CLOSED-LOOP IDENTIFICATION OF PLANTS UNDER MODEL PREDICTIVE CONTROL

by

Elsa de Klerk

Submitted in partial fulfilment of the requirements for the degree

Master of Engineering (Electronic Engineering)

in the

Faculty of Engineering, The Build Environment and Information Technology

UNIVERSITY OF PRETORIA

July 2003

SUMMARY

The open-loop step testing approach, which is, typically, used in process identification for model-based predictive control (MPC), has some disadvantages. Closed-loop system identification is less intrusive than the open-loop approach and can reduce re-identification time considerably. If the process model is identified while the MPC controller is operating, safety, product quality and optimality problems will be avoided.

For above-mentioned reasons, multivariable closed-loop identification for use in MPC is studied. Relevant closed-loop approaches are reviewed and a closed-loop identification methodology for MPC controlled plants is chosen. Simulations and real process data are used to validate and evaluate this methodology.

Many estimation methods fail when applied directly to closed-loop data, because of correlation between the additive output noise and the plant input. However, the prediction error estimation method can still give consistent estimates in the presence of noise. Therefore, this method is utilised.

In open-loop a persistently exciting plant input ensures identifiability. This does not ensure identifiability in closed-loop. For identifiability in closed-loop either the feedback mechanism should be nonlinear or the reference signal should be persistently exciting. The inter-sampling approach, where the plant output is sampled at a higher rate than the control input, can also ensure identifiability. However, a variance simulation study shows that a model identified with this approach is imprecise when structured tests are not performed.

The direct closed-loop system identification approach, where only the plant output and input are used for identification of the plant, is employed. This approach works irrespective of the feedback mechanism, it ensures consistency and optimal precision, and it makes use of standard MATLAB functions.

In the simulation a multivariable MPC controlled plant is identified from closed-loop data. To evaluate the consistency of the methodology, the plant is identified for different controller settings and different added disturbances. Different methods for ensuring identifiability are also investigated. Together with the standard validation tests used in open-loop identification, models are compared to the open-loop identified model and the stability of the closed-loop systems are examined.

The simulation results show that the proposed methodology gives reliable results for the type of system disturbances and constraints used when structured tests are performed that ensure persistently exciting reference signals and good signal-to-noise ratios. The results also show that imprecise models are usually identified when only relying on inter-sampling or the nonlinearity of the controller for identifiability.

A part of an industrial multivariable plant is also identified. No structured tests are performed on the plant. Logged data sets from normal operation are used in the identification. It is concluded that the open-loop identified model is not a good representation of the plant in closed-loop operation at the relevant time. Therefore, the results from the comparison between the open-loop and closed-loop identified models are unreliable.

From this experiment a preliminary conclusion is made that an unsatisfactory model will usually be identified from data measured under normal closed-loop control, because the reference signals are usually not persistently exciting and the signal-to-noise ratios are not good.

Keywords: closed-loop system identification, open-loop, model-based predictive control, multivariable, prediction error estimation method, methodology, identifiability, persistently exciting, nonlinear feedback, inter-sampling, validation, evaluation, variance, simulation, process data, signal-to-noise ratio.

SAMEVATTING

Die opelustraptoetsbenadering, wat tipies gebruik word in aanlegidentifikasie vir modelgebaseerde voorspellingsbeheer (MVB), het sekere nadele. Geslotelusstelselidentifikasie het 'n kleiner impak as die opelusbenadering en kan her-identifikasietyd aansienlik verminder. Indien die prosesmodel geïdentifiseer word terwyl die MVB beheerder in bedryf is, sal probleme met betrekking tot veiligheid, produk kwaliteit en optimaliteit vermy word.

Om bogenoemde rede word multiveranderlike geslotelusidentifikasie vir gebruik met MVB bestudeer. Toepaslike geslotelusbenaderings word hersien en 'n geslotelusidentifikasiemetodiek vir MVB beheerde aanlegte word gekies. Simulasies en prosesdata word gebruik om die metodiek te verifieer en te evalueer.

Baie skattingsmetodes faal wanneer dit direk toegepas word op geslotelusdata as gevolg van die korrelasie tussen die aanleginset en ruis op die aanleguitset. Die foutvoorspellingsskattingsmetode kan egter wel konsekwente skattings gee; dus word dié skattingsmetode geïmplimenteer.

In opelus word identifiseerbaarheid verseker deur 'n aanhoudend stimulerende ("persistently exciting") aanleginset. Dit verseker nie identifiseerbaarheid in geslotelus nie. In geslotelus word identifiseerbaarheid verseker as óf die terugvoermeganisme nie-linieêr is óf die verwysingssein aanhoudend stimulerend is. Die inter-monsteringsmetode, waar die aanleguitset teen 'n hoër spoed as die beheerinset gemonster word, verseker ook identifiseerbaarheid. 'n Simulasiestudie oor variansie dui egter aan dat 'n model geïdentifiseer met dié metode, sonder gestruktureerde toetse, gewoonlik nie presies is nie.

Die direkte geslotelusstelselidentifikasiebenadering, waar slegs die aanleguitset en -inset gebruik word in die identifisering, word gebruik. Hierdie benadering werk onafhanklink van die terugvoermeganisme, dit verseker konsekwente skattings en optimale presiesheid, en dit maak gebruik van standaard MATLAB funksies.

In die simulasie word 'n multiveranderlike MVB beheerde aanleg geïdentifiseer uit geslotelusdata. Om die konsekwensie van die metodiek te evalueer, word die aanleg geïdentifiseer vir verskillende beheerderstellings en ook verskillende toegevoegde versteurings. Verskillende metodes om identifiseerbaarheid te verseker, word ook ondersoek. Saam met die standaard verifieeringstoetse, wat in opelusidentifikasie gebruik word, word modelle ook vergelyk met opelusgeïdentifiseerde modelle en die stabiliteit van die geslotelusstelsels word ondersoek.

Die simulasieresultate toon aan dat die voorgestelde metodiek betroubare resultate lewer vir die tipe versteurings en beheerderstellings wat gebruik is, wanneer gestruktureerde toetse gedoen word wat 'n aanhoudend stimulerende verwysingssein en 'n goeie sein-tot-ruis verhouding verseker. Die resultate toon ook aan dat daar gewoonlik 'n groot variansie in geïdentifiseerde modelle ontstaan wanneer daar slegs op inter-monstering of 'n nie-liniêre beheerder vertrou word vir identifiseerbaarheid.

'n Deel van 'n multiveranderlike industriële aanleg word ook geïdentifiseer. Geen gestruktureede toetse word uitgevoer op die aanleg nie. Data, gemonster tydens normale werking, word gebriuk vir identifikasie. Daar word tot die gevolgtrekking gekom dat die opelusgeïdentifiseerde model nie 'n goeie voorstelling van die aanleg onder geslotelusbeheer, tydens die toepaslike tyd, is nie; dus is die resultate uit die vergelyking tussen die opelus- en geslotelusgeïdentifiseerde modelle onbetroubaar.

Uit die eksperiment kan 'n voorlopige gevolgtrekking gemaak word dat 'n onaanvaarbare model gewoonlik geïdentifiseer sal word uit data gemonster tydens normale geslotelusbeheer, aangesien in dié geval die verwysingsseine gewoonlik nie aanhoudend stimulerend is nie en die sein-tot-ruis verhouding nie goed is nie.

Sleutelwoorde: geslotelusstelselidentifikasie, opelus, modelgebasseerde voorspellingsbeheer, multiveranderlik, foutvoorspellingsskattingsmetode, metodiek, identifiseerbaarheid, aanhoudend stimulerend ("persistently exciting"), nie-linieêre terugvoer, inter-monstering , verifieer, evalueer, variansie, simulasie, prosesdata, sein-tot-ruis verhouding.

Abbreviations

AIC	Akaike's Information Theoretic Criterion
ARMAX	Auto-Regressive Moving Average with External input
ARX	Auto-Regressive with External input
ASYM	Asymptotic Method
BJ	Box-Jenkins
CLOSID	Closed-loop System Identification
CV	Controlled Variable
DMC	Dynamic Matrix Controller
DMK	Dimethyl Ketone
FIR	Finite Impulse Response
FPE	Final Prediction Error
IDARX	Multivariable ARX
IPA	Isopropyl Alcohols
IV	Instrumental Variables
LSE	Least-Squares Estimate
MDL	Minimum Description Length
MIBK	Methyl Iso Butyl Ketone
MIMO	Multiple-Input-Multiple-Output
MISO	Multiple-Input-Single-Output
MPC	Model-based Predictive Control
MPCI	MPC and Identification
MV	Manipulated Variable
OE	Output Error
PDF	Probability Density Function
PE	Persistently Exciting
PEM	Prediction-Error Method
PRBS	Pseudo-Random Binary Signal
QP	Quadratic programme
RBS	Random Binary Signal
RMPCT	Robust Model Predictive Control
SCI	Sasol Chemical Industries
SID	System Identification
SISO	Single-Input-Single-Output
SITB	System Identification Toolbox
SNR	Signal-to-Noise Ration
SSF	Sasol Synthetic Fuels
SVD	Singular Value Decomposition
w.p.	with probability
ZOH	Zero-Order-Hold

CONTENTS

Summary 3

Samevatting 5

Abbreviations 7

1 Introduction 12

1.1 Problem Statement 12

1.2 Literature Review 13

1.2.1 Advantages of Closed-Loop Identification 13

1.2.2 Correlation Problems in Closed-Loop 14

1.2.3 Closed-Loop Identifiability Problem 15

1.2.4 Special-Purpose Closed-Loop Identification Software 16

1.3 Research Objectives 16

1.4 Research Approach 17

1.5 Research Contribution 17

1.6 Organization of the Dissertation 18

2 Process Description 20

2.1 Introduction 20

2.2 Overview of the Industry 20

2.3 Operational Description of the MIBK Process 21

2.4 Controller 24

2.4.1 Model Predictive Controllers in General 24

2.4.2 Dynamic Matrix Controllers 28

2.4.3 The DMCplus Controller implemented on the MIBK plant 28

2.5 Available Process Information 29

2.5.1 Process Model 29

2.5.2 Measured Data 30

2.5.3 Other Relevant Information 31

2.6 Conclusion 31

3 Closed-Loop System Identification Theory 33

3.1 Introduction 33

3.2 System Identification Problem 33

3.3 Closed-Loop Configuration: Assumptions and Notation **36**
3.4 Closed-Loop Correlation Problem **38**
3.5 Informative Closed-Loop Experiments **39**
3.6 Different Approaches to Closed-Loop Identification **41**
3.7 Closed-Loop Identification in the Prediction Error Framework **42**
3.7.1 Assumptions and Notation 42
3.7.2 Prediction Error Estimation Method 44
3.7.3 Family of Model Structures 45
3.7.4 Computing the Estimate 47
3.7.5 Consistency and Identifiability 48
3.7.6 Bias Distribution 50
3.7.7 Asymptotic Variance Distribution 52
3.8 Excitation Signals **53**
3.9 Inter-Sampling Approach **55**
3.9.1 Introduction 55
3.9.2 Model Representation 55
3.9.3 Identifiability Analysis 57
3.10 Conclusion **58**

4 Selection of a Methodology Applicable to MPC Controlled Plants **61**
4.1 Introduction **61**
4.2 Closed-Loop Identification Approach **61**
4.3 Guarantee of Identifiability **63**
4.4 Evaluation of the Inter-Sampling Approach **65**
4.4.1 Set-Up of the Variance Simulation 65
4.4.2 Results of Variance Simulation 67
4.4.3 Discussion of the Variance Simulation Results 73
4.4.4 Conclusion of the Variance Simulation 74
4.5 Model Structure Selection **74**
4.5.1 Model Consistency 74
4.5.2 Compactness and Parametric Structures 75
4.5.3 Numerical Complexity 76
4.5.4 Multivariable Models 76
4.5.5 Model Order Selection 77

4.6 Model Validation ... **78**
4.6.1 Standard Validation with the Closed-Loop Data ... 79
4.6.2 Comparison with the Open-Loop Identified Model ... 81
4.6.3 Examination of the Closed-Loop System ... 82
4.7 System Identification Toolbox ... **83**
4.8 System Identification Steps ... **84**
4.8.1 Experiment Design ... 84
4.8.2 Data Collection ... 85
4.8.3 Model Structure Selection ... 86
4.8.4 Model Estimation ... 87
4.8.5 Model Validation ... 87
4.8.6 Methodology Validation ... 87
4.9 Conclusion ... **88**

5 Validation and Evaluation of the Methodology with Simulations **90**
5.1 Introduction ... **90**
5.2 Simulation Set-Up ... **90**
5.2.1 Plant ... 90
5.2.2 Controller ... 91
5.2.3 Simulation Scenarios ... 92
5.3 Implementation of the Methodology ... **94**
5.3.1 Experiment Design ... 95
5.3.2 Data Collection ... 95
5.3.3 Model Structure Selection ... 98
5.3.4 Model Estimation ... 98
5.3.5 Model Validation ... 99
5.3.6 Methodology Validation ... 99
5.4 Validation Results ... **100**
5.4.1 Expected Results ... 100
5.4.2 Obtained Results ... 101
5.4.3 Discussion ... 125
5.5 Conclusion ... **128**

6 Validation and Evaluation of the Methodology with Real Process Data **130**
6.1 Introduction ... **130**

6.2 Implementation of the Methodology **130**
6.2.1 Experiment Design 130
6.2.2 Data Collection 131
6.2.3 Model Structure Selection 132
6.2.4 Model Estimation 132
6.2.5 Model Validation 133
6.2.6 Methodology Validation 133
6.3 Validation Results **135**
6.3.1 Expected Results 136
6.3.2 Obtained Results 136
6.3.3 Discussion 144
6.4 Evaluation of the Proposed Methodology **147**
6.4.1 Reasons for Success/Failure of the Implemented Methodology 147
6.4.2 Recommendation for Future Implementation of the Methodology 147
6.5 Conclusion **148**

7 Conclusions and Future Research **150**
7.1 Introduction **150**
7.2 Review of the Closed-Loop Techniques **150**
7.3 Selection of a Methodology Applicable to MPC Controlled Plants **151**
7.4 Validation and Evaluation of the Methodology with Simulations **153**
7.5 Validation and Evaluation of the Methodology with Real Process Data **153**
7.6 Direction of Future Research **154**

References **155**

A Closed-Loop System for the Inter-Sampling Simulation Study **160**

B Approximate Realization of Step Response Data **162**

C Residual Analysis of Experimental Data **164**

CHAPTER 1
INTRODUCTION

1.1 PROBLEM STATEMENT

Model-based Predictive Control (MPC) constitutes a class of control algorithms that make direct use of a process model [1]. In recent times, MPC has become one of the dominant methods of advanced industrial process control.

Central to the success of the MPC technique is the derivation of accurate process models. It has also been found that model accuracy degrades with time after the MPC controller has been commissioned. Consequently, controller performance is adversely affected. Therefore, periodic re-identification of the process models and subsequent controller redesign is necessary to ensure optimal long-term controller performance [2].

Industrial project experience has shown that the most difficult and time-consuming work in an MPC project is modeling and identification [3]. In practice, during controller design, the process models are obtained by conducting open-loop step testing on the relevant process unit [4]. Widespread applications of MPC technology call for a more effective and efficient method of multivariable process identification.

The open-loop step testing approach works for stable processes, but the cost is very high. Stepping the manipulated variables may disturb the product quality and the test time is also very long, which occupies much manpower and makes production planning difficult. Furthermore, the tests are done manually, which dictates extremely high commitment of engineers and operators [4]. When the process is nonlinear, ill-conditioned or sensitive then it is also very difficult to carry out open-loop tests. When a process is unstable in open-loop, an open-loop test is also undesirable [5].

Closed-loop System Identification (SID) of the process models may well address some of these issues. This identification technique is less intrusive and may reduce re-identification time considerably [6]. Plant models identified while under closed-loop control also provide better models for controller design than models identified in open-loop [7].

Sasol Chemical Industries (SCI) currently utilise two MPC techniques: Dynamic Matrix Control (DMC) and Robust Model Predictive Control (RMPCT), supplied by AspenTech [8] and Honeywell [9] respectively. A DMC controller is used to control the industrial chemical process, namely the Methyl Iso Butyl Ketone (MIBK) plant in Sasolburg, South Africa. For the above-mentioned reasons it was considered necessary to develop a closed-loop system identification methodology for this plant.

The ability to identify the process model while the existing MPC controller is operating will ensure that safety, product quality and optimality concerns are met [10].

In Section 1.2 an overview of the current activities in the closed-loop SID field is provided. Then, in Section 1.3 the objectives are given and in Section 1.4 the approach to solving this problem is discussed. The contribution of this work is discussed in Section 1.5 and, lastly, in Section 1.6, the organization of the dissertation is explained.

1.2 Literature Review

Closed-loop SID refers to the process in which plant models are identified using data collected from closed-loop experiments, where the underlying process is fully or partly under feedback control [11].

In this section the available literature on closed-loop SID is discussed in terms of the advantages of closed-loop SID, the identifiability and correlation problems of closed-loop SID, and special-purpose software for closed-loop SID.

1.2.1 Advantages of Closed-Loop Identification

It is sometimes necessary to perform identification experiments in closed-loop as the plant may be unstable, or has to be controlled for production, economic, or safety reasons, or contains inherent feedback mechanisms [11]. Also, according to Landau [7], the problem of identifying systems operating under output feedback appears in practice to be one of the most convenient methods to provide good controller design models.

It is, therefore, not surprising that this problem has generated significant interest in the identification literature over the years [6]. The problem offers many possibilities and also some fallacies, and a wide variety of approaches have been suggested, many quite recently.

In the past identification in closed-loop was considered difficult, but it is now considered a very feasible approach that offers a number of practical advantages over open-loop identification [7]:

- validation of the designed controller and on-site re-tuning,
- obtaining better models for controller design,
- controller maintenance,
- iterative identification in closed-loop and controller redesign, and
- controller order reduction.

Controller maintenance involves validation of a controller, and re-tuning of the controller on the basis of a newly identified model. Since closed-loop SID makes it possible to do the validation and identification with the same set of data, it has an advantage over open-loop SID. Also, the model obtained from closed-loop data gives a more accurate image of the current closed-loop behaviour [7].

By identifying a reduced order plant model, which captures characteristics at the critical frequencies for control, the order of a controller can be reduced. A reduced model identified in closed-loop captures these characteristics and this makes closed-loop SID advantageous [7].

The effect of feedback will have an additional advantage if the process is ill-conditioned, meaning that several Controlled Variables (CVs) are strongly correlated. High purity distillation columns, typically controlled by MPC, are often ill-conditioned as the top and bottom compositions are strongly correlated. For the control of ill-conditioned processes, it is important to identify a good estimate of the low-gain direction. In an open-loop test where Manipulated Variables (MVs) are moved independently, the low gain direction has very low power and cannot be estimated accurately from noisy data. In order to amplify the power in the low-gain direction, correlation between MVs are needed. This correlation can be created naturally by feedback control [3].

Closed-loop identification also reduces the disturbances to process operation and eliminate the production of off-specification product. This is because in closed-loop one can specify the amplitude of the set-point movement and the controller will help to keep the CVs within their operation limits. This in turn means that closed-loop tests are less demanding of operators and control engineers [3].

1.2.2 Correlation Problems in Closed-Loop

Unfortunately most of the standard open-loop estimation methods fail when applied directly

Different ways of satisfying the identifiability condition have been analysed and tested. In many traditional methods, identifiability is ensured by adding a Persistently Exciting (PE) test signal into the closed-loop. For this reason many papers on closed-loop SID discuss the topic of experiment design. Many different types of test signals have been designed and tested to ensure PE signals [2, 4, 10, 11, 16, 17]. Some of these will be discussed in Section 4.3.

A novel approach to direct closed-loop identification, called output inter-sampling, has been introduced recently. It is claimed that by using the inter-sampled plant input-output data, traditional restrictive identifiability conditions are removed [18].

1.2.4 Special-Purpose Closed-Loop Identification Software

Special-purpose closed-loop SID software packages exist: Adaptech, for example, developed closed-loop identification software named, WinPIM-CL: Plant Model Identification in Closed-loop Operation [19]. This software uses modified closed-loop Output Error (OE) identification methods and modified IV methods. Furthermore, a closed-loop SID toolbox, called Closed-loop SID (CLOSID), has been developed by Van den Hof *et al.* [20] for the identification of Single-Input-Single-Output (SISO) systems. This toolbox is designed as an add-on to the MathWork's System Identification Toolbox (SITB). It compromises of several closed-loop SID methods.

Closed-loop SID software applicable to plants, controlled by MPC controllers, has also been developed: Butoyi and Zhu [21], who have studied closed-loop identification of large-scale industrial processes for use in MPC, developed closed-loop SID software for model based process control, named Tai-Ji ID: Automatic Closed-Loop Identification. This software uses the Asymptotic Method (ASYM) and is programmemed using MATLAB and the Control System Toolbox. The ASYM method of identification was developed by Zhu [4] to solve industrial identification problems systematically.

In practice, however, most of the industrial control engineers, including those at Sasol, still conduct the standard open-loop step tests on the relevant process units [4] and do not use these closed-loop SID software to obtain the process models.

1.3 Research Objectives

The objectives of this research project are to:

- do a literature review of closed-loop identification techniques,
- develop a closed-loop system identification methodology,
- develop closed-loop system identification software, if necessary, and
- apply the methodology and software to the industrial Methyl Iso Butyl Ketone plant, controlled by a Dynamic Matrix Controller.

1.4 Research Approach

The approach to the research project is to:

- review all the relevant closed-loop identification techniques,
- choose the most appropriate closed-loop identification methodology for plants, controlled by MPC controllers,
- use simulations to validate this methodology,
- use the methodology to identify a part of the MIBK plant from real process data,
- compare the identified model with the open-loop identified process model, and
- evaluate the implemented methodology.

1.5 Research Contribution

Closed-loop SID is a very interesting topic of great practical relevance. Although it has been studied for at least four decades, there are still many open problems to address.

Although identification in closed-loop appears in practice to be one of the most convenient methods to provide good design models and although there is an urgent need for efficient and effective identification methods in control engineering, many industrial control engineers do not even use most of the open-loop identification techniques developed over the last thirty years [4, 22].

Substitution of the open-loop identification technique with the closed-loop identification methodology could save Sasol, as well as other companies in the same field, valuable time and production cost. This will in turn ensure the success of the MPC technique.

A complete discussion of a closed-loop methodology that solves industrial identification problems systematically will encourage and help industrial control engineers to implement

Chapter 6 then discusses the implementation of the methodology proposed in Chapter 4 on the MIBK process data, the validation process that is followed and the results that are obtained. The implemented methodology is thoroughly evaluated and further recommendations are made.

Chapter 7 concludes the dissertation and also discusses the direction of possible future research.

CHAPTER 2

PROCESS DESCRIPTION

2.1 INTRODUCTION

In this chapter the MIBK process that requires a closed-loop SID methodology is described. In order to place this process in context, a brief overview of the industry in which this process is used is given in Section 2.2. To give an idea of the complexity of the plant and the importance of the process, the process itself is discussed in a simplified manner in Section 2.3. The controller, called DMCplus, that controls this plant is then described in Section 2.4. Lastly, other relevant process information, which is needed to design and implement the closed-loop SID methodology, is discussed in Section 2.5. This information includes a description of the original identified process model and the measured data sets that were used in the identification process.

Furthermore, because the plant in question is controlled by an MPC controller, an overview of MPC controllers is included in Section 2.4. The characteristics of these controllers are taken into account when the identification methodology is chosen.

2.2 OVERVIEW OF THE INDUSTRY

The particular process for which the closed-loop SID methodology was developed is used within Sasol Limited. The Sasol group of companies comprises diversified fuel, chemical and related manufacturing and marketing operations. Sasol has interests in oil and gas exploration and production, in crude oil refining and liquid fuels marketing. Its principal feedstocks are obtained from coal. The company converts this coal into value-added hydrocarbons through Fischer-Tropsch process technologies [23].

During their first three decades, Sasol's primary goal was to produce high-quality synthetic fuels from coal for the local market. During the 1990s, Sasol's interests have been shifting towards developing higher-value chemicals for a wider spectrum of applications [23].

Currently, Sasol comprises of several main operating companies. One of these companies is SCI in Sasolburg. This company manufactures more than 200 chemical products, e.g.

ammonia, fertilisers, explosives, linear alpha olefins, carbon and tar products, plastics and a broad range of solvents [23]. These products are derived principally from the benefaction of coal at Sasolburg and from various feedstocks purchased from Sasol Synthetic Fuels (SSF) in Secunda.

2.3 Operational Description of the MIBK Process

A very simplified representation of the MIBK process at SCI in Sasolburg is given in Fig. 2.1. This figure shows that this process consists of two reactors and five distillation columns. For simplicity, the reboilers, which provide the necessary vaporization for the distillation processes, the reflux drums, which hold the condensed vapour from the top of the columns so that liquid (reflux) can be recycled back to the columns, the heat exchangers, valves, pumps and sensors, are not shown in Fig. 2.1 [24].

The purpose of the distillation columns is to separate liquid substances of different densities, using the difference in boiling points. Carefully judged heating evaporates only the lighter substances and subsequent cooling of the resulting gas condenses it back to a liquid [24]. The reactors are industrial plants that provide the right physical conditions for specific chemical reactions to take place [23].

Process Feed: The plant was designed to produce MIBK from the feedstock, Dimethyl Ketone (DMK). DMK, also called acetone, is a manufactured chemical that is found in the natural environment as well. However, industrial processes contribute more acetone to the environment than natural processes. It is a colourless liquid with a distinct smell and taste. It evaporates easily, is flammable and dissolves in water [25].

Export quality acetone, produced by SSF, is received daily by road tanker from Secunda and is off-loaded in Sasolburg. The DMK is then transferred via pump to the surge tank, shown in Fig. 2.1. Unreacted DMK, from the DMK column, is also recycled to this tank [26].

Reactor Section: DMK is fed to the reactors and heated to the reaction temperature with feed-effluent exchangers and steam heaters. The feed is split equally between the two reactors. The reactors are of tubular design, with each tube filled with a catalyst and the feed flowing from top to bottom. The shell side of each reactor is divided into three baffled sections through which tempered water flows. This water controls the reaction temperature by heating or cooling the reaction mixture. The temperature and flow rate of the water to each section are controlled.

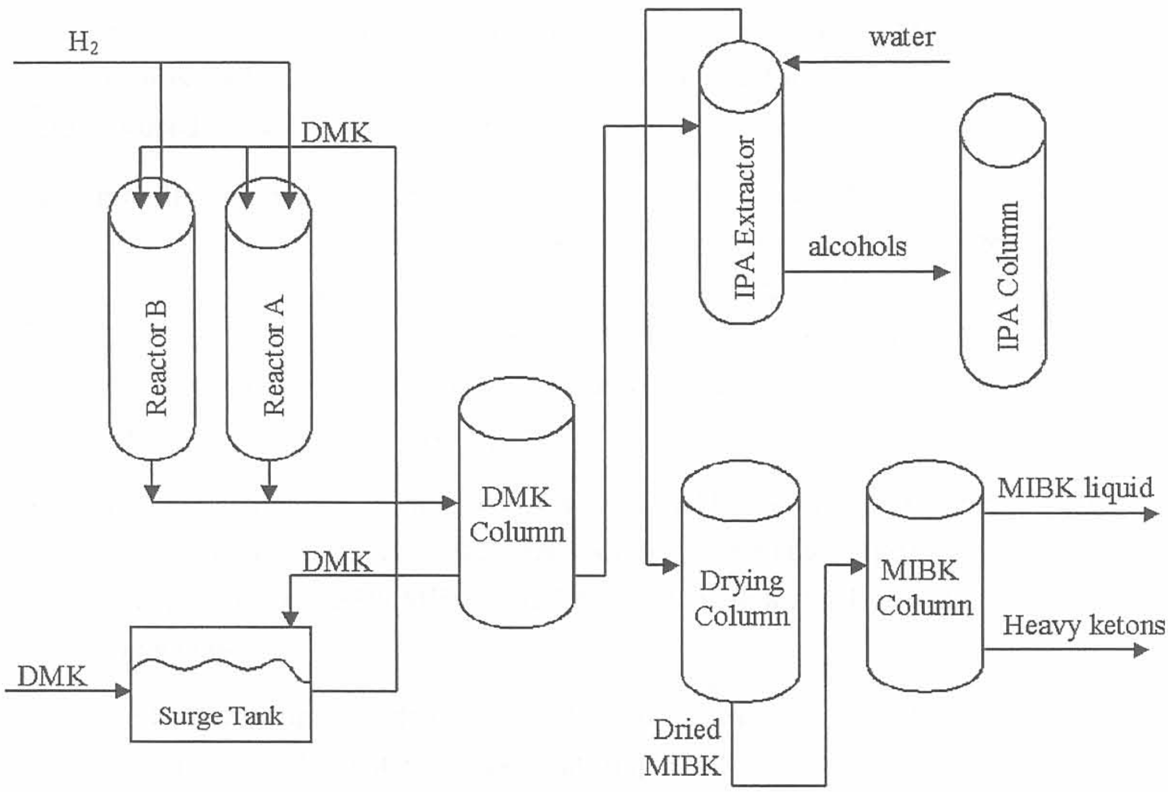

Figure 2.1: A simplified representation of the MIBK process.

Hydrogen is fed via piston compressors to the reactors. The reactor inlets are pressure controlled. The excess hydrogen is flared.

The reaction product collects in the bottom of the reactors. From here it flows under level control to the reactor feed/effluent exchangers, then to the reactor effluent separator and on to the distillation section [26].

Tempered Water System: A closed, circulating tempered water system is used to control the reaction temperatures. Hot water from the reactors is collected in hot water accumulators. The flow from the accumulators is split into two streams. One is heated in a steam heater and sent to the lower and middle sections of the reactors to supply heat for the reaction. The other stream is cooled in a cooling exchanger and then sent to the top and middle sections where it is used to cool down the top zone.

The middle zone water temperature is set via a three-way valve. The flow of the hot water to each zone is set on flow control and these are normally not changed. The flows are set sufficiently high to prevent a major rise in temperature. The inlet temperature sets the heat transfer rate in the reaction sections. There are four temperature indicators in each section. The indicator with the highest temperature is usually selected to be controlled. This temperature increases as the catalyst ages [26].

DMK Column: The DMK column recovers unreacted DMK by withdrawing it as a purified side product from the feed stream. It is then recycled back to the surge tank. The feed to the column is approximately 25% MIBK and 75% DMK with a small percentage of water and other components.

The overhead vapour is condensed and collected in the reflux drum. The reflux is returned to the tower. There is also a steam heater that supplies reboil heat. The condensate from the reboiler is removed.

The bottom product containing the MIBK is fed from the DMK column, under level and flow control, to the Isopropyl Alcohols (IPA) extractor [26].

IPA Extractor: The IPA extractor column extracts alcohols and acids from the MIBK stream by using a counter-current liquid extraction with water as the extractant. Fresh water is supplied under flow control, and is removed as the extract phase from the bottom of the tower and then routed to the IPA column.

The MIBK-rich raffinate phase, which is sent to the drying column, is obtained from the top of the tower under pressure control [26].

IPA Column: The aim of the IPA column is to recover the organics from the water and to recover the MIBK that dissolved in the water [26].

Drying Column: The drying column removes the small amounts of water entrained in the MIBK from the extractor. The bottom product, containing the dried MIBK, is routed to the MIBK column [26].

MIBK Column: The drying column bottom product, which is the feed to the MIBK column, contains essentially pure MIBK with some lighter and heavier components. The MIBK is withdrawn in the MIBK column as a side product under ratio control at approximately 85% of the feed flow. A heavy ketones bottom product is withdrawn from the bottom of the MIBK column. The heavy ketones still contain a small percentage of MIBK [26].

There is a worldwide demand for MIBK and most of the product is, therefore, exported. MIBK, which is a colourless liquid with a pleasant odor, has multiple uses: solvent for paints; varnishes; organic syntheses; extraction process (including extraction of uranium); organic syntheses solvent for protective coating; dewaxing of minerals and oils; synthetic flavoring adjacent; etc. [27].

2.4 Controller

The MIBK plant is controlled by an MPC controller. A general discussion of these controllers is given. DMC controllers are then described and the implementation of the controller on the MIBK plant is discussed.

2.4.1 Model Predictive Controllers in General

MPC is a class of computer control algorithms that explicitly use a process model to predict future plant outputs and compute appropriate control action (new plant input) through on-line (real-time) optimisation of a cost objective over a future horizon, subject to various constraints. Process measurements provide the feedback and, optionally, feedforward element in the MPC structure [1].

Fig. 2.2 shows the structure of a typical MPC system. It shows that a number of options exist for:

- input-output model,

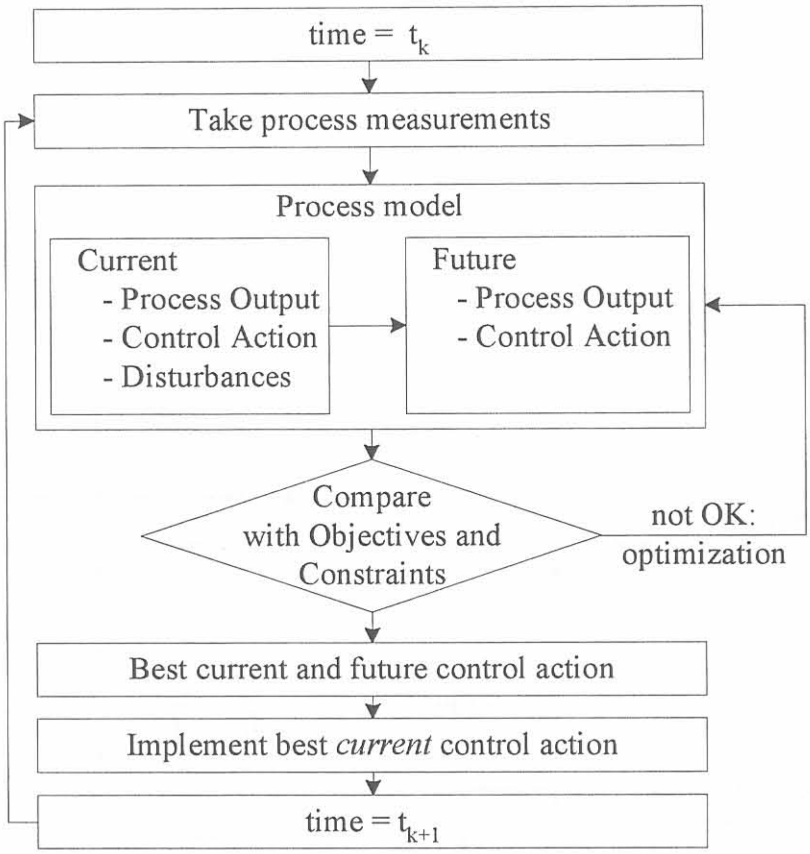

Figure 2.2: The structure of a typical MPC system.

- disturbance prediction,
- prediction and control horizon,
- objective,
- measurement,
- constraints, and
- sampling period (how frequently the on-line optimisation problem is solved) [28].

The possibilities for on-line optimisation are numerous. Fig. 2.2 also shows that the behaviour of an MPC system can be quite complicated, because the control action is determined as the result of the on-line optimisation problem [28].

The MPC control law for the SISO case can most easily be explained by referring to Fig. 2.3. The MIMO case is similar.

For any assumed set of present and future control moves $\Delta u(k)$,$\Delta u(k+1)$,...,$\Delta u(k+i-1)$ the future behaviour of the process outputs $\hat{y}(k+1 \mid k)$, $\hat{y}(k+2 \mid k)$, ..., $\hat{y}(k+h \mid k)$ can

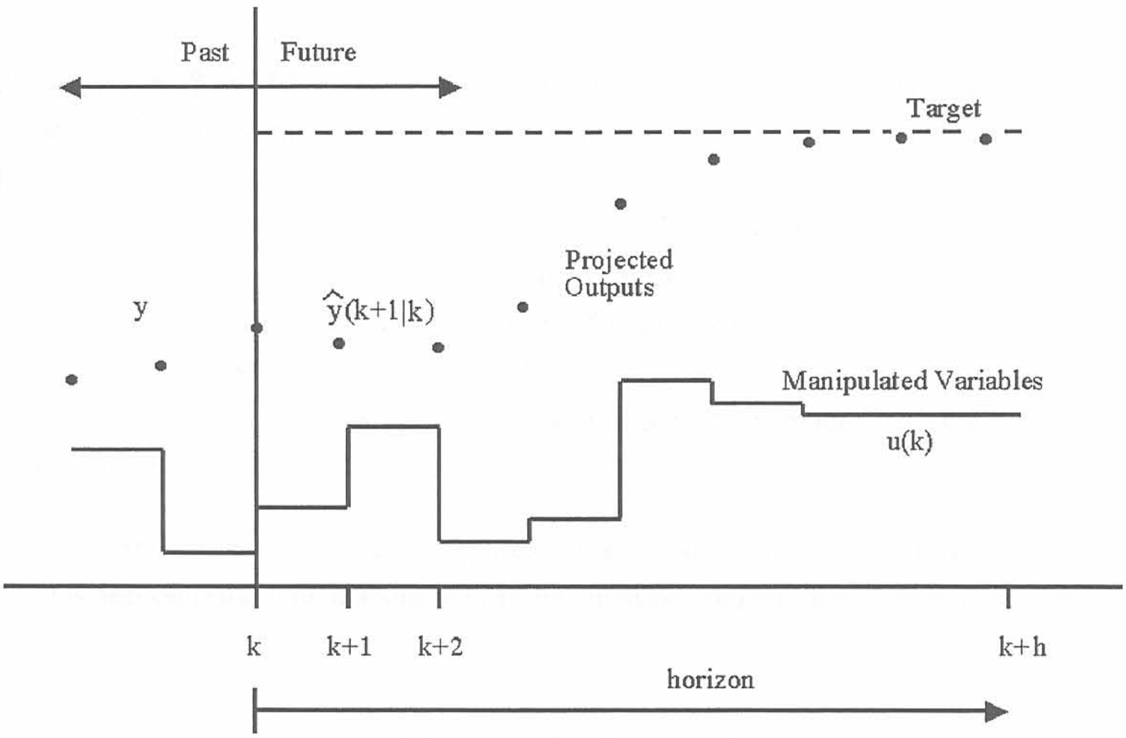

Figure 2.3: The MPC control law [29].

be predicted over a horizon h. The i present and future control moves ($i \leq h$) are computed to minimise a quadratic objective of the form of Eqn. (2.1):

$$\min_{\Delta u(k)\ldots\Delta u(k+i-1)} \sum_{l=1}^{h} \left\|\Gamma_l^y \left[\hat{y}(k+l \mid k) - r(k+l)\right]\right\|^2 + \sum_{l=1}^{i} \left\|\Gamma_l^u \left[\Delta u(k+l-1)\right]\right\|^2, \text{ where} \quad (2.1)$$

$$\|x\|^2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}.$$

In this equation Γ_l^y and Γ_l^u are weighted matrices to penalise particular components of $y(k)$ or $u(k)$ at certain future time intervals. $r(k+l)$ is the, possibly time-varying, vector of future reference values (set-points). Though i control moves are calculated, only the first one, $\Delta u(k)$, is implemented. At the next sampling interval, new values of the measured outputs are obtained, the control horizon is shifted forward by one step and the same computation is repeated. The resulting control law is referred to as *moving horizon* or *receding horizon*. The predicted process outputs $\hat{y}(k+1 \mid k)$, $\hat{y}(k+2 \mid k)$, ..., $\hat{y}(k+h \mid k)$ depend on the measurement $y(k)$, the process model and the assumptions made about the disturbances affecting the outputs [29].

The control action can also be computed subject to hard constraints on the manipulated variables and outputs. The constraints can be: manipulated variable constraints as in Eqn. (2.2)

$$u_{\min}(l) \leq u(k+l) \leq u_{\max}(l); \quad (2.2)$$

or manipulated variable slew rate constraints as in Eqn. (2.3)

$$|\Delta u(k+l)| \leq \Delta u_{\max}(l); \quad (2.3)$$

or output variable constraints as in Eqn. (2.4)

$$y_{\min}(l) \leq \hat{y}(k+l \mid k) \leq y_{\max}(l). \quad (2.4)$$

When hard constraints of this form are imposed, a quadratic programme has to be solved at each time step to determine the control action and the resulting control law is then, generally, nonlinear [29].

2.4.2 Dynamic Matrix Controllers

As mentioned, the DMC controller, which also controls the MIBK plant, is a special type of MPC controller. Key features of the DMC control algorithm include:

- linear step response model for the plant,
- quadratic performance objective over a finite prediction horizon,
- future plant output behaviour specified to follow the set-point as closely as possible, and
- optimal inputs computed as the solution to a Least-Squares Estimate (LSE) problem [1].

The linear step response used by the DMC algorithm relates changes in a process output to a weighted sum of past input changes, referred to as input moves. For the SISO case the step response is represented in Eqn. (2.5):

$$\hat{y}_{k+1} = y_0 + \sum_{j=1}^{h} s_j \Delta u_{k+1-j}. \tag{2.5}$$

The move weights s_j are the step response coefficients and h is the model horizon. Mathematically the step response can be defined as the integral of the impulse response. By using the step response model one can write predicted future output changes as a linear combination of future input moves. The matrix that ties the two together is called a *dynamic matrix* [30].

2.4.3 The DMCplus Controller implemented on the MIBK plant

The marketing strategy for MIBK indicated that every ton of product made can be sold. As a result Sasol's planning objective is to maximise MIBK production at all times.

AspenTech [8] was asked to develop a controller for the MIBK plant with the following objectives in mind:

- maximise production of on-specification MIBK,
- maximise overall yield of MIBK/ton of DMK,
- minimise production of by-products,

- minimise excess hydrogen consumption, and
- minimise energy consumption.

A constrained multivariable DMC controller, called DMCplus, was designed and implemented in September 1998. This controller controls both the reaction and distillation section. Due to slow time constants, a second DMCplus controller was implemented for the slow hexene quality control in the DMK column distillate flow. The second controller has only the distillate flow rate as a manipulated variable. The controllers were joined together with feed-forward variables [26].

2.5 Available Process Information

In this section the original identified process model is described, the measured data sets used in the identification process, is described, and other relevant information regarding the plant and controller is discussed.

2.5.1 Process Model

The process can be modelled as a MIMO system with 56 CVs, and 38 MVs. This model can be represented as a 38x56 matrix with each element representing a SISO model. Luckily this model is partly diagonal (sparse). This means that many of the off-diagonal elements are zero. This property makes it possible to brake down the model into smaller MIMO models that can then be identified separately from the rest of the MVs and CVs.

For this reason only one of these isolated models is used for validation of the developed closed-loop SID methodology in Chapter 4, i.e. an isolated part of reactor A. This part can be modelled as a 4x4 MIMO system [26].

The DMC controller controls the reactor section as follows [26]:

- it tries to keep the temperatures in each zone at their maximum, actuator authority permitting,
- it tries to keep the valves of the condensate flows from saturation,
- it tries to control excess hydrogen flow, and
- it slowly reduces the feed to the reactor if the DMK feed drum level drops too low.

The descriptions and tags of the related CVs considered in the validation process of the proposed closed-loop SID methodology are [26]:

- reactor A zone 1 water inlet temperature valve position (05TIC280A.MV),
- reactor A zone 2 water inlet temperature valve position (05TIC232A.MV),
- reactor A zone 3 water inlet temperature valve position (05TIC279A.MV), and
- reactor A zone 1 water flow valve position (05FIC223A.MV),

The related MVs considered in the validation process are [26]:

- reactor A zone 1 water inlet temperature (05TIC280A.SV),
- reactor A zone 2 water inlet temperature (05TIC232A.SV),
- reactor A zone 3 water inlet temperature (05TIC279A.SV), and
- reactor A zone 1 water flow (05FIC223A.SV),

When AspenTech developed the DMC controller in 1998, open-loop step tests were first conducted in order to identify the MIBK plant [26]. The identified models were stored in step response format. The intention, in this work, is to use these step response models for the comparison and validation of the models identified from the measured process data with the proposed closed-loop SID methodology.

2.5.2 Measured Data

No structured open-loop step tests or any structured closed-loop tests were allowed on the plant. Plant inputs and outputs are however usually logged every 30 seconds. The resulting data sets, logged since 1998, are stored in a database and it was possible to retrieve the desired data.

The logged CVs and MVs of reactor A for October and November 1998 were obtained. This time period was chosen as the controller was commissioned in September 1998, and presumably the plant did not change much from September to November 1998. If this is the case it can be expected that the open-loop identified model from which the controller was designed would be an accurate representation of the plant in October and November. Consequently, the models that were identified by AspenTech before the commissioning of the controller are used to validate the models identified from the closed-loop data, logged during these two months.

Unfortunately, the reference (set-point) values were not logged and, therefore, could not be retrieved. This fact limits the options available for the closed-loop SID methodology.

