, D. M. and Watts, D. G. (1988), *Nonlinear regression analysis and its applications*, John Wiley & Sons, New York.
- Bendat, J. S. (2001), *Nonlinear Systems Techniques and Applications*, John Wiley & Sons, New York.
- Billings, S. A. and Chen, S. (1989), 'Extend model set, global data and threshold model identification of severely non-linear systems', *International Journal of Control* **50**(5), 1897–1923.
- Bittanti, S. and Piroddi, L. (1993), 'Minimum variance control of a class of nonlinear plants with neural networks.', *Artificial Neural Networks, 1993, Third International Conference on* pp. 168–172.
- Bowyer, P. and Danson, F. M. (2004), 'Sensitivity of spectral reflectance to variation in live fuel moisture content at leaf and canopy level', *Remote Sensing of Environment* **92**, 297–308.
- Box, G. E. P., Hunter, W. G. and Hunter, J. (1978), *Statistics for experiments: An introduction to design, data analysis and model building*, Wiley, New York.
- Box, G. E. P. and Jenkins, G. M. (1970), *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco.

- Cacuci, D. G. and Ionescu-Bujor, M. (2004), 'A comparative review of sensitivity and uncertainty analysis of large-scale systems-ii: statistical methods', *Nuclear Science and Engineering* **147**, 204–217.
- Campolongo, F., Saltelli, A., Sorensen, T. and Tarantola, S. (2000), Hitchhiker's guide to sensitivity analysis, in S. A., K. Chan and E. M. Scott, eds, 'Sensitivity Analysis', John Wiley & Sons, West Sussex, England, pp. 15–47.
- Chan, K., Tarantola, S., Saltelli, A. and Sobol', I. M. (2000), Variance-based methods, in A. Saltelli, K. Chan and E. M. Scott, eds, 'Sensitivity Analysis', John Wiley & Sons, West Sussex, England, pp. 167–197.
- Chen, S. and Billings, S. A. (1989), 'Representations of non-linear systems: the nar-max model', *International Journal of Control* **49**, 1013–1032.
- Chen, S. and Billings, S. A. (1992), 'Neural networks for nonlinear dynamic system modeling and identification', *International Journal of Control* **56**(2), 319–346.
- Chen, S., Billings, S. A. and Grant, P. M. (1990), 'Non-linear dynamic system identification using neural networks', *International Journal of Control* **51**(6), 1191–1214.
- Chen, S., Billings, S. A. and Luo, W. (1989), 'Orthogonal least squares methods and their application to nonlinear system identification', *International Journal of Control* **50**(5), 1873–1896.
- Cho, H. and Marcus, S. I. (1987), 'On adaptive control of stochastic bilinear systems', *IEEE Transactions on Automatic Control* **AC-32**(12), 1103–1106.
- Chon, K. H., Korenberg, M. J. and Holstein-Rathlou, N. H. (1997), 'Application of fast orthogonal search to linear and nonlinear stochastic systems', *Annals of Biomedical Engineering* **25**, 793–801.
- Clarke, D. W. and Gawthrop, P. (1975), 'Self-tuning controllers', *Proceedings of the Institution of Electrical Engineers* **122**, 929–934.
- Cliff, M. A., Dever, M. C., Hall, J. W. and Girard, B. (1995), 'Development and evaluation of multiple regression models for prediction of sweet cherry liking', *Food Research International* **28**(6), 583–389.
- Cox, D. C. (1982), 'An analytical method for uncertainty analysis of nonlinear output functions, with applications to fault-tree analysis', *IEEE Transactions on Reliability* **R-31**(5), 465–468.
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G. and Schaibly, J. H. (1973), 'Study of the sensitivity of coupled reaction system to uncertainties in rate coefficients. i theory', *J. Chem. Phys.* **59**(8), 3873–3878.

