


# On the Detection and Selection of Informative Subsequences from Large Historical Data Records for Linear System Identification

David Leonardo Arengas Rojas

Schriftenreihe Mess- und Regelungstechnik der Universität Kassel

Band **11**



kassel  
university   
press

# **Schriftenreihe Mess- und Regelungstechnik der Universität Kassel**

Band 11 / Vol. 11

Herausgegeben von / Edited by  
Univ.-Prof. Dr.-Ing. Andreas Kroll



# **On the Detection and Selection of Informative Subsequences from Large Historical Data Records for Linear System Identification**

David Leonardo Arengas Rojas



This work has been accepted by the Faculty of Mechanical Engineering of the University of Kassel as a thesis for acquiring the academic degree of Doktor der Ingenieurwissenschaften (Dr.-Ing.).

Supervisor: Univ.-Prof. Dr.-Ing. Andreas Kroll  
Co-Supervisor: Univ.-Prof. Dr.-Ing. Olaf Stursberg

Defense day: 15. June 2021



This document – excluding quotations and otherwise identified parts – is licensed under the Creative Commons Attribution-Share Alike 4.0 International License (CC BY-SA 4.0: <https://creativecommons.org/licenses/by-sa/4.0/>).

Bibliographic information published by Deutsche Nationalbibliothek  
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.dnb.de>.

Zugl.: Kassel, Univ., Diss. 2021  
ISBN 978-3-7376-1009-4  
DOI: <https://doi.org/doi: 10.17170/kobra-202201055361>

© 2022, kassel university press, Kassel  
<https://kup.uni-kassel.de>

Printing Shop: Print Management Logistik Service, Kassel  
Printed in Germany

---

## Abstract

---

Performing experiments for system identification of continuously operated plants might be restricted as it can impact negatively normal production or cause safety issues. In such cases, using historical logged data for system identification can become an attractive alternative instead of carrying out new experiments. However, since such plants work normally at operating points that are seldom changed, parameter estimation methods with logged data can suffer numerical problems.

Methods to locate and select informative data sequences is a promising area that can support system identification in processes where performing experiments is constrained. At least three main drawbacks of current approaches can be discussed. Firstly, detection tests used in data selection methods are based on time series models even though, they address dynamical systems where the input sequence should also be considered. In case of processes operating in closed loop, excitation caused by external disturbances is not detected if current approaches only evaluate changes in the set points. Secondly, upper interval bounds can be wrongly defined since the process is described by input-output models that assume white Gaussian noise (WGN) as additive stochastic disturbance. In practical applications, colored noise is more likely to be found than white Gaussian noise (WGN). Thirdly, in current methods model estimation with the retrieved selected intervals is not supported and therefore the quality of selected data for data-driven modeling cannot be practically assessed. In the data selection method proposed in the present thesis, called data selection for system identification (DS4SID), previous drawbacks are addressed and robust tests are designed and implemented. DS4SID can be applied to multivariate processes operating in open or closed-loop. Two tests are proposed for detection and bounding of informative intervals which simplifies the choice of user-defined parameters. A model is computed using a data merging method which can be used for further analysis. The performance of DS4SID is evaluated in a simulated and laboratory multivariate processes. A process unit of the lab-scale factory “ $\mu$ Plant” is used as industry-oriented case study. Models estimated with selected informative intervals are shown to have similar performance than estimates with the entire data set.



---

## Zusammenfassung

---

Die Durchführung von Experimenten zur Erfassung von Daten für die Systemidentifikation bei kontinuierlich betriebenen Prozessanlagen ist oft nur eingeschränkt möglich, entweder weil ein Stillstand unerwünscht ist oder aus Gründen der Betriebssicherheit. In diesem Fall stellt die Auswertung aufgezeichneter Daten eine wertvolle Alternative zur Durchführung neuer Experimente dar. Da die Betriebspunkte nur selten geändert werden, können Verfahren zur Parameterschätzung unter numerischen Problemen leiden. Die Entwicklung neuer Methoden zum Auffinden und Auswählen informativer Datensequenzen ist ein vielversprechendes Forschungsgebiet, das die Systemidentifikation unterstützen kann, bei denen die Durchführung von Experimenten nur begrenzt möglich ist.