2.5.3 Other Relevant Information

Concerning the plant and controller, there are a few other relevant remarks to be considered for identification purposes:

The main controller is tuned for a 180 minute time to steady state with an execution time of once every minute to ensure good control for the fast models [26]. The plant data are sampled twice as fast, which means that the plant input-output data sets are inter-sampled by a factor of two.

Unfortunately, not enough information of the controller is available, which implies that the controller cannot be modelled and, therefore, the controller information cannot be used in the identification or validation process.

No disturbance information is available.

There are legally enforced maintenance shutdowns every 36 months, as well as shutdowns every six months to change the catalyst in the reactors. After a shutdown the plant model is usually not accurate any more and a new identified model is desired to ensure optimal long-term controller performance [26].

2.6 CONCLUSION

The MIBK process which was in need of the closed-loop SID methodology is used within a petro-chemical company, Sasol, and was designed to produce MIBK from the feedstock DMK. The process is implemented with reactors, distillation columns, heat exchangers, valves, pumps, sensors, etcetera. The plant is controlled by a DMC controller, which is a type of MPC controller. The DMC algorithm uses a linear step response model to relate changes in a process output to a weighted sum of past input changes, referred to as input moves.

It is clear from the discussion of MPC controllers that the behaviour of the MPC system can be quite complicated, because control actions are determined from the result of an on-line optimisation problem. Hard constraints are also frequently imposed, which results in a quadratic programme that is solved iteratively to determine the control action. The resulting control law is then generally nonlinear. Chapter 4 discusses why these facts restrict the choices in the closed-loop SID methodology.

An isolated part of a reactor that is part of the multivariable MIBK plant was chosen to be used for the validation of the developed closed-loop SID methodology. No structured tests were performed on the plant and logged data sets from normal operation were obtained for the identification process. These sets only contain the input and output signals of the plant. The reference signals are unknown. It is also not possible to model the controller for which no detailed information is available. These facts limit the available options for the closed-loop SID methodology. Since the input-output data sets are inter-sampled, the output inter-sampling approach can be added to the list of options to be considered in the development of the SID methodology.

CHAPTER 3

CLOSED-LOOP SYSTEM IDENTIFICATION THEORY

3.1 INTRODUCTION

In this chapter the basic closed-loop SID theory is explained. This chapter starts with a discussion of the SID problem in Section 3.2. Then, the closed-loop configuration, together with assumptions and notations, are discussed in Section 3.3. In Section 3.4 the correlation problem is discussed and the question of informative data is addressed in Section 3.5. The different approaches for avoiding the problems with closed-loop SID are then discussed in Section 3.6.

Section 3.7 deals with closed-loop SID in the prediction error framework. The algorithm format, different model structures, and the computation of the estimates are discussed. The relevant statistical properties of this method are also stated. The theory discussed in this section can readily be extended to the multivariable situation, as shown by Forssell and Ljung [12].

Because excitation signals have a very substantial influence on the observed data, these signals are also further discussed in Section 3.8. Lastly, Section 3.9 deals with the theory of the new inter-sampling approach to closed-loop SID, which may also ensure identifiability.

3.2 SYSTEM IDENTIFICATION PROBLEM

The system identification problem is to construct mathematical plant models from experimental data. This is an important problem in statistics and many identification methods, as well as the tools for analysing their properties, have their roots in statistics [6]. According to Ljung [11], to determine a model of a dynamic system, from observed input-output data, the following is required:

Data Record: Data might be available from normal operating records, but it may also be possible to specifically design identification experiments for a process in order to obtain specific information.

Set of Candidate Models (Model Structure): A set of candidate models is obtained by specifying within which collection of models one is going to look for a suitable model.

Criterion to Select a Particular Model in the Set: Given measurement data and a model set, one has to specify in which way the optimal model from the model set is going to be selected. The assessment of model quality is typically based on how models perform when they attempt to reproduce the measured data.

After one has arrived at a particular model, it remains to test whether this model is good enough. Such tests fall under the heading of model validation [11].

The system identification procedure has a natural logic flow: first collect data, then choose a model set, then pick the best model in this set. It is very possible that the model first selected will not pass the model validation tests. One must then go back and revise the various steps of the procedure [11].

Ljung [11] states that a model may be deficient for a variety of reasons:

- the numerical procedure can fail to find the best model according to the criteria,
- the criteria is not well chosen,
- the model set is not appropriate, in that it does not contain a good enough description of the system, or
- the data set is not informative enough to provide guidance in selecting a good model.

The major part of an identification application consists of addressing these problems, in an iterative manner, guided by prior information and the outcomes of previous attempts [11]. This iteration is depicted in Fig. 3.1. Interactive software is required to deal effectively with this problem.

According to Forssell [6], the system identification problem can therefore be divided into the following five subproblems:

- experiment design,
- data collection,
- model structure selection,
- model estimation, and
- model validation.

These steps are also applicable in closed-loop system identification and will be used in Section 4.8 to summarise the developed closed-loop SID methodology.

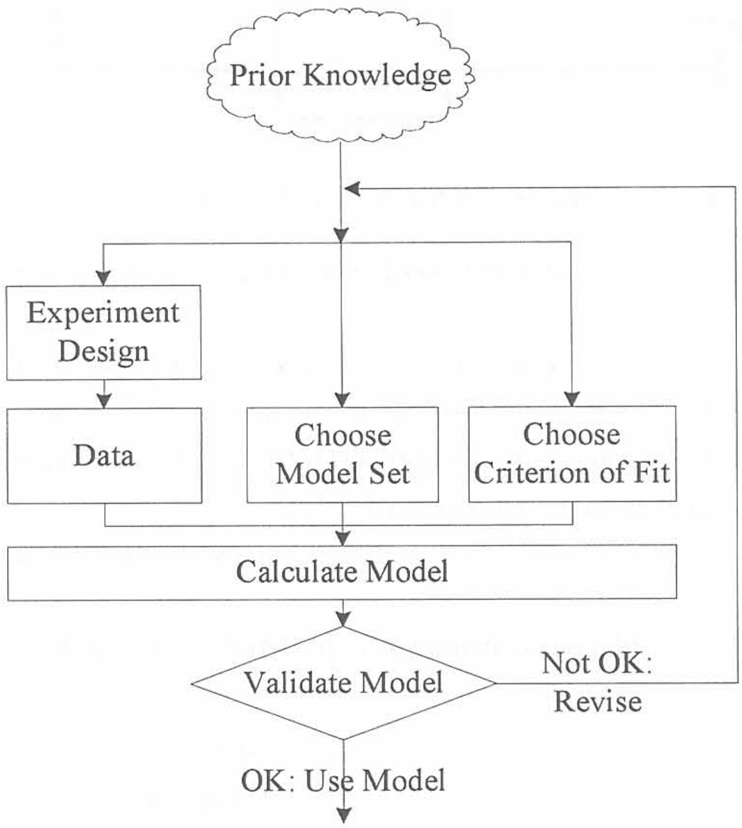

Figure 3.1: The system identification loop.

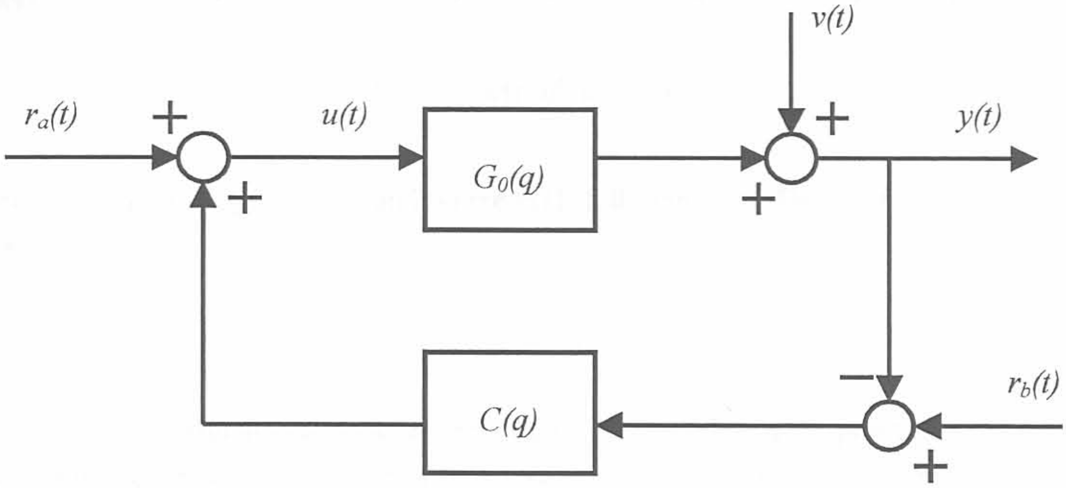

Figure 3.2: Configuration of a system operating in closed-loop

3.3 Closed-Loop Configuration: Assumptions and Notation

When an identification experiment is performed in closed-loop, i.e. the output is fed back to the input by means of some feedback mechanism, it is called closed-loop identification [6]. A typical situation is when a (SISO) system $G_0(q)$ is being controlled by some feedback controller $C(q)$, as in Fig. 3.2 [31]. Note that the input and outputs signals, used in subsequent discussions, are shown in Fig. 3.2.

The following set-up is considered [6]: The system is given by

$$\begin{aligned} y(t) &= G_0(q)u(t) + v(t), \\ v(t) &= H_0(q)e(t), \\ y(t) &\in \Re^{Nx1}, \ u(t) \in \Re^{Nx1}, \ v(t) \in \Re^{Nx1}, \ e(t) \in \Re^{Nx1}. \end{aligned} \tag{3.1}$$

Here $y(t)$ is the output, $u(t)$ the input, $v(t)$ the additive output noise, and $e(t)$ a white noise signal with zero mean and variance λ_0. N is the number of data samples. The symbol q donates the discrete-time shift operator:

$$q^{-1}u(t) = u(t-1). \tag{3.2}$$

Without loss of generality, it is assumed that $G_0(q)$ contains a delay and that the noise model $H_0(q)$ is monic, i.e. $H_0(q) = \sum_{k=0}^{\infty} h(k)q^{-k}$ and $h(0) = 1$ [11], and inversely stable, i.e.

$(H_0(q))^{-1}$ is stable. Furthermore, it can be assumed that the input is generated as

$$u(t) = k(t, y^t, u^{t-1}, r(t)), \tag{3.3}$$

where $y^t = [y(1), ..., y(t)]$, etc., and where $r(t)$, following Van Den Hof [31], is defined as follows:

$$r(t) := r_a(t) - C(q)r_b(t). \tag{3.4}$$

In Fig. 3.2 the signals $r_a(t)$ can be either a reference value, a set-point or a noise disturbance on the regulator output. Similarly, signal $r_b(t)$ can be either a set-point or a measured noise on the output signal. $r(t)$ is a quasi-stationary signal, independent of $v(t)$. $C(q)$ is a given deterministic function such that the closed-loop system is stable.

For now, it can be assumed that the feedback is linear and given by

$$u(t) = r(t) - C(q)y(t). \tag{3.5}$$

The closed-loop equations then become

$$\begin{aligned} y(t) &= S_0(q)G_0(q)r(t) + S_0(q)v(t), \\ u(t) &= S_0(q)r(t) - C(q)S_0(q)v(t), \end{aligned} \tag{3.6}$$

where $S_0(q)$ is the sensitivity function

$$S_0(q) = \frac{1}{1 + G_0(q)C(q)}. \tag{3.7}$$

The closed-loop system can also be rewritten as

$$\begin{aligned} y(t) &= G_0^c(q)r(t) + v_c(t), \\ v_c(t) &= H_0^c(q)e(t), \end{aligned} \tag{3.8}$$

where $G_0^c(q) = S_0(q)G_0(q)$ and $H_0^c = S_0(q)H_0(q)$.

The total expectation operator $\bar{E}$ [11] will also be used in subsequent sections:

$$\bar{E}x(t) = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} Ex(t), \tag{3.9}$$

where Ex, the mathematical expectation of the random vector x, is defined as

$$Ex = \int_{\Re^N} x f_x(t)dt, \tag{3.10}$$

where $f_x(t)$ is the Probability Density Function (PDF) of x.

Lastly, the power spectrum of a signal $s(t)$ will be denoted by $\Phi_s(\omega)$ and the cross spectrum between two signals $s(t)$ and $w(t)$ will be denoted by $\Phi_{sw}(\omega)$.

3.4 Closed-Loop Correlation Problem

Many of the well-known identification methods, which work well in open-loop, fail when applied directly to input-output data obtained under closed-loop control. The reason why these methods fail is the non-zero correlation between the input $u(t)$ and the additive output noise $v(t)$ [6].

These well-known identification methods, which cannot be applied directly to closed-loop data, are listed below with some short explanations of why they fail [6, 11].

Instrumental Variable Method: In open-loop it is common to let the instruments, used in the IV method, be filtered and delayed versions of the input. However, this straightforward approach will fail in closed-loop, because the input is correlated with the output noise.

Subspace Method: An important part of the subspace method is the choice of a suitable multiplication matrix, which is uncorrelated with the noise, in order to eliminate the noise term in the Hankel Matrix equation. This matrix is usually built up of delayed inputs. Again, this approach will fail because of the correlation of the input with the output noise.

Spectral Analysis: When the spectral analysis method is applied in a straightforward fashion, as described by Ljung [11], it will give erroneous results. The estimate of $G(e^{i\omega},\theta)$ will not converge to $G_0(e^{i\omega})$, but to

$$G_*(e^{i\omega},\theta) = \frac{G_0(e^{i\omega})\Phi_r(\omega) - C(e^{-i\omega})\Phi_v(\omega)}{\Phi_r(\omega)+ \mid C(e^{i\omega}) \mid^2 \Phi_v(\omega)}. \tag{3.11}$$

With no external reference signal present, i.e. $\Phi_r(\omega) = 0$, $G_*(e^{i\omega}, \theta)$ will equal the negative inverse of the controller, i.e. $G_*(e^{i\omega}, \theta) = -(C(e^{i\omega}))^{-1}$.

Correlation Analysis: The correlation analysis method will give a biased estimate of the impulse response, because the assumption $\bar{E}u(t)v(t-\tau) = 0$ is violated.

Output Error Models: The OE model will only give a consistent estimate of $G(q)$ if the additive noise is white in the case of closed-loop data, because the noise model is assumed fixed and equal to one.

Thus, the above-mentioned methods are not directly applicable to closed-loop data. Luckily, the Prediction Error Method is directly applicable to closed-loop data, as long as the parameterizing is flexible enough. Ljung [11] states that the PEM estimation method will consistently estimate the system if

- the data are informative, and
- the model set contains the true system (both noise model and dynamics model),

irrespective of whether the data set $\{u(t), y(t)\}$ has been collected under feedback control. It is concluded by Ljung [11] that under the above assumptions, the PEM estimate will have optimal accuracy. The PEM method is described in Section 3.7.

3.5 Informative Closed-Loop Experiments

Forssell [6] says that in order to be able to identify a system it is necessary for the data to be sufficiently rich in such a way that it is possible to uniquely determine the system parameters.

In open-loop the situation is rather transparent and it can be shown that the data set is *informative enough* if the input is persistently exciting of sufficiently high order [11]. This means that the power spectrum of the input signal, $\Phi_u(\omega)$, should be non-zero on at least n points in the interval $-\pi < \omega < \pi$ [11]. Here n is the necessary order of the PE input, and this equals the number of parameters to be estimated.

In closed-loop, the situation is less transparent. The following standard example shows what can happen [6, 11]:

Example: *Proportional Feedback*

Consider the first-order model structure

$$y(t) + ay(t-1) = bu(t-1) + e(t), \tag{3.12}$$

and suppose that the system is under proportional control during the experiment:

$$u(t) = -fy(t). \tag{3.13}$$

Inserting the feedback law into the model gives

$$y(t) + (a+bf)y(t-1) = e(t), \tag{3.14}$$

which is the model of the closed-loop system. From Eqn. (3.14) it can be concluded that *all models* $(\hat{a}, \hat{b})$, subject to

$$\begin{aligned} \hat{a} &= a + \gamma f, \\ \hat{b} &= b - \gamma, \end{aligned} \tag{3.15}$$

with γ an arbitrary scalar, give the same input-output description of the system as the model (a, b) under proportional feedback, as described in Eqn. (3.13). There is consequently no way to distinguish between model (a, b) and all models $(\hat{a}, \hat{b})$. It is of no help to know the regulator parameter f. The experimental condition, as described in Eqn. (3.13), is not informative enough with respect to the model structure of Eqn. (3.12). It is true, though, that the input signal $u(t)$ is PE, since it consists of filtered white noise. Persistence of excitation of the inputs signal $u(t)$ is therefore not a sufficient condition in closed-loop experiments.

If the model structure, given by Eqn. (3.12), is restricted by, for example, constraining b to be 1:

$$y(t) + a(t-1) = u(t-1) + e(t), \tag{3.16}$$

then it is clear that the data, generated as in Eqn. (3.13), are sufficiently informative to distinguish between values of the a-parameter.

Consequently, it could be problematic to obtain relevant information from closed-loop experiments. Conditions for informative data sets must also involve the feedback mechanisms

[11, 31]. An explicit condition, which must be satisfied for the data to be informative, is given in Section 3.7.5.

3.6 DIFFERENT APPROACHES TO CLOSED-LOOP IDENTIFICATION

In Section 3.4 some examples were given of methods that fail when applied directly to closed-loop data. There are some alternative ways to implement the methods to make them applicable to closed-loop situations. According to Forssell [6], a large number of methods have been developed, in particular a number of PEM methods, that are applicable to closed-loop identifications. Depending on what assumptions are made on the nature of the feedback, all closed-loop identification methods can be classified as *direct*, *indirect* and *joint input-output* methods [6, 11]:

Direct Approach: In the direct approach, the basic PEM estimation method is applied in a straightforward manner. No assumptions whatsoever are made on how the data set was generated. The output $y(t)$ of the process and the input $u(t)$ are used in the same way as for open-loop identification.

Indirect Approach: The methods used in the indirect approach assume perfect knowledge of the controller used in the identification experiment. The closed-loop system is identified from reference input $r(t)$ to output $y(t)$. From this identified closed-loop system the open-loop system (plant) is retrieved making use of the known controller $C(q)$. Ljung [11] states that any error in the assumed regulator will directly cause a corresponding error in the estimate of $G(q)$. Because most regulators contain nonlinearities, the indirect identification approach could yield incorrect plant models.

Joint Input-Output Approach: In the joint input-output approach the output $y(t)$ of the process and the input $u(t)$ of the plant are considered as outputs of a system driven by the reference input $r(t)$ and unmeasured noise $v(t)$. Knowledge of the system and the controller is recovered from this joint model. Therefore, in this approach knowledge of the controller is not required, but the controller structure must be known.

The indirect and the joint input-output approaches are typically only used when the feedback law is linear, as in Eqn. (3.5), but can also be applied when the feedback is nonlinear. Of course, in the case of nonlinear feedback, the estimation problems become much more involved [6].

Forssell [6] emphasizes the fact that the direct approach is only applicable when the PEM method and some of the subspace methods are used. The reason for this is the unavoidable correlation between the input and the additive output noise, which, as was shown, rules out most other methods. With the indirect and joint input-output approaches the closed-loop problem is converted into an open-loop problem, because the reference signal, which serves as the input, is uncorrelated with the output noise. Therefore, these approaches are similar to the open-loop methods [6].

In Section 3.7 the PEM methods that fall into these three categories, will be discussed.

3.7 Closed-Loop Identification in the Prediction Error Framework

3.7.1 Assumptions and Notation

The notation, as used by Forssell and Ljung [12], will be used to discuss closed-loop SID in the PEM framework. A parametrised set of models like Eqn. (3.17) is called a model structure and is denoted by M [11].

$$y(t) \;=\; G(q,\theta)u(t) + H(q,\theta)e(t). \tag{3.17}$$

$G(q,\theta)$ will is called the *dynamic model* and $H(q,\theta)$ the *noise model*. $H(q,\theta)$ is assumed monic. The parameter vector θ ranges over a set $D_M \subset \Re^d$ ($d =$ dimension of θ), which is assumed compact and connected [32].

The model structure given by Eqn. (3.17), where $\theta \in D_M$, describes a model set. The true system is contained in the model set if, for some $\theta_0 \in D_M$,

$$G(q,\theta_0) \;=\; G_0(q), \quad H(q,\theta_0) = H_0(q). \tag{3.18}$$

This can also be written as $S \in M$, where S is the true system (noise and plant transfer functions). The case were the true noise properties cannot be described by any noise model found within the model set, but were there exists a $\theta_0 \in D_M$ such that

$$G(q,\theta_0) \;=\; G_0(q), \tag{3.19}$$

can be denoted as $G_0 \in Ģ$, where Ģ is a set of transfer functions obtained in a given structure.

The following notation is used [6]:

$$T(q,\theta) = \begin{bmatrix} G(q,\theta) \\ H(q,\theta) \end{bmatrix}, \tag{3.20}$$

$$T_0(q) = \begin{bmatrix} G_0(q) \\ H_0(q) \end{bmatrix}. \tag{3.21}$$

The notation requires the introduction of the combined signal consisting of $u(t)$ and $e(t)$ in the following way:

$$\chi_0(t) = [u(t) \;\; e(t)]^T. \tag{3.22}$$

The spectrum of $\chi_0(t)$ is given by

$$\Phi_{\chi_0}(\omega) = \begin{bmatrix} \Phi_u(\omega) & \Phi_{ue}(\omega) \\ \Phi_{eu}(\omega) & \lambda_0 \end{bmatrix}. \tag{3.23}$$

The part of the input spectrum that originates from the reference signal $r(t)$ can be written as

$$\Phi_u^r(\omega) \;\; = \;\; \Phi_u(\omega) - |\Phi_{ue}(\omega)|^2 / \lambda_0. \tag{3.24}$$

Using Eqn. (3.6), $\Phi_u^r(\omega)$ can also be expressed as follows,

$$\Phi_u^r(\omega) \;\; = \;\; \left|S_0(e^{i\omega})\right|^2 \Phi_r(\omega), \tag{3.25}$$

were $\Phi_r(\omega)$ is the spectrum of the reference signal. The cross spectrum between $u(t)$ and $e(t)$ is

$$\Phi_{ue}(\omega) \;\; = \;\; -C(e^{i\omega})S_0(e^{i\omega})H_0(e^{i\omega})\lambda_0. \tag{3.26}$$

The spectrum of $v(t)$ is

$$\Phi_v(\omega) = \left|H_0(e^{i\omega})\right|^2 \lambda_0. \tag{3.27}$$

Furthermore, the component of the input spectrum that originates from the noise can be written as [11]

$$\Phi_u^e(\omega) = \left|C(e^{i\omega})\right|^2 |S_0|^2 \Phi_v(\omega). \tag{3.28}$$

3.7.2 Prediction Error Estimation Method

Ljung [11] shows that the one-step ahead predictor for the model structure, given by Eqn. (3.17), can be written as

$$\hat{y}(t|\theta) = H^{-1}(q,\theta)G(q,\theta)u(t) + (1 - H^{-1}(q,\theta))y(t), \tag{3.29}$$

and the prediction error as

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t|\theta) = H^{-1}(q,\theta)(y(t) - G(q,\theta)u(t)). \tag{3.30}$$

Given the model in Eqn. (3.29) and the measured data

$$Z^N = \{y(1), u(1), ..., y(N), u(N)\}, \tag{3.31}$$

the predictor estimate can be determined from [6, 11]:

$$\hat{\theta}_N = \arg \min_{\theta \in D_M} V_N(\theta, Z^N), \tag{3.32}$$

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} l(\varepsilon(t,\theta)). \tag{3.33}$$

Here $l(\cdot)$ is a suitable positive (norm) function. The standard choice is the quadratic norm [6, 11]:

$$l(\varepsilon(t,\theta)) = \frac{1}{2}\varepsilon^2(t,\theta). \tag{3.34}$$

This choice can be combined with some linear, monic and possibly parameterised prefilter $L(q,\theta)$:

$$\begin{aligned} l(\varepsilon(t,\theta)) &= \frac{1}{2}\varepsilon_F^2(t,\theta), \\ \varepsilon_F(t,\theta) &= L(q,\theta)\varepsilon(t,\theta). \end{aligned} \tag{3.35}$$

Without loss of generality the prefilter can be included in the noise model and $L(q, \theta) = 1$ can be assumed, since

$$\epsilon_F(t, \theta) = L(q, \theta)H^{-1}(q, \theta)(y(t) - G(q, \theta)u(t)). \tag{3.36}$$

The above choice of $l(\cdot)$ results in the Least Square Error method [6]. If the criterion function $l(\cdot)$ is equal to $f_e(\cdot)$, the PDF of $e(t)$, the maximum likelihood function is obtained [6]. In the following analysis it is assumed that the quadratic criterion is used.

The estimates $G(q, \hat{\theta}_N)$, and $H(q, \hat{\theta}_N)$ will be written as

$$\hat{G}_N = G(q, \hat{\theta}_N) \text{ and } \hat{H}_N = H(q, \hat{\theta}_N). \tag{3.37}$$

3.7.3 Family of Model Structures

With the PEM approach, the models may have an arbitrary parameterisation [11]. The most simple input-output relationship is obtained by describing it as a linear difference equation:

$$\begin{aligned} & y(t) + a_1 y(t-1) + ... + a_{n_a} y(t - na) \\ = \;& b_1 u(t-1) + ... + b_{n_b} u(t - nb) + e(t). \end{aligned} \tag{3.38}$$

Since the white-noise term $e(t)$ enters as a direct error in the difference equation, the model, given by Eqn. (3.38), is often called an *equation error model* structure [11]. The adjustable parameters are

$$\theta = [a_1 \; a_2 \; ... \; a_{n_a} \; b_1 \; ... \; b_{n_b}]^T. \tag{3.39}$$

The linear difference equation, given by Eqn. (3.38), can be written in the form of Eqn. (3.17), by introducing the following two operators:

$$\begin{aligned} A(q) &= 1 + a_1 q^{-1} + ... + a_{n_a} q^{-n_a}, \text{ and} \\ B(q) &= b_1 q^{-1} + ... + b_{n_b} q^{-n_b}. \end{aligned} \tag{3.40}$$

Eqns. (3.38), (3.17) and (3.40) leads the following expression for $G(q, \theta)$ and $H(q, \theta)$:

$$G(q,\theta) = \frac{B(q)}{A(q)}, \quad H(q,\theta) = \frac{1}{A(q)}. \tag{3.41}$$

This model is called an Auto-Regressive with External Input (ARX) model. In the case when n_a in Eqn.(3.39) is 0, $y(t)$ is a Finite Impulse Response (FIR) model.

The predictor for Eqn. (3.38) can now be given as follows [11]:

$$\hat{y}(t|\theta) = B(q)u(t) + [1 - A(q)]y(t). \tag{3.42}$$

The following vector, called a *linear regressor vector*, is introduced:

$$\varphi(t) = [-y(t-1) \ -y(t-2)... - y(t-n_a) \ \ u(t-1) \ u(t-2)...u(t-n_b)]^T \tag{3.43}$$

With this vector, the predictor can be written as a scalar product between a known vector $\varphi(t)$ and the parameter vector θ. Such a model is called a *linear regression* in statistics [11]:

$$\hat{y}(t|\theta) = \theta^T\varphi(t) = \varphi^T(t)\theta. \tag{3.44}$$

Consider the following generalised model structure [6, 11]:

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t), \tag{3.45}$$

where the $C(q)$-, $D(q)$- and $F(q)$-polynomials are similar to $A(q)$. Sometimes the dynamics from $u(t)$ to $y(t)$ contains a delay of n_k samples, so that some leading coefficients of $B(q)$ are zero; that is, according to Ljung [11]:

$$B(q) = b_{n_k}q^{-n_k} + b_{n_k+1}q^{-n_k-1} + ... + b_{n_k+n_b-1}q^{-n_k-n_b+1}. \tag{3.46}$$

The model structure, given by Eqn. (3.45), contains several common special cases, some of which are listed in Table 3.1.

For the model structures other than ARX and FIR, there is no linear regression. This can be demonstrated for the ARMAX model. Here the regressor vector is

$$\begin{aligned}\varphi(t,\theta) = \ & [-y(t-1)... - y(t-n_a) \ \ u(t-1)...u(t-n_b) \ \ \varepsilon(t-1,\theta) \\ & ...\varepsilon(t-n_c,\theta)]^T\end{aligned} \tag{3.47}$$

Table 3.1: Common model structures.

Polynomials Used in Eqn. (3.45)	**Name of Model Structure**	
$B(q)$	Finite Impulse Response	FIR
$A(q), B(q)$	Auto-Regressive with External input	ARX
$A(q), B(q), C(q)$	Auto-Regressive Moving Average with External input	ARMAX
$B(q), F(q)$	Output Error	OE
$B(q), C(q), D(q), F(q)$	Box-Jenkins	BJ

and the predictor becomes

$$\hat{y}(t|\theta) = \varphi^T(t,\theta)\theta. \tag{3.48}$$

Equation (3.48) is pseudolinear regression, due to the nonlinear effect of θ on the vector $\varphi(t,\theta)$ [11].

3.7.4 Computing the Estimate

Forssell [6] sates that if an FIR or an ARX model is used together with a quadratic criterion function, the standard LSE method is obtained. Then, to find the estimate $\hat{\theta}_N$, one only has to solve a standard LSE problem, which can be done analytically.

For the LSE method the prediction error, given by Eqn. (3.30), becomes

$$\varepsilon(t,\theta) = y(t) - \varphi^T(t)\theta, \tag{3.49}$$

and the criterion function, given by Eqn. (3.33), becomes

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t-1}^{N}\frac{1}{2}\left[y(t) - \varphi^T(t)\theta\right]^2. \tag{3.50}$$

Eqn. (3.50) can be minimised analytically, which gives the LSE, provided the inverse of the term between the square brackets exists [11]:

$$\hat{\theta}_N^{LS} = \arg\min_{\theta\in D_M} V_N(\theta, Z^N) = \left[\frac{1}{N}\sum_{t-1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\frac{1}{N}\sum_{t-1}^{N}\varphi(t)y(t). \tag{3.51}$$

For other parameterisations and criteria, one has to rely on some iterative scheme to find $\hat{\theta}$. The standard choice is to use a search routine of the form

$$\hat{\theta}_N^{(i+1)} = \hat{\theta}_N^{(i)} - \mu_N^{(i)} \left[R_N^{(i)}\right]^{-1} V_N'(\hat{\theta}_N^{(i)}, Z^N), \tag{3.52}$$

where $V_N'(\hat{\theta}_N^{(i)}, Z^N)$ denotes the gradient with respect to θ of the criterion function, given by Eqn. (3.33). $R_N^{(i)}$ is a matrix that modifies the search direction, and $\mu_N^{(i)}$ is a scaling factor that determines the step length.

3.7.5 Consistency and Identifiability

The following consistency and identifiability results for the direct closed-loop SID approach can be found in Forssell's dissertation [6]. Consider the PEM method applied with quadratic criterion

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \frac{1}{2}\varepsilon^2(t,\theta). \tag{3.53}$$

Under mild conditions it is true that

$$V_N(\theta, Z^N) \to \bar{V}(\theta) = \bar{E}\frac{1}{2}\varepsilon^2(t,\theta) \quad \text{with probability (w.p). 1 as } N \to \infty \quad \text{and} \tag{3.54}$$

$$\hat{\theta}_N \to D_c = \arg\min_{\theta \in D_M} \bar{V}(\theta) \quad \text{w.p. 1 as } N \to \infty. \tag{3.55}$$

Using Parseval's relationship [33], one can then write

$$\bar{V}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{2}\Phi_\varepsilon(\omega) d\omega, \tag{3.56}$$

where $\Phi_\varepsilon(\omega)$ is the spectrum of the prediction error. If the data were generated by

$$y(t) = G_0(q)u(t) + H_0(q)e(t), \tag{3.57}$$

it can be shown that

$$\begin{aligned} \varepsilon(t,\theta) &= H^{-1}(q,\theta)(y(t) - G(q,\theta)u(t)) \\ &= H^{-1}(q,\theta)[G_0(q) - G(q,\theta)u(t) + (H_0(q) - H(q,\theta)e(t)) + e(t)] \end{aligned} \tag{3.58}$$

$$= \quad H^{-1}(q,\theta)\tilde{T}^T(q,\theta)\chi_0(t) + e(t),$$

with $\tilde{T}(q,\theta) = T_0(q) - T(q,\theta)$.

If $G(q,\theta)u(t)$ and $G_0(q)u(t)$ depend only on $e(s)$ for $s < t$, and $H(q,\theta)$ and $H_0(q)$ are monic, then the last term of Eqn. (3.58) is independent of the other terms and

$$\Phi_\varepsilon(\omega) \quad = \quad \frac{1}{\mid H(e^{i\omega},\theta)\mid^2}\tilde{T}^T(e^{i\omega},\theta)\Phi_{\chi_0}(\omega)\tilde{T}(e^{-i\omega},\theta) + \lambda_0. \tag{3.59}$$

Together with Eqns. (3.55) and (3.56)

$$\hat{\theta}_N \to D_c = \arg\min_{\theta\in D_M} \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{T}^T(e^{i\omega},\theta)\Phi_{\chi_0}(\omega)\tilde{T}(e^{-i\omega},\theta)\frac{1}{\mid H(e^{i\omega},\theta)\mid^2}d\omega$$

$$\text{w.p. } 1 \text{ as} N \to \infty. \tag{3.60}$$

Forssell [6] states that the result of Eqn. (3.60) shows that the PEM estimate $\hat{T}(q,\hat{\theta}_N)$ will converge to the true transfer function $T_0(q)$, if the parameterisation is flexible enough so that $S \in M$, and if the structure $\tilde{T}^T(e^{i\omega},\theta)$ does not lie in the left nul space of $\Phi_{\chi_0}(\omega)$. Then it can be said that the data are *informative enough* with respect to the chosen model structure [11]. An *informative* data set is one for which the matrix $\Phi_{\chi_0}(\omega)$ is positive definite for almost all frequencies. In the informative case, the limit estimate will always be such that $\tilde{T}^T(e^{i\omega},\theta) = 0$, i.e. $\hat{T}(e^{i\omega})$ will tend to $T_0(q)$ as the number of data points tends to infinity.

If the following factorization is considered:

$$\Phi_{\chi_0}(\omega) \quad = \quad \begin{bmatrix} 1 & \Phi_{ue}(\omega)/\lambda_0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \Phi_u^r(\omega) & 0 \\ 0 & \lambda_0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \Phi_{eu}(\omega)/\lambda_0 & 1 \end{bmatrix}, \tag{3.61}$$

it follows that $\Phi_{\chi_0}(\omega)$ is positive definite for almost all frequencies if and only if $\Phi_u^r(\omega) > 0$ for almost all frequencies. This condition is ensured when there is a nonlinear relationship between the input and the output (see Section 3.5). With a linear feedback law $\Phi_u^r(\omega) = |S_0(e^{i\omega})|^2\,\Phi_r(\omega)$ and because the analytic function $S_0(e^{i\omega})$ can have only a finite number of zeros, the condition becomes $\Phi_r(\omega) > 0$ for almost all frequencies. That is the reference signal $r(t)$ should be persistently exciting [11].

The following will ensure $\Phi_u^r(\omega) > 0$ for almost all ω, i.e. identifiability [31]:

- $r_a(t)$ or $r_b(t)$ should be persistently exciting, i.e. $\Phi_a(\omega) > 0$ or $\Phi_b(\omega) > 0$ for almost all ω, or

- $C(q)$ should be a controller of sufficiently high order, or
- $C(q)$ should be a controller that switches between several settings during the experiment.

3.7.6 Bias Distribution

Direct Approach: Again, for the direct approach, Forssell [6] shows that by inserting the factorization

$$\Phi_{\chi_0}(\omega) = \begin{bmatrix} 1 & 0 \\ \Phi_{ue}(\omega)/\Phi_u(\omega) & 1 \end{bmatrix} \begin{bmatrix} \Phi_u(\omega) & 0 \\ 0 & \Phi_e^r(\omega) \end{bmatrix} \begin{bmatrix} 1 & \Phi_{eu}(\omega)/\Phi_u(\omega) \\ 0 & 1 \end{bmatrix} \quad (3.62)$$

into Eqn. (3.60), gives the following estimate $\hat{\theta}_N$ as the number of data tends to infinity:

$$\hat{\theta}_N \to D_c = \arg\min_{\theta \in D_M} \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ \left| G_0(e^{i\omega}) + B(e^{i\omega}, \theta) - G(e^{i\omega}, \theta) \right|^2 \Phi_u(\omega) \quad (3.63)$$

$$+ \left| H_0(e^{i\omega}) - H(e^{i\omega}, \theta) \right|^2 \Phi_e^r(\omega) \} \frac{1}{|H(e^{i\omega}, \theta)|^2} d\omega \text{ w.p. 1 as } N \to \infty,$$

where

$$\begin{aligned} B(e^{i\omega}, \theta) &= (H_0(e^{i\omega}, \theta) - H(e^{i\omega}, \theta)) \frac{\Phi_{eu}(\omega)}{\Phi_u(\omega)} \\ &= \frac{\lambda_0}{\Phi_u(\omega)} \cdot \frac{\Phi_u^e(\omega)}{\Phi_u(\omega)} \cdot \left| H_0(e^{i\omega}, \theta) - H(e^{i\omega}, \theta) \right|^2 . \end{aligned} \quad (3.64)$$

A variant of this result is given in [12, 31]. Ljung [11] states that the bias $B(e^{i\omega}, \theta)$ will be small in the frequency ranges where either or all of the following holds:

- the noise model is good, i.e. $H_0(e^{i\omega}, \theta) - H(e^{i\omega}, \theta)$ is small,
- the feedback contribution to the input spectrum is small, i.e. $\Phi_u^e(\omega)/\Phi_u(\omega)$ is small, and
- the Signal-to-Noise Ratio (SNR) is good, i.e. $\lambda_0/\Phi_u(\omega)$ is small.

If the system was operating in open-loop, so that $\Phi_{eu}(\omega) = 0$, then the bias term $B(e^{i\omega}, \theta)$ equals zero regardless of the noise model $H(e^{i\omega}, \theta)$ [11].

Indirect Approach: For the indirect approach the closed-loop system can be parameterised in terms of the open-loop system as follows

$$y(t) = \frac{G(q,\theta)}{1+G(q,\theta)C(q)}r(t) + H_*(q)e(t), \tag{3.65}$$

where $H_*(q)$ is a fixed noise model. The indirect method is equivalent to the direct method with the noise model $H(q,\theta) = (1+G(q,\theta)C(q))H_*(q)$, as covered in [12]. If the model structure is as given by Eqn. (3.65) then Forssell [6] shows that

$$\hat{\theta}_N \rightarrow D_c = \arg\min_{\theta \in D_M} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\frac{G_0(e^{i\omega}) - G(e^{i\omega},\theta)}{1+G(e^{i\omega},\theta)C(e^{i\omega})}\right|^2 \frac{|S_0(e^{i\omega})|^2\,\Phi_r(\omega)}{|H_*(e^{i\omega})|^2}d\omega \tag{3.66}$$

$$\text{w.p. 1 as } N \rightarrow \infty$$

This means that the indirect method can give unbiased estimates of $G_0(e^{i\omega})$ if the parameterisation of $G(e^{i\omega},\theta)$ is flexible enough, that is, $G_0 \in G$, even with a fixed noise model. This is the main advantage of the indirect method.

Joint Input-Output Approach: It is shown in [12], that, for the joint input-output approach, if the model structure is:

$$y(t) = G(q,\theta)\hat{u}(t) + H_*(q)e(t), \quad \hat{u}(t) = \hat{S}_N(q)r(t), \quad \tilde{u}(t) = u(t) - \hat{u}(t) \tag{3.67}$$

then

$$\hat{\theta}_N \rightarrow D_c = \arg\min_{\theta \in D_M} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|G_0(e^{i\omega}) + \tilde{B}(e^{i\omega}) - G(e^{i\omega},\theta)\right|^2 \frac{\Phi_{\hat{u}}(\omega)}{|H_*(e^{i\omega})|^2}d\omega \tag{3.68}$$

$$\text{w.p. 1 as } N \rightarrow \infty,$$

where

$$\tilde{B}(e^{i\omega}) = G_0(e^{i\omega})\Phi_{\tilde{u}\hat{u}}(\omega)\Phi_{\hat{u}}^{-1}(\omega). \tag{3.69}$$

Thus, an unbiased estimation of $G_0(e^{i\omega})$ can be obtained, regardless of the noise model used, as long as the correlation between $\hat{u}$ and $\tilde{u}$ is negligible.

3.7.7 Asymptotic Variance Distribution

Direct Approach: For the direct approach, Ljung [11] derives the following equations for asymptotic variance of the estimated transfer function:

$$Cov\, \hat{G}_N \sim \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u^r(\omega)}, \tag{3.70}$$

where n is the model order and N is the number of data samples. The denominator of Eqn. (3.70) contains the spectrum of that part of the input that originates from the reference signal $r(t)$. The open-loop expression contains the *total* input spectrum in the denominator [11]. It is the signal-to-noise ratio, where *signal* is the part of the plant input that derives from the injected reference, that determines the quality of the open-loop transfer function estimate. From this perspective, the part of the input that originates from the feedback has no information value when estimating $G(q)$ [11].

The purpose of feedback is to make the sensitivity function S_0 small, especially at frequencies with disturbances. Feedback will thus decrease the information content of measured data at these frequencies, but it will also allow one to inject more input power at certain frequency ranges, without increasing the output power [11].

Ljung [11] stresses that the result of Eqn. (3.70) is asymptotic when the orders of both $G(q)$ and $H(q)$, as well as N, tends to infinity.

Indirect and Joint Input-Output Approaches: Suppose that the input is generated as in Eqn. (3.5), i.e. linear feedback is present. Forssell and Ljung [12] show that when the transfer function estimate $\hat{G}_N$ is found using either the indirect or the joint input-output approach, Eqn. (3.70) also holds for these approaches.

Forssell [6] shows that as far as the *asymptotic variance of the parameter estimates* is concerned, i.e. in the finite model order case, the indirect and joint input-output approaches are typically less precise than the optimal direct method that meets the Cramèr-Rao bound. In the direct approach the noise in the loop is used to reduce the variance, while in the other closed-loop approaches, the part of the input signal that originates from the noise is not used in the identification process. However, this effect of the noise vanishes as more and more parameters are used to estimate the noise model $H(q)$ [12].

Please note that if the *variance of* a model is small the model is *precise* and if the bias of a model is small the model is *accurate*.

3.8 EXCITATION SIGNALS

An experiment is *informative enough* if it generates a data set that is informative enough. In open-loop this is achieved when the input $u(t)$ is PE of a certain order, i.e. so that it contains sufficiently many distinct frequencies. As mentioned in Section 3.5, in closed-loop the experiment may still be non-informative for the above-mentioned situation, and the easiest way to ensure an *informative* experiment is to ensure that $r_a(t)$ or $r_b(t)$ is PE [31].

The excitation signals determine the operating point of the system, as well as which parts and modes of the system are excited during the experiment. The user's freedom in choosing these signals' characteristics may vary considerably according to the application. In the process industry it is not always possible to manipulate a system during normal production [11].

Two different aspects are associated with the choice of excitation signals. One concern is the second-order properties of the signal, such as its spectrum. The other concern is the *shape* of the signal. Excitation signals can be sums of sinusoids, filtered white noise, pseudorandom signals, or binary signals (assuming only two values), and so on [34]. Typical signals are discussed below.

Step or Pulses: The step and pulse signals are often used for transient analysis. Typical information acquired, are the largest and smallest frequencies and the static gain of the process. These signals can be used to determine which type of excitation signals to consider in the final experiment. However, when a plant has a low frequency bandwidth, these signals will suffice, since the frequency components that make up the step excite the plant sufficiently to obtain an accurate model with low uncertainty around the crossover frequency region [34].

White Noise: General use is made of signals that are broad-banded, i.e., they have a spectrum that covers a wide frequency range. This generally guarantees that the signals are sufficiently exciting. One such signal is discrete white noise. White noise has a flat frequency spectrum. Such a signal can be generated from white Gaussian noise, filtered through a linear filter. Virtually any signal spectrum can be achieved in this way [34].

Random Binary Signals: A Random Binary Signal (RBS) is a random process that assumes only two values. It has the same spectral properties as white noise and can be generated in a number of different ways. Since the signal is binary, it has maximum signal power under amplitude bound constraints [35].

Pseudo-Random Binary Signals: A Pseudo-Random Binary Signal (PRBS) is a periodic, deterministic signal with white-noise-like properties. Typically the clock frequency of the PRBS should be about 2.5 times faster than the bandwidth to be covered by the signal [35].

Multisinusoidal Signal: Instead of distributing the input power over a relatively wide range of frequencies, one can also choose to concentrate on a relatively small number of frequency points. This can be done by using a periodic excitation signal, e.g. a multisinusoidal signal. This kind of periodic excitation is beneficial in, e.g. nonparametric model identification, and in situations where accurate knowledge of the plant is required at a limited number of frequencies [34].