- Cukier, R. I., Levine, H. B. and Shuler, K. E. (1978), ‘Nonlinear sensitivity analysis of multi-parameter model systems’, *Journal of Computational Physics* **26**, 1–42.
- Davis, P. J. and Rabinowitz, P. (1984), *Methods of numerical integration*, 2 edn, Academic Press, New York.
- Desborough, L. D. and Harris, T. J. (1992), ‘Performance assessment measures for univariate feedback control’, *Canadian Journal of Chemical Engineering* **70**, 1186–1197.
- Desborough, L. D. and Harris, T. J. (1993), ‘Performance assessment measures for univariate feedforward/feedback control’, *Canadian Journal of Chemical Engineering* **71**, 605–616.
- Diaz, H. and Desrochers, A. A. (1988), ‘Modeling of nonlinear discrete-time systems from input-output data’, *Automatica* **5**, 629–641.
- Doyle III, F. J., Packard, A. and Morari, M. (1989), ‘Robust controller design for a nonlinear cstr’, *Chemical Engineering Science* **44**, 1929–1947.
- Doyle III, F. J., Pearson, R. K. and Ogunnaike, B. A. (2002), *Identification and Control Using Volterra Models*, Springer, London.
- Draper, D., Pereira, A., Prado, P., Saltelli, A., R. Cheal, S. E., Mendes, B. and Tarrantola, S. (1999), ‘Scenario and parametric uncertainty in gesamac: a methodological study in nuclear waste disposal risk assessment’, *Computer Physics Communications* **117**, 142–155.
- Francos, A., Elorza, F. J., Bouraoui, F., Bidoglio, G. and Galbiati, L. (2003), ‘Sensitivity analysis of distributed environmental simulation models: understanding the model behavior in hydrological studies at the catchment scale’, *Reliability Engineering & System Safety* **79**, 205–218.
- Frey, H. C. and Patil, S. R. (2002), ‘Identification and review of sensitivity analysis methods’, *Risk Analysis* **22**, 553–578.
- Gao, J. and Budman, H. M. (2005), ‘Design of robust gain-scheduled pi controllers for nonlinear process’, *Journal of Process Control* **15**, 807–817.
- Ghaoui, L. E. (1995), ‘State-feedback control of system with multiplicative noise via linear matrix inequalities’, *Systems & Control Letters* **24**, 223–228.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J., eds (1996), *Markov Chain Monte Carlo in practice*, Chapman and Hall, London.
- Goodwin, G. C., McInnes, B. and Long, R. S. (1982), ‘Adaptive control algorithms for wastewater treatment and ph neutralization’, *Optimal Control Applications and Methods* **3**, 443–459.



- Granger, C. W. J. and Teräsvirta, T. (1993), *Modeling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
- Graybill, F. A. (1983), *Matrices with application in statistics*, 2 edn, Wadsworth, Belmont, CA.
- Grimble, M. J. (2005), ‘Non-linear generalized minimum variance feedback, feedforward and tracking control’, *Automatica* **41**, 957–969.
- Gunawan, R., Cao, Y., Petzold, L. and Doyle III, F. (1995), ‘Sensitivity analysis of discrete stochastic systems’, *Biophysical Journal* **88**, 2530–2540.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.
- Harber, R. and Unbehauen, H. (1990), ‘Structural identification of nonlinear dynamic systems - a survey on input/output approaches’, *Automatica* **26**(4), 651–677.
- Harris, T. J. (1989), ‘Assessment of control loop performance’, *Canadian Journal of Chemical Engineering* **67**, 856–861.
- Harris, T. J., Boudreau, F. and MacGregor, J. F. (1996), ‘Performance assessment of multivariable feedback controllers’, *Automatica* **32**, 1505–1518.
- Harris, T. J., Seppala, C. T. and Desborough, L. D. (1999), ‘A review of performance monitoring and assessment techniques for univariate and multivariate control system’, *Journal of Process Control* **9**(1), 1–17.
- Harris, T. J., Seppala, C. T., Jofriet, P. J. and Surgenor, B. W. (1996), ‘Plant-wide feedback control performance assessment using an expert system framework’, *Control Engineering Practice* **4**(9), 1297–1303.
- Harris, T. J. and Yu, W. (2007), ‘Controller assessment for a class of non-linear systems’, *Journal of Process Control* **17**, 607–619.
- Heald, J. P. and Stark, J. (2000), ‘Estimation of noise levels for models of chaotic dynamical systems’, *Journal of Physics Review Letter* **84**, 2366–2369.
- Helton, J. C. and Davis, F. J. (2002), ‘Illustration of sampling-based methods for uncertainty and sensitivity analysis’, *Risk Analysis* **22**, 591–622.
- Helton, J. C. and Davis, F. J. (2003), ‘Latin hypercube sampling and the propagation of uncertainty in analysis of complex systems’, *Reliability Engineering and System Safety* **81**(1), 23–69.
- Helton, J. C., Johnson, J. D., Sallaberry, C. J. and Storlie, C. B. (2006), ‘Survey of sampling-based methods for uncertainty and sensitivity analysis’, *Reliability Engineering and System Safety* **91**, 1175–1209.
- Hernández, E. and Arkun, Y. (1996), ‘Stability of nonlinear polynomial arma models and their inverse’, *International Journal of Control* **63**, 885–906.