Es gibt mindestens drei wesentliche Nachteile der derzeitigen Methoden: Erstens arbeiten Datenselektionsverfahren zur Erkennung von Ausreißern nur auf einzelnen Zeitreihen, obwohl sie aus dynamischen Systemen stammen, bei denen auch die Eingangssignale berücksichtigt werden sollten. Zweitens können die oberen Intervallgrenzen falsch definiert werden, wenn das angenommene Rauschmodell nicht der Realität entspricht. Drittens wird in derzeitigen Datenselektionsverfahren keine Modellparameterschätzung durchgeführt und daher kann der Wert der selektierten Daten nicht praktisch bewertet werden. Mit der in dieser Arbeit vorgeschlagenen Datenselektionsmethode mit der Bezeichnung DS4SID werden die bisherigen Nachteile behoben und robuste Tests vorgestellt. Das zweistufige Verfahren detektiert Ausreißer und bestimmt die oberen Intervallgrenzen informativer Datensequenzen, was die Festlegung von Entwurfsparametern vereinfacht. Mit Hilfe des Zusammenführens der selektierten Datensequenzen wird ein Modell berechnet, welches für die weitere Auswertung verwendet werden kann. Die Leistungsfähigkeit von DS4SID wird mit Hilfe von Simulationsstudien und einer industrienahen Fallstudie eines realen Mehrgrößensystems überprüft. Dabei wurde eine Prozessinsel der Modellfabrik “ $\mu$ Plant” genutzt. Es konnte gezeigt werden, dass Modelle, die mit Hilfe von selektierten Daten berechnet worden sind, vergleichbar sind mit Modellen, die mit dem gesamten Datensatz berechnet wurden.





---

## Acknowledgments

---

The present thesis is the result of my research work in the Department of Measurement and Control (Fachgebiet Mess- und Regelungstechnik MRT) at the University of Kassel, Germany. My scientific research was initially supported by a cooperation program between the Fundación para el futuro de Colombia (COLFUTURO) and the German Academic Exchange Service (DAAD). The Department of Measurement and Control supported me as research assistant during the last stage of my work. I would like to thank to my supervisor Univ.-Prof. Dr.-Ing. Andreas Kroll for his valuable scientific advice and support throughout my thesis. I am extremely grateful for the fruitful discussions and numerous suggestions that helped improving the quality of this work.

I express my acknowledgement to Prof. Dr.-Ing. Olaf Stursberg for reviewing the present work. Prof. Dr. rer. nat. Bernhard Sick as well as Dr. rer. nat. Hanns-Jakob Sommer were part of the examination board and their comments helped improving the quality of this work.

The scientific exchange with Dr.-Ing Salman Zaidi was greatly appreciated. I wish to extend my special thanks to Dipl.-Ing. Axel Dürrbaum whose feedback and technical support contributed to enhance this work. The assistance provided by Johann Pankratz was greatly appreciated. Special thanks to M.Sc. Johannes Mette, André Borkowski and M.Sc. Alex Knaub for their active participation and commitment in setting up the process island where some of the experiments were performed.

I would like to express my special thanks to my parents, Cecilia and Luis, my sisters, Kelly and Tatiana, and my brother Luigi who supported me during the entire process and for their motivating words. Su apoyo y motivación hicieron posible alcanzar este importante logro. I am grateful to my grandmother Celina for her effort and resilience and my grandparents Abel and Elida for their support. I would like to express a deep sense of gratitude to Radhia, Omar and Syrine for the unforgettable and funny moments that helped keep the correct balance in life while doing this research. I specially thank my wife Ghada for her endless love and support that made it possible to finish this work.