Periodic Excitation: Forssell [6] shows that there are several advantages in using periodic excitation signals, such as PRBS or multisinusoidal signals. The most obvious advantages are the data reduction resulting from the averaging of the measured signals over the periods. Another advantage is that with periodic excitation it is possible to separate the *signal* from the output *noise* and therefore the noise properties can be estimated separately.

Theoretically, in the absence of noise, any arrangement of n non-zero points in the frequency domain of the excitation signal will be sufficient to uniquely determine n parameters. However, with the slightest bit of noise, the whole fit can become disastrous. In the more practical case were there is noise and only a finite amount of data points, a *good* arrangement of n non-zero points in the frequency domain is required for a proper fit. It will also help to include more than n frequency points in the spectrum. Even then, the arrangement of these points is still relevant, e.g. 20 points around the resonance frequency of a second order system are much better than 100 points in the roll-off region. Also, because of correlation, the effective number of points might be less [36].

In both the bias and variance expressions, the input spectrum appears as a weighting term [11]. The bias and variance will be less where the input spectrum is large. Therefore, it is best to put the energy of the input in that region where one wants to model the plant accurately. It is also advisable to choose a signal with a spectrum that is non-zero at frequencies where the system has a good SNR [36].

In off-line situations it is always a good idea to check the spectrum of a signal before using it as an excitation signal, to see if it is acceptable.

3.9 INTER-SAMPLING APPROACH

3.9.1 Introduction

In the restrictive situation were the reference signal is not PE, it usually would mean that the controller should be of a sufficient high order or time-varying in order to ensure identifiability of the plant. However, Sun *et al.* [18] show in their paper that the information obtained by inter-sampling the plant output during the control interval can extend the identification capability.

Output inter-sampling, which is also referred to as output over-sampling, is a very useful technique in blind identification problems [37]. It is also effective in closed-loop SID. The inter-sampling approach is actually the direct closed-loop SID approach, which uses the output inter-sampling technique to address the problem of identifiability [18]. Inter-sampling the output in this context possibly has the same effect as a nonlinear controller [36].

3.9.2 Model Representation

In their proof Sun *et al.* [18] consider a strictly proper linear plant with time delay τ, which is managed by a digital controller whose holding period is T. The control input to the plant is then a piecewise signal with a holding period T. The closed-loop configuration is as shown in Fig. 3.2, where $y(t)$ is the plant output sampled at a sampling interval T. The discrete model of the plant $G(q)$, identified from data sampled at an interval of T, can be expressed as

$$G(q) = \frac{B(q)}{A(q)} = \frac{b_{\tau_T} q^{-\tau_T} + \ldots + b_{\tau_T+n_b-1} q^{-\tau_T-n_b+1}}{1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}}, \tag{3.71}$$

and the controller $C(q)$ can be expressed as

$$C(q) = \frac{L(q)}{P(q)} = \frac{l_{\tau_l} q^{-\tau_T} + \ldots + l_{\tau_l+n_l-1} q^{-\tau_l-n_l+1}}{1 + p_1 q^{-1} + \ldots + p_{n_p} q^{-n_p}}, \tag{3.72}$$

where τ_T is the discrete-time delay. The model in Eqn. (3.71) is referred to as the T-model. Please note that the notation used in the proof of Sun *et al.* [18] has been changed to match the notation, used in this work.

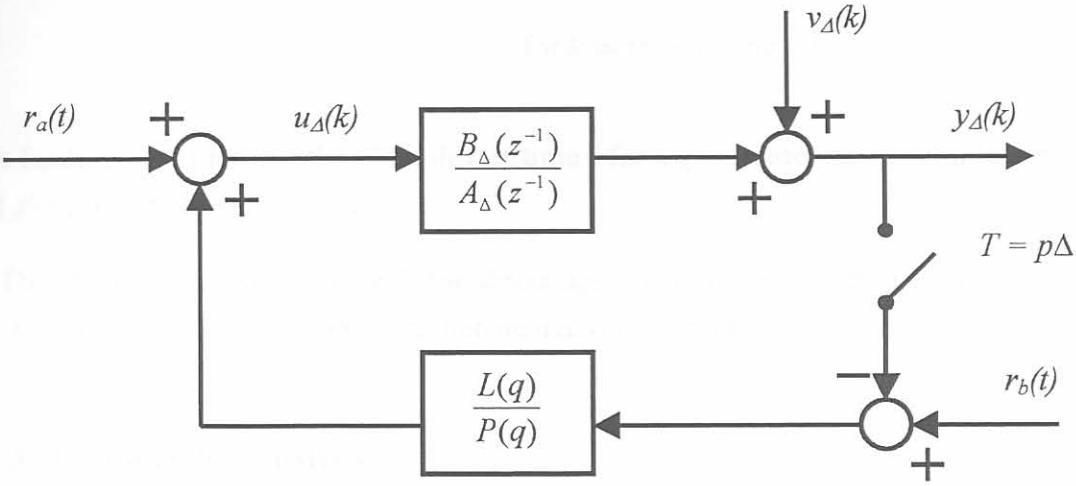

Figure 3.3: Closed-loop identification by output inter-sampling.

Let $y_\Delta(k)$ in Fig. 3.3 be the plant output sampled at a short sampling interval $\Delta = T/p$, where p is an integer, while the control interval is still T. Therefore, the output is sampled p times faster than the rate at which control action is taken. The controller only uses the plant output of every $T = p\Delta$ for control, i.e. $u_\Delta(k) = u_\Delta(k + 1\Delta) = ... = u_\Delta(k + (p-1)\Delta)$. The plant Δ-model with sampling interval Δ is given by

$$\frac{B_\Delta(z^{-1})}{A_\Delta(z^{-1})} = \frac{b_{\Delta,\tau_\Delta} z^{-\tau_\Delta} + ... + b_{\Delta,\tau_\Delta + n_{b\Delta} - 1} z^{-\tau_\Delta - n_{b\Delta} + 1}}{1 + a_{\Delta,1} z^{-1} + ... + a_{\Delta,n_{a\Delta}} z^{-n_{a\Delta}}}, \tag{3.73}$$

where z^{-1} is a shift operator, like q, and τ_Δ is the corresponding delay time.

The polynomials $A(q)$ and $B(q)$ of the T-model can be uniquely determined from the Δ-model as follows [18]:

$$\begin{aligned} A(q) &= \det(I - A^p q^{-1}) \\ B(q) &= q^{-\alpha(\tau_\Delta - 1)} c^T \mathrm{adj}(I - A^p q^{-1}) \left(\sum_{i=0}^{\beta(-\tau_\Delta)} A^i b q^{-1} + \sum_{i=\beta(-\tau_\Delta)+1}^{p-1} A^i b q^{-2} \right), \end{aligned} \tag{3.74}$$

where A, b, and c are the state-space realization of the Δ-model and *det* and *adj* are the matrix determinant and adjoint matrix respectively. The adj$U$ is defined as the matrix formed from the transpose of matrix U after all elements have been replaced by their cofactors [39].

The functions $\alpha(\cdot)$ and $\beta(\cdot)$ are defined as

$$\alpha(k) \;=\; t, \;\; \beta(k) = j, \quad \text{for } k = tp + j \;\; \text{and } 0 \leq j < p. \tag{3.75}$$

The function $\alpha(k)$ returns the old index of time t for a given inter-sampled index of time k and $\beta(k)$ returns the remainder.

The Δ-model is estimated with the direct approach, as described in Section 3.7, from $u_\Delta(k)$ and $y_\Delta(k)$. The T-model can then be determined making use of Eqn. (3.74).

3.9.3 Identifiability Analysis

The identifiability can now be considered under the restrictive situation where $r_a(t) = 0$ and $r_b(t) = 0$. $u(t)$ can then be expressed by

$$u(t) = \frac{-L(q)A(q)}{P(q)A(q) + L(q)B(q)} v(t). \tag{3.76}$$

Let the transfer function of the plant Δ-model be given by

$$y_\Delta(k) \;=\; \frac{B_\Delta(z^{-1})}{A_\Delta(z^{-1})} u_\Delta(k) + v_\Delta(k). \tag{3.77}$$

The prediction error, given by Eqn. (3.30), then becomes

$$\varepsilon_\Delta(k, \hat{\theta}_\Delta) \;=\; \frac{\hat{\Gamma}_\Delta(z^{-1})}{\hat{\Pi}_\Delta(z^{-1})} (\hat{A}_\Delta(z^{-1}) y_\Delta(k) - \hat{B}_\Delta(z^{-1}) u_\Delta(k)). \tag{3.78}$$

Here $\Gamma_\Delta(z^{-1})$ and $\Pi_\Delta(z^{-1})$ are coprime, i.e. have no common factors [11], minimum phase polynomials. The gradient vector of $\varepsilon_\Delta(k, \hat{\theta}_\Delta)$, the regressor, is then

$$\varphi_\Delta(k, \hat{\theta}_\Delta) = -\frac{d\varepsilon_\Delta(k, \hat{\theta}_\Delta)}{d\hat{\theta}_\Delta}. \tag{3.79}$$

The following conditions are assumed:

- the true system model is in the model set defined by Eqn. (3.77) and Eqn. (3.78),
- the closed-loop is stable and $\tau_T \geq 1$,
- $\varepsilon_\Delta(k)$ is a stationary signal with zero mean and finite fourth-order moments [38], and is independent of $r_1(t)$ and $r_2(t)$,
- $A_\Delta(z^{-1})$, $\Gamma_\Delta(z^{-1})$ and $\Pi_\Delta(z^{-1})$ are minimum phase and any pair of them is coprime,

- $\Pi_\Delta(z^{-1})/\Gamma_\Delta(z^{-1})$ and $B_\Delta(z^{-1})/A_\Delta(z^{-1})$ are smooth functions with respect to the true parameters.

By making use of the fact that $k = tp + j$ and that $v_\Delta(k)$ is a stationary signal Sun *et al.* [18] prove that $\varphi_\Delta(k, \hat{\theta}_\Delta)$ has cyclostationarity in k. With the aid of cyclostationarity, the correlation matrix of the regressor $\varphi_\Delta(k, \hat{\theta}_\Delta)$ has full-rank, even in closed-loop.

By making use of the convergence theorem of the PEM algorithm [11] for $p \geq 2$, and the fact that the correlation matrix of the regressor has full-rank, Sun *et al.* [18] prove that $\hat{\theta}_\Delta$ converges to a minimum such that

$$V_N(\hat{\theta}_\Delta) = \frac{1}{p}\sum_{j=0}^{p-1} E\varepsilon_\Delta^2(tp + j, \hat{\theta}_\Delta) = E\varepsilon_\Delta^2(k) \quad \text{w.p. 1 as } N \to \infty. \tag{3.80}$$

In other words the closed-loop Δ-model is identifiability. Moreover, following Eqn. (3.74), the T-model is also identifiable [18]. The proof does not depend on the property of $r_a(t)$, $r_b(t)$, or the controller structure.

Sun *et al.* [18] also mention that by increasing p, more information can be acquired, but may contribute little when Δ is very small. Therefore, the estimation error of the Δ-model will not always decrease much by increasing p. On the other hand, the convergence error from the Δ-model to the T-model increases with p. Both the estimate variance of the Δ-model and the model conversion error should, therefore, be considered.

According to Sun *et al.* [18] the main drawback of this method is still, as with the normal direct approach, that the estimation accuracy depends on the noise model estimation.

3.10 Conclusion

In this chapter it is shown that the basic problem with closed-loop data is that it is typically less informative about the open-loop system, and that many estimation methods may fail when applied in a direct way to closed-loop data, because of the correlation between the additive output noise and the input to the plant. Researchers have studied different ways of avoiding the problems associated with closed-loop identification, which lead to a characterization of the possible closed-loop SID methods into three categories: the direct, the indirect and the joint input-output approach.

From analysing the literature it is concluded that the PEM estimation method, applied in the direct fashion with a noise model that can describe the true noise properties, gives consistent estimates and optimal accuracy. Therefore, the PEM method is currently regarded as the prime estimation method and this estimation method will, thus, also be used in the proposed methodology.

The *identifiability* analysis, found in literature, shows that the open-loop condition for informative experiments, namely that the input should be PE of sufficiently high order, does not ensure identifiability in closed-loop. In closed-loop only a part of the input originates from the reference signal. The other part of the input, which originates from the output, is correlated with noise and PE of this part will not ensure identifiability. So, it is concluded that, in closed-loop, the spectrum of that part of the input that originates from the reference signal should be non-zero. This means that there should be either a nonlinear relationship between the input and the output or the reference signal should be PE.

From the identifiability analysis of the new *inter-sampling method*, where the plant output is sampled at a higher rate than the control input, it can be concluded that this approach can also ensure identifiability. Thus, when the traditional identifiability conditions are not satisfied, the direct closed-loop identification scheme can also be considered. The precision of this method is discussed in the next chapter.

By reviewing the available bias and variance analysis of the different approached the following can be concluded:

Bias: In the direct approach, if the noise model cannot describe the true noise properties, a bias will result, but will be small in frequency ranges where either or all of the following holds:

- the noise model is good, i.e. $H_0(e^{i\omega}, \theta) - H(e^{i\omega}, \theta)$ is small,
- the feedback contribution to the input spectrum is small, i.e. $\Phi_u^e(\omega)/\Phi_u(\omega)$ is small, and
- the SNR is good, i.e. $\lambda_0/\Phi_u(\omega)$ is small.

However, the indirect approach can give unbiased estimates of $G_0(e^{i\omega})$ if the parameterisation of $G(e^{i\omega}, \theta)$ is flexible enough, even with a fixed noise model. Also the bias in the joint input-output approach does not depend on the noise model.

Variance of Estimated Transfer Function: The asymptotic variance of the estimated transfer function is equal for the direct, indirect and joint input-output approaches, if the feedback is linear. This variance is, however, worse than the variance obtained in open-loop SID, since the denominator of the variance expression contains the spectrum of only that part of the input that originates from the reference signal, while the open-loop expression has the total input spectrum in the denominator.

Variance of Parameter Estimates: When the asymptotic variance of the parameter estimates, i.e. the finite model order case, is considered, the direct approach is optimal, since it meets the Cramèr-Rao bound. The other closed-loop approaches are less precise, since they do not meet this bound.

In the next chapter, the reviewed closed-loop SID theory, as well as other criteria, is considered when a closed-loop SID methodology for MPC controlled plants is chosen.

CHAPTER 4

SELECTION OF A METHODOLOGY APPLICABLE TO MPC CONTROLLED PLANTS

4.1 INTRODUCTION

In this chapter the most appropriate SID methodology is chosen for the identification of plants, controlled by MPC controllers, from measured closed-loop data.

The merits and weakness of the different methods and options regarding closed-loop system identification are considered. A methodology is selected by taking into account the characteristics of MPC controllers and industrial plants. The selection of an identification approach is discussed in Section 4.2 and the way in which identifiability can be guaranteed is discussed in Section 4.3.

One of the options for the guarantee of identifiability is the novel inter-sampling approach. An evaluation, through a simulation study, of the model variance is performed. The aim is to determine the value of this new approach. Therefore the results and conclusions from this study are presented in Section 4.4.

In Section 4.5 the selection of a type of model structure is discussed and the selected validation method is discussed in Section 4.6.

For the sake of completeness this discussion is ended with Section 4.7 in which a motivation is given for the use of the SID toolbox instead of custom-written algorithms or extra software, in the implementation of the selected identification methodology. The selected closed-loop SID methodology is then summarised in a step by step description of the method in Section 4.8.

4.2 CLOSED-LOOP IDENTIFICATION APPROACH

Common assumptions in closed-loop identification are that the existing controller is linear and the process is single variable. Practical MPC controllers are usually nonlinear because, invariably, plant inputs and outputs are constrained. Plants are often also multivariable, making many closed-loop identification results not suitable for MPC applications [3]. This is also the case for the MIBK plant, since there are constraints on the plant inputs and outputs, making the controller nonlinear, and the plant is, of course, also multivariable.

As stated in Section 3.6, the indirect and the joint input-output approaches are typically only used when the feedback law is linear. Lakshminarayanan *et al.* [40] emphasize this fact by stating that irrespective of the identification algorithm, consistent estimation from the joint input-output method is obtainable only when the controller is linear and time-invariant. Although a modification of these methods can sometimes be applied when the regulator is nonlinear, this will cause considerably more work. The MPC controller is also very complex compared to other controllers, since the MPC algorithm computes the control action at each sampling interval using on-line optimisation [28]. This fact makes the indirect and the joint input-output approaches even less attractive for MPC.

There is one variant of the joint input-output approach that does work with nonlinear controllers. This is the projection method, devised by Forssell and Ljung [14]. However, this method depends on the derivation of noncasual FIR models, which are nonparametric. According to Zhu *et al.* [41], parametric models are better for use in industrial process identification, since these models are more accurate and precise, require shorter test time, and can be more user friendly than nonparametric models. Therefore, for industrial processes, which are the processes usually controlled by MPC controllers, nonparametric models are not ideal, which, in turn, makes the projection method not ideal.

According to Ljung [11], the direct approach should be seen as the natural approach to closed-loop data analysis. The main reasons are:

- this method works regardless of the complexity of the controller,
- no special algorithms and software are required,
- consistency and optimal accuracy are obtained if the model structure contains the true system, including the noise properties, and
- unstable systems can be handled without problems, as long the closed-loop system is stable and the predictor is stable.

For a stable predictor any unstable poles of the plant model $G(q)$ must be shared by the noise model $H(q)$. Parametric models like ARX and ARMAX satisfy this constraint [11].

The only drawback with the direct approach is that a good noise model is needed in order to prevent a bias in the estimate $\hat{G}_N$ [6, 17]. Thus, as stated by Lakshminarayanan *et al.* [40], if the process model and disturbance model structures are completely known, then direct identification is the obvious choice as it can provide model parameters with the least variance and with considerably less effort.

In the case of closed-loop SID of plants controlled by MPC controllers, the identification will usually be re-identification, since the first identification takes place before the MPC controller is designed and implemented. This, in turn, implies that the model and disturbance model structures are known when closed-loop SID is performed. Therefore, the direct closed-loop identification method is the best choice when MPC controllers are involved.

When the available data sets for the MIBK plant are considered then the direct method is actually the only choice. The reason being that there are no measured reference signals available and not enough information to model the controller, as required to employ the indirect, joint input-output and projection method.

4.3 Guarantee of Identifiability

The purpose of feedback is to make the sensitivity function $So(q)$ small, especially at frequencies where the disturbance signals $v(t)$ have energy. As stated in Section 3.7.7, feedback will worsen the measured data's information about the system at these frequencies [11].

However, for the direct approach an experiment will still be informative in each of the following situations [31]:

- the reference signal is persistently exciting,
- the controller is of sufficiently high order, or
- the controller switches between several settings during the experiment.

Although a multivariable MPC changes its structure as it deals with changes in the active constraints, this cannot guarantee a given number of changes in the controller settings, since these changes are unpredictable under normal operation. Making deliberate changes to a multivariable MPC may have unforeseen consequences for process operation, and it does not provide the same control over the SNR as adding external test signals [10]. Therefore, the best way to satisfy the identifiability condition is to ensure that the reference signal is persistently exciting. For plants controlled by MPC controllers, this has been done in many different ways:

Doma *et al.* [10] satisfy this condition by adding as many different statistically independent external test signals as there are independent variables to the reference signals. Since simultaneous application of the test signals is, according to them, undesirable for complex multivariable processes, because of unforeseen interaction with unknown multivariable disturbances, they add the external tests signals one variable at a time.

However, data from a single variable test may not contain sufficient information about the multivariable character of the process, i.e. ratio between the different variables. Multivariable testing can solve this problem [3]. Therefore, this type of testing is preferred on condition that MIMO model estimation is also possible. This will be further discussed in Section 4.5.4.

The ASYM method, developed by Zhu [4], ensures informative data by making use of optimal test signals designed to minimise the sum of the squares of the simulated error. The desired spectra of these test signals are realised by PRBS signals or filtered white noise.

An approach to simultaneous constrained MPC and Identification (MPCI) was developed by Shouche *et al.* [2]. In this approach, a persistent identification criterion is used as an additional constraint in the standard on-line optimisation of MPC.

Case studies were done by applying (1) MPCI, (2) MPC with external dithering signals (PRBS signals were added to control signals) and (3) MPC with no external input (constant references). The first two cases gave good results due to the process excitation by the controller in the MPCI case, and the external dithering signals in the MPC case. However, the case study of the MPC with no excitation, gave very poor parameter estimates due to the lack of information about the dynamics of the process [2].

Therefore, although the MPC controller is nonlinear, time-varying and complex, which in general yield informative experiments [11], a persistently exciting signal $r(t)$ will guarantee this. The type of excitation signal needed, depends on the characteristics of the plant. For choice of excitation signals, refer to Section 3.8.

In the case of the MIBK plant, data sets from the normal operation of the system were used for the identification of the plant. Since no knowledge of the reference signals was available, it is possible that these signals were not PE and that the system is, therefore, not identifiable from this data.

However, another option is to inter-sample the plant inputs and outputs, since this is also claimed to ensure identifiability - even when there is no persistent excitation signal [18]. The question of identifiability has only a *yes* or *no* answer. Another important question is how precise the model identified from the inter-sampled data is. The model may be identifiable, but still have such a large variance that it is unacceptable [36], as illustrated in the variance simulation study in the next section.

4.4 EVALUATION OF THE INTER-SAMPLING APPROACH

A simulation study was done to determine how precise a model identified from closed-loop inter-sampled data is. In this study the variance of the identified model was evaluated. An estimation of the variance was obtained from a Monte Carlo simulation.

4.4.1 Set-Up of the Variance Simulation

In the Monte Carlo simulation the closed-loop system, described in Addendum A, was simulated to determine the closed-loop response signals, with the seed of the added noise different in each run. A total of 50 runs were performed. For each run in the simulation the resulting input and output signals of the plant were logged. From each of these fifty sets of input-output signals a plant Δ-model, described in Section 3.9.2, was identified with the direct closed-loop SID approach and the PEM estimation method. From these Δ-models the T-models, also described in Section 3.9.2, were determined, making use of the *d2d* MATLAB function that implements Eqn. (3.74).

From these models the estimation error was determined with Eqn. (4.1)

$$ERR = \frac{1}{50}\sum_{l=1}^{50}\left(\left|\hat{a}_i^{(l)} - \bar{a}_i\right|^2 + \left|\hat{b}_i^{(l)} - \bar{b}_i\right|^2\right), \tag{4.1}$$

where $\hat{a}_i^{(l)}$ and $\hat{b}_i^{(l)}$ are the parameter estimates for run l of the simulation and $\bar{a}_i$ and $\bar{b}_i$ are the mean parameter estimates.

The variance of the identified models was also evaluated, by determining the variance in the step, impulse, Bode magnitude and Bode phase responses of the fifty identified models. The variance of the step responses (*step_var*) was determined making use of Eqn. (4.2) and the *var* function in MATLAB, which determines the variance, i.e. square of the standard deviation for a set of values.

$$step\_var = \frac{1}{L}\sum_{t=1}^{L} var(step\_resp(t)), \text{ where} \tag{4.2}$$

$$step\_resp(t) = \begin{bmatrix} step^{(1)}(t) \\ step^{(2)}(t) \\ \vdots \\ step^{(50)}(t) \end{bmatrix}, \text{ and where}$$

$step^{(l)}(t)$ is the step response value at time index t of the model, identified in run l of the simulation, and L is the time span over which the step responses was determined. Eqn. (4.2) determines the variance of the step response at each time instant t and then determines the average variance from these values. The variance of the impulse responses was determined similarly. Also, the variances of the Bode magnitude and phase responses were determined similarly, with t substituted with the frequency index and L substituted with the frequency span of these responses.

A number of cases are evaluated in which different parameters were varied:

- case 0 (verification case): the reference signal was stepped to 10 and the integer p, which determines the number of times the signals are inter-sampled, was varied,

Reference signal equal to zero:

- case 1: p was varied,
- case 2: similar to case 1, p was varied , but the number of samples N was kept constant,
- case 3: p was kept constant, but the noise power was varied,

Reference signal stepped from 0 to 0.001:

- case 4: p was varied,
- case 5: p was kept constant, but the noise power was varied,
- case 6: p was kept constant, but the bandwidth of the plant was varied (a different closed-loop system, also described in Addendum A was used), and
- case 7: for the plant, simulated in open-loop, p was varied.

For these case the variances and estimation errors were determined by doing the Monte Carlo simulation for different values in these parameters.

In case 0 (verification case), identifiability was ensured by stepping the reference signal to a significantly larger value. This case verifies the software: it shows that under normal conditions, i.e. PE reference signal and a good SNR, the software can be used to identify an accurate and precise model.

In the first number of cases, the reference signal was made zero to ensure that this signal is not PE. In these cases, p was varied to determine how the precision of a model is influenced

Table 4.1: Variance in Responses and Estimation Error in Parameters

	Step	Impulse	Bode Magnitude	Bode Phase	ERR
$p = 1$	4.9108×10^{114}	2.0656×10^{115}	52.3338	3.2036×10^{3}	0.0835

by the number of times the output is inter-sampled, when identifiability is not ensured with a PE reference signal or a nonlinear controller. The number of samples N was then kept constant to see if the influence of p is only due to the increase in data samples. Furthermore, the noise power was varied to determine what the influence of the SNR is on the precision of the model when the reference signal is zero.

In the other cases the signals were stepped from zero to a very small value to ensured a PE reference signal and a large (bad) SNR. In these cases p was varied to determine how the precision of a model is influenced by the number of times the output is inter-sampled, when identifiability is ensured with a PE reference signal, but with the SNR very large. The noise power was also varied to determine what the influence of the SNR is on the precision of the model. Lastly, the bandwidth of the plant was varied to determine if the influence of p is dependent on the sampling frequency $\frac{1}{T}$ relative to the plant bandwidth. The influence of p was also evaluated in open-loop in order to verify the closed-loop results.

4.4.2 Results of Variance Simulation

4.4.2.1 Case 0 (Verification Case)

In the verification case a satisfactory, i.e. accurate and precise, model was identified for any value of p. The identified model agrees very well with the true model in both the time domain, i.e. step and impulse responses, and frequency domain, i.e. phase and magnitude responses. These results verify the simulation software for the variance analysis.

4.4.2.2 Case 1

When $p = 1$ the data were not inter-sampled, i.e. the plant output was sampled at the same rate as the control input. For $r(t) = 0$ and $p = 1$ unstable models were identified and the estimation error, as well as the variance in the step, impulse, Bode magnitude and Bode phase responses are large, as can be seen in Table 4.1.

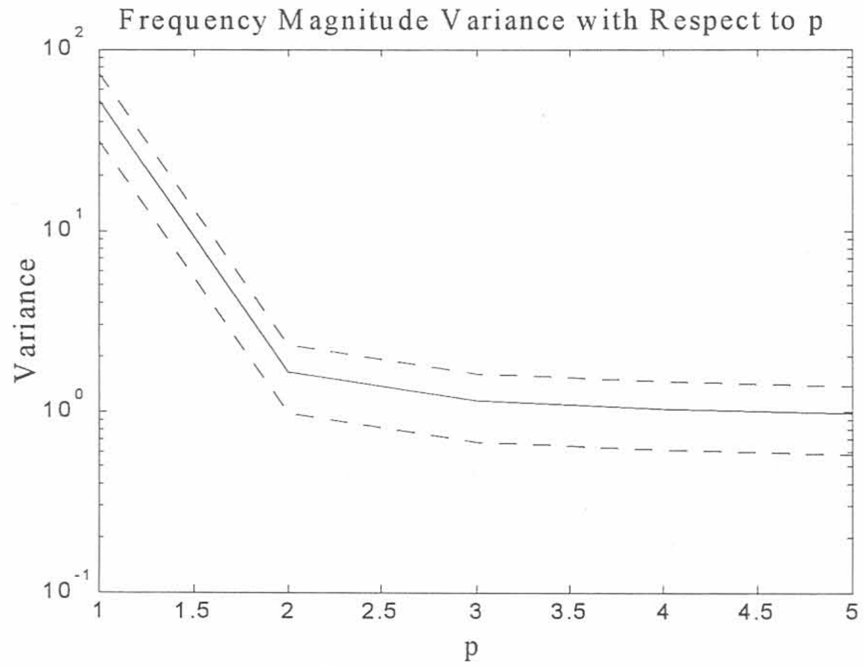

Figure 4.1: The variance in the Bode magnitude response for different values of p. The dashed lines represent the 95% confidence bounds.

For $p = 2$ most of the identified models were stable and the estimation error, as well as the variance in the step , impulse and Bode magnitude, are significantly smaller, as shown for the magnitude response in Fig. 4.1. Figure 4.2 shows that the reduction in variance for the Bode phase response only starts to happen at $p = 3$. The variance as well as the bias in the model are still very large compared to values obtained in the verification case.

For $p > 2$, the estimation error and variances for most of the responses are slightly reduced from those obtained for $p = 2$. Figures 4.1 and 4.2 show how the variances are reduced for the Bode magnitude and Bode phase responses. In the figures the 95% confidence bounds of the determined *variance values* are also shown. These confidence bounds are determined as in [42]:

$$\text{confidence bound} = \textit{variance value} \pm 2 * \sigma, \quad \text{where} \tag{4.3}$$

$$\sigma = \sqrt{\frac{2}{\text{simulation runs} - 1}} * \textit{variance value}, \quad \text{simulation runs} = 50. \tag{4.4}$$

4.4.2.3 Case 2

The results obtained in case 2 are similar to the results of case 1. However, the estimation

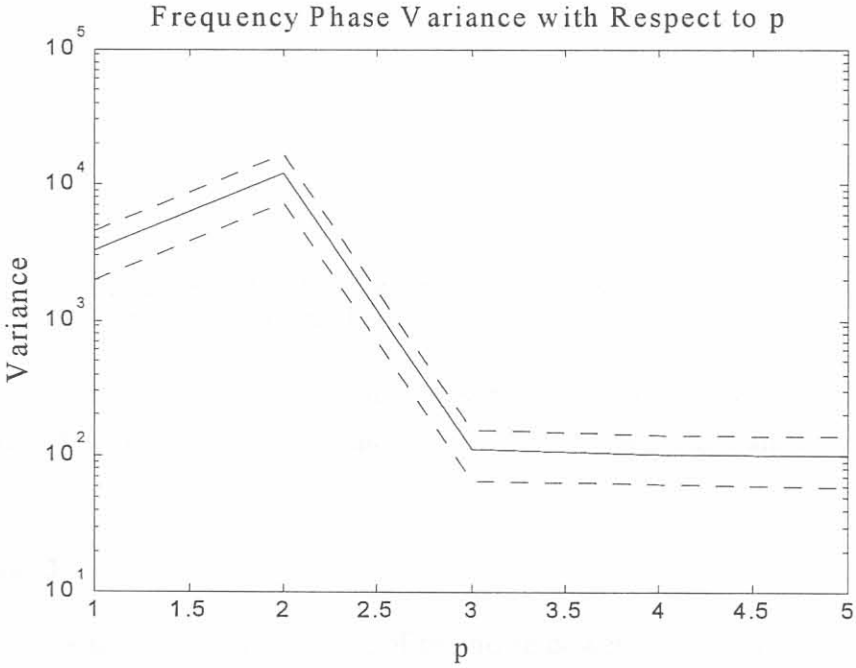

Figure 4.2: The variance in the Bode phase response for different values of p. The dashed lines represent the 95% confidence bounds.

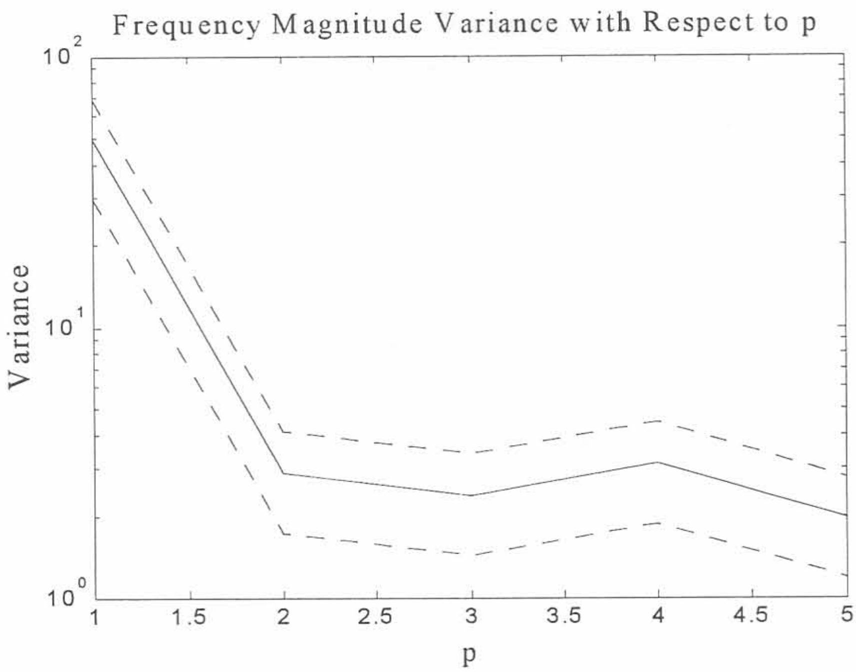

Figure 4.3: The variance in the Bode magnitude response for different values of p, with N constant. The dashed lines represent the 95% confidence bounds.

error and variances do not decrease as much as in case 1, but stay almost constant at higher values. Figure 4.3 shows how the variance of the Bode magnitude change.

4.4.2.4 Case 3

For a constant p and $r(t) = 0$, the size of the noise power does not have any effect on the size of the variances for the step, impulse, Bode magnitude and Bode phase responses.

4.4.2.5 Case 4

In case 4 the estimation error and variances of the responses are much smaller than in case 1, but still unsatisfactory. There is also not a big difference between the variance for $p = 1$ and $p \geq 2$. Again, the variance in the Bode magnitude response is shown in Fig. 4.4. For most of the responses the variances stay almost constant or even increase.

4.4.2.6 Case 5

When the reference was stepped, there was a small transient effect in the plant output. For

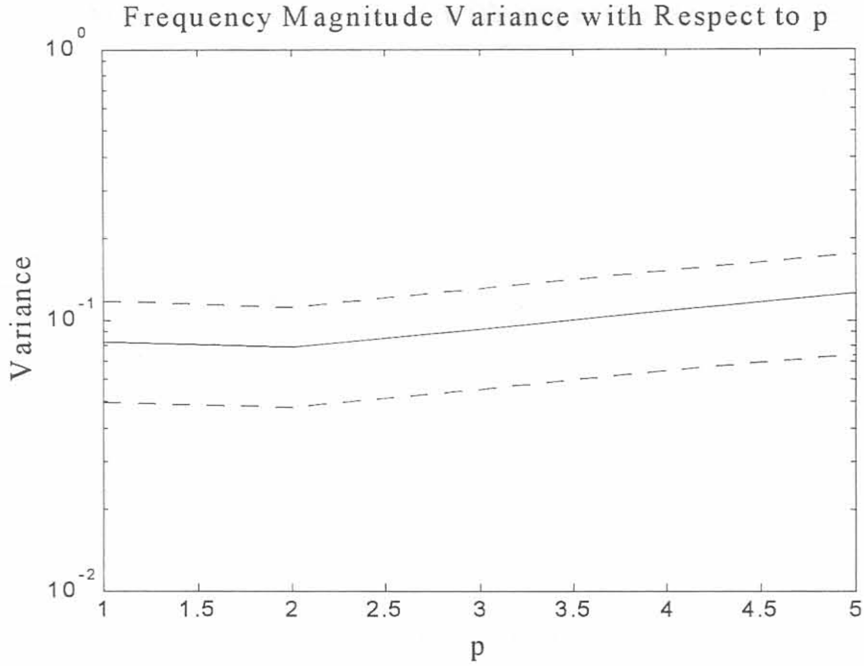

Figure 4.4: The variance in the Bode magnitude response for different values of p, with the reference signal stepped to 0.001. The dashed lines represent the 95% confidence bounds.

a very small noise power the transient effect was still visible, which resulted in a small variance, i.e. a precise estimation of the plant. As the noise power increases, the SNR worsens and the variance increases, as shown in Figs. 4.5 and 4.6. In Fig. 4.5 the 95% confidence bounds are too large to be shown.

4.4.2.7 Case 6

The original sampling frequency was made much slower than ten times the plant bandwidth. Repeatedly, the plant bandwidth was decreased by changing the time constant, and a new controller was designed (see Addendum A), but the execution time of the controller, as well as original sampling time T, was kept constant. The estimation error and variances of the responses for different values of p were determined.

When the original sampling frequency is much smaller than ten times the plant bandwidth, an increase in p decreased the estimation error, as well as the variances in the responses. However, as the original sampling frequency comes close to ten times the plant bandwidth, an increase in p does not significantly affect the variance any more.

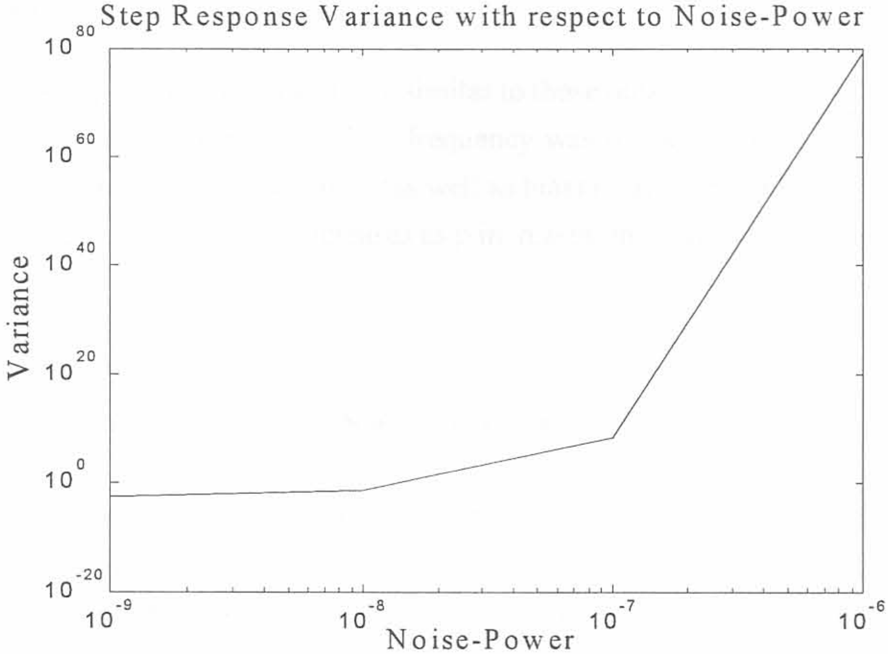

Figure 4.5: The variance in the step responses, with respect to the noise power, for a small step in the reference input.

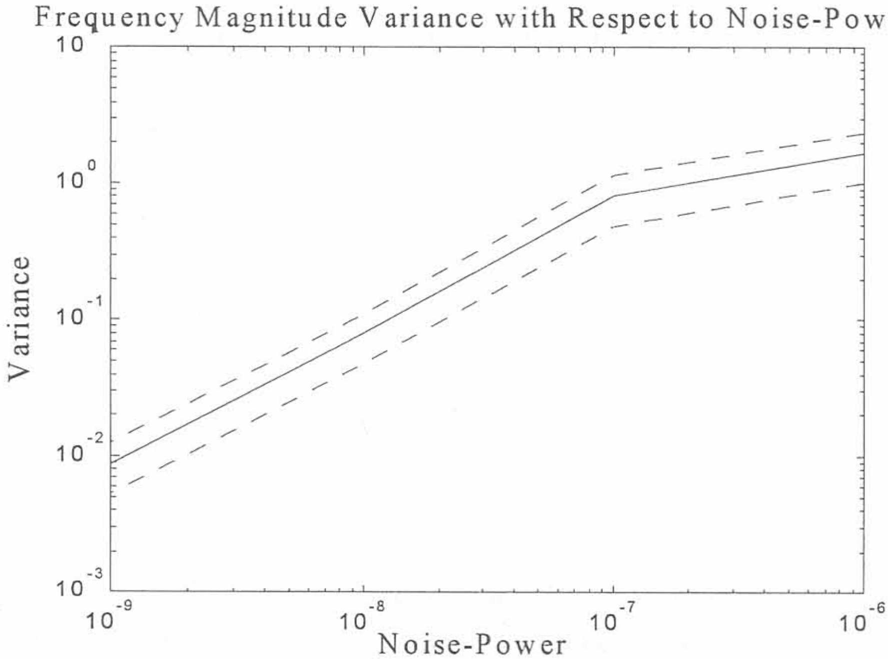

Figure 4.6: The variance in the Bode magnitude responses, with respect to the noise power, for a small step in the reference input. The dashed lines represent the 95% confidence bounds.

4.4.2.8 Case 7

For the open-loop system, the results are similar to those obtained in cases 4, 5 and 6. Firstly, N was kept constant and the sampling frequency was in the range of ten times the plant bandwidth. Again, the model variance (as well as bias) is large and an increase in p does not decrease the variance. When N increases as p increases, the variance improves only slightly [36].

4.4.3 Discussion of the Variance Simulation Results

From the fact that the estimation error and variances in case 1 decrease significantly from $p = 1$ to $p = 2$ or 3, it is concluded that even when the controller is of a lower order than the plant and the reference is not excited, inter-sampling the input and output of the plant at least once ($p = 2$) does make the plant identifiable. However, these results also show that identifiability does not guarantee a small variance or a small bias. Furthermore, this case shows that an increase in p does not have a significant effect in decreasing the variance. The small decrease in the variance can mostly be contributed to the fact that the number of samples N increases as p increases, because in case 2 where N is kept constant, the variance does not decrease as much as in case 1, but stays almost constant at a higher value.

For some responses an increase in variance occurs at certain instances. These increases can be contributed to the convergence error from the Δ-model to the T-model that increases with p [18]. For p very large, i.e. a small Δ, an increase in the variance can also be expected, because of the fact that a very small sampling time causes most least squares algorithms to become numerically unstable [11].

For a zero reference signal, there is no transient effect present in the plant output. Case 3 shows that in this case an increase or decrease in the noise power does not effect the variance and the variance is, in general, very large. The SNR refers to the ratio between the plant input signal and the noise. The noise power has no effect, since in this situation, the input signal only originates from the noise and thus the SNR is not relevant. However, when even a very small test signal is added tot the reference, as in cases 4 and 5 where the reference is stepped from 0 to 0.001, the variance is much smaller than obtained in cases 1, 2 and 3. In this situation the size of the noise power and thus the SNR has an effect on the size of the variance and the variance is only acceptable when the SNR is good. This agrees with the variance expression in Eqn. (3.70), where it is the signal-to-noise ratio, where *signal* is the part of the plant input that derives from the injected reference, that determines the quality of the open-loop transfer function estimate.

Case 6 shows that when the original sampling frequency $\frac{1}{T}$ is very slow in comparison with the plant bandwidth, i.e. much less than ten times the plant bandwidth (e.g. 50 times), then an increase in p will decrease the variance significantly. This agrees with the rule that the sampling frequency should lie in the range of ten times the plant bandwidth [11].

The open-loop simulation in case 7 confirms the closed-loop results, namely that an increase in p does not decrease the variance significantly.

4.4.4 Conclusion of the Variance Simulation

It is concluded that, although the inter-sampling approach ensures identifiability, it does not ensure a precise model with a small variance (or bias). Structured tests, where external test signals are added to the reference inputs, should be performed to ensure good SNRs.

Also, increasing the sampling frequency more than twice from the control action rate, will only reduce this variance significantly if the original sampling frequency $\frac{1}{T}$ is much slower than ten times the plant bandwidth.

4.5 Model Structure Selection

The choice of a suitable type of model structure is a crucial step in the identification process in order to obtain a good and useful model [6]. For closed-loop identification, the choice of the model structure depends on three often conflicting issues [4]:

- the consistency of the model in closed-loop identification,
- the compactness of the model, and
- the numerical complexity in estimating the parameters.

These and related issues will now be discussed.

4.5.1 Model Consistency

Since the direct SID approach is chosen in Section 4.2, the only estimation method that delivers consistent results when implemented on closed-loop data, is the PEM estimation method. With the PEM approach, the models may have arbitrary parameterisation.

The following model structures are all special cases of the more general PEM model family:

- FIR,
- ARX,
- ARMAX,
- OE, and
- BJ.

For all of these structures, the parameters are determined by minimizing the sum of squares of the prediction error [3]. Therefore, any of these structure will be consistent in direct closed-loop SID, except OE that will fail when the input noise is not white [11].

As already explained in Section 3.4, OE models, as well as the nonparametric models obtained from instrumental variables, spectral analysis and many subspace methods, are not consistent in closed-loop.