- Homma, T. and Saltelli, A. (1996), ‘Importance measures in global sensitivity analysis of nonlinear models’, *Reliability Engineering and System Safety* **52**, 1–17.
- Hoo, K. A., Piovoso, M. J., Schnelle, P. D. and Rowan, D. A. (2003), ‘Process and controller performance monitoring: overview with industrial applications’, *International Journal of Adaptive Control and Signal Processing* **17**, 635–662.
- Hopperstad, O. S., Leira, B. J., Remseth, S. and Tromborg, E. (1999), ‘Reliability-based analysis of a stretch-bending process for aluminium extrusions’, *Computer and Structures* **71**, 63–75.
- Huang, B. (2002), ‘Minimum variance control and performance assessment of time variant processes’, *Journal of Process Control* **12**, 707–719.
- Huang, B. and Shah, S. L. (1999), *Performance Assessment of Control Loops: Theory and Applications*, Springer, London.
- Huang, B., Shah, S. L. and Kwok, E. K. (1997), ‘Good, bad or optimal? performance assessment of multivariable processes’, *Automatica* **33**, 1175–1183.
- Hwang, S. Y. and Basawa, I. V. (2005), ‘Explosive random-coefficient ar(1) processes and related asymptotics for least-squares estimation’, *Journal of Time Series Analysis* **26**(6).
- Ionescu-Bujor, M. and Cacuci, D. G. (2004), ‘A comparative review of sensitivity and uncertainty analysis of large-scale systems-i: deterministic methods’, *Nuclear Science and Engineering* **147**, 189–203.
- Ishigami, T. and Homma, T. (1990), An importance quantification technique in uncertainty analysis for computer models, in ‘Proceedings of the ISUMA ’90, First International Symposium on Uncertainty Modeling and Analysis’, University of Maryland, USA, pp. 398–403.
- Jansen, M. J. W., Rossing, W. A. H. and Daamen, R. A. (1994), Monte carlo estimation of uncertainty contributions from several independent multivariate sources, in G. J. and G. V. Straten, eds, ‘Predictability and nonlinear modeling in natural sciences and economics’, Kluwer, Dordrecht, pp. 334–343.
- Jazwinski, A. H. (1970), *Stochastic process and Filtering Theory*, Academic Press, New York.
- Jelali, M. (2006), ‘An overview of control performance assessment technology and industrial applications’, *Control Engineering Practice* **14**, 441–466.
- Jimenes, J. C. and Ozaki, T. (2002), ‘Linear estimation of continuous-discrete linear state space models with multiplicative noise’, *Systems & Control Letters* **47**, 91–101.