---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	State of the Art . . . . .	4
1.3	Contributions . . . . .	9
1.4	Thesis Outline . . . . .	10
1.5	Publications . . . . .	11
<b>2</b>	<b>Change Detection</b>	<b>13</b>
2.1	Stochastic Processes . . . . .	13
2.2	Parametric Modeling of Time Series . . . . .	18
2.3	Power Spectral Density . . . . .	23
2.4	Detection Problem . . . . .	26
2.5	Discussion . . . . .	30
<b>3</b>	<b>System Identification of Parametric Models</b>	<b>31</b>
3.1	Discrete-Time Systems . . . . .	31
3.2	Parametric Models . . . . .	32
3.2.1	Single-Input Single-Output models . . . . .	37
3.2.2	Multi-Input Multi-Output models . . . . .	39
3.3	Parameter Estimation . . . . .	42
3.3.1	Least Squares Method . . . . .	43
3.3.2	Instrumental Variable Methods . . . . .	44
3.3.3	Prediction Error Methods . . . . .	47
3.4	Recursive Parameter Estimation . . . . .	50
3.4.1	Recursive QR Least Squares Method . . . . .	50
3.4.2	Overdetermined Recursive Instrumental Variables Method . . . . .	53
3.5	The CUSUM Test for Input-Output Models . . . . .	54
3.6	Persistence of Excitation . . . . .	55
3.7	Discussion . . . . .	58
<b>4</b>	<b>Combination of Data Sets for System Identification</b>	<b>61</b>
4.1	Data Merging Problem . . . . .	61

4.2	Weighted Identification Methods . . . . .	63
4.3	Least Squares Approach for Data Merging . . . . .	64
4.4	Instrumental Variables Approach for Data Merging . . . . .	65
4.5	Multiple-cost Approach for Data Merging . . . . .	66
4.6	Discussion . . . . .	67
<b>5</b>	<b>Data Selection Method for System Identification (DS4SID)</b>	<b>69</b>
5.1	Identifying Data Sets Predominantly at Steady-state . . . . .	69
5.2	General Overview of DS4SID . . . . .	76
5.3	Determining Lower Interval Bounds . . . . .	79
5.4	Determining Upper Interval Bounds . . . . .	82
5.5	Combination of Informative Intervals . . . . .	86
5.6	DS4SID Algorithm . . . . .	87
5.7	Choice of Design Parameters . . . . .	92
5.8	Discussion . . . . .	93
<b>6</b>	<b>Case Studies</b>	<b>97</b>
6.1	Performance Assessment . . . . .	97
6.1.1	Data Reduction Ratio . . . . .	97
6.1.2	Goodness of Fit . . . . .	98
6.2	Simulation Case Study: Binary Distillation Column . . . . .	98
6.2.1	Process Description . . . . .	98
6.2.2	Controller Design . . . . .	101
6.2.3	Performed Experiments . . . . .	103
6.2.4	Results and Discussion . . . . .	108
6.3	Industry-oriented Case Study: The Process Unit II . . . . .	111
6.3.1	Process Description and Control Scheme . . . . .	111
6.3.2	Noise Analysis . . . . .	116
6.3.3	Performed Experiments . . . . .	118
6.3.4	Results and Discussion . . . . .	120
6.4	Discussion . . . . .	123
<b>7</b>	<b>Conclusions and Outlook</b>	<b>125</b>
7.1	Conclusions . . . . .	125
7.2	Outlook . . . . .	128
	<b>Appendix</b>	<b>135</b>
	<b>A Experiments in the Process Unit II</b>	<b>135</b>

---

<b>B Evaluation of Model Residuals</b>	<b>139</b>
B.1 Output Flow Rate . . . . .	139
B.2 Level of the first Reactor . . . . .	145
<b>Bibliography</b>	<b>147</b>

## Acronyms

<b>Acronym</b>	<b>Description</b>
AR	autoregressive
ARMA	autoregressive moving average
ARMAX	autoregressive moving average with external input
ARX	autoregressive with external input
CSTR	continuously stirred tank reactor
CUSUM	cumulative sum
DS4SID	data selection for system identification
EIV	extended instrumental variables
FIR	finite impulse response
GLRT	generalized likelihood ratio test
HMI	human machine interface
i.i.d.	independent and identically distributed
IMC	internal model control
IPMM	instrumental product moment matrix
IV	instrumental variables
IVM	instrumental variables method
K-S	Kolmogorov-Smirnov
LiP	linear-in-the-parameters
LSE	least squares estimate
LSM	least squares method
LTI	linear time-invariant
MA	moving average
MFD	matrix fraction description
MIMO	multi-input multi-output
MISO	multi-input single-output
MLE	maximum likelihood estimator
MOD	model-on-demand