4.5.2 Compactness and Parametric Structures

Parametric models, such as the ARX and the ARMAX, are much more compact, since they need fewer parameters to describe a plant's dynamic behaviour than nonparametric models, such as FIR models (according to Zhu *et al.* [41] FIR models are nonparametric, but some may argue that FIR models are parametric). A more compact model will be more accurate and precise, provided that the parameter estimation algorithm converges to a global minimum and the model order is selected properly. Zhu *et al.* [41] show that the bias of parametric models is smaller than the bias caused by truncation in FIR models, since parametric models have infinite length (e.g. infinite impulse response) and no truncation is necessary. Parametric models also reduce the model variance of nonparametric models.

Also, for the same model precision, a parametric model requires up to 75% less testing time when compared to nonparametric models; or, put in another way, for the same test data, a parametric model can be much more precise [41].

Therefore, a parametric model is a better choice in the modeling of industrial processes, which are the kind of processes that are controlled by MPC controllers. Since many of these industrial processes may also be unstable, it seems natural to choose from these parametric models the ARX or ARMAX model structure, because these models give stable predictors [11].

4.5.3 Numerical Complexity

As stated in Section 3.7.4, when the ARX and FIR models are used together with a quadratic criterion function, the standard linear LSE method is obtained. This means that parameter estimates can be found by solving only a standard LSE problem, which can be done analytically. Typically, for other parameterisations and criteria, one has to rely on some iterative search scheme to find the best estimate [6].

The linear LSE method is numerically simple and reliable, and problems with local minima or convergence do not occur [41].

Although the ARMAX model is more compact than the ARX model, the numerical complexity in the parameter estimation of the ARMAX model is much higher than for the ARX model, since nonlinear optimisation routines are needed. Therefore, the ARMAX optimisation routines often suffer from local minima and convergence problems when identifying multivariable processes [4].

4.5.4 Multivariable Models

Multivariable systems are often more challenging to model [11]. As mentioned, data from a single variable test may not contain sufficient information about the multivariable character of the process and multivariable testing can solve this problem [3]. However, in multivariable testing all the reference signals can be excited simultaneously. This makes SISO model estimation, where only the parameters of one SISO transfer functions can be estimated at a time, undesirable. Multiple-Input-Single-Output (MISO) model estimation, where partial models of a system's behaviour are constructed from all the input channels and one output channel, is more desirable, but MIMO model estimation of all the transfer functions is the most desirable. The reason for this is that models for prediction and control will be able to produce better results if constructed for all outputs simultaneously. This follows from the fact that knowing the set of al previous output channels gives a better basis for prediction than just knowing the past outputs in one channel [11].

Multivariable ARX (IDARX) models in MATLAB are parametric models with several inputs and several outputs that can be estimated using a standard MATLAB function. Here it is allowed for n_a, the order of polynomial $A(q)$, n_b, the order of polynomial $B(q)$ and n_k, the time delay, to contain one row for each output number and one column for each input number. For most other model structures several MISO models have to be identified, which can then, afterwards, be combined into a multivariable model [43].

4.5.5 Model Order Selection

For the purpose of control, it is important to select the model order so that the process model from inputs to outputs is accurate. In the time domain, this requires for the simulation error or predicted output error of the model to be minimal [4]. See Section 4.6.1.1 for a discussion of the simulation error and predicted output error.

For models of the ARX type, various orders and delays can be efficiently studied with the command *arxstruc* in MATLAB. The ARX model can be fitted to the validation data set for many different model structure orders. For each of these models, the sum of squared prediction errors is computed, as they are applied to the validation data set. The structure (order) that has the smallest loss function (best fit) for the validation set can then be selected. Such a procedure is known as *cross-validation* and is a good way to approach the model order selection problem. It is usually a good idea to visually inspect how the fit changes with the number of estimated parameters [43].

A good idea is to first establish a suitable value for the delay by testing second-order models with different delays. Use the delay that yields the best fit. All combinations of ARX models, with different orders for polynomials $A(q)$ and $B(q)$, with delays around the chosen value can be inspected to make sure [43].

However, if the model is validated on the same data set from which it is estimated the fit always improves as the flexibility of the model structure increases. One needs to compensate for this automatic decrease of the loss functions. There are several approaches. The best known technique is Akaike's Final Prediction Error (FPE) criterion and his closely related Information Theoretic Criterion (AIC). Both simulate the cross-validation situation, where the model is tested on another data set [11].

The FPE is formed as

$$FPE = \frac{1+\frac{d}{N}}{1-\frac{d}{N}}V, \tag{4.5}$$

where d is the total number of estimated parameters and N is the length of the data record. V is the loss function (quadratic fit) for the structure in question. The AIC is formed as

$$AIC = \log(V(1 + 2\frac{d}{N})). \tag{4.6}$$

According to Akaike's theory [43], the model with the smallest FPE, or AIC should be chosen.

A related criterion is Rissanen's Minimum Description Length (MDL) approach, which selects the structure that allows the shortest over-all description of the observed data [43].

If substantial noise is present, the ARX models may also need to be of high order to describe simultaneously the noise characteristics and the system dynamics. The reason being that for ARX models the disturbance model $1/A(q)$ is directly coupled to the plant model $B(q)/A(q)$ [43].

There are different ways to go about selecting the model order. According to Forssell [6], one typically starts by checking if the model of lowest order is sufficiently good. The model order is then increased until a model that passes the validation test is found. If the resulting model order is too high, since a high order is often not desirable for control design [44], model reduction can be considered. As a general rule of thumb, one then knows that the variance error will dominate the bias error for this model [6].

However, Leskens *et al.* [44] first identify a very high-order ARX model and then, afterwards, reduce the model order. Zhu *et al.* [3] also use this method in the ASYM approach.

It is decided to follow the more typical approach of first identifying a lower order model and then increasing the order, depending on the validation results.

4.6 Model Validation

Before a model can be delivered to the user, it has to pass some validation test. Model validation can loosely be said to deal with the question of whether the best model is also *good enough* for its intended use [6].

What is a good model? In the linear case, the answer is that the model fit has to be good in certain frequency ranges, typically around the cross-over frequency. In general, one can say that the desired control performance dictates the required quality of the model fit: the higher the demands on control performance are, the better model fit is required [6].

The methods chosen for the validation of the identified model are discussed under different headings. Some of the validation methods form part of the validation step in the proposed methodology and other methods are used to validate the methodology itself:

- the methods that fall under *standard validation* form part of the proposed methodology and can always be used to validate the identified models, and
- the methods that fall under *comparison with open-loop identified model* and *examination of closed-loop system* are used to validate the proposed methodology and does not form part of the methodology itself.

4.6.1 Standard Validation with the Closed-Loop Data

Common validation tools are *residual analysis* and *cross-validation*, where the model is simulated using *validation* data and where the output is compared to measured output data [6]. The *validation* data are a set of the measured closed-loop data that is different from the *estimation* data set. The aim is to validate the model using data that were not used to fit the model, since it is not so surprising that such a model will reproduce the estimation data. Therefore, this is a much more stringent test than using the same data for fitting and validation [11].

If a limited amount of data is available about two thirds of the region should be used to fit the data and one third should be used to validate the model.

4.6.1.1 Simulation and Prediction

In the simulation and prediction test, it is determined how well the model is capable of reproducing the validation data. One can work with k-step ahead model predictions $\hat{y}_k(t \mid m)$ as the basis for comparison. This means that $\hat{y}_k(t \mid m)$ is computed from past data,

$$u(t-1), ...u(1), y(t-k), ..., y(1), \tag{4.7}$$

using the model m. The case when k equals ∞ correspond to the use of past inputs only, i.e., pure simulation [11]. This is a very stringent test.

For control applications, the predicted output over a time span corresponding to the dominant plant model time constant, will be an adequate variable to look at [11]. In the case of MPC controllers, the model output is predicted over a model horizon [30]. Thus, the model has to predict the output accurately, i.e. reproduce the validation data, only for a number of steps ahead, equal to the model horizon h. A model that satisfies this condition will be satisfactory for the design of an MPC controller. In re-identification the model horizon of the old controller is known and this value can be used for validation.

The model can then be evaluated by a visual inspection of plots of $y(t)$ and $\hat{y}_k(t \mid m)$, or by the numerical value of fit $J_k(m)$ for N number of samples:

$$J_k(m) = \frac{1}{N}\sum_{t=1}^{N} |y(t) - \hat{y}_k(t \mid m)|^2 . \qquad (4.8)$$

4.6.1.2 Residual Analysis Test

The *leftovers* from the modeling process - the part of the data that the model could not reproduce - are the *residuals*,

$$\varepsilon(t) \quad = \quad y(t) - \hat{y}_k(t \mid m). \qquad (4.9)$$

These contain information about the quality of the model and can by analysed to draw conclusions about the validity of the model [11]:

Whiteness Test: An *auto-correlation* of the error signal, $R_\varepsilon^N(\tau)$, determines whether the error signal is white noise or not. If it is white then the model is an unbiased estimator, which means that model parameter estimates will be *true* on average [11]. If the auto-correlation function of the error signal does not fall significantly outside the 99% confidence region, except for $R_\varepsilon^N(\tau) = 1$, then one can reasonably assume that the error is white noise. The errors can also be plotted and visually inspected to see if outliers are present or if it represents white noise.

Independence of Residuals and the Past Inputs: The independence of residuals and past inputs can be determined from the cross-correlation between these signals, $R_{\varepsilon u}^N(\tau)$. An appealing way to carry out this test is to plot the *cross-correlation* function. The confidence limits of this function will be horizontal lines. For a good model, the cross-correlation function should not fall significantly outside the 99% confidence region. If negative correlation is present, it does not mean that the model structure is deficient, only that output feedback occurs - the current error influences the future input [11]. A rule of thumb is that a slowly varying cross-correlation function outside the confidence region is an indication of too few poles, while sharper peaks indicate too few zeros or incorrectly estimated delays [43].

If these cases are satisfied, the model is, statistically speaking, a very good model.

4.6.1.3 Model Reduction

One method that tests if a model is a simple and appropriate system description, is to apply some model reduction technique to it. If a model order can be reduced without affecting the input-output properties very much, then the original model is unnecessarily complex [11]. Söderström [45] has developed this idea for pole-zero cancellation. If the pole-zero plot, including confidence intervals, indicate pole-zero cancellation in the dynamics, this suggests that a lower order model can be used. In the case of ARX when pole-zero cancellations occur, the extra poles are usually introduced to describe the noise. In this case another model structure should be tried [11].

4.6.2 Comparison with the Open-Loop Identified Model

The proposed closed-loop SID methodology will be used for re-identification of the plant. Since the open-loop SID method was previously used, the closed-loop identified models should be compared to models obtained from open-loop tests.

To validate the proposed identification methodology, the plant model in question should, therefore, also be estimated from data obtained from open-loop tests. If the model identified from the open-loop test data is accurate and precise, it can serve as a reference model for evaluating a model obtained from the closed-loop data. If the models are comparable, it is an indication that the proposed identification methodology is valid.

Since the models are black-box models, the consistency of the input-output behaviour of the models should be evaluated [11]. This can be done as follows:

Firstly, these models can be compared visually in the time and frequency domain. The Bode phase and magnitude responses as well as the step and impulse responses of both models can be compared. Similarly the pole-and-zero plots and Nyquist plots can be compared.

Secondly, the fit of the measured output with the pure simulated output and with the h-step predicted output of the open-loop identified model can be compared with the closed-loop identified model's fit.

Lastly, a residual analysis test can be done for the open-loop identified model. These results can then be compared with the results obtained for the closed-loop identified model.

These tests should be performed using the same validation data set. A good comparison will indicate that the identification methodology is comparable in accuracy to open-loop SID.

It is possible that the closed-loop identified model can be more accurate than the open-loop identified model. Therefore, in the simulation case where the true model is known, the open-loop identified model and the closed-loop identified model should also be compared, by evaluating how close these models are to the true model. The accuracy of these models can be determined by visually and numerically comparing them to the true model in the time and frequency domain. A typical numerical value that can then be compared for the two estimated models is the relative norm

$$\begin{aligned} freqfit &= \left\| \frac{\left\|\hat{G}_N(\omega)\right\|_2^2 - \left\|G_0(\omega)\right\|_2^2}{\left\|G_0(\omega)\right\|_2^2} \right\|^2, \quad \text{where} \\ \left\|G(\omega)\right\|_2^2 &= \sup_x \frac{\left\|G(\omega)x\right\|^2}{\left\|x\right\|^2}, \quad \text{and} \\ \left\|x\right\|^2 &= \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}. \end{aligned} \tag{4.10}$$

The model that gives the smallest value for Eqn. (**??**) has the best fit in the chosen frequency range. A similar value can be computed in the time domain, e.g. the relative norm for the step response coefficients

$$stepfit = \left(\sum_{i=1}^{n} \left| \frac{\hat{s}_i - s_{0i}}{s_{0i}} \right|^2\right)^{1/2}. \tag{4.11}$$

4.6.3 Examination of the Closed-Loop System

The goal of model validation is to test whether the model is good enough for its purpose and to provide advice for possible re-identification if the identified model is not valid for its intended use [3]. The purpose of the model in question is to design an acceptable MPC controller. Therefore, the final proof rests in demonstrating the acceptable performance of the controller.

A good validation test is therefore to see how the MPC controller, designed from the closed-loop identified model, controls the true plant. If the regulator, based on the identified model, gives satisfactory control, the model is a *valid* one [11].

Often, it will be impossible, costly, or dangerous to test all models with respect to their intended use in practice [11]. Therefore, this validation method is not included in the proposed methodology. However, when the true model is known, a good validation test of the methodology is to simulate how the MPC controller, designed from the identified model, controls the true plant. The stability of the resulting closed-loop system should be analysed. The system is stable if all the poles of the discrete closed-loop system are inside or on the unit-circle.

4.7 SYSTEM IDENTIFICATION TOOLBOX

Nowadays there are well-supported and user friendly tools available for the identification of linear systems on the basis of experimental data. In particular, the Mathwork's System Identification Toolbox (SITB), version 6.1, which is equipped with a graphical user interface, can be mentioned. This toolbox enables the user to identify and validate models in different types of model structures. Additionally, there is users' support in terms of graphical tools for model evaluation, as well as support for e.g. bookkeeping of identified models [20].

In this SITB there are only limited possibilities to identify models on the basis of data that are obtained under closed-loop experimental conditions. This particular experimental situation often requires special treatment, in the sense that besides input and output signals of a plant, measured external excitation signals can be involved, as well as some, possibly known, controllers that are implemented on the system. In these cases an add-on to the SITB, such as the CLOSID toolbox is needed [20].

Luckily, in the direct closed-loop SID approach no reference signals, or controller information, are involved in the identification process, only the plant inputs and outputs. Thus, the SITB, without any custom-written algorithms or extra software, is sufficient and suitable for the direct closed-loop SID approach. Therefore, no new software has been developed for this methodology.

The functions of the MATLAB SITB, used in the implementation of the methodology, are discussed in Section 5.3 and Section 6.2.

4.8 System Identification Steps

The appropriate closed-loop SID methodology, for plants controlled by MPC controllers, can now be summarised by addressing each of the five SID subproblems, discussed in Section 3.2. Each of these subproblems can be seen as a step in the SID procedure. The five steps, together with a summary of the chosen methods for validation of the methodology, are given.

The direct closed-loop SID approach should be used when implementing these SID steps. This approach will deliver the best SID results for plants controlled by nonlinear controllers, such as constrained MPC controllers, since this method is applicable to systems with arbitrary feedback mechanisms.

4.8.1 Experiment Design

Experiment design involves issues like choosing what signals to measure, choosing the sampling time, and choosing excitation signals. Once these issues have been settled, the actual identification experiment can be performed and process data be collected.

4.8.1.1 Signals to be Measured

Since the direct closed-loop SID approach is used, only the inputs and outputs of the plant have to be measured. Since this methodology will usually be implemented for re-identification purposes, the signals to be used as inputs and outputs would have been chosen earlier on and would therefore be known at this stage.

4.8.1.2 Sampling Time

The sampling frequency should lie in the range of ten times the plant bandwidth [11]. Since the MPC controller is already implemented, the measurement devices are usually already in place and the sampling time predetermined. Usually, the sampling time of these devices is the same as the execution time of the controller.

If possible, the inputs and outputs should be inter-sampled in such a way that it is sampled twice as fast as the execution time of the controller. Although the identified model may still have a very large variance, at least identifiability will be ensured when applying the inter-sampling method.

4.8.1.3 Excitation Signals

If possible, one should always try to add a persistently exciting signal to the reference input, either $r_a(t)$ or $r_b(t)$, since this is the easiest way to ensure identifiability. The type of signal is determined by the plant dynamics.

4.8.2 Data Collection

4.8.2.1 Collection

The collection of the data should be straightforward, since in the case of industrial plants, controlled by MPC controllers, the inputs and outputs of the plant are constantly logged. Usually there is some type of database present from which the desired data can be retrieved.

4.8.2.2 Preprocessing

The raw data that have been collected from identification experiments are not likely to be suitable for immediate use in identification algorithms. There are several possible deficiencies in the data that should be attended to:

- high frequency (above the frequency of interest to the system dynamics) disturbances in the data record,
- occasional burst and outliers, missing data and non-continuous data records, and
- drifts and offsets, low-frequency disturbances, possibly of a periodic character [11].

In off-line applications, the data sets should always be plotted first to inspect them for these deficiencies. Standard preprocessing methods used for open-loop data can also be used for closed-loop data. In the MATLAB SITB there are many routines to plot data, filter data, and remove trends in data. The preprocessing methods used in this specific application are discussed in Section 5.3 and Section 6.2.

Since a prediction error model is employed, it is especially important to remove trends, drifts and outliers. This will prevent the discrepancy in signal levels to dominate the criterion of fit, which could mask the dynamic properties [11].

4.8.2.3 Time Delay

At this stage, the time delay of the model should also be selected. This can be estimated by making use of visual inspection of the data inputs and outputs as well as knowledge of the plant and the previous model. After the model has been estimated, the choice of time delay can be validated. If the validation results, especially the residual analysis, is unsatisfactory, the time delay should be re-estimated and the model re-identified. It is easy to estimate many models for different time delays in the MATLAB SITB. From these models the one with the best fit can be selected.

4.8.3 Model Structure Selection

4.8.3.1 Type of Structure

The ARX type model structure is the best choice of model structure, provided that the noise model is accurate for the process to be modelled, which will usually be the case in re-identification.

This step should be done with care. The plant in question should always be considered carefully before a final choice of the model structure is made. If the ARX structure does not deliver good validation results, then other structures can easily be determined and compared, making use of the standard validation methods in MATLAB.

4.8.3.2 Order Selection

The selection of the model order forms part of the model structure selection. In the case of re-identification an older version of the model is usually available. The order of the older model can be used - at least for a start. After the model has been estimated, this choice of order can be validated. If the validation results are unsatisfactory another order should be chosen and the model re-identified. Actually, many models for different orders, from low to high, can then be estimated and the best order can be selected by comparing loss functions. In the case where no validation data are available, one should make use of the AIC, FPE or MDL model structure selection criteria. The procedure is described in Section 4.5.5.

4.8.4 Model Estimation

Given a suitable model structure and measured data, one can turn to the actual estimation of the model parameters. Since there exist special-purpose software tools that are very efficient and easy to use for model estimation, this step is perhaps the most staightforward one [6].

The estimation method should, of course, be the PEM estimation method, since it delivers consistent results for closed-loop data. A multiple-output model can be estimated with the standard MATLAB SITB command, *idarx,* and a MISO or SISO model can be estimated with the *arx* command. These function use the standard LSE method to estimate the best parameters for an ARX model structure.

4.8.5 Model Validation

For model validation the standard validation methods, outlined in Section 4.6, together with a measured closed-loop validation data set, not used in the model estimation step, can be used. These are:

- simulation and prediction,
- residual analysis, and
- model reduction - check for pole-zero cancellation.

If the model fails the validation test, some, or all, of the above steps have to be reconsidered and repeated until a model that passes the validation test is found.

4.8.6 Methodology Validation

The methodology validation step does not form part of the proposed methodology itself, but is included, because, in this work, it forms the final step in the selection of an appropriate methodology. In this step the methodology is validated. In the case where the results obtained from this step are unsatisfactory, the whole methodology should be re-evaluated. The methods employed for the validation, also outlined in Section 4.6, are:

- a visual comparison of the closed-loop identified model with the open-loop identified reference model in both the time and frequency domain,
- comparison of the pure simulation and prediction fit of the closed-loop identified model with the fit obtained for a model identified in open-loop,
- comparison of the residuals of the closed-loop identified model with those obtained for a model identified in open-loop,
- a comparison of the open-loop and closed-loop identified models, by evaluating how close these models are to the true model, and
- examination of the closed-loop system's stability for the controller designed from the new model.

4.9 Conclusion

By reviewing the relevant literature one can see that there are many options available, regarding identification approaches, guarantees of identifiability, model structures and model validation techniques. From all these options the most appropriate ones were chosen to be used in the closed-loop identification of plants controlled by MPC controllers. These choices are mainly based on: the theory regarding closed-loop SID; characteristics of MPC controllers, e.g. nonlinearity; characteristics of industrial plants, e.g. multivariable; keeping the methodology relatively uncomplicated; and results obtained by other researchers for similar cases.

The methodology was, thus, developed by making selections, based on certain criteria, from the many available options in closed-loop system identification.

Closed-Loop SID Approach: The direct closed-loop SID approach was chosen for the proposed methodology. The direct approach is actually the only option for the MIBK process, since the reference signals, as well as the controller settings, of the MIBK process are not available and the indirect and joint input-output approaches can, therefore, not be implemented. However, it is concluded that the direct approach will deliver the best SID results for plants controlled by nonlinear controllers, such as constrained MPC controllers, since this method works regardless of the complexity of the controller. This approach also ensures consistency and optimal precision and simplifies the development of the methodology, since it makes use of the standard functions in the MATLAB SITB and does not require any custom-written algorithms or extra software.

Guarantee of Identifiability: It can be concluded that, although the MPC controller is nonlinear, time-varying and complex, which in general yield informative experiments, this is not a guarantee for identifiability. The reason being that the changes in a multivariable MPC structure, as it deals with changes in the active constraints, are unpredictable under normal operation and can thus not guarantee a given number of changes in the controller settings. A PE reference signal will, however, guarantee identifiability.

Inter-Sampling: From the variance simulation study on the inter-sampling approach it is concluded that, although this method ensures identifiability, without structured tests that ensure good SNRs, the variances of the identified models are very large and the precision is thus unsatisfactory.

Model Structure: The ARX type model structure was chosen. This is the best choice of model structure, provided that the noise model is accurate for the process to be modelled, since:

- it utilises the consistent closed-loop PEM estimation method,
- it is parametric, which in turn ensures compactness and accuracy,
- it can handle unstable systems without problems, since the predictor is stable,
- its parameters can be determined from a numerically simple and reliable LSE method that does not suffer from problems with local minima or convergence, and
- MATLAB allows for ARX structures to be estimated from MIMO data.

Model Validation: Since the aim of this closed-loop SID methodology is to substitute the open-loop SID method for re-identification, the identified models should be compared to models obtained from open-loop tests. Thus, together with the standard *validation* tests used in open-loop SID, the methodology should also be validated by comparing the closed-loop identified models to the open-loop identified models and by doing an examination of the resulting closed-loop systems. The chosen validation tests are summarised in Sections 4.8.5 and 4.8.6. It is also recommended that the models should be validated with validation data, i.e. data not used to fit the model.

The proposed methodology is summarised in terms of the five SID steps in Section 4.8.

In the following two chapters the chosen methodology is validated as well as evaluated. In the next chapter the results obtained from implementing the method on data acquired from a simulation of a 2x2 MIMO plant, controlled by an MPC controller, are discussed.

CHAPTER 5

VALIDATION AND EVALUATION OF THE METHODOLOGY WITH SIMULATIONS

5.1 INTRODUCTION

A simulation study was done in which the proposed closed-loop identification methodology, see Section 4.8, was successfully implemented. In this study a multivariable plant, controlled by an MPC controller, was identified from simulated closed-loop data. In order to evaluate the consistency of the identification methodology, the plant was identified for different settings in the controller, as well as for different added disturbances [46, 47]. Different methods for ensuring informative experiments were also considered.

In this chapter this simulation is described. First of all, in Section 5.2, the set-up, regarding the type of plant used, the chosen controller settings and all the cases considered, are discussed. Section 5.3 follows this discussion with an explanation of how each of the identification steps is implemented in MATLAB.

The obtained models are validated making use of the chosen methods described in Section 4.6. The validation process is also described in Section 5.3 in terms of the MATLAB commands. In Section 5.4 the expected validation results are discussed and the obtained results are given and also discussed. Finally, it is concluded in Section 5.5 that the proposed methodology can deliver consistent and satisfactory identified models of MIMO plants, controlled by MPC controllers.

5.2 SIMULATION SET-UP

5.2.1 Plant

The plant, which was used in the simulation, is a linear multivariable plant with two inputs and two outputs. The closed-loop configuration of the plant and controller is shown in Fig. 5.1. This plant, also given in Eqn. (5.1), is a benchmark example used in many of the MPC toolbox examples [29]. This is a scaled down version of what can be found in industry and it aids in demonstrating the basic identification steps of the proposed methodology for identifying multivariable plants, controlled by MPC controllers.

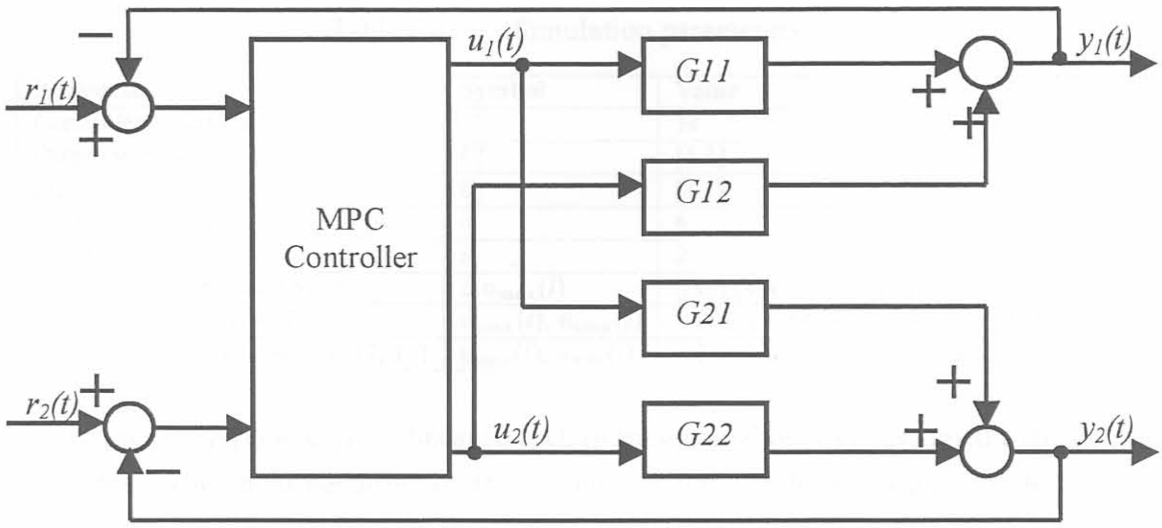

Figure 5.1: Closed-loop configuration of a multivariable plant controlled by an MPC controller.

$$\begin{bmatrix} y_1(s) \\ y_2(s) \end{bmatrix} = \begin{bmatrix} \frac{12.8e^{-s}}{16.7s+1} & \frac{-18.9e^{-3s}}{21.0s+1} \\ \frac{6.6e^{-7s}}{10.9s+1} & \frac{-19.4e^{-3s}}{14.4s+1} \end{bmatrix} \begin{bmatrix} u_1(s) \\ u_2(s) \end{bmatrix} \tag{5.1}$$

Since the MPC toolbox requires discrete-time models, the discrete step response model of the plant was obtained with the MPC toolbox function *tfd2step,* using a sampling time of 1min. The discretised MIMO transfer function was approximated with 90 step response coefficients.

5.2.2 Controller

For open-loop stable plants, the nominal stability of the closed-loop system depends on the prediction horizon h, the number of control moves i and the weighting matrices Γ_l^y and Γ_l^u. No precise conditions on h, i, Γ_l^y and Γ_l^u exist that guarantee closed-loop stability. In general, decreasing i relative to h makes the control action less aggressive and tends to stabilise a system. For $h = \infty$, nominal stability of the closed-loop system is guaranteed for any finite i, and time-varying input and output weights. More commonly Γ_l^u is used as a tuning parameter. Increasing Γ_l^u always has the effect of making the control action less aggressive [29].

In general, one must choose the horizons and weights by trial-and-error, using simulations to judge the effectiveness of these choices. In this simulation equal weightings were chosen

Table 5.1: Simulation parameters.

Parameter	**Symbol**	**Value**
Controller execution time	T	1s
Output weights	Γ_l^y	[1 1]
Input weights	Γ_l^u	[1 1]
Prediction horizon	h	6
Control moves	i	2
Saturation limits (cases 6, 9)	$\Delta u_{\max}(l)$	$\lvert\Delta u_1(k+l)\rvert \leq 0.1,\ \lvert\Delta u_2(k+l)\rvert \leq 0.05$
Input constraints (case 10)	$u_{\max}(l), u_{\max}(l)$	$-0.5 \leq u_i(k+l) \leq 0.5,\ i=1,2$
Output constraints (cases 11, 12, 13)	$y_{\max}(l), y_{\min}(l)$	$-1.5 \leq \hat{y}_i(k+l) \leq 1.5,\ i=1,2$

for $u_1(t)$, $u_2(t)$, $y_1(t)$ and $y_2(t)$, because each of these variables were assumed to be of equal importance. The equal penalties on the outputs reflects the desire to track both set-points accurately.

Because there is delay in the system, i was chosen to be smaller than h. This also makes the control action less aggressive and ensures stability. Because there was no *ringing*, i.e. damped oscillation, in the manipulated variables for the chosen parameters, it was not necessary to implement *blocking*. When blocking is implemented $\Delta u_i(t)$ is kept zero for a number of steps, e.g. $\Delta u_i(t)$ is kept zero for first two steps $u(k+2) = u(k+1) = u(k)$ [29].

Table 5.1 gives the final parameters chosen for the MPC controller. The MPC toolbox function *mpccon* was used for the design of the unconstrained controllers and the function *cmpc* was used for the design of the constrained controllers. These functions make use of quadratic optimisation, see Eqn. (2.1). When the function *mpccon* is used, the quadratic programme (QP) problem is solved analytically and this results in a linear controller. The *cmpc* function solves the QP problem iteratively, which results in a nonlinear controller. Furthermore, the default state estimate was used to calculate the controller gain matrix, since this results in a DMC controller [29].

5.2.3 Simulation Scenarios

Data sets from the following different cases, summarised in Table 5.2, were used for identification. In most of these cases, structured tests were performed by adding external test signals to the reference inputs. These signals are discussed in Section 5.3.1. In the other cases no reference inputs were added, i.e. $r_i(t) = 0$. Either no disturbance was added, or a unit disturbance was added at $t = 1$ and removed at $t = 2$ (pulse disturbance), or a constant disturbance of one unit was added from $t = 1$ onwards (step disturbance).

Unconstrained Control Law:

- case 1: no disturbances to the system,
- case 2: pulse disturbance added to $u_1(1)$,
- case 3: output pulse disturbance added to $y_1(1)$,
- case 4: step disturbance added to $u_1(1)$,
- case 5: output step disturbance added to $y_1(1)$,
- case 6: saturation limits on the manipulated variables,
- case 7: no disturbances to the system, **but with** $r_i(t) = 0$,
- case 8: output step disturbance added to $y_1(1)$, **but with** $r_i(t) = 0$,
- case 9: saturation limits on the manipulated variables and output step disturbance added to $y_1(1)$, **but with** $r_i(t) = 0$ **and the output inter-sampled**,

Constrained Control Law:

- case 10: enforced hard bounds on the manipulated variables
- case 11: enforced hard bounds on the output variables,
- case 12: enforced hard bounds on the output variables, **but with** $r_i(t) = 0$, and
- case 13: enforced hard bounds on the output variables with output step disturbance added to $y_1(1)$, **but with** $r_i(t) = 0$.

Note that the hard bounds are fundamentally different from the saturation limits. The hard bounds are defined relative to the beginning of the prediction horizon, which moves as the simulation progresses. Therefore, at each sampling period k, the hard constraints apply to a block of calculated moves that begins at sampling period k and extends for the duration of the input horizon i. The saturation constraints, on the other hand, are relative to the fixed point, $t = 0$, the start of the simulation [29].

The function *mpcsim* was used to simulate the different closed-loop responses for the unconstrained cases 1-9. In cases 10-13, where hard bounds were implemented, the function *cmpc* was used to generate the controllers and to simulated the closed-loop responses, because it determines optimal changes of the manipulated variables subject to constraints [29]. In all the cases a simulation time of 90s was used.

Cases 1-6 and 10-11 are used to evaluate the consistency of the proposed identification methodology. In these cases the plant was identified for different settings in the controller

Table 5.2: Case scenarios.

Case	Uncon-strained	Saturation Limits	Constrained	Input Disturbance	Output Disturbance	PE Reference Input	Inter-Sampled
1	√					√	
2	√			√		√	
3	√				√	√	
4	√			√		√	
5	√				√	√	
6	√	√				√	
7	√						
8	√				√		
9	√	√			√		√
10			√ (input)			√	
11			√ (output)			√	
12			√ (output)				
13			√ (output)		√		

(cases 1, 6, 10, 11), as well as for different added disturbances (cases 1-5). In these cases, **structured tests** were performed by adding external test signals to the reference inputs. These structured tests ensured good SNRs and PE reference signal, which in turn ensured identifiability. Here, again, SNR refers to the ratio between the noise and the plant input signal.

In cases 7-9, and 12-13 no structured tests were performed. The reference signals were zero and thus not PE. Here other methods to ensure identifiability are considered. In cases 7 and 8, no identifiability condition was satisfied: references were not PE; controllers were linear; and outputs were not inter-sampled. In case 9 the outputs were inter-sampled and the plant was thus identifiable. In case 12 and 13 the plant was also identifiable, since the controllers were constrained and thus nonlinear. In case 13 a disturbance was added, to evaluate the influence of the SNR with nonlinear feedback.

5.3 Implementation of the Methodology

The proposed methodology, as described in Section 4.8, was implemented in MATLAB. This implementation is discussed in terms of the five SID steps, as well as the methodology validation step.

5.3.1 Experiment Design

Signals to be Measured: The manipulated variables, $u_1(t)$ and $u_2(t)$, as well as the output variables, $y_1(t)$ and $y_2(t)$, were measured. These values were obtained as outputs of the *mpcsim* and *cmpc* functions.

Sampling Time: A sampling time of 1min, which is equal to the execution time of the controller, was used in cases 1-8 and 10-13.

The standard MPC toolbox function *mpcsim* was modified to allow inter-sampling of the output. This inter-sampling was implemented in case 9.

Excitation Signals: In cases 1-6 and 10-11 **structured tests** were performed by adding external test signals to the reference inputs, $r_1(t)$ and $r_2(t)$, to guarantee informative data. For each of these structured test cases two different trials were run where different test signals were used. In the first trial $r_1(t)$ was stepped at time zero and $r_2(t)$ was kept constant for 45s after which it was also stepped.

In the second trial a multivariable PRBS signal was generated making use of the MATLAB function *idinput*. The uncorrelated PRBS signals were added to the reference inputs. The period of these PRBS signals was taken as a tenth of the slowest time constant, which resulted in a period of 2min. The PRBS signals gave slightly better results than only step references. Therefore, the results obtained in the second trail of each case are discussed. A representation of the PE reference signals is given in Fig. 5.2. Some typical resulting input and output signals are given in Fig. 5.3.

In cases 7-9 and 12-13, where other methods to ensure identifiability are considered, **no structured tests** were performed and the reference inputs were kept at zero. In case 7, with the linear controller and no disturbances, the resulting inputs and outputs were zero. In cases 8, 9 and 13 where there were added disturbances, all the inputs and outputs look similar to Fig. 5.4. For case 12, with the nonlinear controller and no disturbances, the inputs and outputs are shown in Fig. 5.5.

5.3.2 Data Collection

Collection: The data collection was straightforward, since these values are the outputs that were generated by the *mpcsim* and *cmpc* functions.

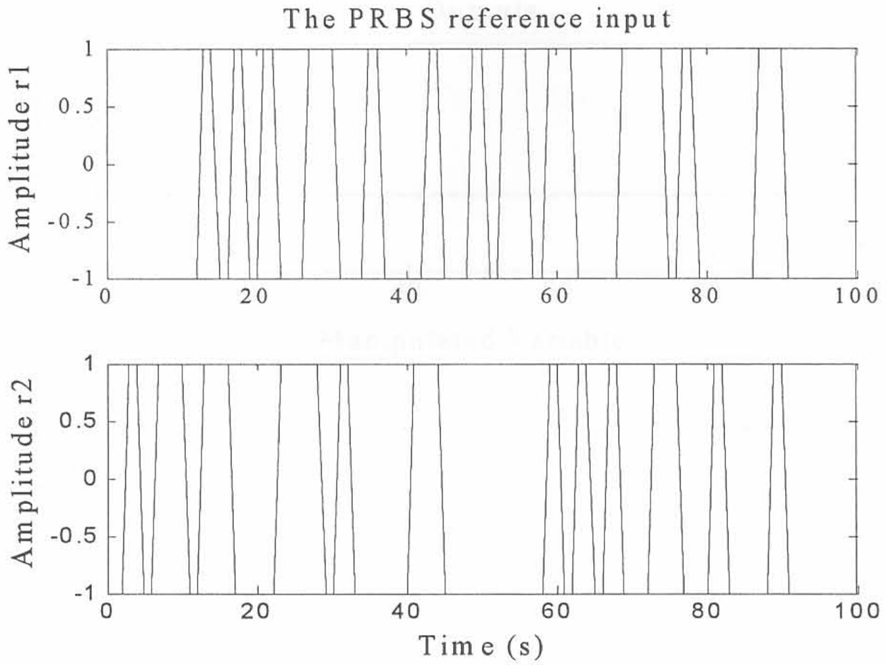

Figure 5.2: The persistently exciting PRBS reference signals used in trial 2.

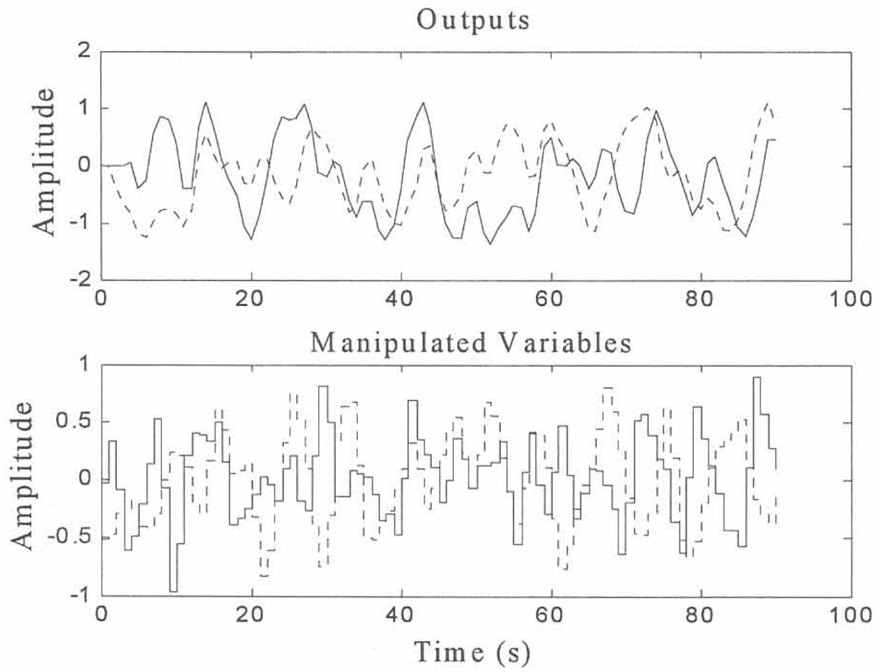

Figure 5.3: The resulting inputs and outputs in case 11 for a PRBS reference input. The dotted lines represent u_1 and y_1 and the solid lines represent u_2 and y_2.

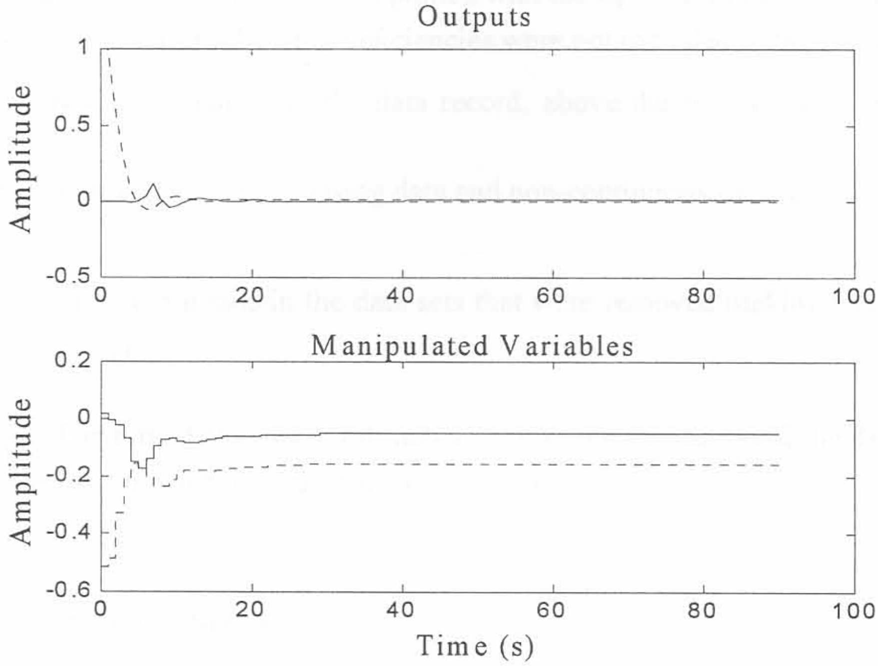

Figure 5.4: The resulting input and output signals in case 13 with $r_i(t) = 0$ and a disturbance. The dotted lines represent u_1 and y_1 and the solid lines represent u_2 and y_2.

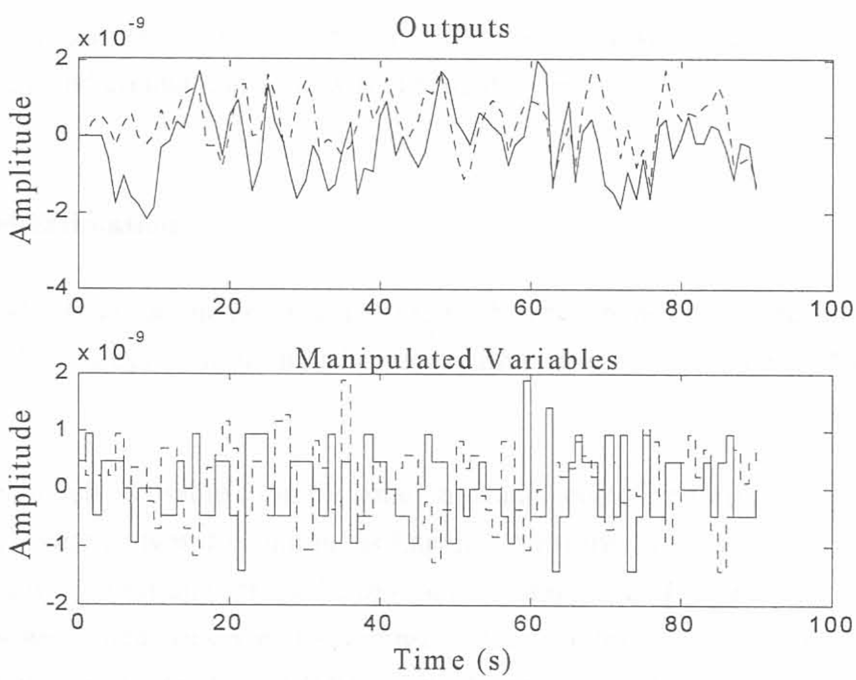

Figure 5.5: The resulting input and output signals in case 12 with $r_i(t) = 0$. The dotted lines are u_1 and y_1 and the solid lines are u_2 and y_2.

Preprocessing: The data sets were first plotted with the *idplot* command in order to inspect them for deficiencies. The following deficiencies were not included in the simulation:

- high frequency disturbances in the data record, above the frequency of interest to the system dynamics, and
- occasional burst and outliers, missing data and non-continuous data records.

However, there were trends in the data sets that were removed making use of the MATLAB function *dtrend*.

Time Delay: The time delays were estimated from visual inspection of the data inputs and outputs, as well as knowledge of the true plant and the previous model.

5.3.3 Model Structure Selection

Type of Structure: The multivariable ARX type model structure was used.