- Kampen, N. G. V. (1992), *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam.
- Keviczky, L., Vajk, I. and Hethéssy, J. (1979), A self-tuning extremal controller for the generalized hammerstein model, *in* ‘Proceedings of the 5th IFAC Symposium on Identification and System Parameter Estimation’, Darmstadt, West Germany, pp. 1147–1152.
- Kleijnen, J. P. C. and Helton, J. C. (1999), ‘Statistical analyses of scatterplots to identify important factors in large-scale simulations, 1: review and comparison of techniques’, *Reliability Engineering and System Safety* **65**, 147–185.
- Knapp, T. D. and Budman, H. M. (2000), ‘Robust control design of non-linear processes using empirical state affine models’, *Int. J. Control* **73**(17), 1525–1535.
- Ko, B. S. and Edgar, T. F. (2000), ‘Performance assessment of cascade control loops’, *AIChE Journal*.
- Koda, M., McRae, G. J. and Seinfeld, J. H. (1979), ‘Automatic sensitivity analysis of kinetic mechanisms’, *International Journal of Chemical Kinetics* **11**, 427–444.
- Koh, T. and Powers, E. J. (1985), ‘Second-order volterra models filtering and its application to nonlinear system identification’, *IEEE Transactions on Acoustics, Speech and Signal Processing* **ASSP-33**(6), 1445–1456.
- Korenberg, M. J. (1988), ‘Identifying nonlinear difference equation and functional expansion representations: the fast orthogonal algorithm’, *Annals of Biomedical Engineering* **16**, 123–142.
- Kozub, D. J. (1996), Controller performance monitoring and diagnosis: experiences and challenges, *in* ‘Proceedings of the chemical process control conference’, Lake Tahoe, USA, pp. 83–96.
- Kurtz, M. J. and Henson, M. A. (1997), ‘Input-output linearizing control of constrained nonlinear processes’, *J. of Process Control* **7**, 3–17.
- Leontaritis, I. J. and Billings, S. A. (1985a), ‘Input-output parametric models for nonlinear systems, part i: deterministic nonlinear systems’, *International Journal of Control* **41**(4), 303–328.
- Leontaritis, I. J. and Billings, S. A. (1985b), ‘Input-output parametric models for nonlinear systems, part ii: stochastic nonlinear system’, *International Journal of Control* **41**(4), 329–344.
- Leontaritis, I. J. and Billings, S. A. (1987), ‘Model selection and validation methods for non-linear systems’, *International Journal of Control* **45**, 311–341.
- Li, K., Peng, J. X. and Irwin, G. W. (2005), ‘A fast nonlinear model identification method’, *IEEE Trans. on Automatic Control* pp. 1211–1216.