---

<b>Acronym</b>	<b>Description</b>
MPC	model predictive controller
NARMAX	nonlinear autoregressive moving average models with exogenous inputs
NARX	nonlinear autoregressive exogenous model
NBJ	nonlinear Box-Jenkins
NRMSE	normalized root mean square error
OBF	orthonormal basis functions
OLS	ordinary least squares
ORIV	overdetermined recursive instrumental variables method
p.e.	persistently exciting
P&ID	process and instrumentation diagram
PDF	probability density function
PEM	prediction-error identification methods
PRBS	pseudorandom binary sequence
PSD	power spectral density
RGA	relative gain array
RIV	recursive instrumental variables
RLS	recursive least squares
S-W	Shapiro-Wilk
SISO	single-input single-output
SNR	signal-to-noise ratio
SVD	singular value decomposition
WGN	white Gaussian noise





## Latin Symbols

<b>Symbol</b>	<b>Description</b>
$b_f$	parameters of a Laguerre model
$\text{cov}(x_1, x_2)$	covariance between random variables $x_1$ and $x_2$
$\det \mathbf{X}$	determinant of the matrix $\mathbf{X}$
$\text{diag}(\cdot)$	diagonal matrix constructed with $(w[1], \dots, w[N])$
$E(\cdot)$	expectation operator
$G_0(q^{-1}, \Theta)$	“true” transfer function from $u$ to $y$ parametrized by $\Theta$
$\hat{G}_N(q^{-1})$	transfer function from $u$ to $y$ estimated with $\mathbf{Z}^N$
$g[k]$	test statistic in the cumulative sum (CUSUM) algorithm
$\mathcal{H}_0$	null-hypothesis in a statistical test
$\mathcal{H}_1$	alternative hypothesis in a statistical test
$H(q^{-1}, \hat{\Theta}_s)$	transfer function from $e$ to $y$ parametrized by $\hat{\Theta}_s$
$\mathbf{I}_N$	identity matrix of size $N \times N$
$\mathbf{I}(\hat{\Theta})$	information matrix associated with a model parametrized by $\hat{\Theta}$
$J(\Theta)$	cost function used to compute the estimate $\hat{\Theta}$
$k$	discrete time argument
$k_\zeta$	discrete index $k$ from which delayed input and/or output values are considered for the overdetermined recursive instrumental variables method (ORIV)
$M_0$	growing time window for computation of $\mathcal{H}_0$ in a detection test
$M_1$	fixed size sliding window for computation of $\mathcal{H}_1$ in a detection test
$m$	iteration index
$\mathcal{M}(\Theta)$	model parametrized by the vector $\Theta$
$N$	number of observations in an experiment
$n$	model order

<b>Symbol</b>	<b>Description</b>
$n_a$	model order of the denominator of a transfer function $G(q^{-1}, \Theta) = B(q^{-1}, \theta_b)/A(q^{-1}, \theta_a)$
$n_b$	model order of the numerator of a transfer function $G(q^{-1}, \Theta) = B(q^{-1}, \theta_b)/A(q^{-1}, \theta_a)$
$n_k$	input-output delay of a transfer function $G(q^{-1}, \Theta) = q^{-(n_k-1)}B(q^{-1}, \theta_b)/A(q^{-1}, \theta_a)$
$n_p$	number of model parameters
$p_x(x)$	probability density function (PDF) of the random variable $\mathbf{x}$
$p_{x_1, x_2}(x_1, x_2)$	joint PDF of the random variables $x_1$ and $x_2$
$p_{\mathbf{x}}(\mathbf{x}; \theta)$	likelihood function of $\mathbf{x}$
$q^{-1}$	backward shift operator
$\{r[k]\}$	reference sequence of a dynamic system operated in closed loop
$\hat{R}_N(\hat{\Theta})$	sample covariance matrix of a stochastic model parametrized by $\hat{\Theta}$
$s[k]$	log-likelihood ratio increment
$T(\mathbf{x})$	detection test evaluated on the random vector $\mathbf{x}$
$t$	continuous time argument
$\text{tr } \mathbf{X}$	trace of the matrix $\mathbf{X}$
$T_s$	sampling time
$\{u[k]\}$	input sequence of a dynamic system
$V_N(\Theta)$	minimizing function for parameter estimation
$\mathbf{W}$	weighting matrix
$x(t)$	continuous deterministic variable at time $t$
$\{x(t)\}$	sequence of continuous deterministic variables
$x[k]$	discrete variable at time $k$
$\{\mathbf{x}(t)\}$	vector of random variables
$X(z)$	$z$ -transform of a sequence $\{x[k]\}$
$\mathbf{X}$	deterministic matrix