Order Selection: Since a plant model is necessary to design an MPC controller it was assumed that an old model already existed, which gave an indication of the model order. Therefore, the known order, namely first-order, of the model was used and it was not necessary to estimate and compare models with different orders.

5.3.4 Model Estimation

The LSE PEM estimation method was used to fit the chosen models to the estimation range of the data. The multiple-output models were estimated with the standard MATLAB SITB command, *idarx*.

To get the models in usable form for simulation and controller design, the discrete identified *th* models were convert to continuous Laplace transform models. First the ARX difference equation was determined from the *th* models making use of the *th2arx* command. The ARX models were then converted to continuous Laplace transfer functions with the *d2cm* command. The Zero-Order-Hold (ZOH) method, together with a sampling time of 1s and 0.5s for the inter-sampled model, were used for the transformation from discrete-time to continuous-time.

5.3.5 Model Validation

The standard validation methods, together with a measured closed-loop validation data set, not used for model estimation, were used. The validation data set is a part of the data set obtained from the simulation in case 6 where PRBS reference inputs were used.

Simulation and Prediction: The *compare* command was used to compute, for each identified model, both the *pure simulated* output ($\hat{y}_{\infty}(t \mid m)$) and the 6-step ahead predicted output ($\hat{y}_6(t \mid m)$). The command was also used to determine the numerical values of fit $J_k(m)$ between these outputs and the measured outputs (simulated "true" output). In MATLAB this value of fit is represented as a percentage. The higher the percentage, the better the fit.

Residual Analysis: The *resid* command was used to calculate and display the auto-correlation function of the residuals (test for whiteness), as well as the cross-correlation between the residuals and the plant input (test for independence). The residuals were also plotted for a simple visual inspection.

Model Reduction: Since first-order models were estimated in this simulation, the models were not inspected for order reduction.

5.3.6 Methodology Validation

In the methodology validation step, the proposed closed-loop SID methodology is validated.

Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model: The *step* command was used to plot the step responses and the *impulse* command was used to plot the impulse responses for each of the SISO transfer functions of the different identified models, including the open-loop identified model. The obtained plots are visually compared. For the frequency domain comparison the *bode* command was used to plot both the amplitude and phase responses for each of the SISO transfer functions, of the different identified models, including the open-loop identified model. These responses are also visually compared.

Simulation and Prediction Fit Comparison with the Open-Loop Identified Model: The *compare* command was used to determine the percentage of fit for the pure simulated and 6-step ahead predicted outputs of the model identified in open-loop. These values are compared with those obtained for the closed-loop identified models.

Residual Comparison with the Open-Loop Identified Model: Again, the *resid* command was used to calculate and display the auto-correlation function of the residuals, as well as the cross-correlation between the residuals and the plant input for the model identified from open-loop data. These functions are visually compared with the ones obtained for the closed-loop identified models.

Comparison of Both Models with the True Model: Equation (??) was used to determine the numerical value of fit in the frequency domain, *freqfit,* between the true model and the open-loop identified model, as well as with each of the closed-loop identified models. For each model, the total *freqfit* value was taken as the sum of the *freqfit* values of each SISO transfer function. The obtained numerical values are compared. In the same manner, Eqn. (4.11) was used to determine the numerical value of fit, *stepfit,* for the step responses of the models. Large values indicate unsatisfactory models.

Closed-Loop System Examination: Again, the discrete step response models of the identified models were obtained with the MPC toolbox function *tfd2step* using a sampling time of 1s. The transfer functions were approximated with 90 step response coefficients. The *mpccon* command was then used to design the new controllers. These controllers were used to control the true plant. The *mpcsim* command was used to simulate the closed-loop responses.

Furthermore, the *mpccl* command was used to determine the closed-loop models. The commands *smpcpole* and *max* were used to determine the maximum poles in order to see if the closed-loop systems are stable. The *smpcpole* command computes all the discrete poles and the *max* command then determines which pole has the largest absolute value. If the maximum absolute value is equal to or smaller than one, all the poles are on or inside the unit circle.

5.4 Validation Results

In this section the expected validation results are discussed and the obtained validation results are then shown and also discussed.

5.4.1 Expected Results

It was expected that, irrespective of the constraints and added disturbances, satisfactory models would be identified in all the cases where structured tests were performed that ensured

PE reference inputs and good SNRs. The direct closed-loop SID approach was used together with the PEM estimation method and this estimation method works if the data set is informative and the model set contains the true system, irrespective of correlation with the additive output noise. Informative data sets were ensured with the PE reference inputs and it was also known that the chosen model set (first-order ARX) contained the "true" model. Therefore, it was a reasonable assumption that these models would describe the plant accurately and that the controllers designed from these models would be able to control the plant.

It was also expected that if the noise model of the ARX structure is not an accurate description of the true noise model, the identified models might contain a bias. A possible bad fit in the low frequency regions was also expected, since the ARX model structure penalises the high-frequency errors more than low-frequency errors [44].

A possible difference between the models identified from open-loop and closed-loop data was also expected, since the frequency weighting for these two types of models are very different [11] and the plant may exhibit different dynamics under presence of the controller than in open-loop [6]. For the same reason it was expected that the closed-loop identified model might even produce a better controller, as the input weighting in this case favours the cross-over frequency region relevant for the controller [48]. Therefore, it was also be expected that inaccurate models in the low and high frequency regions could still result in good controllers. A bias in these frequency regions could therefore be irrelevant.

In general, unsatisfactory models were expected in the cases where structured tests were not performed, since in these cases PE reference signals did not enure identifiability and good SNRs were also not ensured. Also, for the inter-sampling method that ensures identifiability an imprecise model was expected. The reason being that, although the system was identifiable, it is shown in Section 4.4 that with $r_i(t) = 0$ a large variance can still result. However, for a constrained control law and good SNRs the identification of an accurate and precise model was expected, since a nonlinear controller ensures identifiability.

5.4.2 Obtained Results

In all cases where structured tests with PE reference signals and good SNRs were used, the identified models correspond very well with the model obtained from open-loop data and ensured good controller performance. The only case where no structured test was performed and a good model was still identified, is case 12. Here the controller was constrained and no disturbance was added to the system. The following cases delivered satisfactory models:

- case 1,
- case 2,
- case 3,
- case 4,
- case 5,
- case 6,
- case 10,
- case 11, and
- case 12 ($r_i(t) = 0$ and constrained controller).

The models identified in the above-mentioned cases are very similar and, consequently, the obtained validation results also look very similar. These satisfactory results are classified as **class A** results. The results from all of these cases cannot be shown. Therefore, in each section, the results of only one or two of these cases are given as a representation of all **class A** results.

All the cases that delivered unsatisfactory results had $r_i(t) = 0$. These cases are:

- case 7 (with no disturbance),
- case 8 (with disturbance),
- case 9 (output inter-sampled), and
- case 13 (constrained controller with disturbance).

In case 7, with the linear controller and no disturbances, the resulting inputs and outputs were zero and, therefore, no model could be identified. Thus, no validation results are given for this case. The results obtained in cases 8, 9 and 13 are unsatisfactory and also similar. These unsatisfactory results are classified as **class B** results and, in each section, the results of only one or two of these cases are given as a representation of all **class B** results.

5.4.2.1 Simulation and Prediction

In Figs. 5.6 and 5.7 the pure simulated and 6-step ahead predicted output for the model identified in case 11 are compared with the measured output of the validation data set. In the simulation the *measured* output is the output computed with the *true* model, while the simulated and predicted outputs are computed with the *identified* model. These figures show that,

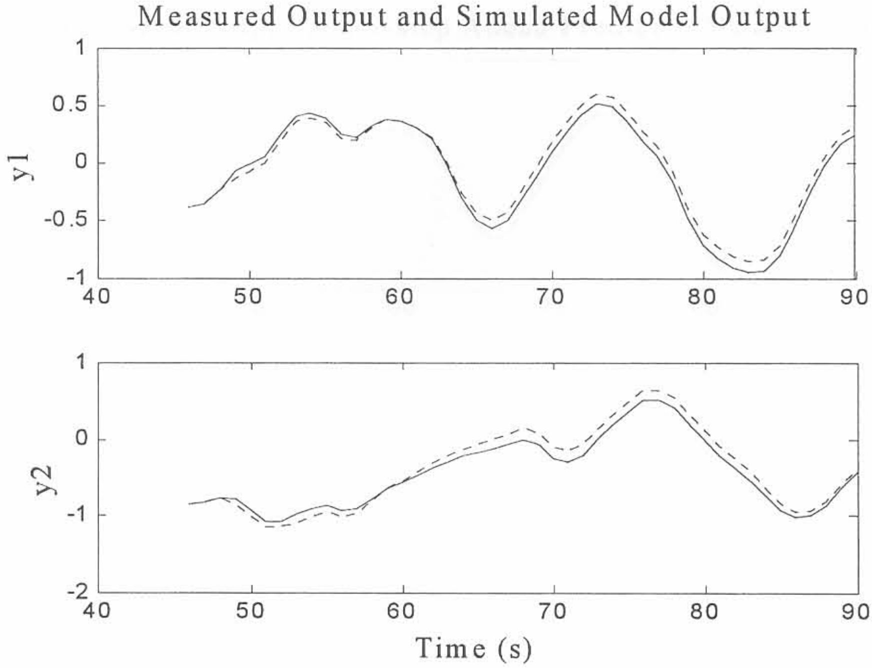

Figure 5.6: The measured (solid line) and pure simulated output (dotted line) for the model identified in case 11.

although the 6-step ahead prediction is better than the pure simulation, both the predicted outputs and pure simulated outputs follow the true output closely. These are representative of the **class A** results.

The computed numerical values of fit for the models identified in all the different cases are shown in Figs. 5.8, 5.9, 5.10 and 5.11. All the **class A** models are capable of reproducing the validation data with an average of 81.3% fit for pure simulation and 89.6% for 6-step ahead prediction. These figures show that all the **class B** cases obtain much smaller percentages of fit.

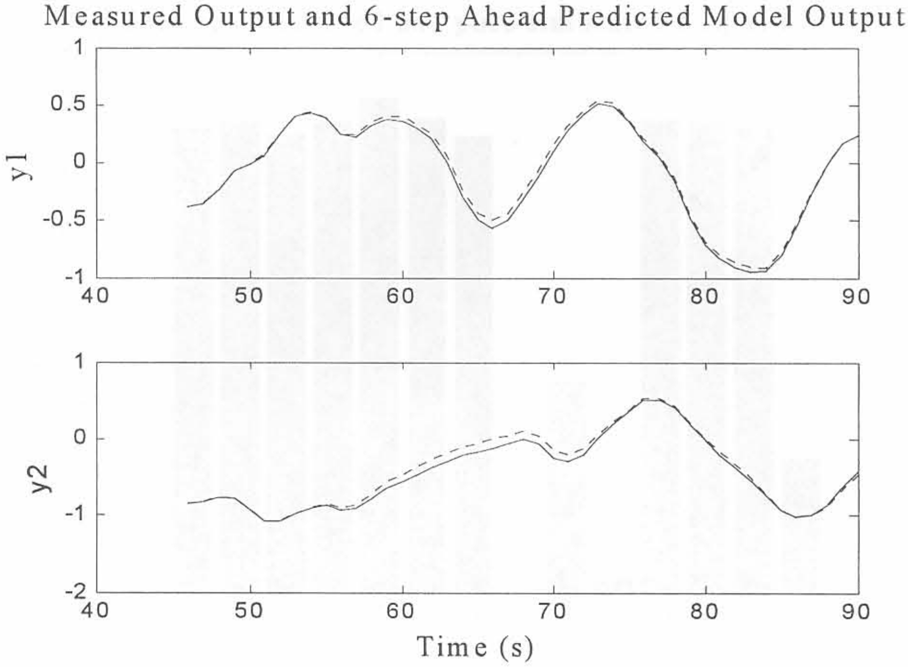

Figure 5.7: The measured (solid line) and 6-step ahead predicted output (dotted line) for the model identified in case 11.

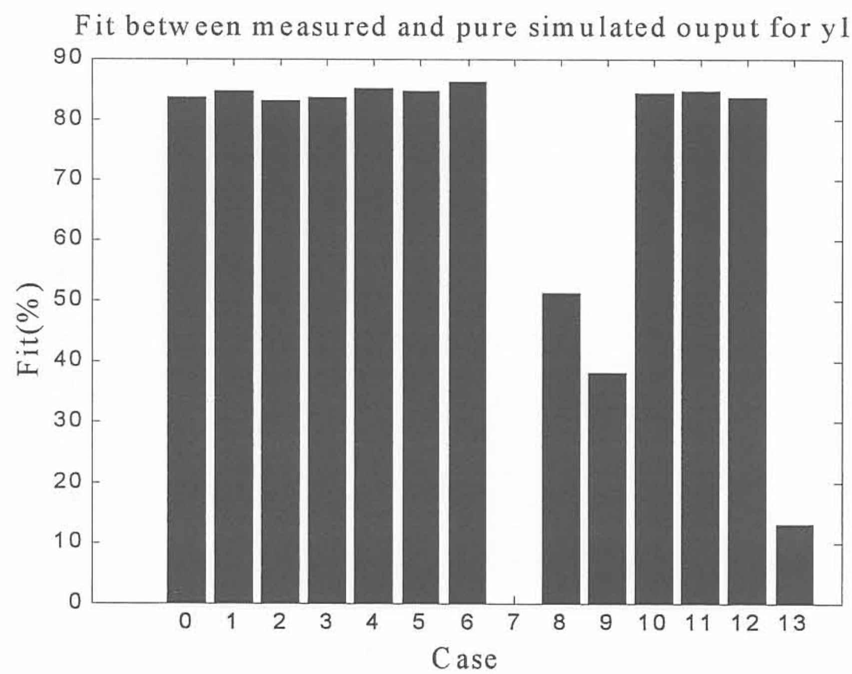

Figure 5.8: Fit between measured and pure simulated output y_1 for different cases and for the model identified from open-loop data (case 0).

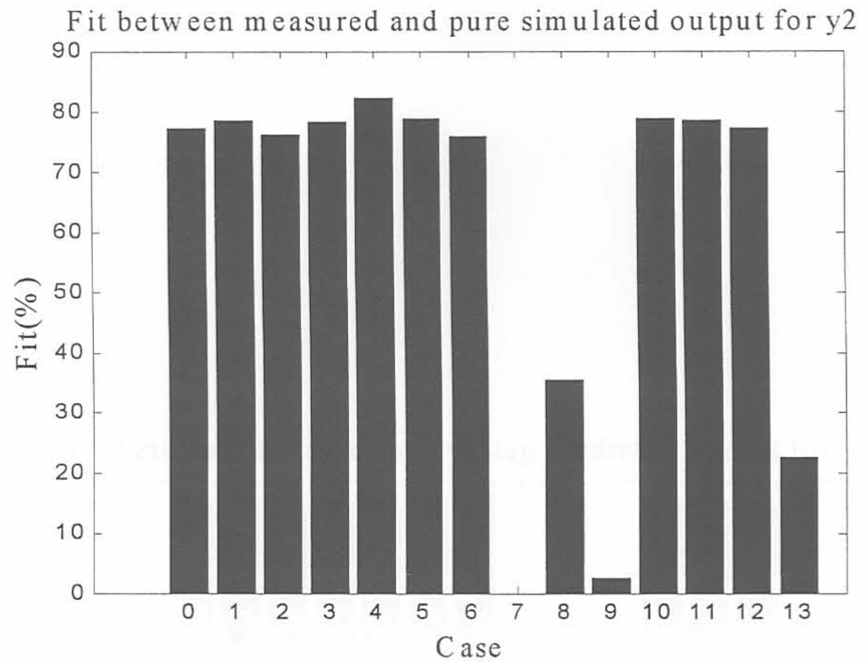

Figure 5.9: Fit between measured and pure simulated output y_2 for different cases and for the model identified from open-loop data (case 0).

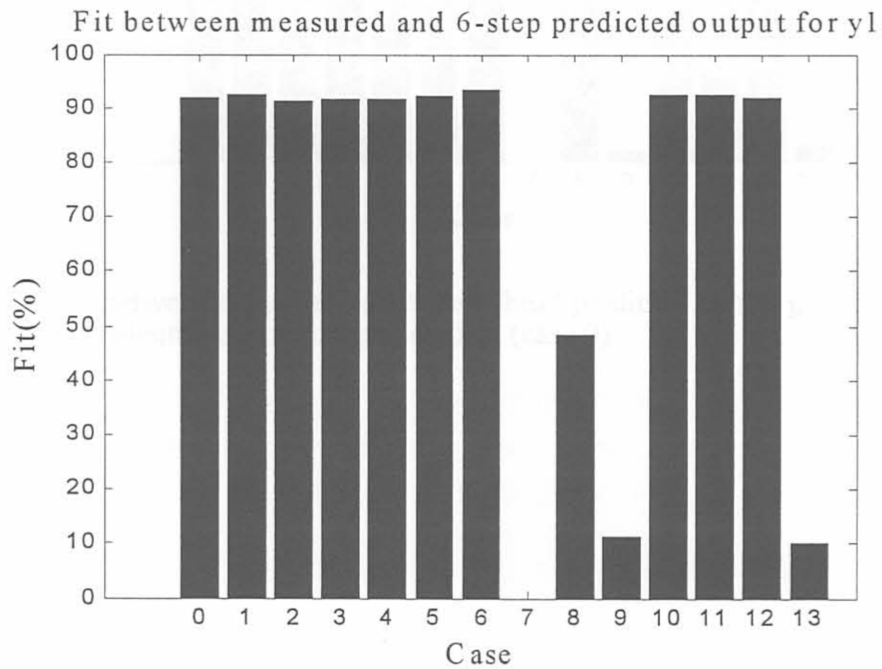

Figure 5.10: Fit between measured and 6-step ahead predicted output y_1 for different cases and for the model identified from open-loop data (case 0).

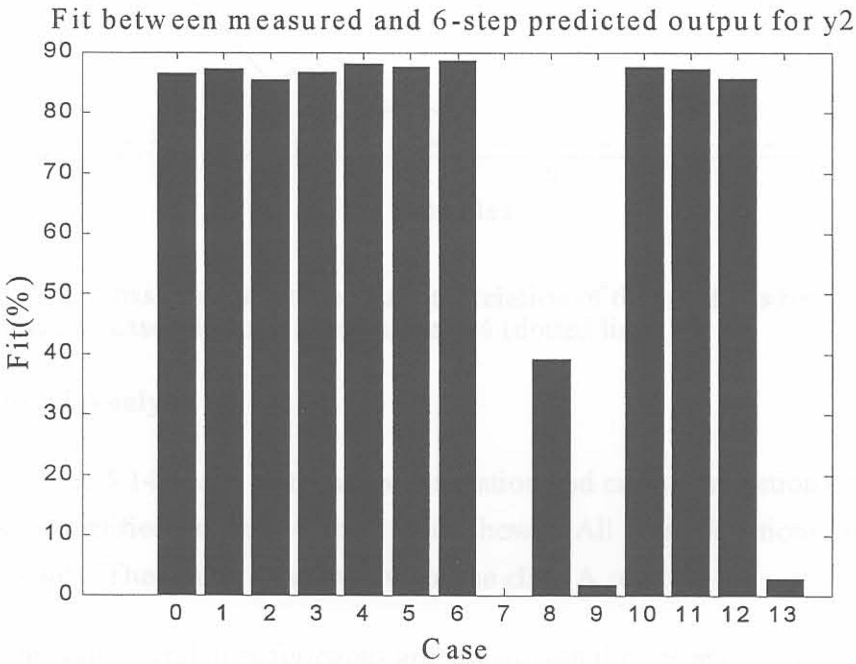

Figure 5.11: Fit between measured and 6-step ahead predicted output y_1 for different cases and for the model identified from open-loop data (case 0).

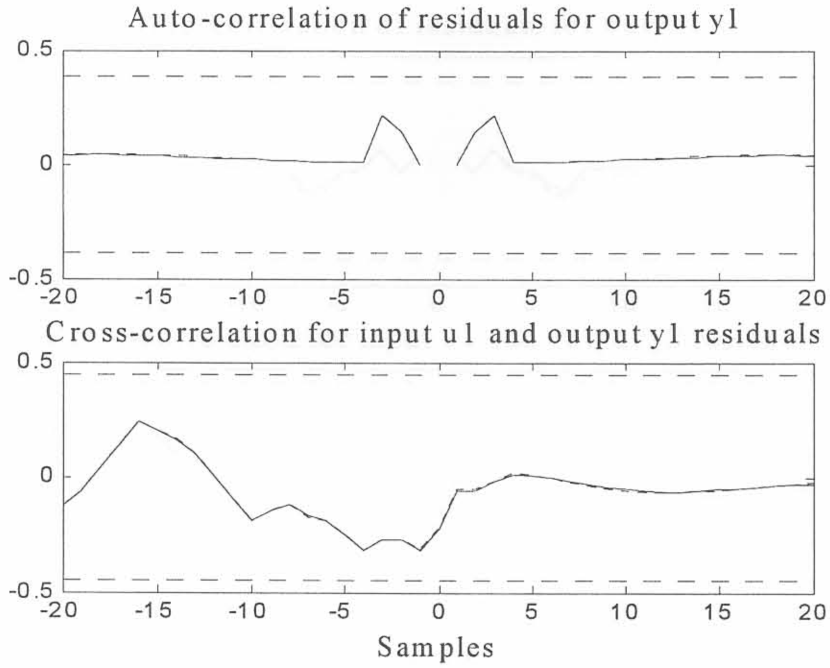

Figure 5.12: The cross-correlation and auto-correlation of the residuals for y_1 and u_1 of the models identified in case 11 (solid line) and case 4 (dotted line).

5.4.2.2 Residual Analysis

In Figs. 5.12, 5.13, 5.14 and 5.15 the auto-correlation and cross-correlation of the residuals for the models identified in cases 4 and 11 are shown. All these functions stay within the confidence bounds. These are representative of the **class A** results.

Note that the auto-correlation functions are scaled with the variance. Therefore, the value at lag = 0 is 1. However, this is not indicated to keep the scale of the figures between 0.5 and -0.5 which is the interesting region [36].

In Figs. 5.16, 5.17, 5.18 and 5.19 the auto-correlation and cross-correlation of the residuals for the models identified in cases 9 and 13 are shown. These functions do not always stay within the confidence bounds, but go outside the bounds. These are representative of the **class B** results.

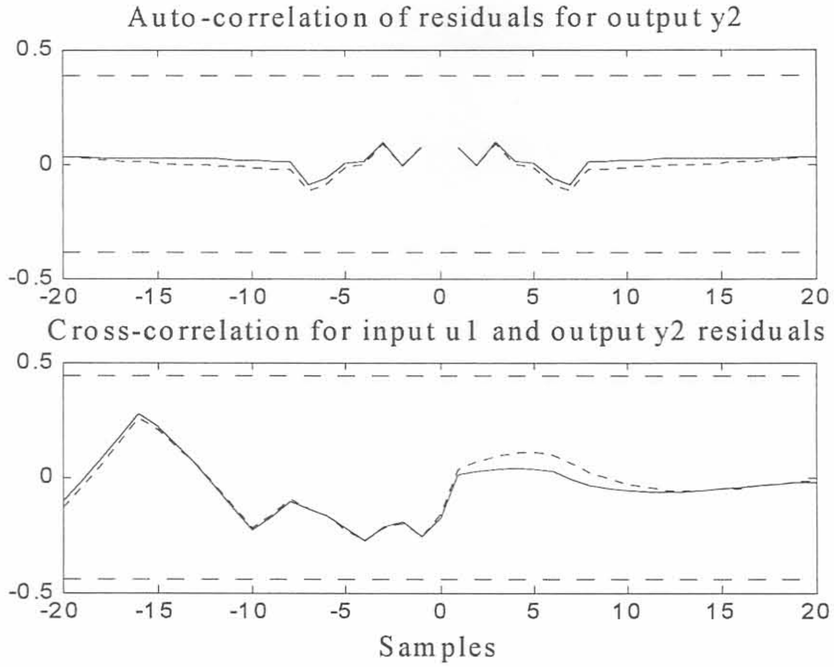

Figure 5.13: The cross-correlation and auto-correlation of the residuals for y_2 and u_1 of the models identified in case 11 (solid line) and case 4 (dotted line).

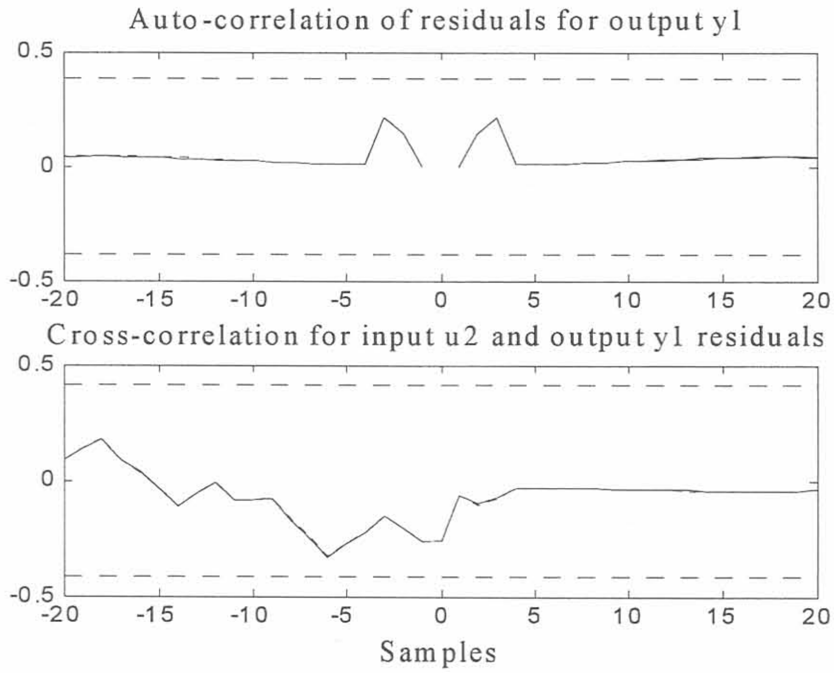

Figure 5.14: The cross-correlation and auto-correlation of the residuals for y_1 and u_2 of the models identified in case 11 (solid line) and case 4 (dotted line).

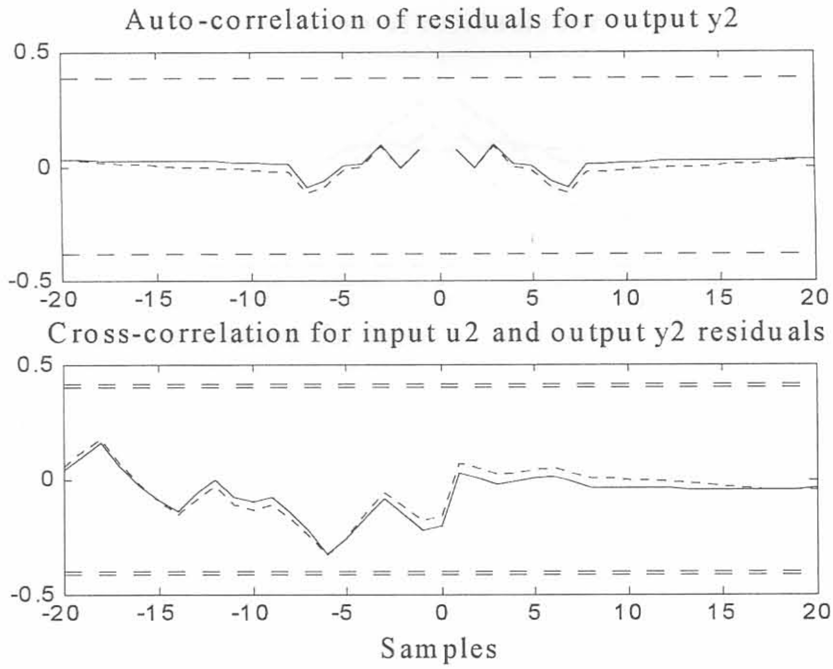

Figure 5.15: The cross-correlation and auto-correlation of the residuals for y_2 and u_2 of the models identified in case 11 (solid line) and case 4 (dotted line).

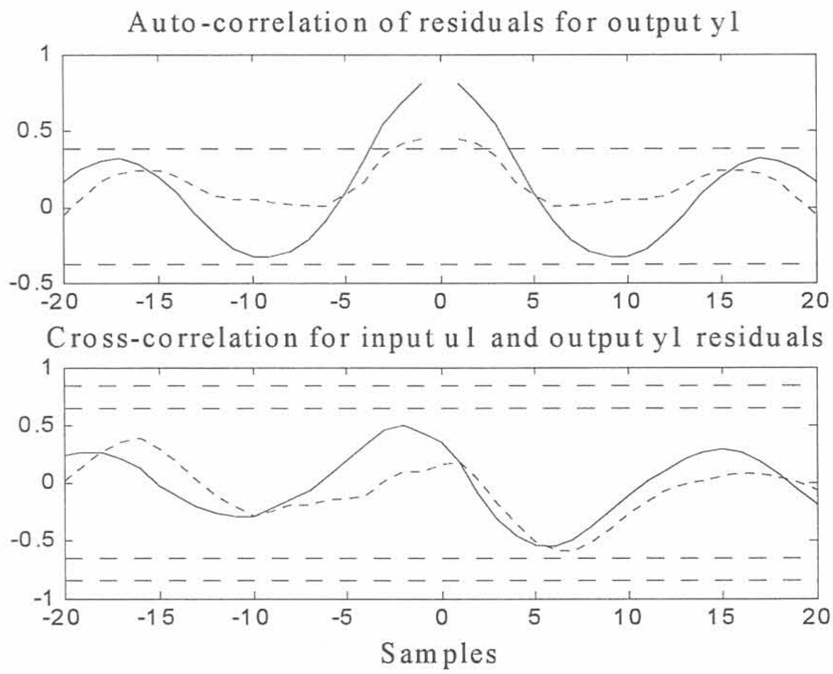

Figure 5.16: The cross-correlation and auto-correlation of the residuals for y_1 and u_1 of the models identified in case 13 (solid line) and case 9 (dotted line).

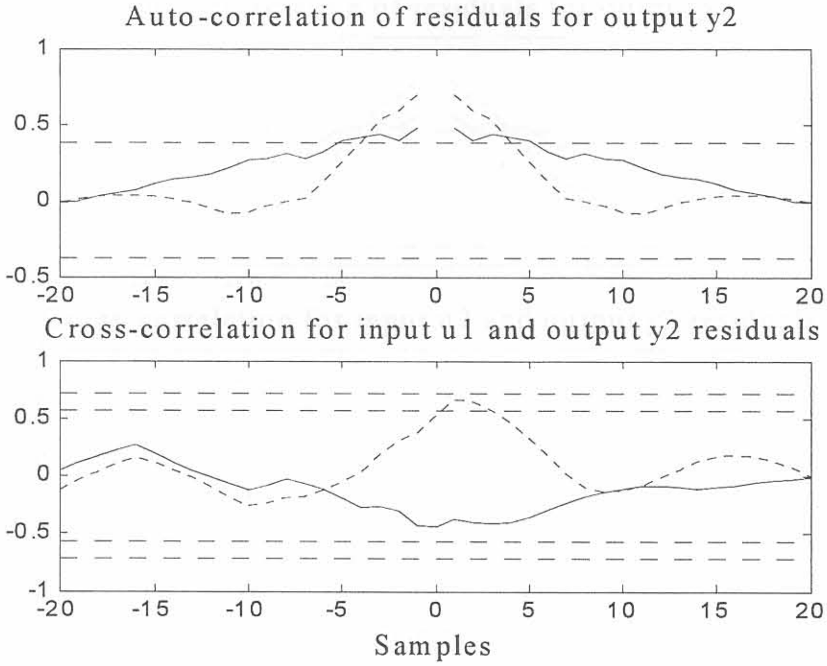

Figure 5.17: The cross-correlation and auto-correlation of the residuals for y_2 and u_1 of the models identified in case 13 (solid line) and case 9 (dotted line).

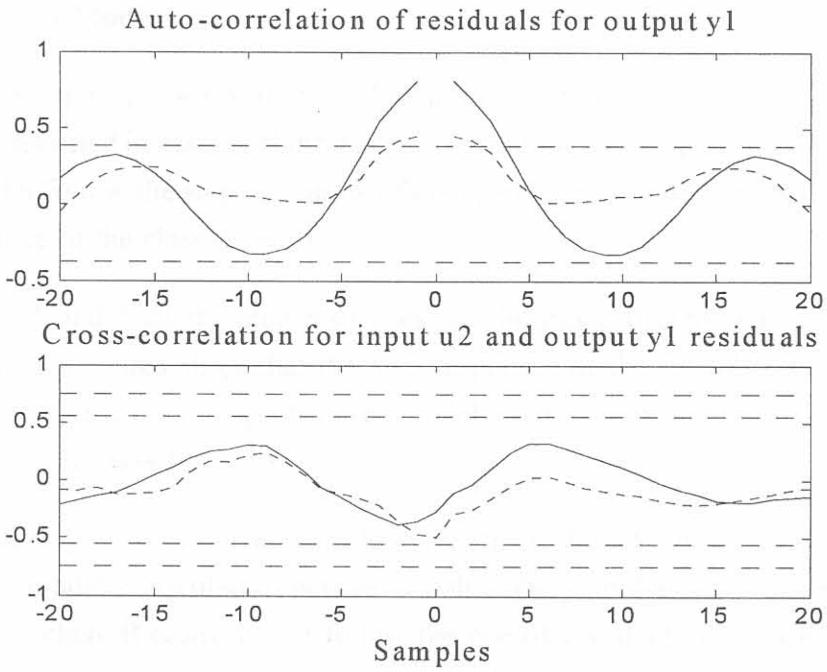

Figure 5.18: The cross-correlation and auto-correlation of the residuals for y_1 and u_2 of the models identified in case 13 (solid line) and case 9 (dotted line).

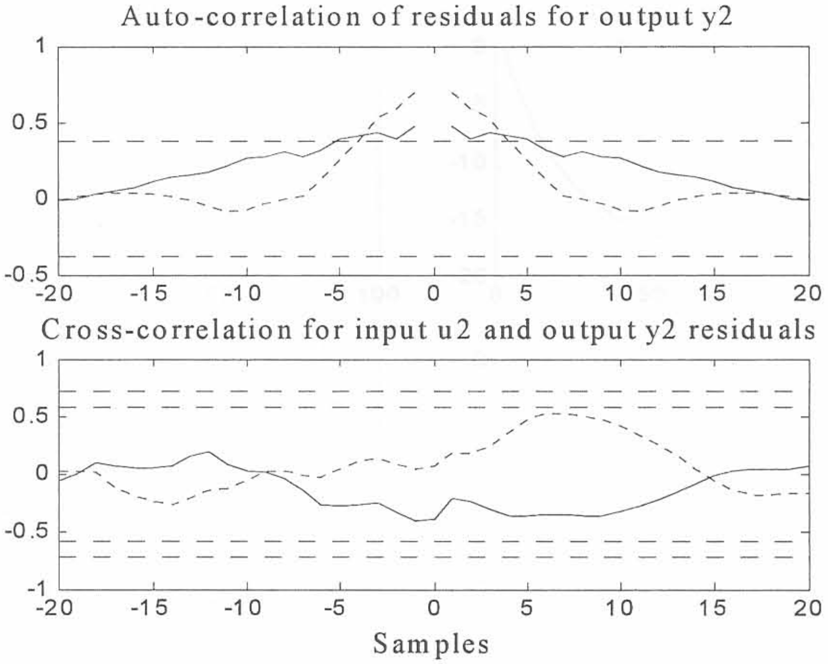

Figure 5.19: The cross-correlation and auto-correlation of the residuals for y_2 and u_2 of the models identified in case 13 (solid line) and case 9 (dotted line).

5.4.2.3 Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model

In Fig. 5.20 the step responses of the model identified from the open-loop data are compared with a model identified in case 3. This figure shows that the step responses of the closed-loop identified model follow the step responses of the open-loop identified model closely. These are representative of the **class A** results.

In Figs. 5.21 and 5.22 the step responses for the models identified in cases 8 and 9 are shown. These figures show that the step responses of the models identified in these cases do not follow the open-loop identified model's step responses in Fig. 5.20. These are representative of the **class B** results.

The impulse responses of the models identified in the **class A** cases also follow the open-loop identified model's impulse responses closely. The impulse responses of the models identified in the **class B** cases do not follow the open-loop identified model's impulse responses.

In Figs. 5.23, 5.24, 5.25 and 5.26 the model identified from open-loop data and the model identified in case 6 are compared in the frequency domain. The frequency responses of the

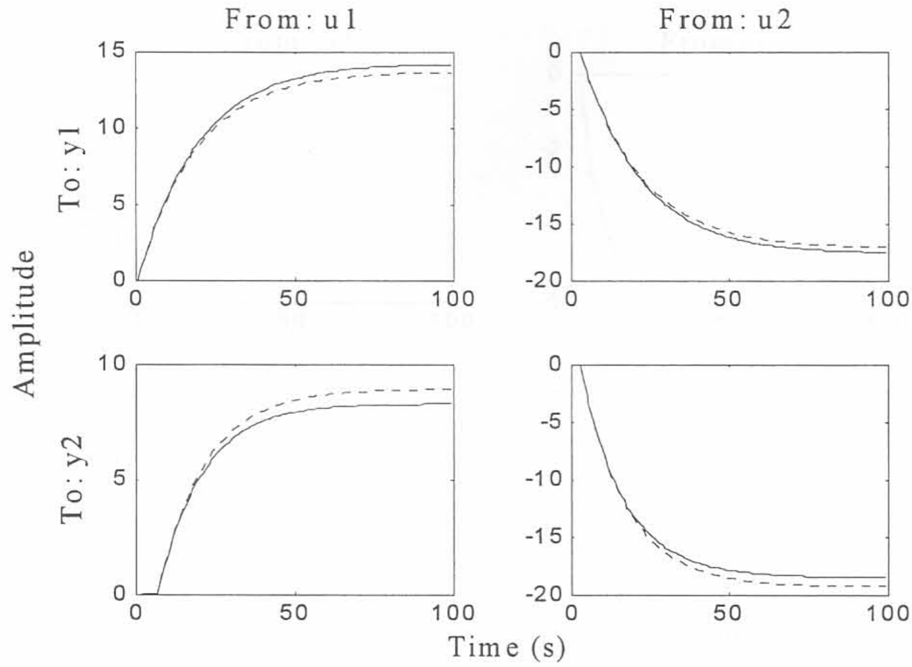

Figure 5.20: The step responses for the model identified from open-loop data (solid line) and the model identified in case 3 (dotted line).

closed-loop identified model agree very well with that of the open-loop identified model in the low, high and cross-over frequency regions. These are representative of the **class A** results.

In Figs. 5.27, 5.28, 5.29 and 5.30 the frequency responses of the open-loop identified model is compared with that of the model identified in case 13. The plots show that this closed-loop identified model does not compare very well with the open-loop identified model in the frequency domain. This is also the case for the other **class B** models.

5.4.2.4 Simulation and Prediction Fit Comparison with the Open-Loop Identified Model

The computed numerical values of fit for the models identified in all the different cases are compared with the numerical value of fit of the model identified from open-loop data in Figs. 5.8, 5.9, 5.10 and 5.11. The open-loop identified model is indicated with the label *open*. The graphs show that the percentages of fit in all the **class A** cases are comparable to the open-loop identified model's percentages of fit, while the percentages of fit in all the **class B** cases are much smaller.

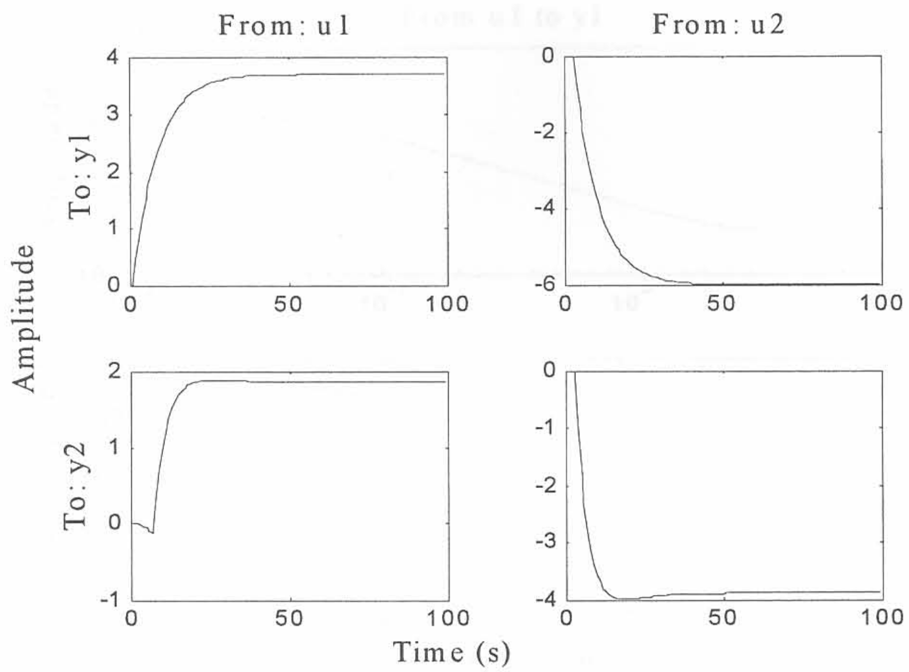

Figure 5.21: The step responses for the model identified in case 8.

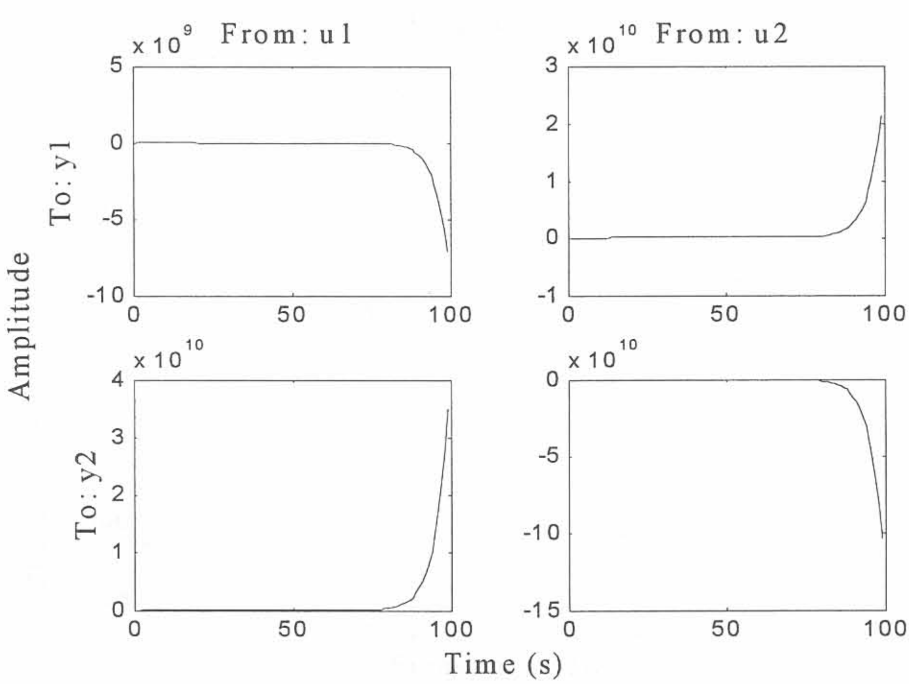

Figure 5.22: The step responses for the model identified in case 9.

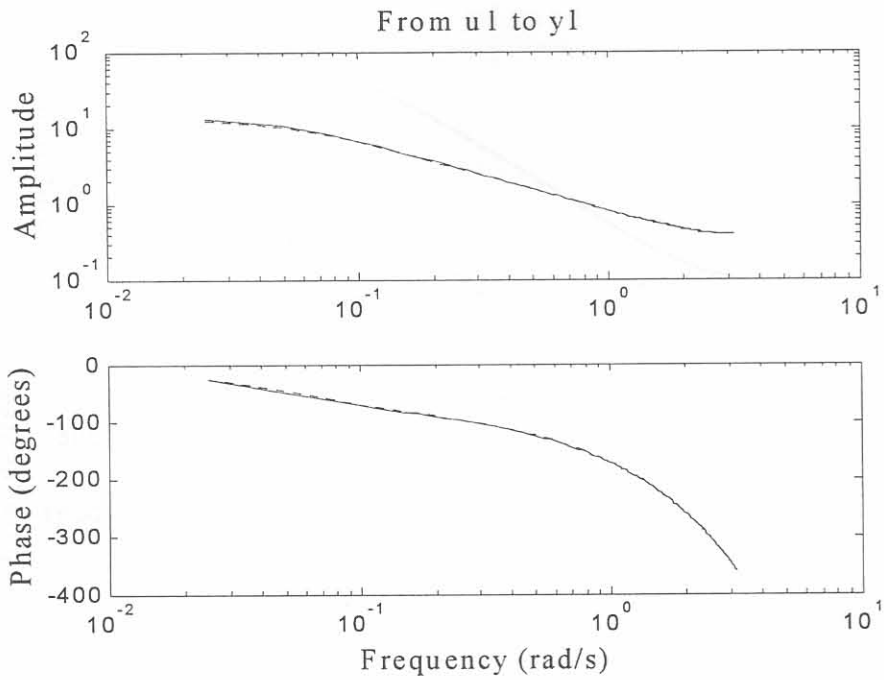

Figure 5.23: The bode plots for the model identified from open-loop data (solid line) and of the model identified in case 6 (dotted line).