- Ljung, L. (1987), *System identification: Theory for the user*, Prentice-Hall, New York.
- Lütkepohl, H. (1991), *Introduction to Multiple Time Series analysis*, Springer-Verlag, Berlin.
- Lynch, C. B. and Dumont, G. A. (1996), ‘Control loop performance monitoring’, *IEEE Transactions on Control Technology* **4**, 185–192.
- Madár, J., Abonyi, J. and Szeifert, F. (2005), ‘Genetic programming for the identification of nonlinear input-output models’, *Industrial and Engineering Chemistry Research* **44**(9), 3178–3186.
- Maner, B. R. and III, F. J. D. (1997), ‘Polymerization reactor control using autoregressive-plus volterra-based mpc’, *American Institute Chemical Engineering Journal* **43**, 1763–1784.
- McCarthy, M. A., Burgman, M. A. and Ferson, S. (1995), ‘Sensitivity analysis for models of population viability’, *Biological Conservation* **73**(2), 93–100.
- McMay, M. D. (1995), ‘Evaluating prediction uncertainty’, Los Alamos National Laboratories in NUREG/CR-6311, LA-12915-MS.
- McMay, M. D. (1996), ‘Variance-based methods for assessing uncertainty importance’, NUREG-1150 analysis, LA-UR-96-2695.
- McMay, M. D., Beckman, R. J. and Conover, W. J. (1979), ‘A comparison of three methods for selecting values of input variables in the analysis of output from a computer code’, *Technometrics* **21**, 239–245.
- McNabb, C. A. and Qin, S. J. (2005), ‘Projection based mimo control performance monitoring: Ii- measured disturbance and setpoint changes’, *Journal of Process Control* .
- Nicholls, D. F. and Quinn, B. G. (1982), *Random Coefficient Autoregressive Models: An Introduction*, Lecture Notes in Statistics, 11, Springer, New York.
- Parzen, E. (1962), *Stochastic Processes*, Holden Day, San Francisco.
- Pearson, R. K. (1999), *Discrete-Time Dynamic Models*, Oxford University Press, New York.
- Pearson, R. K. and Ogunnaike, B. A. (1997), Nonlinear process identification, *in* M. Henson and D. Seborg, eds, ‘Nonlinear Process Control’, Prentice Hall, Upper Saddle River, N.J., pp. 11–102.
- Pearson, R. K., Ogunnaike, B. A. and III, F. J. D. (1996), ‘Identification of structurally constrained second-order volterra models’, *IEEE Transactions on Signal Processing* **44**, 2837–2845.

- Petersen, I. R. and James, M. R. (1996), ‘Performance analysis and controller synthesis for nonlinear systems with stochastic uncertainty constraints’, *Automatica* **32**(7), 959–972.
- Planas, C. and Depoutot, R. (2002), Sensitivity analysis for signal extraction in economic time series, *in* S. A., K. Chan and E. M. Scott, eds, ‘Sensitivity Analysis’, John Wiley & Sons, West Sussex, England, pp. 293–327.
- Pourahmadi, M. (1986), ‘On stationarity of the solution of a doubly stochastic model’, *Journal of Time Series Analysis* **7**, 123–131.
- Qin, S. J. (1998), ‘Control performance monitoring - a review and assessment’, *Computer & Chemical Engineering* **23**, 173–186.
- Rabitz, H., Alis, O. F., Shorter, J. and Shim, K. (1999), ‘Efficient input-output model representations’, *Computer Physics Communications* **117**, 11–20.
- Ramsey, J. B. (1996), ‘If nonlinear models cannot forecast, what use are they?’, *Studies in Nonlinear Dynamics and Econometrics* **1**(2), 65–86.
- Saisana, M., Saltelli, A. and Tarantola, S. (2005), ‘Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators’, *Journal of Royal Statistical Society A* **168**(2), 307–323.
- Sales, K. R. and Billings, S. A. (1990), ‘Self-tuning control of non-linear armax models’, *International Journal of Control* **51**, 753–769.
- Saltelli, A. (2002), ‘Sensitivity analysis for importance assessment’, *Risk Analysis* **22**(3), 579–590.
- Saltelli, A., K. C. and Scott, E. M., eds (2000), *Sensitivity analysis*, Wiley, New York.
- Saltelli, A., Ratto, M., Tarantola, S. and Campolongo, F. (2005a), ‘Sensitivity analysis for chemical models’, *Chemical Reviews* **105**(7), 2811–2827.
- Saltelli, A., Ratto, M., Tarantola, S. and Campolongo, F. (2005b), ‘Sensitivity analysis for chemical models’, *Chemical Reviews* **105**, 2811–2827.
- Saltelli, A., Ratto, M., Tarantola, S. and Campolongo, F. (2006), ‘Sensitivity analysis practices: strategies for model-based inference’, *Reliability Engineering and System Safety* **91**, 1109–1125.
- Saltelli, A. and Tarantola, S. (2002), ‘On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal’, *Journal of the American Statistical Association* **97**(459), 702–709.
- Saltelli, A., Tarantola, S. and Chan, K. (1999), ‘A quantitative model-independent method of global sensitivity analysis of model output’, *Technometrics* **41**(1), 39–56.

- Saltelli, A., Tarantola, S. and Ratto, M. (2004), *Sensitivity Analysis in Practice*, Wiley, New York.
- Schaibly, J. H. and Shuler, K. E. (1973), ‘Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. part ii, applications’, *Journal of Chemical Physics* **59**, 3879–3888.
- Schetzen, M. (1980), *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley & Sons, New York.
- Seppala, C. (1999), Dynamic analysis of variance methods for monitoring control system performance, PhD thesis, Queen’s university, Department of Chemical Engineering.
- Seppala, C. T., Harris, T. J. and Bacon, D. W. (2002), ‘Time series methods for dynamic analysis of multiple controlled variables’, *Journal of Process Control* **12**, 257–276.
- Siefert, M., Peinke, J. and Friedrich, R. (2003), ‘On a quantitative method to analyze dynamical and measurement noise’, *Europhysics Letters* **61**, 466–.
- Sobol’, I. M. (1969), ‘Multidimensional quadrature formulas and haar functions [in russian]’, Nauka, Moscow.
- Sobol’, I. M. (1993), ‘Sensitivity estimates for nonlinear mathematical models’, *Matematicheskoe Modelirovanie*, 2, 112-118 (in Russian), translated in *Mathematical Modeling and Computational Experiments*, 1, 407-414.
- Sobol’, I. M. (2001), ‘Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates’, *Mathematics and Computer in Simulation* **55**, 271–280.
- Stachurski, J. (2003), ‘Economic dynamical systems with multiplicative noise’, *Journal of Mathematical Economics* **39**, 135–152.
- Stanfelj, N., Marlin, T. E. and MacGregor, J. F. (1993), ‘Monitoring and diagnosing process control performance: the single-loop case’, *Industrial & Engineering Chemistry Research* **32**, 301–314.
- Stein, M. (1987), ‘Large sample properties of simulations using latin hypercube sampling’, *Technometrics* **29**, 143–151.
- Stone, M. (1971), *Approximate calculation of multiple integrals*, Prentice-Hall, Englewood Cliffs.
- Svoronos, S., Stephanopoulos, G. and Aris, R. (1981), ‘On bilinear estimation and control’, *International Journal of Control* **34**, 615–684.
- Tan, A. H. (2006), ‘Wiener-hammerstein modeling of nonlinear effects in bilinear systems’, *IEEE Trans. on Automatic Control* **31**, 648–652.

- Terasvirta, T., Digk, D. V. and Medeiros, M. C. (2005), 'Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: a re-examination', *International Journal of Forecasting* **21**, 755–774.
- Thornhill, N. F., Cox, J. W. and Paulonis, M. (2003), 'Diagnosis of plant-wide oscillation through data-driven analysis and process understanding', *Control Engineering Practice* **11**, 1481–1490.
- Thornhill, N. F., Oettinger, M. and Fedenezuk, M. S. (1999), 'Refinery-wide control loop performance assessment', *Journal of Process Control* **9**, 109–124.
- Tienungoon, S., Ratkowsdy, D. A., McMeekin, T. A. and Ross, T. (2000), 'Growth limits of listeria monocytogenes as a function of temperature, ph, nacl and lactic acid', *Applied and Environmental Microbiology* **66**(11), 4979–4987.
- Tong, H. (1990), *Non-linear Time Series*, Oxford University Press, New York.
- Treichl, T., Hofmann, S. and Schröder, D. (2002), 'Identification of nonlinear dynamic miso systems on fundamental basics of the volterra theory', PCIM'02, Nürnberg, Germany.
- Tyler, M. L. and Morari, M. (1995), Performance assessment for unstable and nonminimum-phase systems, *in* 'Proc IFAC workshop on-line fault detection supervision chemical process industries', Newcastle-upon-Tyne, U.K., pp. 200–205.
- Varma, A., Morbidelli, M. and Wu, H. (1999), *Parametric sensitivity in chemical systems*, Cambridge university press, Cambridge.
- Weyl, H. (1938), 'Mean motion', *American Journal of Mathematics* **60**, 889–896.