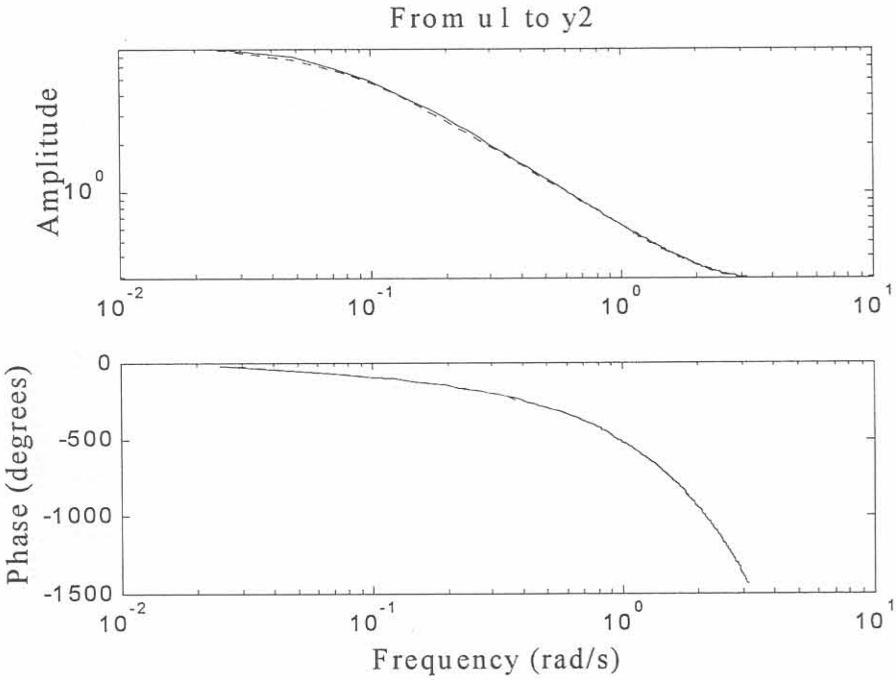

Figure 5.24: The bode plots for the model identified from open-loop data (solid line) and of the model identified in case 6 (dotted line).

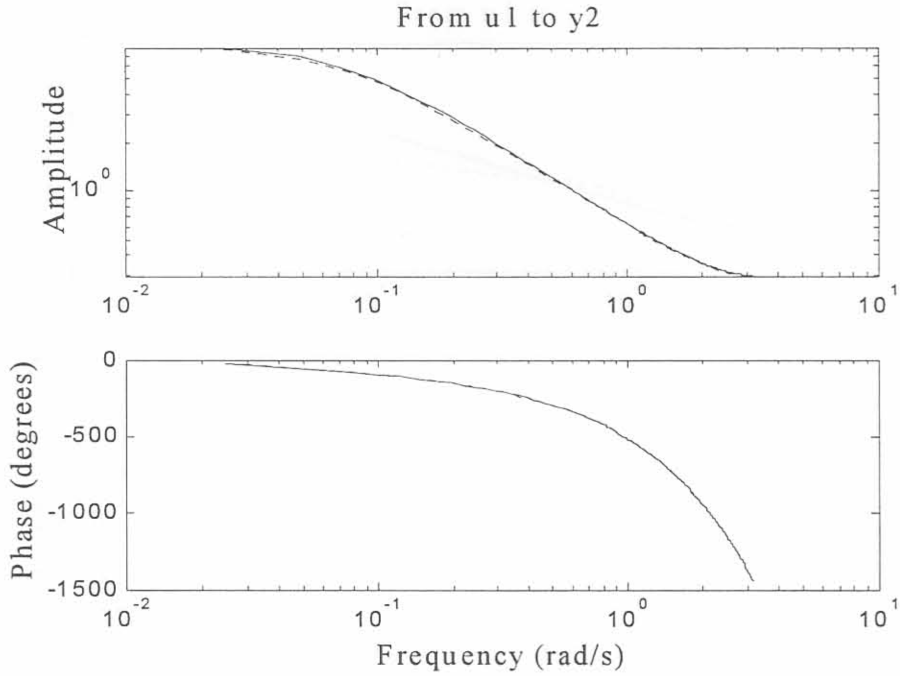

Figure 5.25: The bode plots for the model identified from open-loop data (solid line) and of the model identified in case 6 (dotted line).

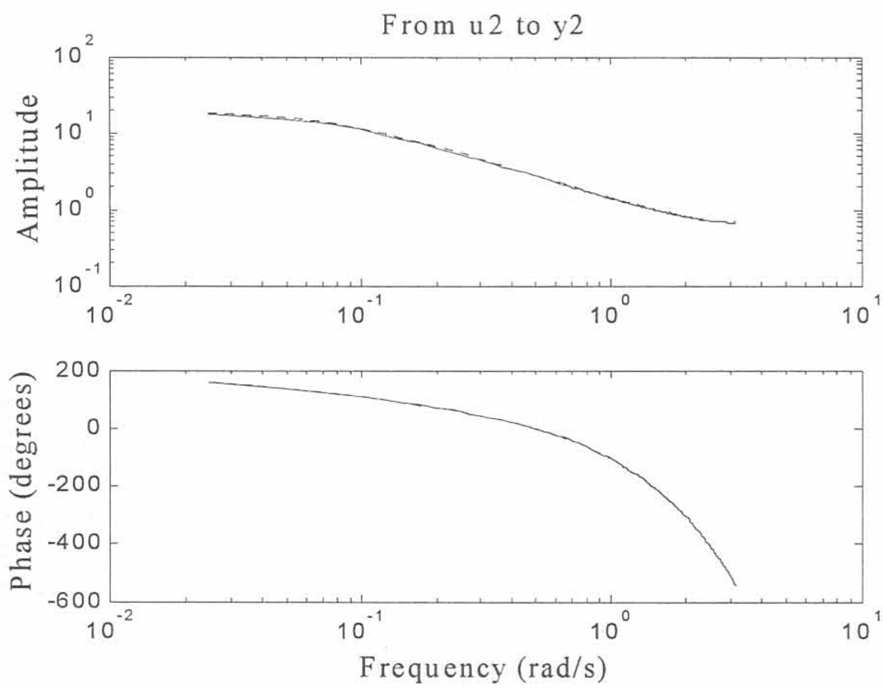

Figure 5.26: The bode plots for the model identified from open-loop data (solid line) and of the model identified in case 6 (dotted line).

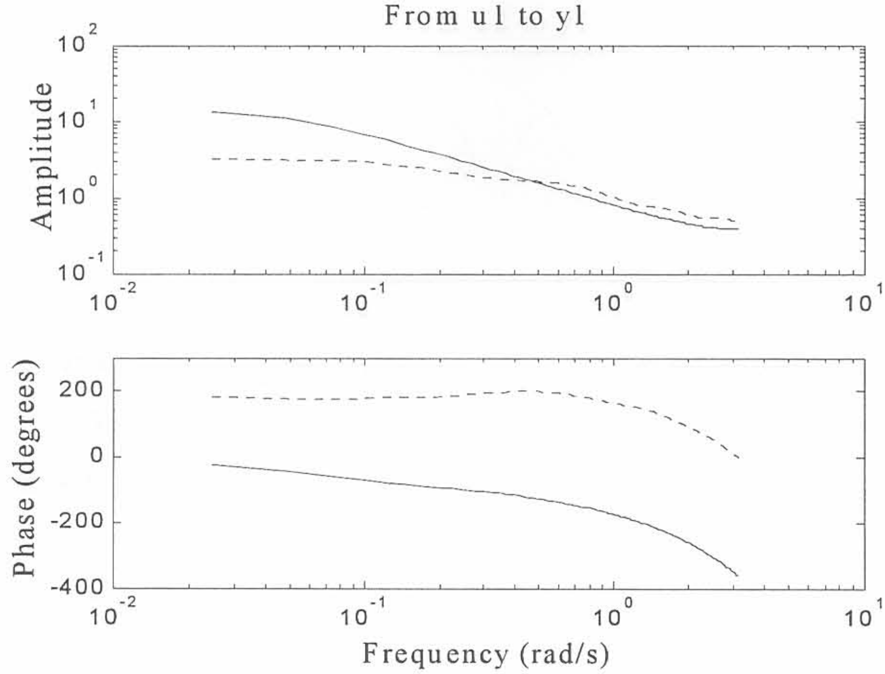

Figure 5.27: The bode plots for the model identified from open-loop data (solid line) and of the model identified in case 13 (dotted line).

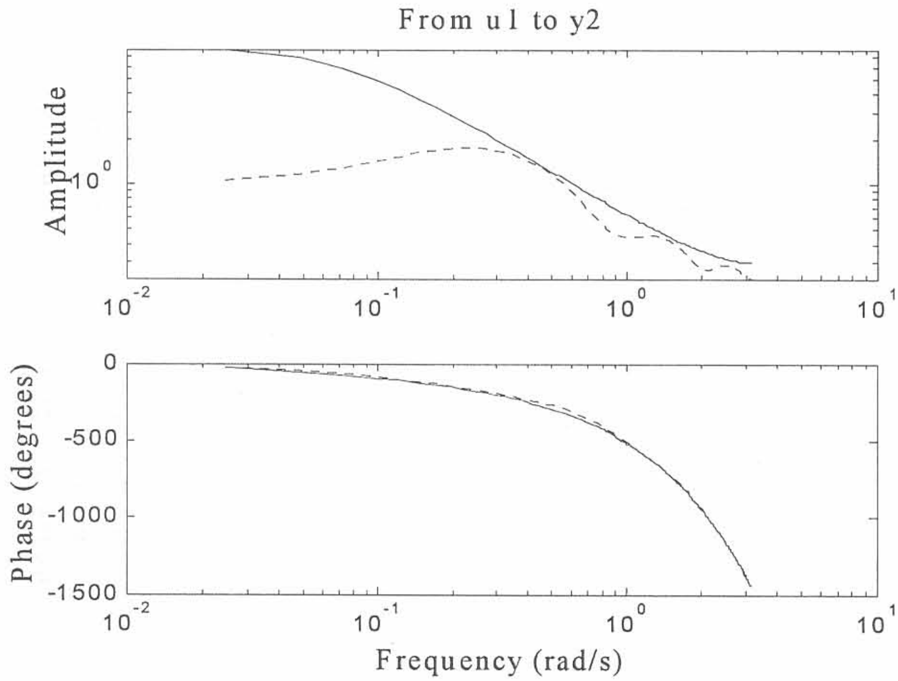

Figure 5.28: The bode plots for the model identified from open-loop data (solid line) and of the model identified in case 13 (dotted line).

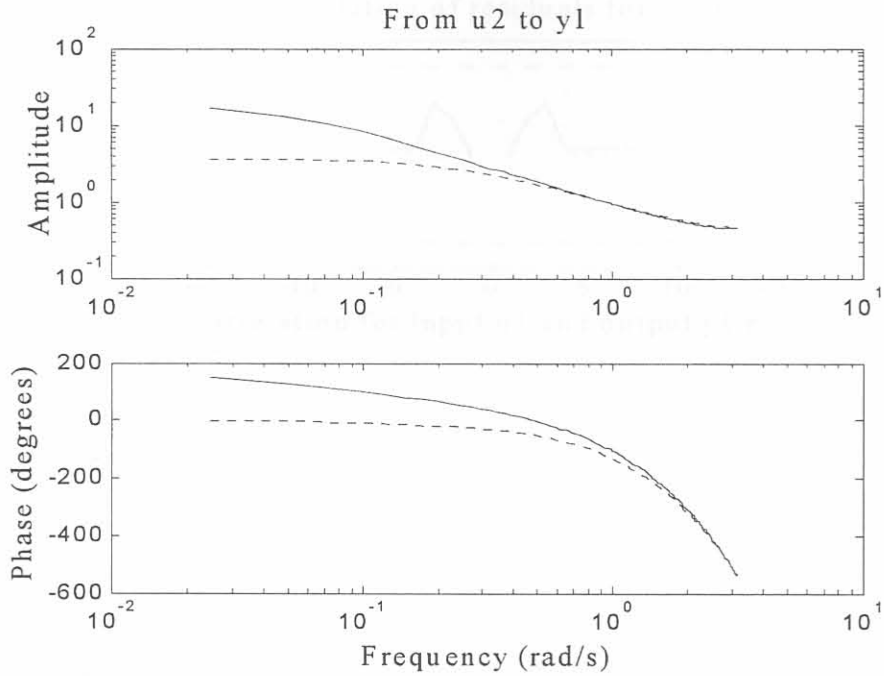

Figure 5.29: The bode plots for the model identified from open-loop data (solid line) and of the model identified in case 13.

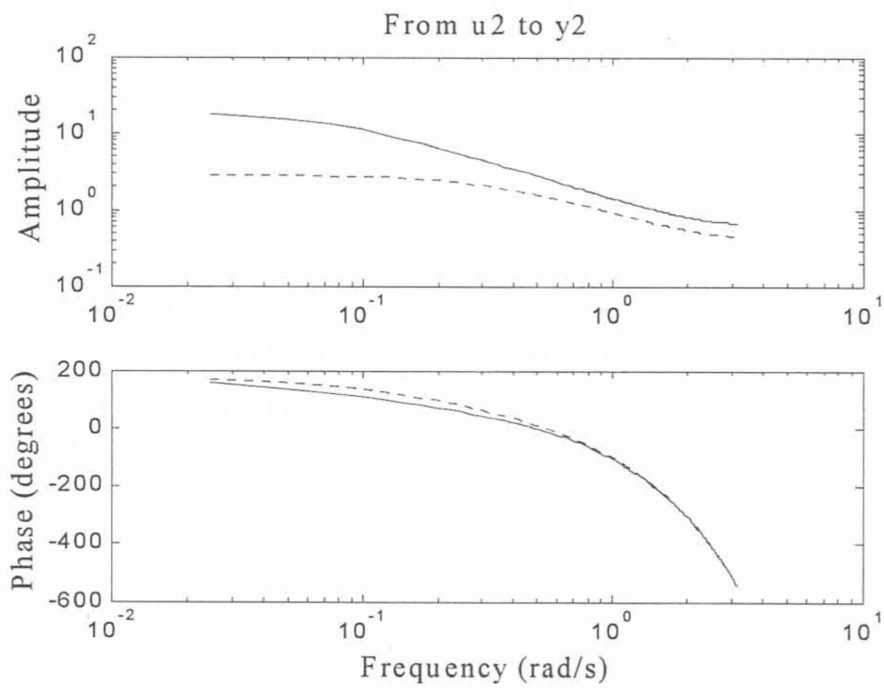

Figure 5.30: The bode plots for the model identified from open-loop data (solid line) and of the model identified in case 13 (dotted line).

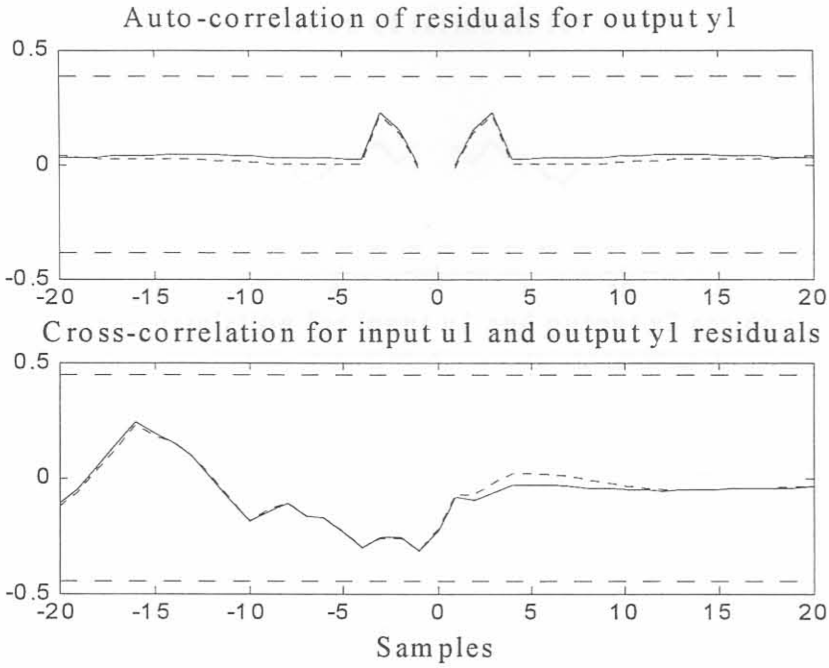

Figure 5.31: Comparison of the cross-correlation and auto-correlation of the residuals for y_1 and u_1 of the models identified from the open-loop data (solid line) and case 6 (dotted line).

5.4.2.5 Residual Comparison with the Open-Loop Identified Model

In Figs. 5.31, 5.32, 5.33 and 5.34 the auto-correlation and cross-correlation of the residuals for the model identified from the open-loop data and the models identified from the closed-loop data in case 6 are compared. The figures show that for the **class A** models and the open-loop identified model, these functions are comparable.

The auto-correlation and cross-correlation functions of the open-loop identified model and the functions obtained in the **class B** cases, shown in Figs. 5.16, 5.17, 5.18 and 5.19, can also be compared. This comparison shows that these functions for the **class B** models are not comparable to the open-loop identified model's functions.

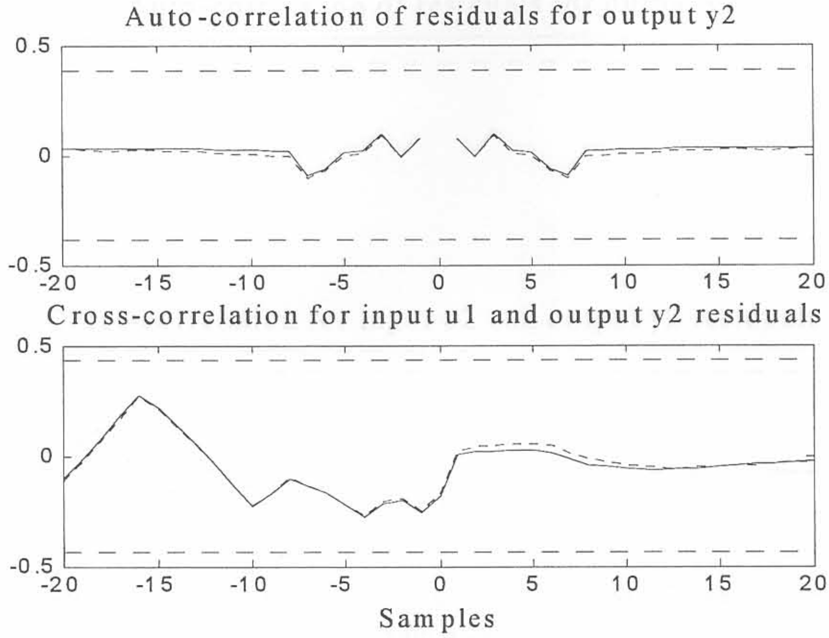

Figure 5.32: Comparison of the cross-correlation and auto-correlation of the residuals for y_2 and u_1 of the models identified from the open-loop data (solid line) and case 6 (dotted line).

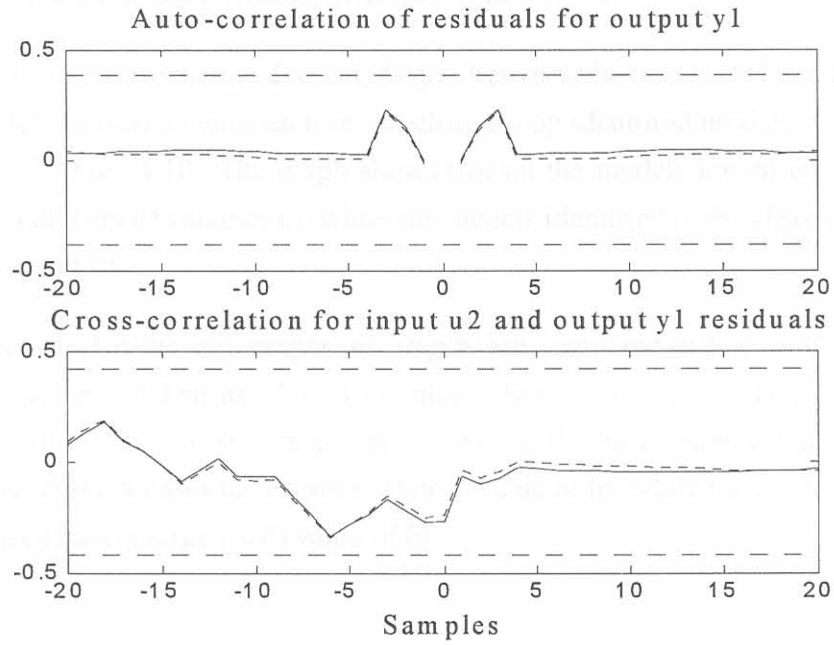

Figure 5.33: Comparison of the cross-correlation and auto-correlation of the residuals for y_1 and u_2 of the models identified from the open-loop data (solid line) and case 6 (dotted line).

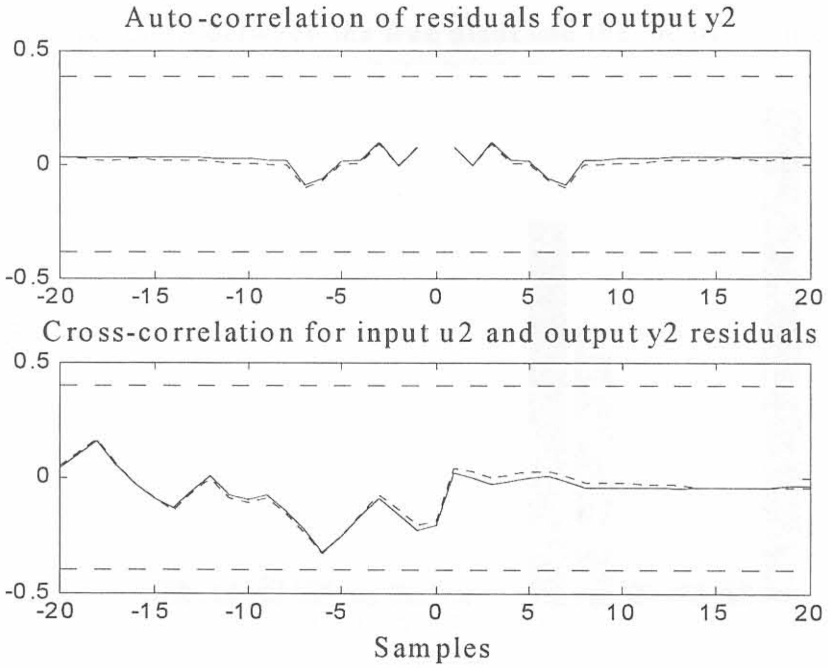

Figure 5.34: Comparison of the cross-correlation and auto-correlation of the residuals for y_2 and u_2 of the models identified from the open-loop data (solid line) and case 6 (dotted line).

5.4.2.6 Comparison of Both Models with the True Model

The values of fit in the frequency domain, *freqfit*, between the true model and the open-loop identified model, as well as with each of the closed-loop identified models, were computed and are plotted in Fig. 4.10. The graph shows that all the models identified in the **class A** cases have a small (good) value of fit, while the models identified in the **class B** cases have a large (bad) value of fit.

The values of fit for the step responses, *stepfit*, are compared in Fig. 5.36. In Fig. 5.36 the maximum value is taken as 150. The values obtained in cases 9 and 13 are actually 3.7944×10^{14} and 1.0918×10^{22} respectively. Again, the graph shows that all the models identified in the **class A** cases have a small (good) value of fit, while the models identified in the **class B** cases have a large (bad) value of fit.

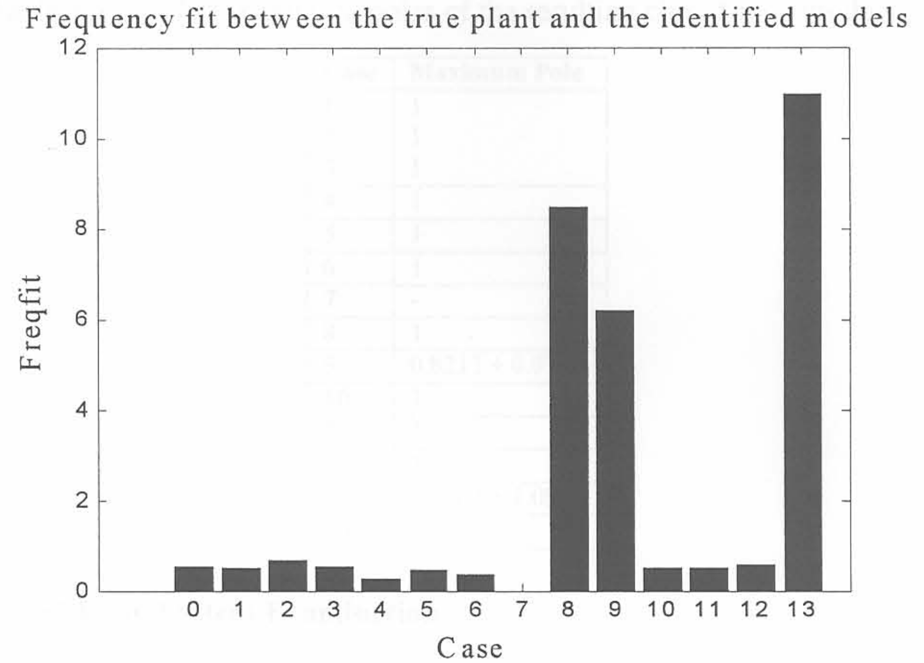

Figure 5.35: The fit in the frequency domain between the true plant and the identified models. Case 0 is the open-loop identified model.

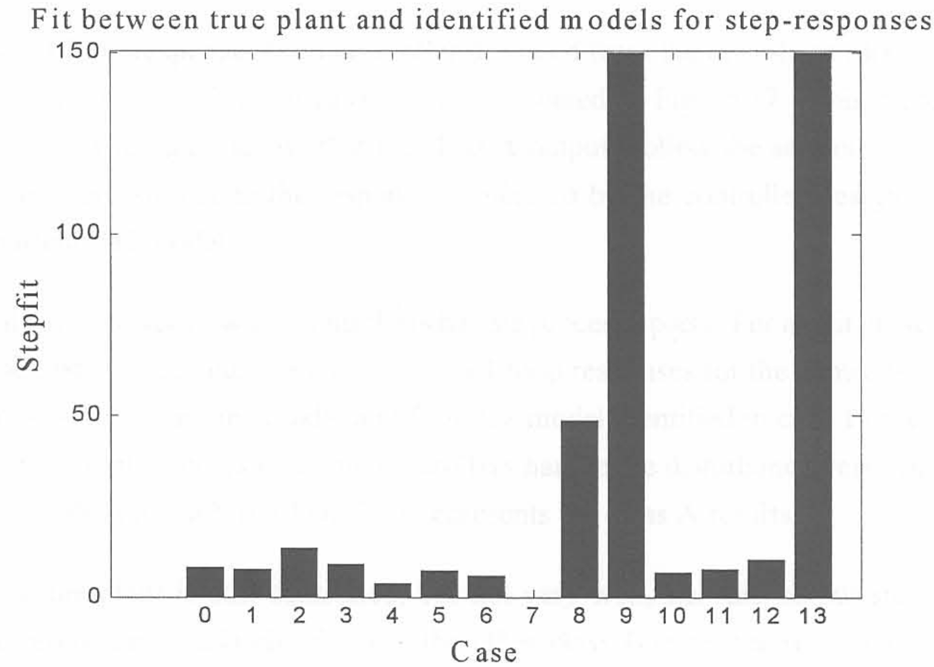

Figure 5.36: The fit in the time domain between the step-responses of the true plant and the identified models. Case 0 is the open-loop identified model.

Table 5.3: The maximum poles of the resulting closed-loop models

Case	Maximum Pole
1	1
2	1
3	1
4	1
5	1
6	1
7	-
8	1
9	0.8213 + 0.9758i
10	1
11	1
12	1
13	0.9462 + 1.0921i
open	1

5.4.2.7 Closed-Loop System Examination

The maximum poles of the resulting closed-loop function are given in Table 5.3. In all the **class A** cases the maximum poles are on the unit circle and the closed-loop systems are thus marginally stable. In two of the **class B** cases the maximum poles are outside the unit circle and the closed-loop systems are thus unstable.

The closed-loop responses for a controller designed from the open-loop identified model and for a controller identified in case 11, are compared in Fig. 5.37. This represents the **class A** results. The figure shows that the **class A** outputs follow the set-points and that the responses are very similar to the responses generated by the controller designed from the open-loop identified model.

The real test is to see how the controllers handle process upsets. For a unit pulse change at $t = 1$ in the disturbance adding to u_1, the closed-loop responses for the controllers designed from the open-loop identified model and from the model identified in case 11 are compared in Fig. 5.38. The plot shows that both controllers handle the disturbance very similarly and the set-point values are still reached. This represents the **class A** results.

In case 8 the plant is still controlled, but not very well, i.e. the output signals do not follow the reference signals closely. For the other **class B** cases the resulting closed-loop responses are also unsatisfactory, because the outputs do not follow the reference signals and the systems become unstable.

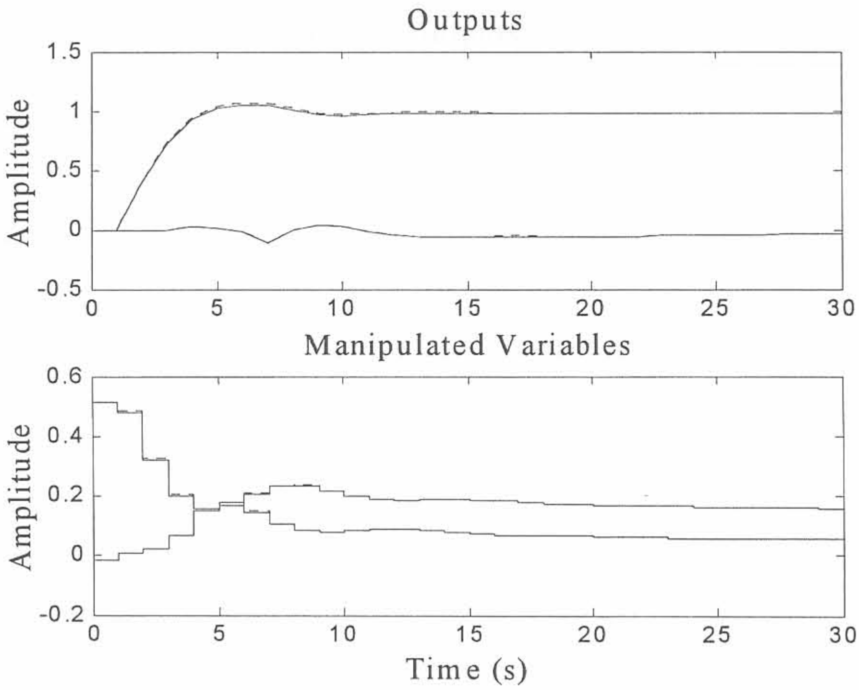

Figure 5.37: The closed-loop response for a controller designed from the open-loop identified model (solid line) and for a controller designed from the model identified in case 11 (dotted line) with $r_1(t) = 1$ and $r_2(t) = 0$.

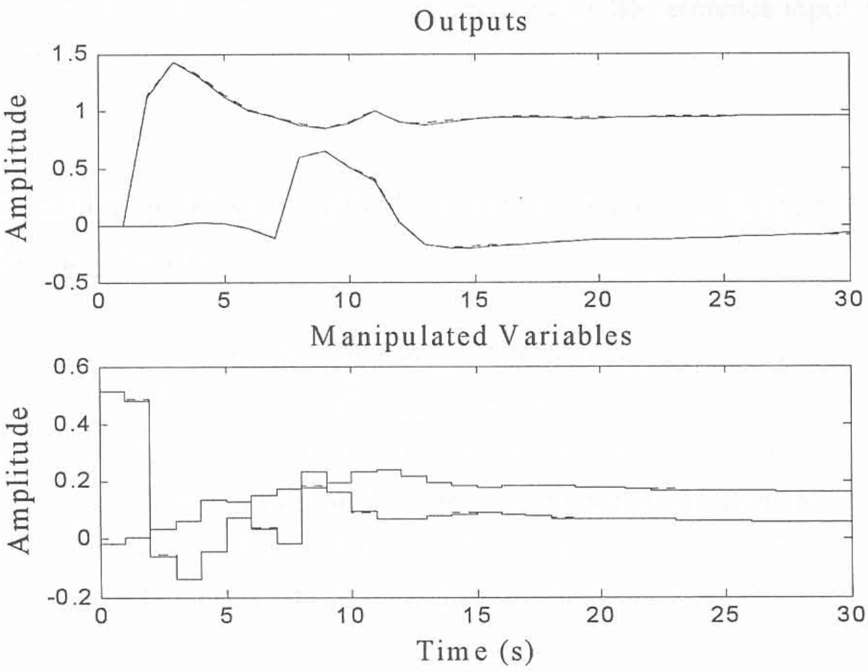

Figure 5.38: The closed-loop response for a controller designed from the open-loop identified model (solid line) and for a controller designed from the model identified in case 11 (dotted line) with $r_1(t) = 1$ and $r_2(t) = 0$. A pulse disturbance was added at $t = 1$ to u_1.

5.4.3 Discussion

As mentioned, two trials were run with different test signals. The PRBS signals gave slightly better results than the step inputs and therefore these results are shown. A possible cause for this slight difference can be that the step input is not PE enough. For a step to be PE, the step should be preceded by zeros for a length of time approximately equal to the impulse response of the true system [36]. In the first trial, $r_1(t)$ was stepped at time zero and therefore did not satisfy this condition.

The results obtained in trial 2 (persistently exciting PRBS reference inputs) are now further discussed.

5.4.3.1 Simulation and Prediction Analysis and Comparison with the Open-Loop Identified Model

In all the cases where structured tests were performed, the identified models are capable of reproducing the validation data satisfactorily. The percentages of fit obtained for the open-loop identified model are also comparable with these values. Thus, according to the simulation and prediction analysis, all these models are satisfactory and also comparable to the open-loop identified model. This result indicates that the proposed SID methodology works irrespective of the disturbances and constraints that were used.

In the case where a nonlinear controller was used, a good percentage of fit results only when there were no added disturbance as in case 12. This model is thus also satisfactory. Case 7 is similar to case 12, except that the controller is linear. In this case no model could be identified from the zero input and output signals that resulted. These results confirm the fact that nonlinear feedback ensures identifiability.

Poor results are obtain in the other cases where the reference inputs were zero (**class B**). The percentages of fit are much lower than the percentages of fit obtained for the open-loop identified model. In case 8 this can be contributed to the fact that the system is not identifiable. However, in case 9 where output inter-sampling was used and in case 13 where a nonlinear controller was used, the systems were identifiable. Here imprecise models still resulted because of the bad SNR.

5.4.3.2 Residual Analysis and Comparison with the Open-Loop Identified Model

The auto-correlation and cross-correlation of the errors with the outputs, indicate that the closed-loop identified models in all the structured test cases, as well as in the zero reference case with the nonlinear controller and good SNRs (case 12), describe the plant accurately. In these cases the auto-correlation functions are between the confidence bounds, which mean the errors are white and the models unbiased. A visual inspection of the errors confirms this result. The cross-correlation functions are also between the confidence bounds, which indicate that the error signals and u_i are independent. These results are similar to that of the model identified from the open-loop data. Again, this indicates that the proposed SID methodology works irrespective of the disturbances and constraints that were used and it confirms that nonlinear feedback ensures identifiability.

The residual analysis also indicates that the other models, identified in the cases when $r_i(t) = 0$, are deficient, since these cross-correlation functions and the auto-correlation functions go outside the confidence bounds. This confirms that the other methods that ensure identifiability do not necessarily deliver accurate models.

5.4.3.3 Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model

Since, in the structured test cases, the step responses of the closed-loop identified models and the open-loop identified model agree very well, it can be concluded that these models are approximately equal in the time domain. Thus, the proposed SID methodology deliver satisfactory models. Also the model identified in case 12 and the open-loop identified model compare well in the time domain, which again confirms the fact that nonlinear feedback ensures identifiability.

The comparison in the frequency domain shows that all the closed-loop models identified from the structured tests and case 12 are equal in accuracy in the low, high and cross-over frequency regions to the open-loop identified model. Again, this indicates that the proposed SID methodology works irrespective of the disturbances and constraints that were used and it confirms that nonlinear feedback ensures identifiability.

The other models identified with no PE reference signal present, did not agree very well with the open-loop identified model in either the time or frequency domain. Therefore, these models do not give the correct description of the plant. This confirms that the other methods that ensure identifiability do not necessarily deliver accurate models.

5.4.3.4 Comparison of Both Models with the True Model

The computed values of fit between the identified models and the true model, show that models estimated from the structured tests closed-loop data represent the true model just as well as the open-loop identified model. Thus, the proposed SID methodology delivers satisfactory models.

The computed values of fit, between the identified model in case 12 and the true model, also shows that this closed-loop estimated model represents the true model just as well as the open-loop identified model. Again, it confirms that nonlinear feedback ensures identifiability and that good SNRs are necessary.

These values of fit are much worse (larger) for the models identified in the other cases where structured tests were not performed. From this it is apparent that the other models, identified from experimental data with no reference signals present, are not comparable with the open-loop identified model and are also very different from the true plant. Again, this confirms that the other methods that ensure identifiability do not necessarily deliver accurate models.

5.4.3.5 Closed-Loop System Examination

The closed-loop responses show that the closed-loop models identified from the structured test data, as well as the model identified in case 12, are good enough for their purpose, namely closed-loop MPC control. In these cases the resulting closed-loop systems are stable. These responses also show that the open-loop and closed-loop identified models result in very similar controllers. Again, this indicates that the proposed SID methodology works irrespective of disturbances and constraints that were used and it confirms that nonlinear feedback ensures identifiability.

In case 8, where $r_i(t) = 0$, the plant is still controlled, but not very well. This can be due to the fact that the model may still be relatively accurate in the cross-over frequency region and therefore still results in a stable closed-loop system. For the other two cases with $r_i(t) = 0$, the resulting closed-loop responses are unsatisfactory, because, firstly, the outputs do not follow the reference signals and, secondly, the closed-loop systems are unstable. Again, this confirms that the other methods that ensure identifiability do not necessarily deliver accurate models.

5.4.3.6 Synopsis

As expected, irrespective of the disturbances and constraints that were used, satisfactory models are obtained in all the cases where structured tests, which ensured PE reference signals and good SNRs, were performed. Thus, the proposed SID methodology works for the type of disturbances and constraints used.

There is also no significant bias in the models identified from the structured test data. The reason can be that the noise model is good and that the SNRs were also acceptable. Furthermore, there is also no significant bad fit in the low frequency regions, which may sometimes be expected from ARX models.

There is also no significant difference between the models identified from the open-loop data and the closed-loop structured test data and therefore the plant probably did not exhibit very different dynamics in closed-loop than in open-loop [6]. For this reason these closed-loop identified models produced equivalent controllers to the one produced by the open-loop identified model.

Then, in the cases where the reference signals were zero, unsatisfactory models were obtained. For cases 7 and 8 the reason is that the data were not informative enough. In case 9, where the inter-sampling method ensured identifiability, an imprecise model was identified, because it is shown in Section 4.4 that with $r_i(t) = 0$ a large variance will result. The only exception, as expected, is case 12 with the constrained controller and no disturbance, since in this case the nonlinearities in the controller ensured identifiability and the good SNR ensured a small variance in the model. The changes in the input and output signals were very small. Therefore, when the disturbance was added in case 13, the SNR became unacceptable and a large variance resulted in the identified model. With the reference signals zero, the SNRs are only good for *very* small disturbances.

5.5 CONCLUSION

From these simulation results, it can be concluded that the proposed closed-loop system identification methodology gives reliable results, i.e. accurate and precise models, for MIMO plants controlled by MPC controllers, for the type of system disturbances and constraints that were used, as long as the reference signals are PE and the SNRs (ratios between noise and plant input signals) are good.

When PE reference signals are used, the proposed closed-loop SID methodology and open-loop SID method deliver comparable identification results. When the reference signals are not PE, or when they result in bad SNRs, the methodology should be reconsidered.

Other methods that ensure identifiability, e.g. inter-sampling and nonlinear controllers, also do not guarantee precise models if the SNR is not good, which can happen when no structured tests are performed. Structured tests should be conducted to ensure good SNRs, and the easiest way to ensure identifiability is to make the reference signals PE.

These validation results hold for the ideal case, at least. In the next chapter the implementation of the methodology on real process data is discussed.

CHAPTER 6

VALIDATION AND EVALUATION OF THE METHODOLOGY WITH REAL PROCESS DATA

6.1 INTRODUCTION

The proposed methodology was implemented on real closed-loop process data to try and identify a part of a reactor in the MIBK process. The process description of the MIBK plant can be found in Chapter 2.

In Section 6.2 the implementation and validation procedure is described in terms of each of the identification steps. This procedure is very similar to the procedure followed in the simulation, see Section 5.3. In Section 6.3 the expected validation results are discussed and the final results obtained are then given and also discussed. In Section 6.4 the implemented methodology is thoroughly evaluated and further recommendations, with regards to successful implementation of this methodology, are made.

Finally, it is concluded in Section 6.5 that, although it is possible to identify satisfactory models from closed-loop data for controller design, structured tests that ensure good SNR and PE reference signals should still be performed.

6.2 IMPLEMENTATION OF THE METHODOLOGY

The proposed methodology, as described in Section 4.8, was again implemented in MATLAB. This implementation is briefly discussed in terms of the five SID steps, as well as the methodology validation step. The similarities to Section 5.3 are mentioned and the differences are emphasized.

6.2.1 Experiment Design

Signals to be Measured: The measured CVs, $y_i(t)$, and MVs, $u_i(t)$, of a part of reactor A for October and November 1998 were obtained. These signals are described in Section 2.5.1 and the signal tags and labels are given in Table 6.1.

Table 6.1: Signal tags and labels.

Variable No.	Tag	Label
1	05TIC280A.MV	y_1
2	05TIC232A.MV	y_2
3	05TIC279A.MV	y_3
4	05FIC223A.MV	y_4
5	05TIC280A.SV	u_1
6	05TIC232A.SV	u_2
7	05TIC279A.SV	u_3
8	05FIC223A.SV	u_4

Sampling Time: The plant data were sampled every 30s, which is twice as fast as the controller execution time of 1min. Therefore, the plant input and output data sets were inter-sampled by a factor of two. In MATLAB the data sets were also resampled, making use of the *resample* function, to obtain a sampling time of 1min. The plant was identified from these data sets, as well as from the inter-sampled data sets. Thus, the influence of the inter-sampling could also be seen.

Excitation Signals: No structured closed-loop tests were allowed on the plant and the reference (set-point) values were also not logged and could thus not be retrieved. Therefore, it is not known wether these signals were PE.

6.2.2 Data Collection

Collection: The desired data sets were retrieved from a database on which the relevant data sets are stored.

Preprocessing: No high frequency disturbances or bursts and outliers in the data record were observed. There were, however, some missing data sets and at some stages the data records were non-continuous. A routine was written to search for the sections in the data records where the data sets were not sampled every 30s. These sections were not considered for estimation or validation.

Again, the trends in the data sets were removed, making use of the MATLAB function *dtrend.*

The continuous data sets were visually inspected to select the ranges where the MVs were the most excited. Models were estimated from different sections in the data. For estimation the range that gave the best results is 9 November 1998 13:00:00 - 10 November

1998 17:12:00. The models were validated with different data ranges. The validation results for the data range from 10 November 1998 17:13:30 - 11 November 1998 15:16:00 are shown in Section 6.3.

Time Delay: From visual inspection of the data, as well as from the known open-loop identified models, it was concluded that all the time delays were for practical purposes approximately zero. Just to make sure of these values, second-order models with different delays were compared by making use of the estimation data set and the *compare* function. The best fit selected the delays. From these tests, very small time delays (relative to the sampling time) of 30s were selected.

6.2.3 Model Structure Selection

Type of Structure: Again, the multivariable ARX type model structure was used.