## Appendix A. Proof of Theorem 1

The expression for  $Y_{t+b}$  is written down by inspection from Eq.(3.19). By successively substituting for all values of  $Y_{t+b-i}, i = 1..(b-1)$ , it is observed that:

$$Y_{t+b} = \delta^{-1}(q^{-1})f(Y_t^*, U_t^*) + \sum_{j=0}^{\infty} \tau_j(f_D(a_{t+b-1-j}^*) + a_{t+b-j}) \quad (\text{A-1})$$

where  $\tau_j$  is the  $j^{\text{th}}$  impulse coefficient of  $[\delta(q^{-1})\phi(q^{-1})]^{-1}\nabla^d$ . It is necessary that  $\delta(q^{-1})$  and  $\phi(q^{-1})$  be stable so  $\tau_j, j \geq 0$  forms a convergent series. In this representation the output disturbance is:

$$D_{t+b} = \sum_{j=0}^{\infty} \tau_j(f_D(a_{t+b-1-j}^*) + a_{t+b-j}) \quad (\text{A-2})$$

Firstly, let us consider the 1-step conditional prediction of the disturbance  $D_t$  given an information set  $I_t$ :

$$\hat{D}_{t+1/t} = \sum_{j=0}^{\infty} \tau_j(f_D(a_{t-1-j}^*) + \mu_a) \quad (\text{A-3})$$

since  $\tau_0 = 1$ . The b-step ahead forecast is:

$$\begin{aligned} \hat{D}_{t+b/t} &= E\left\{\sum_{j=0}^{\infty} \tau_j(f_D(a_{t+b-1-j}^*) + a_{t+b-j})|I_t\right\} \\ &= E\left\{\sum_{j=0}^{b-1} \tau_j(f_D(a_{t+b-1-j}^*) + a_{t+b-j})|I_t\right\} + \sum_{j=b}^{\infty} \tau_j(f_D(a_{t+b-1-j}^*) + a_{t+b-j}) \end{aligned} \quad (\text{A-4})$$

Now in the above equation  $E\{a_{t+k}|I_t\} = \mu_a, k = 1..b$  and  $E\{a_{t-k}|I_t\} = a_{t-k} = D_{t-k} - D_{t-k/t-k-1}, k \geq 0$ . It is now necessary to evaluate terms of the form  $E\{f_D(a_{t+k}^*)|I_t\}, k = 1..b-1$ . Each of these terms requires evaluation of the integral:

$$E\{f_D(a_{t+k}^*)|I_t\} = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_k f_D(a_{t+k}^*)p(a_{t+k}, \dots, a_{t+1})da_{t+k} \cdots da_{t+1} \quad (\text{A-5})$$

where  $p(a_{t+k}, \dots, a_{t+1})$  is the joint distribution of  $a_{t+k} \cdots a_{t+1}$ . In the case of linear disturbances where  $\{a_t\}$  is a white noise sequence with mean zero and constant variance, the expectation term in Eq.(A-4) is zero.

Combining these results with Eq.(A-1)

$$\hat{Y}_{t+b/t} = \delta^{-1}(q^{-1})(f(Y_t^*, U_t^*)) + \hat{D}_{t+b/t} \quad (\text{A-6})$$



and

$$\begin{aligned} e_{t+b/t} &= Y_{t+b} - \hat{Y}_{t+b/t} \\ &= \sum_{j=0}^{b-1} \tau_j (f_D(a_{t+b-1-j}^*) - E\{f_D(a_{t+b-1-j}^*)|I_t\} + a_{t+b-j} - \mu_a) \end{aligned} \quad (\text{A-7})$$

The prediction error is feedback invariant thus proving the theorem. We also notice from Eq.(A-7) that  $E\{e_{t+b/t}\} = 0$ .