Order Selection: The old step-response models obtained from the open-loop tests were examined. Since these step response models are nonparametric with 180 parameters each, they did not supply much information regarding a suitable model structure. The state-space models obtained from these step response models are of a very high order and therefore did not give a good indication of the appropriate model order. Therefore, the multivariable ARX models were fitted to the estimation data set for different model structure orders, starting at a low order. For each of these models, the sum of squared prediction errors were computed with the *compare* function, as they were applied to the estimation data set. The percentage of fit did not improve significantly from the second-order model onwards (up to an eight-order model was tested). Therefore, a second-order model was chosen.

Since only single-output data sets are handled by the *arxstruc* and *selstruc* routines, these functions could not be used in the structure selection.

6.2.4 Model Estimation

Again, similar to the simulation in Chapter 5, the *idarx* command, which uses the LSE PEM estimation method, was used to fit the chosen models to the estimation data. The model was also transformed from discrete-time to continuous-time.

6.2.5 Model Validation

Simulation and Prediction: As in the simulation, the pure simulated and 6-step ahead predicted outputs, as well as the percentage of fit for the identified models were computed with the *compare* command. The prediction horizon of the real controller was not known, therefore, an arbitrary horizon of 6 steps was assumed.

Residual Analysis: Again, the *resid* command was used to calculate and display the auto-correlation function of the residuals, as well as the cross-correlation between the residuals and the plant inputs.

Model Reduction: The *minreal* function can be used to cancel pole-zero pairs in a model transfer function and thus to reduce the model order if necessary. This function was used to try and cancel pole-zero pairs in the identified model. There were no pole-zero pairs to cancel and the model order is thus not too high.

6.2.6 Methodology Validation

Preparation of the Open-Loop Identified Models: The AspenTech DMCplus model file of the open-loop identified MIBK plant was imported into an EXCEL spreadsheet. The model is in a step response format, with each of the SISO transfer functions made up of 180 step response coefficients. These data sets were sorted and the step response coefficients of the relevant SISO transfer functions were imported into MATLAB. In MATLAB a function was written to construct state-space models from the step response coefficients, making use of singular value decomposition (SVD). This function is described in Addendum B. SISO models with orders between 19 and 59 were obtained. The state-space transfer functions could and were then reduced to fourth-order models. The SISO step responses of this model are shown in Fig. 6.1. This figure shows that the resulting open-loop identified model of the selected part of reactor A consists of ten non-zero SISO transfer functions and that the transfer functions from u_1 to y_2, y_3 and y_4, from u_2 to y_4, from u_3 to y_4 and from u_4 to y_2 are zero.

The closed-loop identified model was compared with both the full order and the reduced order models. Since the results are very similar, the validation results obtained for the fourth-order model are shown in this chapter.

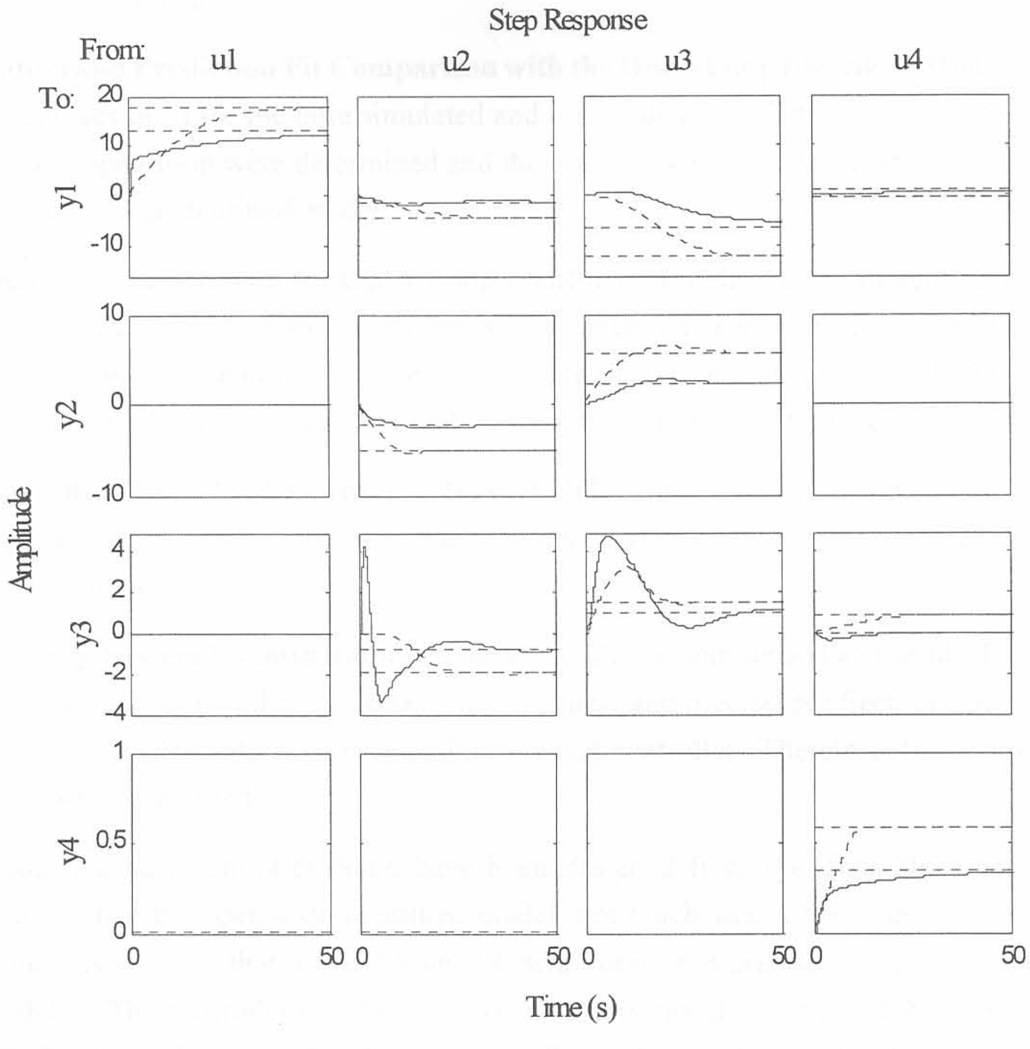

Figure 6.1: The SISO step responses of the open-loop identified (dotted lines) and closed-loop identified (solid lines) models. The horizontal dotted lines are the steady sate values of these responses.

Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model: Again, the *step* command was used to plot the step responses and the *impulse* command was used to plot the impulse responses for each of the SISO transfer functions of the models. The *bode* command was also used to plot both the amplitude and phase for each of the SISO transfer functions of the models. The plots for the open-loop and closed-loop identified models are visually compared.

Simulation and Prediction Fit Comparison with the Open-Loop Identified Model: Again, the percentages of fit for the pure simulated and 6-step ahead predicted output of the model identified in open-loop were determined and these values are compared with those obtained for the closed-loop identified model.

Residual Comparison with the Open-Loop Identified Model: Again, the auto-correlation function of the residuals, as well as the cross-correlation between the residuals and the plant inputs for the model identified from open-loop data were computed and these functions are visually compared with the ones obtained for the closed-loop identified model.

Comparison of Both Models with the True Model: Since this is not a simulation and the "true" plant model is thus not known, the identified models could not be compared with the "true" plant model.

Closed-Loop System Examination: From the simulation comparison and residual analysis the open-loop model could not be accepted as accurate and precise, see Section 6.3.3. There was also no available information regarding the real controller. Therefore, the closed-loop system was not examined.

Although a new controller could have been designed from the closed-loop identified model to control the open-loop identified model, not much would have been gained from this. The reason being that it is not known how accurate and precise the open-loop identified model is. The controller may be able to control this model, but it is not the "true" plant. Also, it should be determined if the real controller, redesigned from the closed-loop identified model for the original design parameters, is able to control the plant. This cannot be determined, since the controller parameters are not known and the real controller can thus not be modelled.

6.3 Validation Results

In this section the expected validation results are discussed, the obtained results are shown and finally, the obtained results are discussed and compared with the expected results.

6.3.1 Expected Results

In Chapter 5 it is concluded that the proposed closed-loop SID methodology and open-loop SID method deliver comparable identification results only when structured tests, which ensure PE reference signals and good SNRs, are performed. Therefore, it was expected that an unsatisfactory model, which is not a good description of the plant, would be identified, because structured tests were not performed and the reference inputs were thus probably not PE.

It is also concluded in Chapter 5 that when the reference signals are not PE, or result in bad SNRs for the plant input signals, the methodology should be reconsidered, since, in general, methods that ensure identifiability, e.g. inter-sampling and nonlinear controllers, still do not guarantee satisfactory models with small variances. Therefore, even though the measured data sets were informative, since the data signals were inter-sampled, a large variance in the identified model was still expected.

It was also expected that even if satisfactory structured tests were performed, with the noise model of the ARX structure not an accurate description of the true noise model, the identified model would contain a bias. A possible bad fit in the low frequency regions was also expected, since the ARX model structure penalises the high-frequency misfit behaviour more than low-frequency misfit behaviour.

Even if a good model were identified from closed-loop, a possible difference between the models identified from open-loop and closed-loop data was still expected, since the frequency weighting for these two types of models are different and the plant may have exhibited closed-loop dynamics different form the open-loop.

6.3.2 Obtained Results

The validation results for the second-order model, estimated from the closed-loop data, measured from 9 November 1998 13:00:00 to 10 November 1998 17:12:00, are given in this section. This model is the most accurate identified model. Models were also estimated from other ranges of the measured data sets. Although there are similar characteristics in these models, the models vary considerably for the step responses, bode plots, percentage of fit, etcetera.

When only the estimation data are used in the validation all the results are good. However, the validation results for other ranges in the measured sets of data are less attractive, as

Table 6.2: Percentage of fit between measured and predicted outputs for chosen validation range.

Signal	# Steps ahead Prediction	Open-loop identified model	Closed-loop identified model (inter-sampled)
y_1	∞	55.27%	32.22%
	6	-	91.28%
y_2	∞	-4.306%	58.41%
	6	-	93.21%
y_3	∞	49.13%	34.09%
	6	-	10.55%
y_4	∞	-39.76%	-71.21%
	6	-	74.28%

shown. The validation results for the data range 10 November 1998 17:13:30 - 11 November 1998 15:16:00 are given in this section. The model was also validated with some other data ranges from October and November. For each of these sets different results were obtained.

For the resampled data, a model was also identified. This model, obtained from data that were not inter-sampled, caused MATLAB to warn that *the matrix is close to singular*. Although this model and the model obtained from the inter-sampled data have similar characteristics, it is worse in accuracy for the step responses, bode plots, percentage of fit, etcetera.

6.3.2.1 Simulation and Prediction Analysis and Comparison with the Open-Loop Identified Model

The percentage of fit between the measured and pure simulated output signals, as well as the 6-step ahead predicted output signals, for the open-loop and closed-loop identified models are compared in Table 6.2. Since the open-loop identified model is in state-space form, the 6-step ahead predicted outputs were not computed. In Figs. 6.2, 6.3 and 6.4 the pure simulated, predicted and measured outputs are also shown. These results show that the open-loop and closed-loop identified model have very low percentages of fit. With the 6-step ahead prediction, this fit improves significantly.

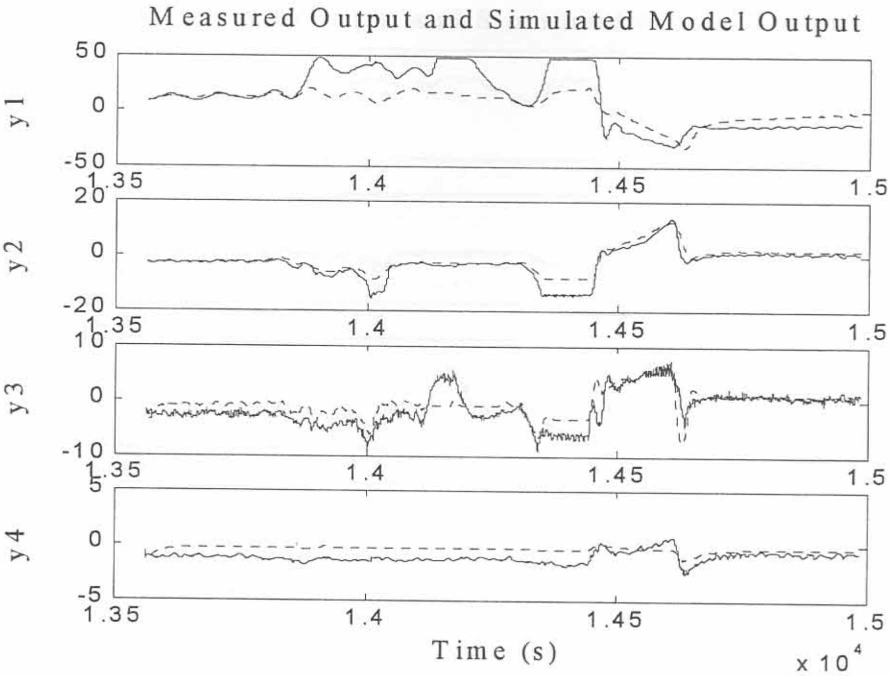

Figure 6.2: A comparison of the measured outputs (solid lines) and the simulated outputs of the closed-loop identified model (dotted lines).

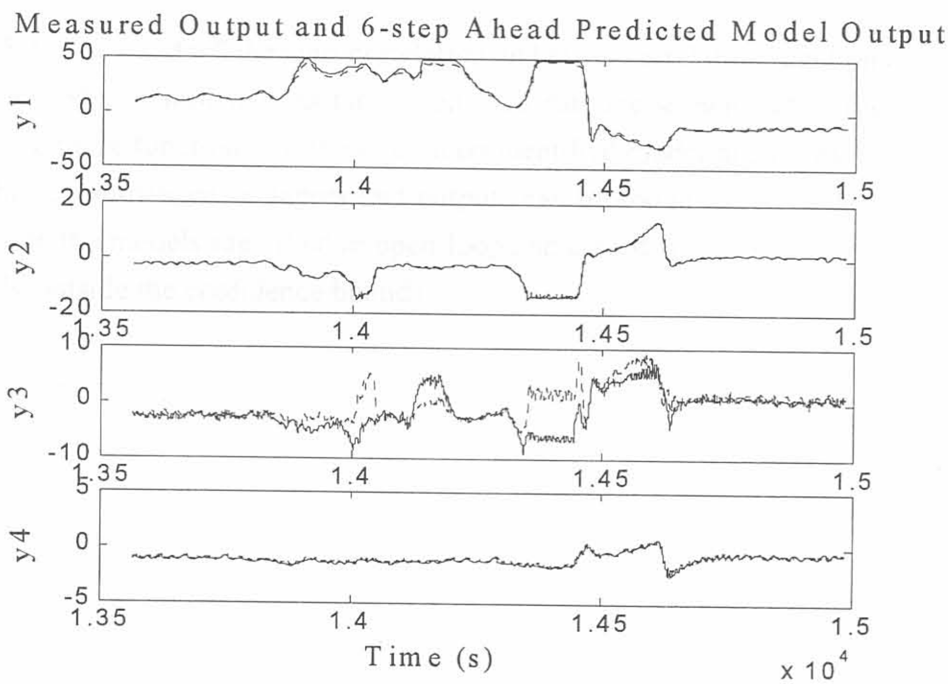

Figure 6.3: A comparison of the measured outputs (solid lines) and the 6-step ahead predicted outputs of the closed-loop identified model (dotted lines).

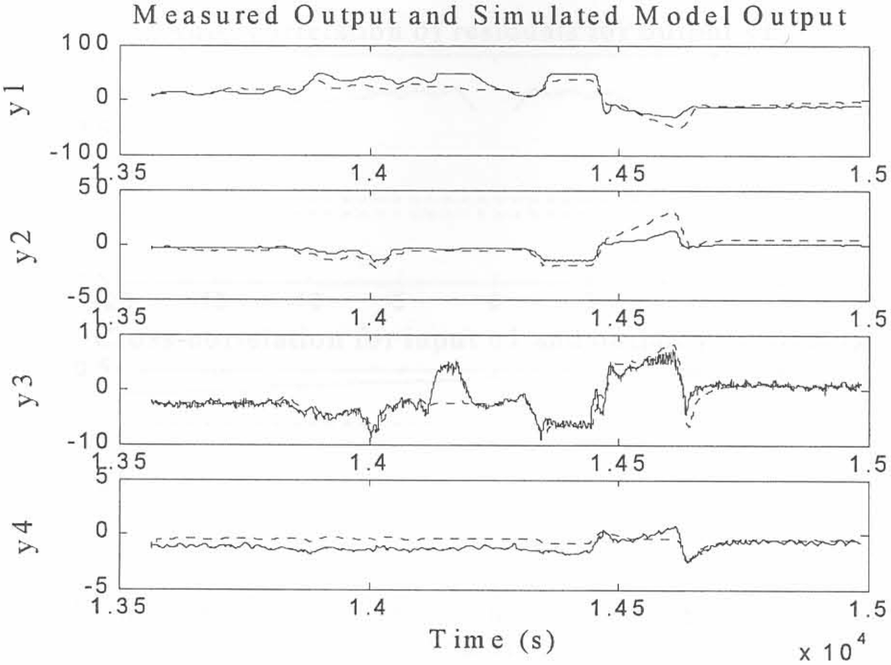

Figure 6.4: A comparison of the measured outputs (solid lines) and the simulated outputs of the open-loop identified model (dotted lines).

6.3.2.2 Residual Analysis and Comparison with the Open-Loop Identified Model

In Figs. 6.5, 6.6, 6.7 and 6.8 the auto-correlation and cross-correlation functions of the residuals for the model identified from the closed-loop data are shown and in Figs. 6.9, 6.10, 6.11 and 6.12 these functions for the open-loop identified model are shown. The functions for the other combinations of inputs and outputs can be found in Addendum C. In these plots, for both the models identified in open-loop and closed-loop, most of the functions go significantly outside the confidence bounds.

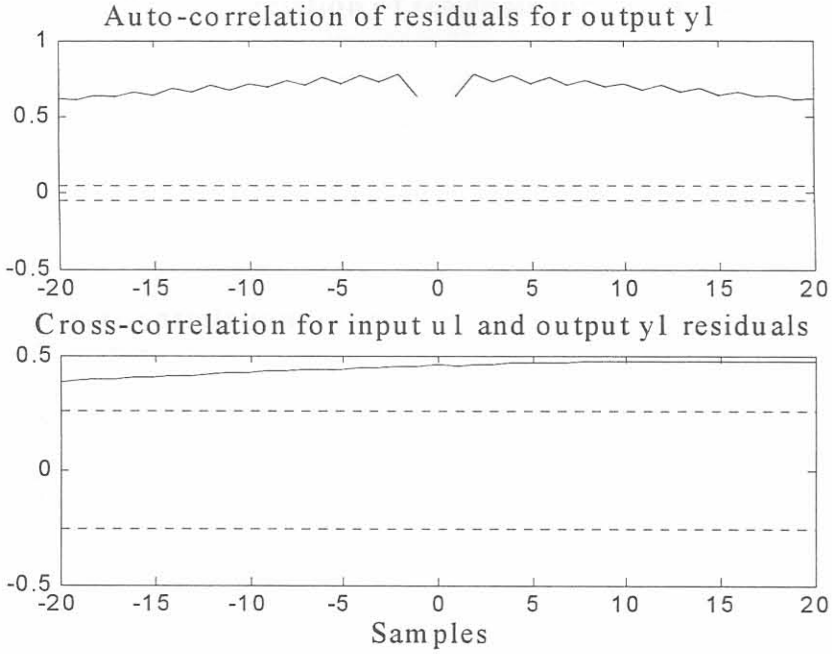

Figure 6.5: The cross-correlation and auto-correlation of the residuals for y_1 and u_1 of the closed-loop identified model.

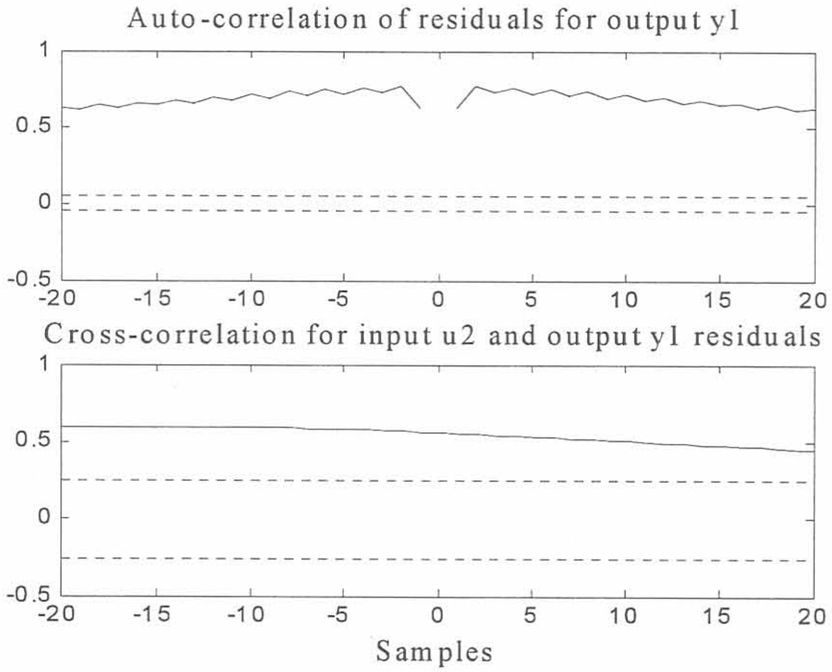

Figure 6.6: The cross-correlation and auto-correlation of the residuals for y_1 and u_2 of the closed-loop identified model.

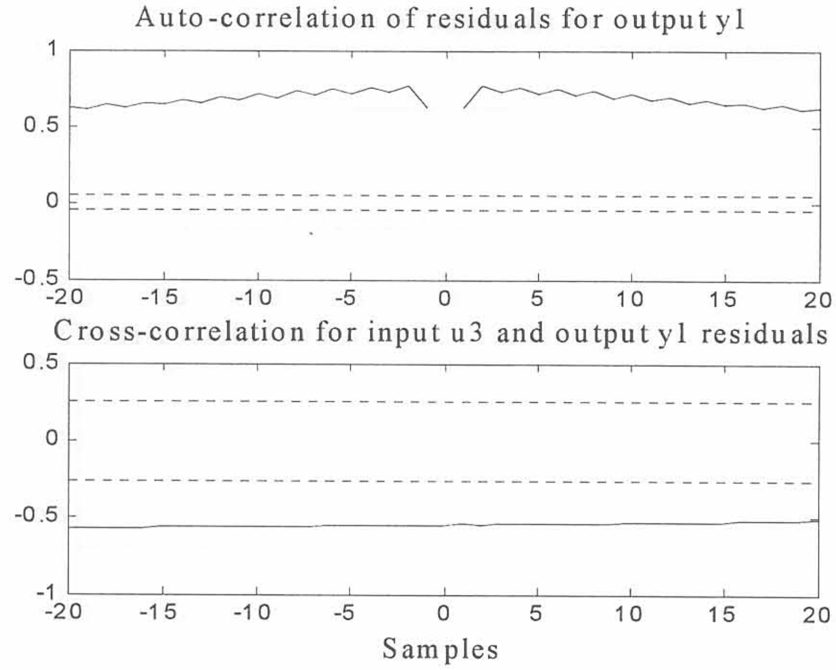

Figure 6.7: The cross-correlation and auto-correlation of the residuals for y_1 and u_3 of the closed-loop identified model.

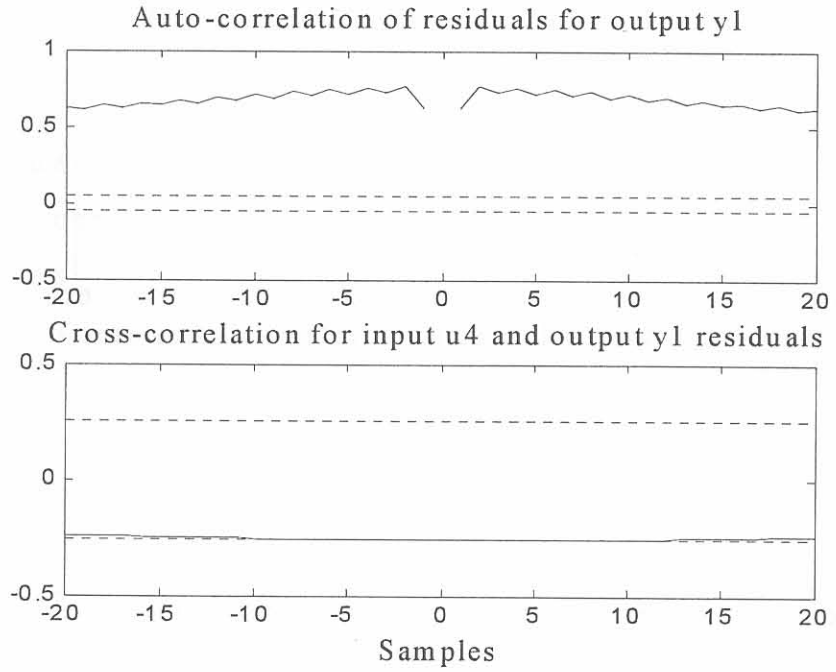

Figure 6.8: The cross-correlation and auto-correlation of the residuals for y_1 and u_4 of the closed-loop identified model.

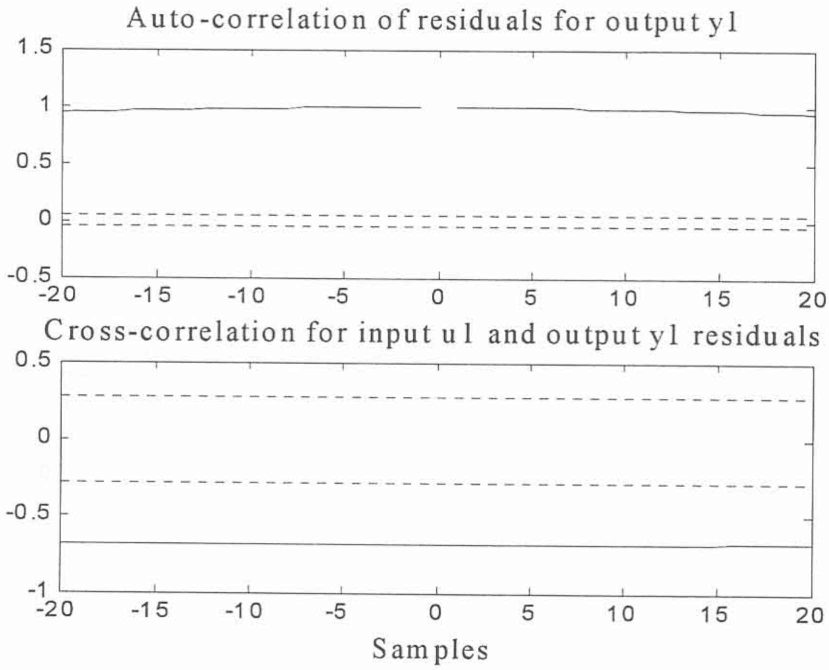

Figure 6.9: The cross-correlation and auto-correlation of the residuals for y_1 and u_1 of the open-loop identified model.

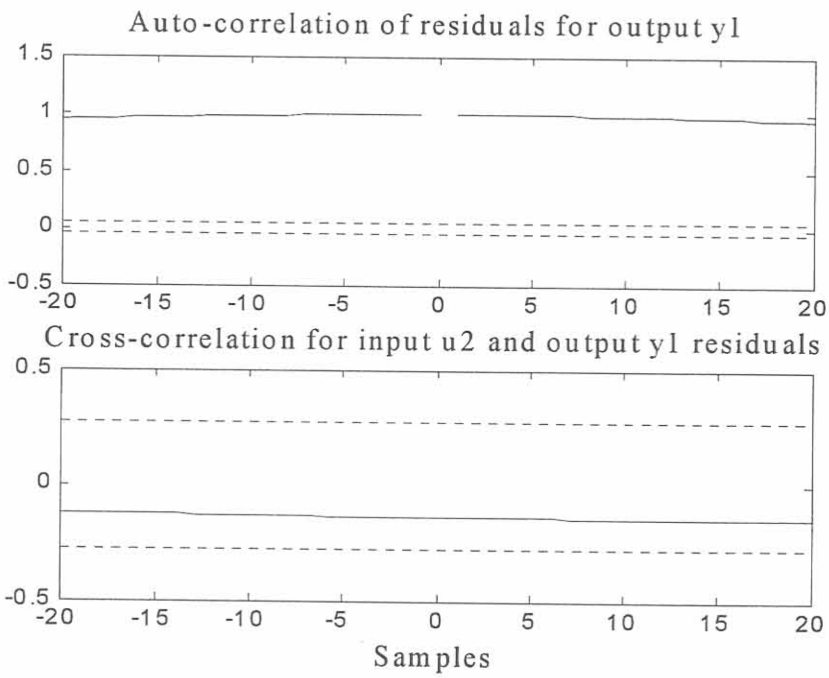

Figure 6.10: The cross-correlation and auto-correlation of the residuals for y_1 and u_2 of the open-loop identified model.

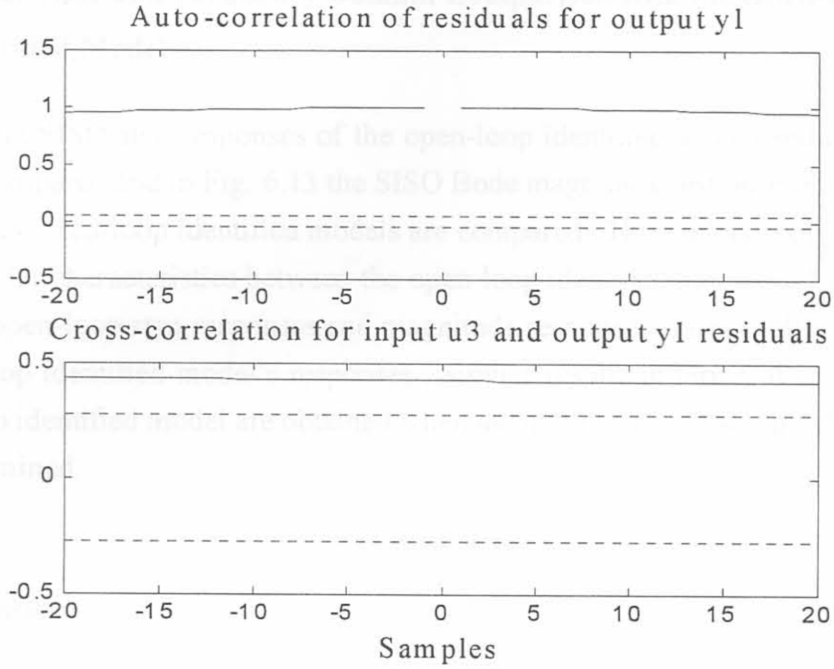

Figure 6.11: The cross-correlation and auto-correlation of the residuals for y_1 and u_3 of the open-loop identified model.

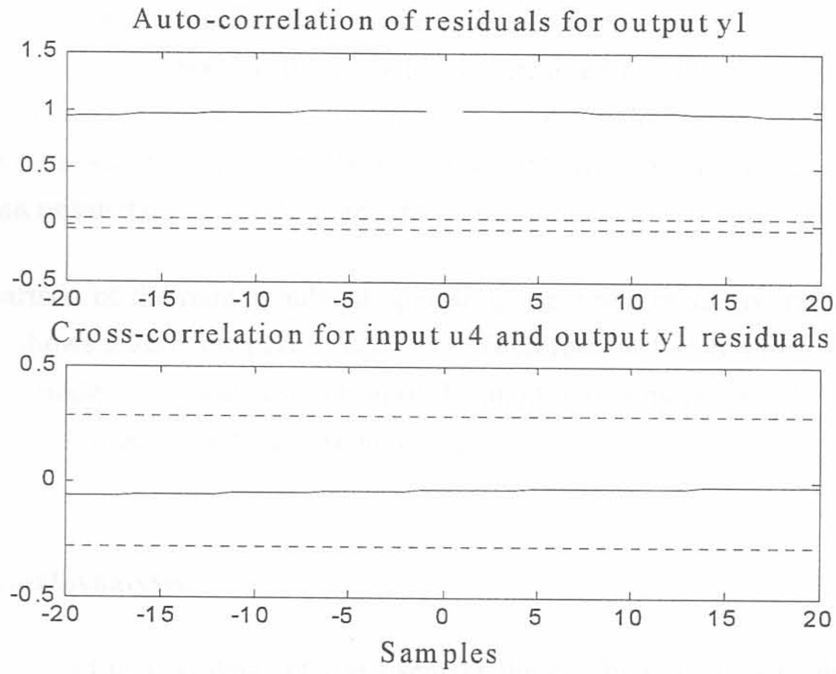

Figure 6.12: The cross-correlation and auto-correlation of the residuals for y_1 and u_4 of the open-loop identified model.

6.3.2.3 Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model

In Fig. 6.1 the SISO step responses of the open-loop identified and closed-loop identified models are compared and in Fig. 6.13 the SISO Bode magnitude responses of the open-loop identified and closed-loop identified models are compared. These plots show that, although there are similar characteristics between the open-loop identified and closed-loop identified models, the open-loop step responses and magnitude responses are not followed closely by the closed-loop identified model's responses. Similar results in terms of comparison with the open-loop identified model are obtained when the impulse responses and the Bode phase plots are examined.

6.3.3 Discussion

6.3.3.1 Simulation and Prediction

When the closed-loop data from which the model was estimated, are reproduced (pure simulation and 6-steps ahead prediction), all the results are quite good, i.e. high percentage fits are obtained. This shows that the closed-loop identified model is able to reproduce the data from which it is estimated and the estimation process is thus satisfactory. However, the model do not satisfactorily reproduce the validation data (low percentage of fit). Therefore, as expected, an unsatisfactory model is identified from the measured closed-loop data.

The comparison of the pure simulated open-loop identified model outputs and the measured outputs shows a very low percentage of fit. Therefore, it does not look as if this open-loop identified model is a good description of the plant operating in closed-loop, during the time in which the closed-loop data were measured.

6.3.3.2 Residual Analysis

Again, when the estimation data set was used for the residual analysis of the closed-loop identified model, all the results are good, i.e. the functions are inside the confidence bounds. This shows that the closed-loop identified model is able to reproduce the data from which it was estimated. However, for the validation data, the functions go significantly outside the

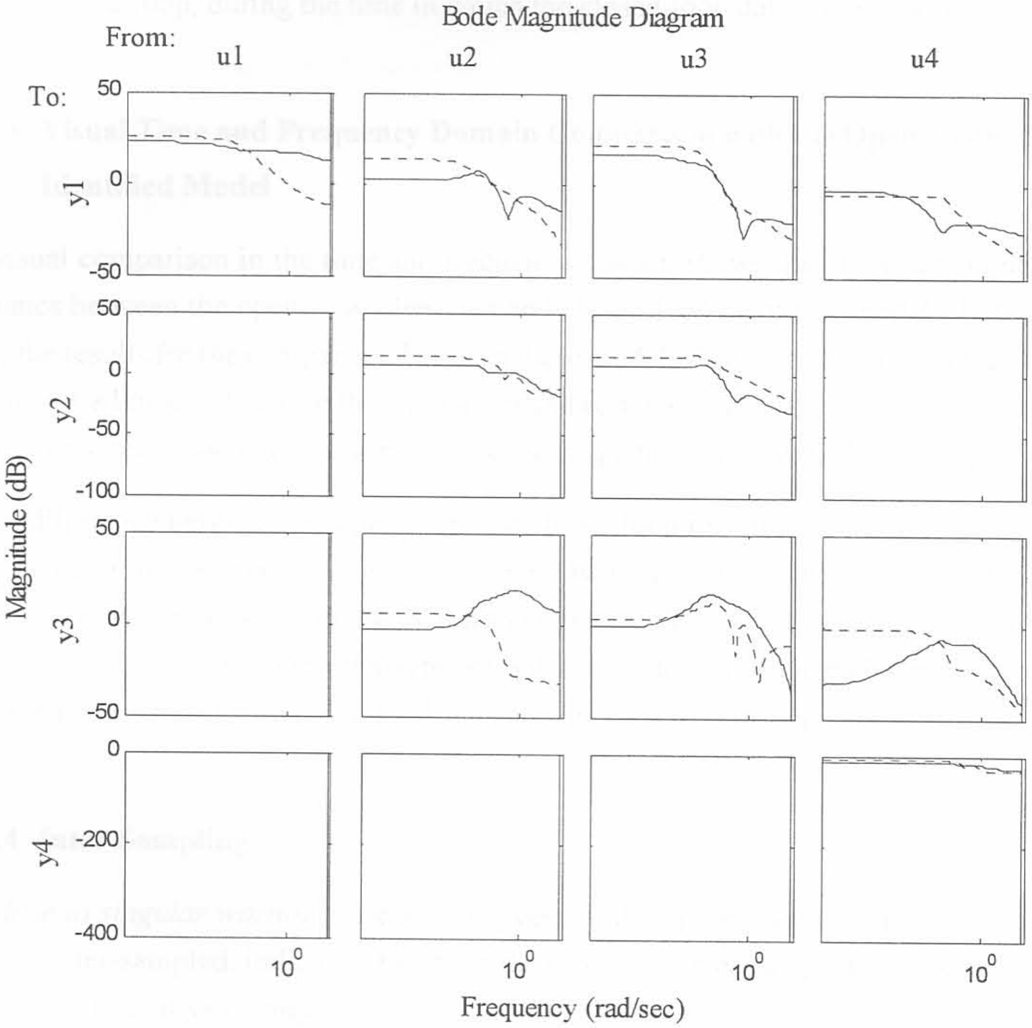

Figure 6.13: The SISO Bode magnitude plots of the open-loop identified (dotted lines) and closed-loop identified (solid lines) models.

confidence bounds. Again, this result shows that an unsatisfactory model is identified from the measured closed-loop data.

The residual analysis of the open-loop identified model also shows that the auto-correlation and cross-correlation functions go outside the confidence bounds. Thus, these results also indicate that the open-loop identified model is not a good description of the plant operating in closed-loop, during the time in which the closed-loop data were measured.

6.3.3.3 Visual Time and Frequency Domain Comparison with the Open-Loop Identified Model

The visual comparison in the time and frequency domain shows that there are similar characteristics between the open-loop identified and closed-loop identified models, but unfortunately the results for the comparison between these models do not say much about the closed-loop identified model, because the simulation and residual analysis show that the open-loop identified model is not a good representation of the plant operating in closed-loop.

The difference between the open-loop and closed-loop identified models can be because of the fact that the frequency weighting for these two types of models are very different and that the plant may have exhibited very different dynamics in closed-loop than in open-loop. Another reason may be that the characteristics of the plant changed in the short time between when the plant was identified in open-loop and when the closed-loop data were measured.

6.3.3.4 Inter-Sampling

The *close to singular warning*, which was given for the model identified from the data that were not inter-sampled, indicates that if the data were not inter-sampled, the data would not have been informative enough.

Therefore, inter-sampling did make the data informative enough. However, the fact that the models, estimated from different ranges in the measured data sets, vary considerably, shows that an imprecise model resulted, as expected.

6.3.3.5 Synopsis

Not much can be learned from the comparison between the open-loop and closed-loop identified models, since the open-loop model is not able to satisfactory reproduce the measured closed-loop data.

From the simulation and prediction, as well as residual analysis, it is concluded that an unsatisfactory model for controller design was identified from data measured under normal closed-loop control. Although the data were informative enough, the absence of structured tests that ensure PE reference signals and good SNRs resulted in an undesirable large variance in the closed-loop identified model. This result emphasizes the requirement for structured tests, which ensure PE reference signals and good SNRs.

6.4 Evaluation of the Proposed Methodology

6.4.1 Reasons for Success/Failure of the Implemented Methodology

From the experiment results it cannot be concluded that the proposed closed-loop SID methodology failed or succeeded, only that a satisfactory model was not identified when the methodology was implemented on the measured process data.

For the simulation data, it is shown that the methodology does work, but that satisfactory results depend on the excitation properties of the reference signals, as well as the SNR.

Since structured tests were not performed on the plant, the excitation properties of the reference signals and the SNRs could not be controlled. Thus, the most probable reason for the unsatisfactory identification of the model in the experiment is that, although the data were informative enough, since it was inter-sampled, the reference signals were probably not PE and the SNRs not good, which resulted in an undesirable large variance in the closed-loop identified model.

6.4.2 Recommendation for Future Implementation of the Methodology

From the results obtained it can be concluded that data, measured when the plant is under normal closed-loop control, are not necessarily informative enough and the SNRs are not necessarily good. Thus, structured test should be conducted. Even with structured tests the proposed identification technique is less intrusive and can reduces re-identification time considerably. These structured tests must ensure that the SNRs are good. In order to ensure informative data, it is easy to use these structured tests to ensure that the reference signals are PE of a sufficiently high order. Care should be taken to design appropriate test signals for the relevant plant.

Typical knowledge available from closed-loop experiments is:

1. measurements of $y(t)$ and $u(t)$,
2. knowledge about excitation properties of $r_a(t)$ and $r_b(t)$,
3. measurements of $r_a(t)$ and $r_b(t)$, and
4. knowledge of $C(q)$.

At least knowledge on (1) and (2) should be obtained in order to implement the proposed methodology satisfactorily. The knowledge of the excitation properties of the reference signals (2) will help in choosing informative data sets with good SNR for identification. The extra knowledge of either (3) or (4) will allow for consistent SID of $G_0(q)$, irrespective of the noise model $H_o(q)$ [31], since then the indirect or joint input-output approach can also be used.

It is also advisable that when future implementation of the proposed methodology is planned, a record should be kept of all previously identified models, i.e. record of time delays, orders, etcetera, as well as all changes to the plant. This extra information will simplify the implementation of the methodology considerably, since one will have an idea of what the structure looks like and how the process changed.

Furthermore, preprocessing of the data is also very important, since the data are not likely to be in shape for immediate use in identification algorithms.

6.5 Conclusion

The results obtained in this experiment validate the proposed methodology's requirement for structured tests that ensure PE reference signals and good SNRs. From this experiment it can be concluded that, an unsatisfactory model will most probably be identified from data measured under normal closed-loop control.

According to the simulation results in Chapter 5, the proposed closed-loop system identification methodology will give reliable results for MIMO plants controlled by MPC controllers, for the type of system disturbances and constraints used in the simulation, as long as structured tests that ensure PE reference signals and good SNRs are used.

Thus, a preliminary conclusion from this experimental results is that structured tests are needed to ensure that the proposed closed-loop SID methodology delivers reliable results. However, in future, structured test, with PE reference signals, on a real process are still needed, to validate the simulation results.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

7.1 INTRODUCTION

From the current practice it is concluded that the open-loop step testing approach, which is typically used for process identification in Model-based Predictive Control (MPC), has a number of disadvantages. The closed-loop system identification (SID) approach is less intrusive than the open-loop SID approach and may reduce re-identification time considerably. The ability to identify the process model while the existing MPC controller is operating will mean that safety, product quality and optimality requirements are met.

In this work, for the above-mentioned reasons, multivariable closed-loop SID for use in MPC was studied. The approach to the research project was to, first, review all the relevant closed-loop identification techniques. Second, the most appropriate closed-loop SID methodology for plants controlled by MPC controllers was chosen. Third, simulations were used to validate this methodology. Fourth, the methodology was employed to identify part of the MIBK plant from real process data, and a known process model was compared to the identified model. Lastly, the implemented methodology was evaluated by inspecting the results obtained from these simulations in Chapter 6 and the experiment in Chapter 7.

The conclusions reached after these four stages are now discussed. The conclusions arrived at after the third and fourth stages of the research approach are also an evaluation of the proposed methodology.

7.2 REVIEW OF THE CLOSED-LOOP TECHNIQUES

From the available closed-loop theory it is concluded that the basic problem with closed-loop data is that it is, typically, less informative about the open-loop system and that many estimation methods fail when applied in a direct way to closed-loop data, because of the correlation between the noise and the input to the plant. It is concluded that the prediction error method (PEM) applied in the direct fashion, with a noise model that can describe the true noise properties, still gives consistent estimates and optimal accuracy. Therefore, the PEM method was chosen for the proposed methodology.

The *identifiability* analysis shows that the open-loop condition for informative experiments, namely that the input should be persistently exciting (PE) of sufficiently high order, does not ensure identifiability in closed-loop. It is concluded that, in closed-loop, the spectrum of that part of the input that originates from the reference signal should be non-zero. This means that there should be either a nonlinear relationship between the input and the output or the reference signal should be PE.

From the identifiability analysis of the new *inter-sampling method*, where the plant output is sampled at a higher rate than the control input, it is concluded that this approach can also ensure identifiability.

The review of the available bias and variance analysis of the different approached lead to the following conclusions:

Bias: In the direct approach a bias will result, if the noise model cannot describe the true noise properties. This bias will be small in frequency ranges where either or all of the following holds: the noise model is good, the feedback contribution to the input spectrum is small, and the signal-to-noise ratio (SNR) is good. However, the indirect and joint input-output approaches can give unbiased estimates of $G_0(e^{i\omega})$, even with a fixed noise model.

Variance of Estimated Transfer Function: The asymptotic variance of the estimated transfer function is equal for the direct, indirect and joint input-output approaches, if the feedback is linear. However, this variance is worse than the variance obtained in open-loop SID, since the denominator of the variance expression contains only the spectrum of that part of the input that originates from the reference signal, while the open-loop expression has the total input spectrum in the denominator.

Variance of Parameter Estimates: When the asymptotic variance of the parameter estimates, i.e. the finite model order case, is considered, the direct approach is optimal, since it meets the Cramèr-Rao bound. The other closed-loop approaches are less precise, since they do not meet this bound.

7.3 Selection of a Methodology Applicable to MPC Controlled Plants

It is apparent from the relevant literature that there are many options available, regarding identification approaches, guarantees of identifiability, model structures and model validation techniques. From all the options the most appropriate ones were chosen for the proposed closed-loop identification methodology for MPC controlled plants. These choices are mainly based on: the theory regarding closed-loop SID; characteristics of MPC controllers; characteristics of industrial plants; keeping the methodology relatively uncomplicated; and results from similar cases.

Closed-Loop SID Approach: The direct closed-loop SID approach was chosen, since this approach is applicable to systems with arbitrary feedback mechanisms and it ensures consistency and optimal precision. Furthermore, it simplifies the development of the methodology, because it makes use of the standard functions in the MATLAB system identification toolbox and does not require any custom-written algorithms or extra software. Thus, it is not necessary to develop any new closed-loop SID software.

Guarantee of Identifiability: It is concluded that a PE reference signal should be used in order to guarantee identifiability. The changes in a multivariable MPC structure, as it deals with changes in the active constraints, are unpredictable under normal operation and can thus not guarantee a given number of changes in the controller settings. Therefore, although MPC controllers are nonlinear, time-varying and complex, which in general yield informative experiments, this should not be taken as a guarantee for identifiability. The option of *inter-sampling* the plant output to ensure identifiability was also considered. However, from the variance simulation study it is concluded that, without structured tests that ensure good SNRs, a model identified with this approach has unsatisfactory precision.

Model Structure: It is concluded that the ARX type model structure is the best choice of model structure, provided that the noise model is accurate for the process to be modelled. The reasons being: it utilises the numerically simple and reliable least square estimation (LSE) method that is a PEM estimation method and is thus consistent in closed-loop; it is parametric, which ensures compactness and accuracy; it can handle unstable systems, since it has a stable predictor; and MATLAB allows for MIMO model estimation of ARX structures.

Model Validation: It is concluded that the standard validation tests used in open-loop SID should also be used to validate closed-loop identified models. Furthermore, it is concluded that the methodology should be validated by comparing the closed-loop identified models to the open-loop identified models and by doing an examination of the closed-loop systems. The validation methods, summarised in Sections 4.8.5 and 4.8.6, were chosen. Together with these methods, a measured closed-loop validation data set is used.

The proposed methodology is further summarised in terms of the five SID steps in Section 4.8.

7.4 Validation and Evaluation of the Methodology with Simulations

In the simulation study a multivariable plant, controlled by an MPC controller, was identified from simulated closed-loop data. In order to evaluate the consistency of the identification methodology, the plant was identified for different settings in the controller as well as for different added disturbances. Different methods to ensure identifiability were also considered.

From these simulation results, it is concluded that the proposed closed-loop system identification methodology gives reliable results, i.e. accurate and precise models, for multivariable plants controlled by MPC controllers, for the type of system disturbances and constraints used in the simulation, as long as the reference signals are PE and the SNRs (ratios between noise and plant input signals) are good.

It is concluded that when PE reference signals are used, the proposed closed-loop SID methodology and open-loop SID method deliver comparable identification results. When the reference signals are not PE, or when they result in bad SNRs, the methodology should be reconsidered. Other methods that ensure identifiability, e.g. inter-sampling and nonlinear controllers, also do not guarantee precise models, if the SNR is not good, which is possible when no structured tests are performed.

Structured tests should be conducted to ensure good SNRs, and the easiest way to ensure identifiability is to make the reference signals PE.

7.5 Validation and Evaluation of the Methodology with Real Process Data

A part of a reactor, which is part of the multivariable MIBK plant at Sasol, which was designed to produce MIBK from the feedstock DMK, was chosen for the validation of the chosen closed-loop SID methodology. No structured tests could be performed on the plant. Thus, logged data sets from normal operation were used instead. Only the input and output signals of the plant were known. The reference signals, as well information regarding the controller settings, the noise and the disturbances, were not available.

Unfortunately, it is concluded that the available open-loop identified model is not a good representation of the plant in closed-loop operation at the relevant time, since the open-loop identified model is not able to satisfactory reproduce the measured closed-loop data. Therefore, not much could be learned from the validation tests where the open-loop and closed-loop identified models are compared.

From the standard validation tests, it is concluded that an unsatisfactory model for controller design was identified from the data measured under normal closed-loop control. The most probable reason for this is that, although the data were informative enough, since it was inter-sampled, the reference signals were probably not PE and the SNRs were not good, which resulted in an undesirable large variance in the closed-loop identified model.

From this experiment a preliminary conclusion is that an unsatisfactory model will usually be identified from data measured under normal closed-loop control, because when structured tests are not performed PE reference signals and good SNRs are not guaranteed.

7.6 Direction of Future Research

Much more research is needed on the challenging topic of closed-loop system identification. Especially, in model validation, the quantification of the model errors in a form that is more suitable for controller design is still an open research question [3].

Closed-loop identification techniques can also be used in MPC controller monitoring and maintenance [3]. When sufficient industrial experience of closed-loop identification is obtained, the next logical step is adaptive MPC.

An adaptive MPC technique, called MPCI, has already been proposed. The basis philosophy of MPCI is the inclusion of PE constraints in on-line optimisation. In the technique, proposed by Nikolaou and Eker, process outputs are free to move away from set-points, as long as they remain within specification bounds. Process inputs, on the other hand, are constrained to excite the process as much as possible, for the generation of maximum parameter information, while process outputs violate specification bounds as little as possible. There are many facets of MPCI that still need to be studied, along the lines of the rich literature on SID [48].

As mentioned, only a preliminary conclusion can be made from the experimental results, namely that structured tests that ensure informative experiments are needed to ensure successful implementation of the proposed closed-loop SID methodology. Therefore, in future, structured tests, with PE reference signals, on a real process are needed to validate the results obtained in the simulation study, namely that structured tests are needed.

When the necessity of PE reference signals has been shown, it will be desirable to do further studies on implementation of MPCI, where the PE condition is included in the online optimisation.

REFERENCES

[1] JM Maciejowski, *Predictive control with constraints, 1st ed.*, Prentice Hall, New York, USA, ch. 1, 2002.

[2] M Shouche, H Genceli, P Vuthandam and M Nikolaou, "Simultaneous constrained model predictive control and identification of DARX processes", *Automatica*, Vol. 34, no. 12, pp. 1521-1530, December 1998.

[3] Y Zhu and F Butoyi, "Case studies on closed-loop identification for MPC", *Control Engineering Practice*, Vol. 10, no. 4, pp. 403-417, April 2002.

[4] Y Zhu, "Multivariable process identification for MPC: the asymptotic method and its applications", *Journal of Process Control*, Vol. 8, no. 2, pp. 101-115, April 1998.

[5] U Forssell and L Ljung, "Identification of unstable systems using output error and Box-Jenkins model structures", Proc. of the 37th IEEE Conference of Decision and Control, Tampa, FL, pp. 3932-3937, 1998.

[6] U Forssell, "Closed-loop identification methods, theory, and applications", Ph.D. Dissertation no. 566, Dept. of Elect. Eng., Linköping Univ., Linköping, Sweden, 1999.

[7] ID Landau, "Identification in closed loop: a powerful design tool (better design models, simpler controllers)", *Control Engineering Practice*, Vol. 9, no. 1, pp. 51-65, January 2001.

[8] AspenTech website: http://www.aspentech.com, January 2003.

[9] Honeywell website: http://www.honeywell.com, January 2003.

[10] MJ Doma, PA Taylor and PT Vermeer, "Closed-loop identification of MPC models for MIMO processes using genetic algorithms and dithering one of the variables at a time: application to an industrial distillation tower", *Computers & Chemical Engineering*, Vol. 20, no. 976, pp. 1034-1040, May 1996.

[11] L Ljung, *System identification theory for the user, 2nd ed.*, Prentice Hall, New Jersey, USA, 1999.

[12] U Forssell and L Ljung, "Closed-loop identification revisited", *Automatica*, Vol. 35,

no.7, pp.1215-1241, July 1999.

[13] L Ljung and T McKelvey, "Subspace identification from closed-loop data", *Signals Processing*, Vol. 52, pp. 209-215, 1996.

[14] U Forssell and L Ljung, "Projection method for closed-loop identification", *IEEE Transactions on Automatic Control*, Vol. 45, no. 11, pp. 2101-2106, December 2000.

[15] I Gustafsson, L Ljung, and T Söderstrom, "Identification of processes in closed-loop - identifiability and accuracy aspects", *Automatica*, Vol. 13, pp. 59-75, 1977.

[16] U Forssell and L Ljung, "Some results on optimal experiment design", *Automatica*, Vol.36, no.5, pp. 749-756, May 2000.

[17] E de Klerk and IK Craig, "An assignment to teach closed-loop system identification", *Proc. of the 15th IFAC Triennial World Congress*, Barcelona, Spain, July 2002.

[18] L Sun, H Ohmori and A Sano, "Output intersampling approach to direct closed-loop identification", *IEEE Transactions on Automatic Control*, Vol. 46, no. 12, pp. 1936 - 1941, December 2001.

[19] Adaptech software website: http://www.adaptech.com/ english/ winpim-cl-e, September 2002.

[20] PMJ Van den Hof, RA de Callafon and ET Donkelaar, "CLOSID - a MATLAB toolbox for closed-loop system identification", *Journal A*, Vol. 40, no. 3, pp. 19-24, 1999.

[21] Mathworks website: http://www.mathworks.com/ products/ connections/ product_ main, September 2002.

[22] E de Klerk and IK Craig, "A laboratory experiment to teach closed-loop system identification", *Accepted for publication in the IEEE Transactions on Education*, March 2003.

[23] J Collings, *Mind over matter: The Sasol story: a half-century of technological innovation*, Sasol, Johannesburg, South Africa, 2002.

[24] HZ Kister, *Distillation operation*, McGraw-Hill Publishing Company, New York, USA, 1990.

[25] Agency for toxic substances and disease registry website: http://www.atsdr.cdc.gov/

tfacts21. html, September 2002.

[26] NK Narotam, L Munusamy and TM Manne, Sasol, SCI, Private Communications, 2002.

[27] Speclab website: http://www.speclab.com/ compound/ c10810.htm, September 2002.

[28] M Nikolaou, "Model predictive controllers: a critical synthesis of theory and industrial needs", *Advances in Chemical Engineering*, Vol. 26, Academic Press, 2001.

[29] M Morari and NL Ricker, *Model predictive control toolbox: for use with MATLAB*, The MathWorks Inc., Mass, USA, 1995.

[30] DE Seborg, TF Edgar and DA Mellichamp, *Process dynamics and control*, John Wiley & Sons, New York, USA, ch. 12, 27, 1989.

[31] PMJ Van den Hof, "Closed-loop issues in system identification", *Annual Reviews in Control*, Vol. 22, pp. 173-186, 1998.

[32] HT Croft, KJ Falconer and RK Guy, *Unsolved problems in geometry*, Springer-Verlag, New York, USA, p. 2, 1991.

[33] IS Gradshteyn and IM Ryzhik, *Tables of Integralds, series, and products, 6th ed.*, Academic Press, San Diego, CA, p. 1101, 2000.

[34] K Godfrey, *Petrubation signals for system identification*, Prentice Hall, Hemel Hempstead, UK, 1993.

[35] J Schoukens, R Pintelon and H van Hamme, "Identification of linear dynamic systems using piecewise constant excitation: use, misuse and alternatives", *Automatica*, Vol. 30, no. 7, pp. 1153-1169, 1994.

[36] PMJ Van den Hof, XJA Bombois, SG Douma, MFP Tolsma and R Bos, Delft University of Technology, Faculty of Applied Physics, Signals, Systems and Control, Private Conversations, May - July 2002.

[37] EW Bai, "A blind approach to the Hammerstein Wiener model identification", *Automatica*, Vol. 38, pp. 967 979, 2002.

[38] A Papoulis, *Probability, random variables, and stochastic processes, 2nd ed.*, McGraw-Hill, New York, USA, pp. 145-149, 1984.

[39] NS Nise, *Control system engineering 2nd ed.*, The Benjamin/Cummings Publishing

Company, Inc., New York, USA, pp. 755-756, 1995.

[40] S Lakshminarayanan, G Emoto, S Ebara, K Tomida and SL Shah, "Closed loop identification and control loop reconfiguration: an industrial case study", *Journal of Process Control,* Vol. 11, pp. *587-599, 2001.*

[41] Y Zhu, E Arrieta, F Butoyi and F Cortes, "Parametric versus nonparametric models in MPC process identification", *Hydrocarbon Processing*, Vol. 79, no. 2, 2000.

[42] ER Dougaenrty, *Probability and statistics for the engineering, computing and physical science,* Prentice Hall, New Jersey, USA, 1990.

[43] L Ljung, *System identification toolbox: for use with MATLAB*, The Math Works Inc., Mass, USA, 1995.

[44] M Leskens, LBM van Kessel and PMJ Van den Hof, "MIMO closed-loop identification of an MSW incinerator", *Control Engineering Practice*, Vol. 10, pp.315-326, 2002.

[45] T Söderström, "Test for pole-zero cancellation in estimated models", *Automatica*, Vol. 11, pp. 193-197, 1975.

[46] E de Klerk and IK Craig, "Closed-loop system identification of a MIMO plant controlled by a MPC controller", *Proc. of the 6th IEEE Africon Conference*, George, South Africa, pp. 85-90, October 2002.

[47] E de Klerk and IK Craig, "Multivariable closed-loop system identification of plants under model predictive control", *To be presented at the 13th IFAC Symposium on System Identification*, Rotterdam, the Netherlands, August 2003.

[48] PMJ van den Hof, "Identification of experimental models for control design", *IEEE Instrumentation and Proc. Measurement Tecnology Conference*, Budapest, Hungary, May, 2001.

[48] M Nikolaou and SA Eker, "Adaptive control through on-line optimisation: the MPCI paradigm", *Applied Mathematics and Computer Science*, Vol. 9, no. 1, pp.101-128, 1999.

[49] JB van Helmont, AJJ van der Weiden and H Anneveld, "Design of optimal controller for a coal fired Benson boiler based on modified approximate realization algo-

rithm", In R Whalley (Ed.), *Application of multivariable system techniques (AMST 90),* Elesvier Publ. Comp., London, pp. 313-320, 1990.

[50] BL Ho and RE Kalman, "Effective construction of linear state-variable models from input-output functions", *Regelungstechnik*, Vol. 14, pp 545-548, 1966.

[51] PMJ van den Hof, *System identification,* Lecture Notes tn3111, Delft University of Technology, January 2001.

ADDENDUM A

CLOSED-LOOP SYSTEM FOR THE INTER-SAMPLING SIMULATION STUDY

The closed-loop system, which are used in the simulation study of the variance of models identified from inter-sampled closed loop data, is shown in Fig. A.1.

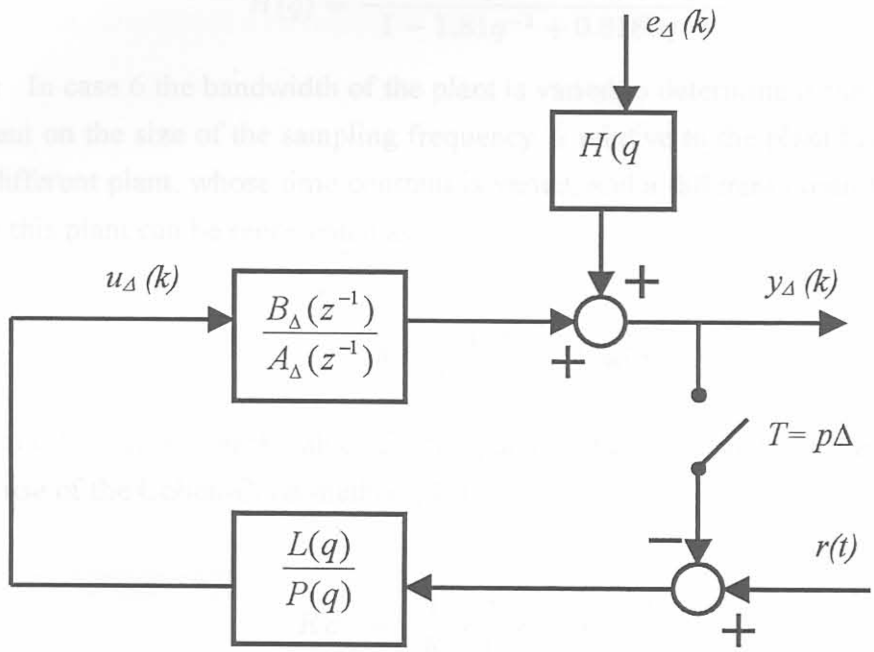

Figure A.1: The closed-loop system that was used in the simulation study.

A second order plant, which is directed by a feedback controller, is considered. In the Laplace form the true plant can be described with

$$G(s) = \frac{s+2}{s^2+2s+1}. \tag{A.1}$$

With the control interval T equal to $0.1s$, the discrete-time T-model, as described in Section 3.9.2, is

$$\frac{B(q)}{A(q)} = \frac{0.09984q^{-1} - 0.08173q^{-2}}{1 - 1.81q^{-1} + 0.8187q^{-2}}. \tag{A.2}$$

The plant is controlled by a second-order controller,

$$C(q) = \frac{L(q)}{P(q)} = \frac{4 - 7.323q^{-1} + 3.356q^{-2}}{1 - 1.67q^{-1} + 0.6703q^{-2}}. \tag{A.3}$$

The noise is generated by a white random process through a stable filter with the same poles as the plant and a zero at $s = 0$. A unit noise variance is used. In discrete-time form of this transfer function is

$$H(q) = \frac{0.004679q^{-1} + 0.004377q^{-2}}{1 - 1.81q^{-1} + 0.8187q^{-2}}. \tag{A.4}$$

Case 6: In case 6 the bandwidth of the plant is varied to determine if the influence of p is dependent on the size of the sampling frequency $\frac{1}{T}$ relative to the plant bandwidth. In this case a different plant, whose time constant is varied, and a different controller are used . In Laplace this plant can be represented as

$$G(s) = \frac{K/\tau}{1 + 1/\tau} e^{-\theta}, \text{ with} \tag{A.5}$$

$K = 1$ and $\theta = 1$. For each value of τ the parameters of the controller are determined by making use of the Cohen-Coon method [30]:

$$\begin{aligned} Kc &= \frac{1}{K} * \frac{\tau}{\theta} * (0.9 + \frac{\theta}{12\tau}), \\ \tau_i &= \frac{\theta * (30 + 3 * \theta)}{(9 + 20 * \frac{\theta}{\tau})}, \\ C(s) &= K_c(1 + \frac{1}{\tau_i s}). \end{aligned} \tag{A.6}$$

Here the noise is also generated by a white random process through a stable filter with the same poles as the plant and a zero at $s = 0$.

ADDENDUM B

APPROXIMATE REALIZATION OF STEP RESPONSE DATA

The function that was used to construct approximated state-space models from step response coefficients, makes use of the modified algorithm of Kung, described by Van Helmont, *et al.* [49], for step response data [51]:

- For a finite sequence of step response coefficients $\{s(t)\}_{t=1,\ldots N}$ a matrix $R_{n_r n_c}$, with $n_r + n_c = N$, is constructed as in equation (B.1)

$$R_{n_r n_c} = \begin{bmatrix} s(1)-s(0) & s(2)-s(0) & \cdots & s(n_c)-s(0) \\ s(2)-s(1) & s(3)-s(1) & \cdots & s(n_c)-s(1) \\ \vdots & \vdots & \vdots & \vdots \\ s(n_r)-s(n_r-1) & \cdots & \cdots & s(n_r+n_c-1)-s(n_r-1) \end{bmatrix}. \tag{B.1}$$

- Singular Value Decomposition is then applied: $R_{n_r n_c} = U\Sigma V^T$.
- The decrease in the singular values with growing index is evaluated and a number n of significant singular values is chosen.
- The approximated rank n matrix is then constructed: $R(n) = U_n \Sigma_n V_n^T$ with

$$\begin{aligned} U_n &= U \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \\ V_n &= V \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \text{ and} \\ \Sigma_n &= \begin{bmatrix} I_n & 0 \end{bmatrix} \Sigma_n \begin{bmatrix} I_n \\ 0 \end{bmatrix}. \end{aligned} \tag{B.2}$$

- The Ho-Kalman algorithm [50] is then applied to construct the state-space realization:

$$\begin{aligned} C &= \text{first } p \text{ (number of outputs) rows of } U_n \Sigma_n^{1/2}, \\ B &= \text{first } m \text{ (number of inputs) rows of } \Sigma_n^{1/2} V_n^T, \\ A &= \Sigma_n^{-1/2} U_n^T \cdot R^{\uparrow} \cdot V_n \Sigma_n^{-1/2}, \text{ and} \\ D &= s(1), \end{aligned} \tag{B.3}$$

where $R^{\uparrow}$ is $R(n)$ shifted over one block row upwards.

The continuous state-space matrices is computed with the MATLAB function *d2cm*, making use of the known sampling time and the ZOH method. The state-space model,

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), ...x(0), \\ y(t) &= Cx(t) + Du(t), \end{aligned} \tag{B.4}$$

is then computed with the MATLAB function *ss*.

ADDENDUM C

RESIDUAL ANALYSIS OF EXPERIMENTAL DATA

In Figs. C.1, C.2, C.3, C.4, C.5 and C.6 the auto-correlation and cross-correlation functions of the residuals, for the model identified from the real closed-loop data are shown and in Figs. C.7, C.8, C.9, C.10, C.11 and C.12 these functions for the open-loop identified model are shown. The functions for the other combinations of inputs and outputs can be found in Section 6.3.2.2.

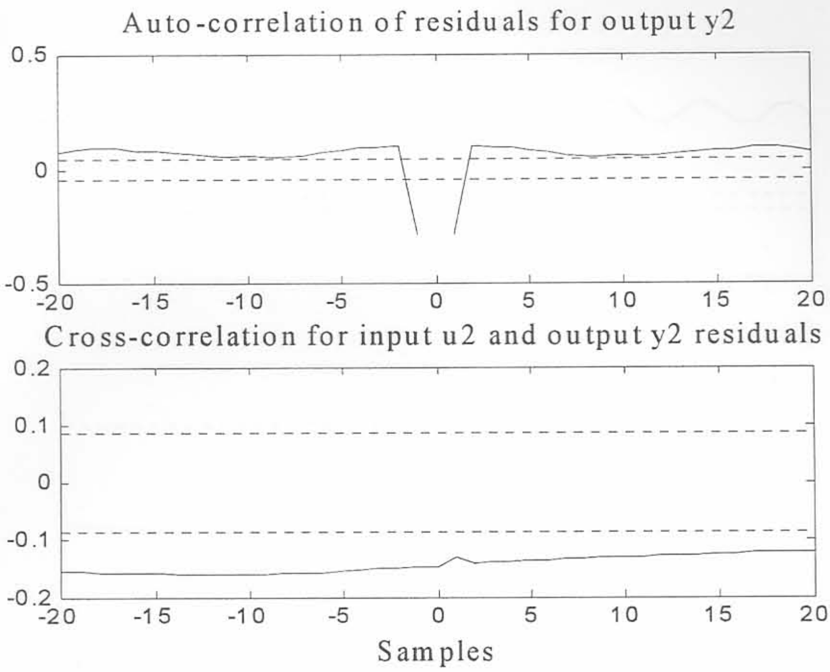

Figure C.1: The cross-correlation and auto-correlation of the residuals for y_2 and u_2 of the closed-loop identified model.

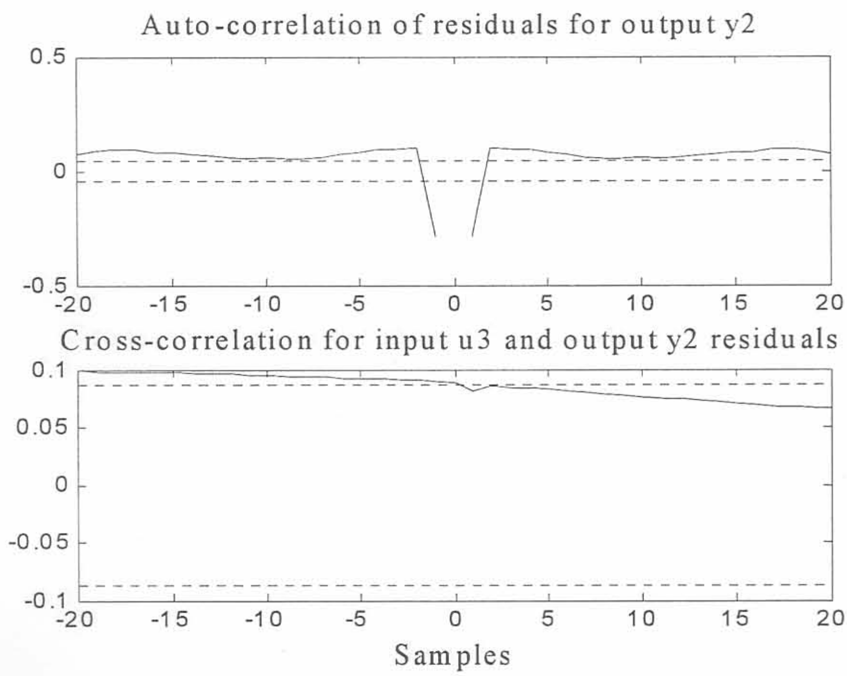

Figure C.2: The cross-correlation and auto-correlation of the residuals for y_2 and u_3 of the closed-loop identified model.

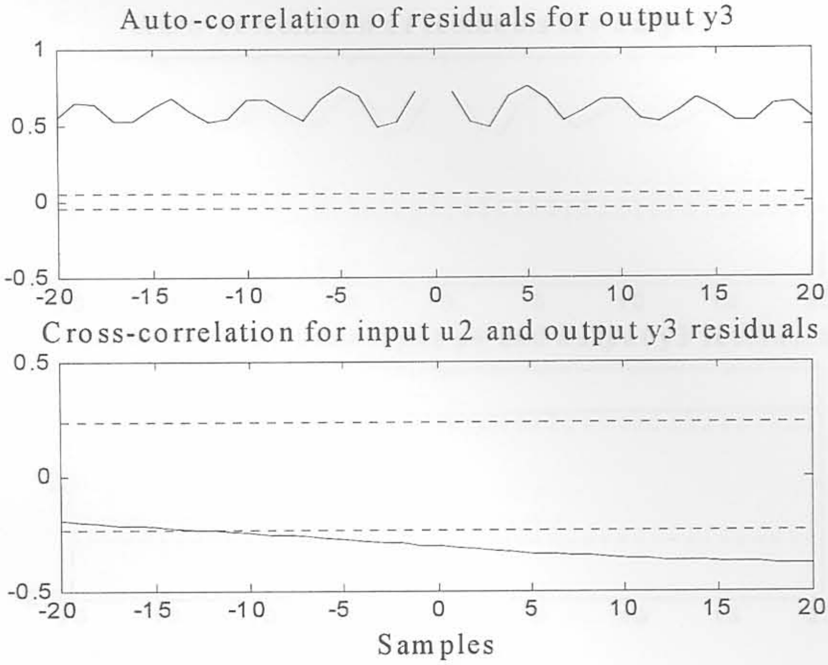

Figure C.3: The cross-correlation and auto-correlation of the residuals for y_3 and u_2 of the closed-loop identified model.

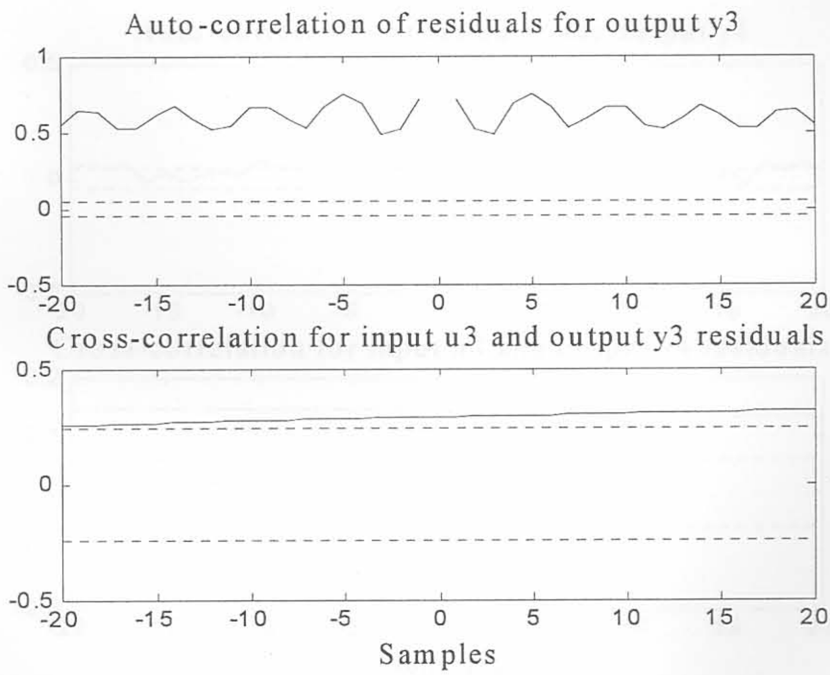

Figure C.4: The cross-correlation and auto-correlation of the residuals for y_3 and u_3 of the closed-loop identified model.

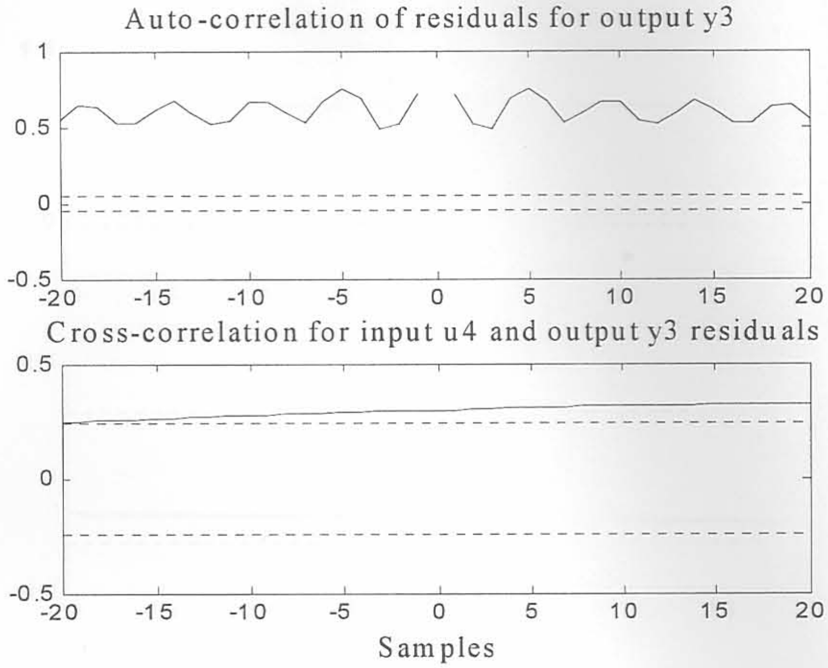

Figure C.5: The cross-correlation and auto-correlation of the residuals for y_3 and u_4 of the closed-loop identified model.

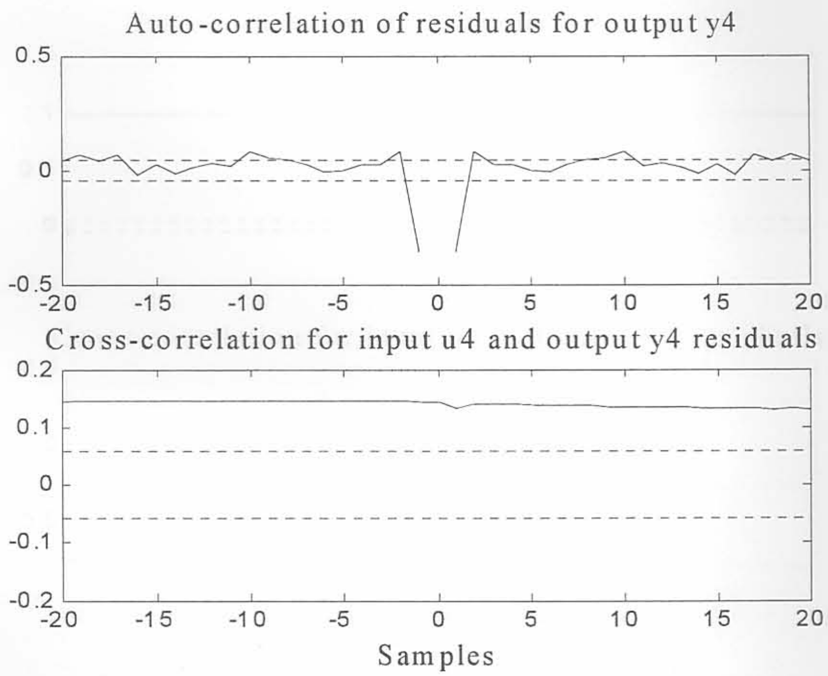

Figure C.6: The cross-correlation and auto-correlation of the residuals for y_4 and u_4 of the closed-loop identified model.

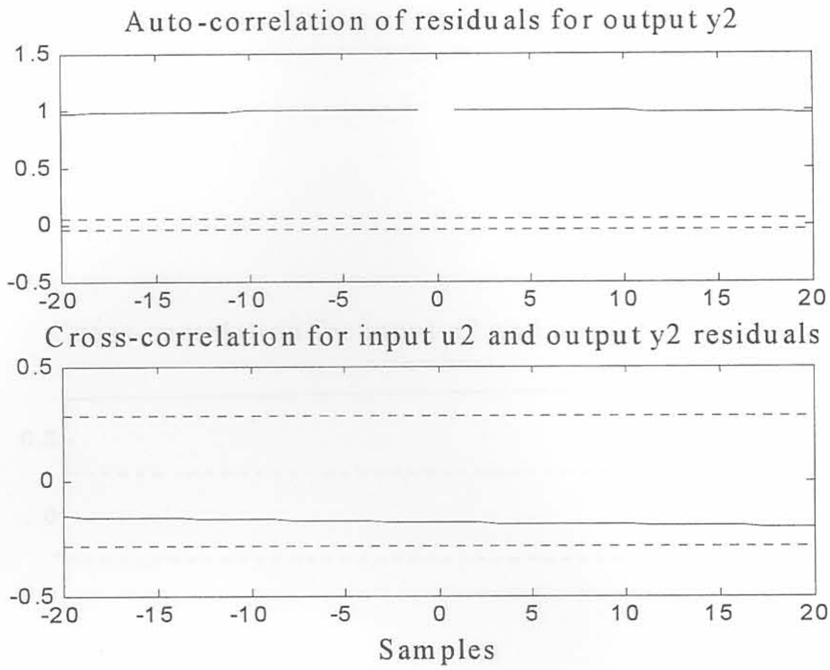

Figure C.7: The cross-correlation and auto-correlation of the residuals for y_2 and u_2 of the open-loop identified model.

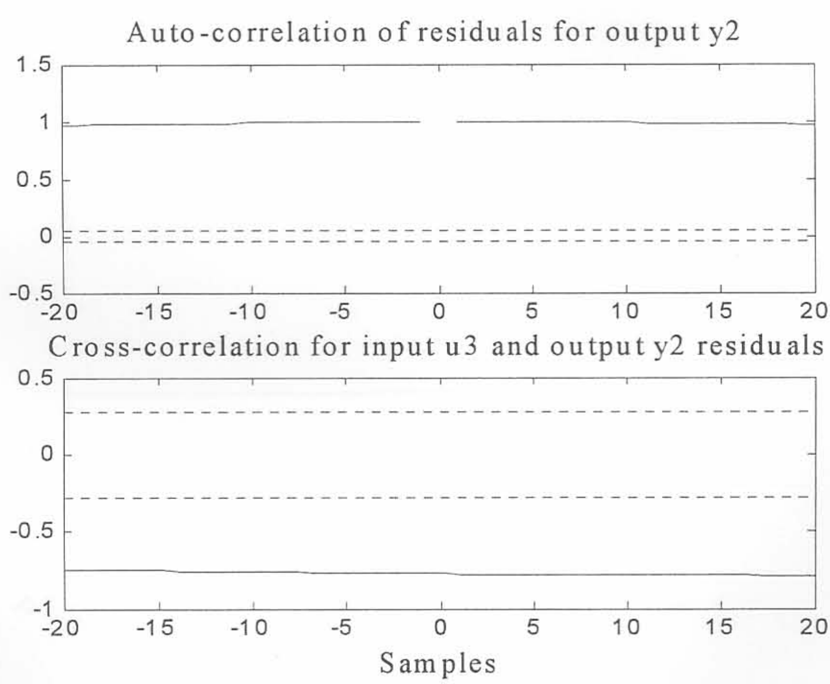

Figure C.8: The cross-correlation and auto-correlation of the residuals for y_2 and u_3 of the open-loop identified model.

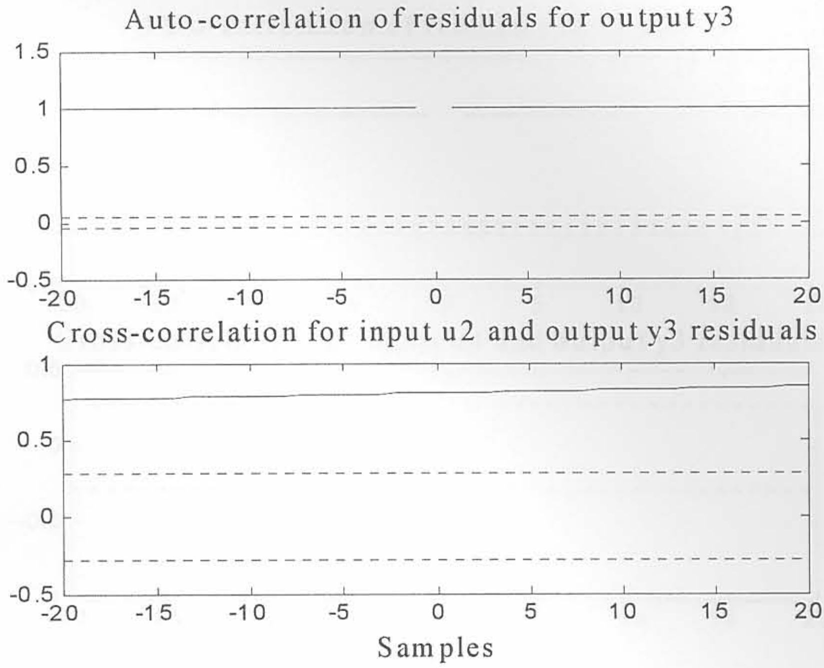

Figure C.9: The cross-correlation and auto-correlation of the residuals for y_3 and u_2 of the open-loop identified model.

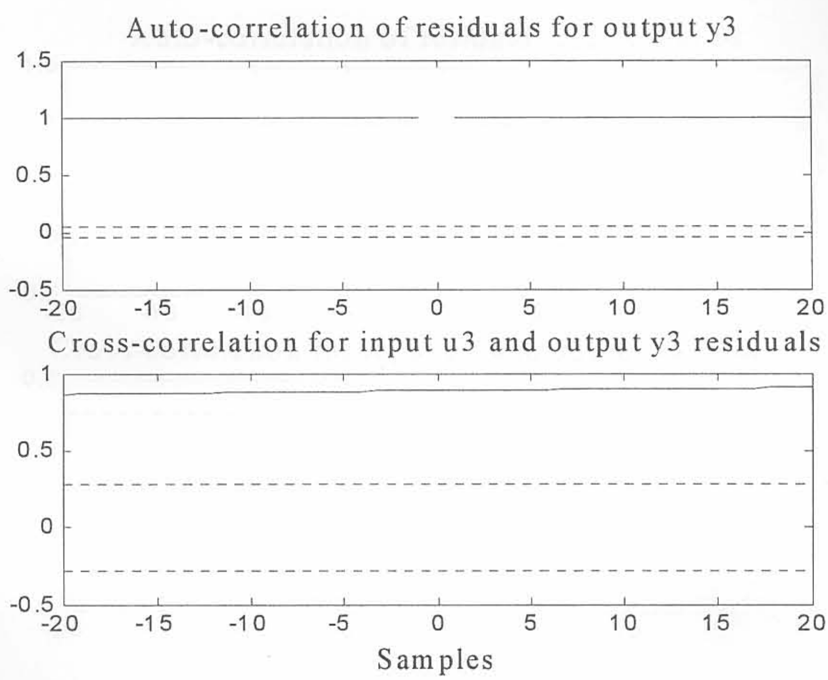

Figure C.10: The cross-correlation and auto-correlation of the residuals for y_3 and u_3 of the open-loop identified model.

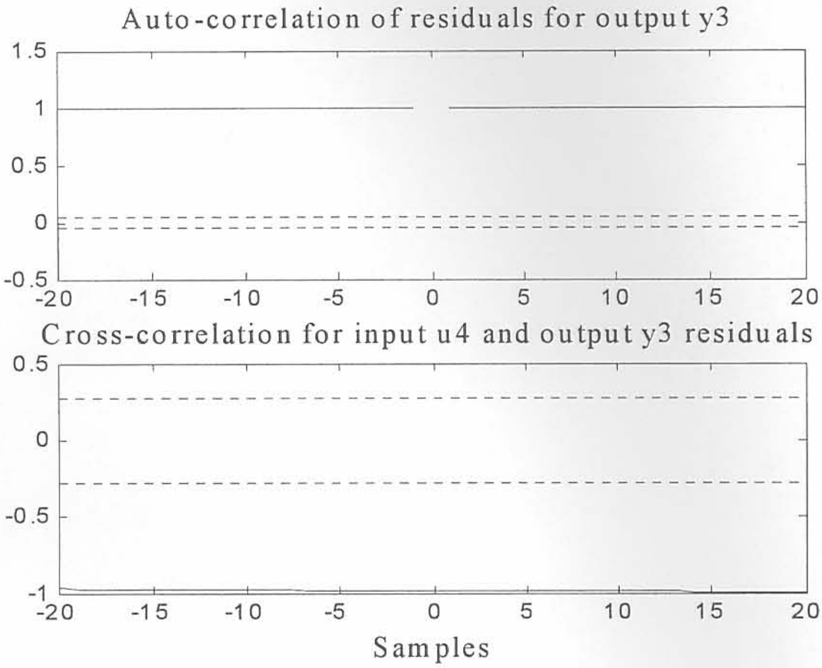

Figure C.11: The cross-correlation and auto-correlation of the residuals for y_3 and u_4 of the open-loop identified model.

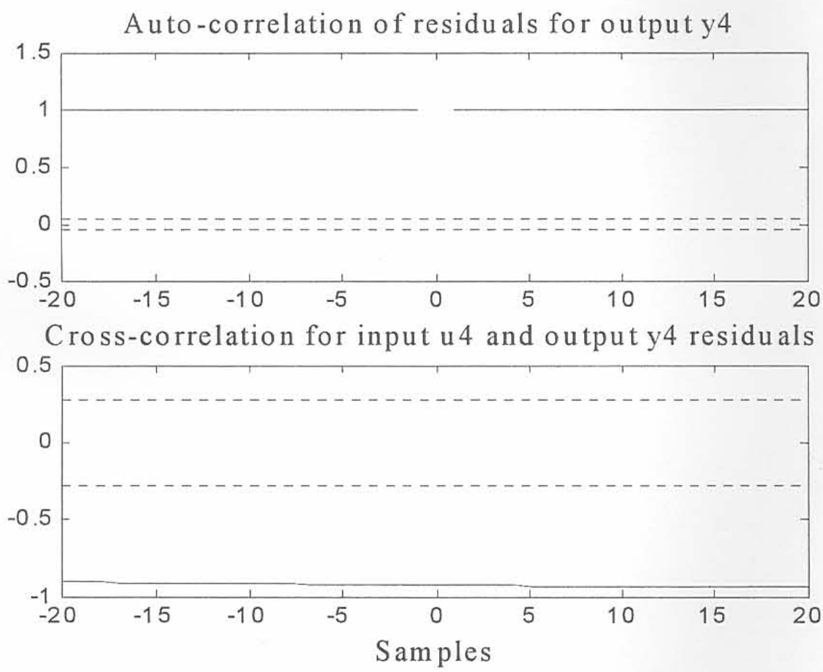

Figure C.12: The cross-correlation and auto-correlation of the residuals for y_4 and u_4 of the open-loop identified model.