---

Symbol	Description
$\mathbf{X}$	stochastic matrix
$\mathbf{y}^T$	transpose of the vector $\mathbf{y}$
$\{y[k]\}$	output sequence of a dynamic system
$\hat{y}(k k-1; \hat{\Theta})$	one-step-ahead prediction of $y[k]$
$\{\mathbf{Z}\}_{k=k_{i1}}^{k_{i2}}$	informative interval from $k = k_{i1}$ to $k = k_{i2}$



## Greek Symbols

Symbol	Description
$\varepsilon[k]$	model residuals
$\gamma_c$	threshold for change detection in the CUSUM algorithm
$\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$	reciprocal of the condition number corresponding to the auxiliary information matrix
$\hat{\mu}$	estimated sample mean
$\mathbf{v}$	measurement noise
$\phi(\omega)$	spectral density
$\hat{\sigma}^2$	sample variance
$\Theta_0$	“true” parameter vector
$\Theta$	parameter vector
$\hat{\Theta}$	parameter vector estimated with $Z^N$
$\sigma_e^2$	variance of $\{\mathbf{e}[k]\}$
$\varphi[k]$	regression vector
$\Phi$	regression matrix
$\mathcal{N}(0, \sigma_e^2)$	random variable normally distributed with zero mean and variance $\sigma_e^2$
$\xi$	dominating pole of a dynamic system
$\zeta[k]$	instrumental vector



# CHAPTER 1

---

## Introduction

---

The data selection method proposed in the present thesis can be used to support system identification in processes where performing experiments for data-driven modeling is constrained. In the present chapter, motivation and contribution of this research work and the state of the art are presented.

### 1.1 Motivation

System identification involves mainly three aspects: data set, model and estimation method. Data sets can be obtained, for instance, by performing designed experiments or from logged records. In the case of performing an experiment, design decisions such as type of excitation signal, sampling time and duration of the experiment are made by the user. The goal is to perform an experiment which results in maximally informative data *i.e.* data sets from which the dynamics of a system can be modeled. However, performing experiments can sometimes be restricted in continuously operated plants since it may impact negatively production and it is time consuming. Alternatively, use of logged process data is an attractive option. Such process records are, indeed, large data sets that have often been logged for years. Decaying prices of storage devices and sensors facilitates data logging resulting in large data bases. In a refinery, for instance, measurements are recorded with sampling intervals between 1 s and 60 s from around 60000 sensors which results in approximately 300 GB per year [8]. In [12], databases of 2.5 GB from a process in a chemical plant were used for data selection to support system identification. A continuously operated process in a lab-scale plant, that is introduced in [Chapter 6](#), can generate databases of around 1 GB from approximately 65 sensors with a sampling interval of 250 ms.

The latter examples introduce the data selection problem for system identification which consist in searching and extracting specific data intervals from historical records to support model estimation instead of collecting measurements from new experiments.

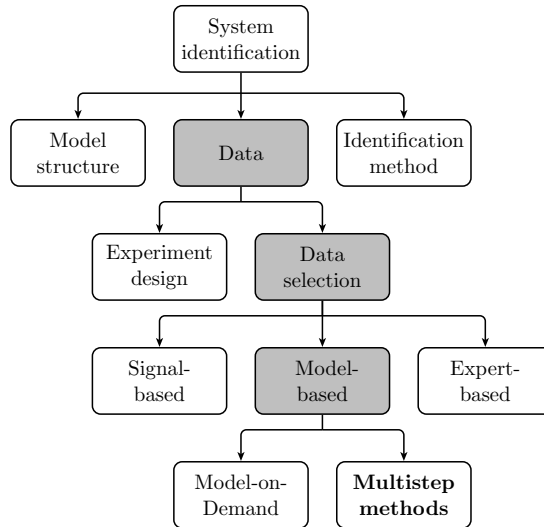


Data selection for identification was proposed as a promising research field as reported in [39, 58]. In both publications, the authors propose to adapt identification methods for “mining” data sequences that contain useful information on the system dynamics. The previously described situations are frequently found, for instance, in continuously operated plants, that represent an interesting application field for data selection methods.

This research work is focused on selection of informative data sequences from large databases to support system identification. The motivation of searching informative data intervals is to evaluate the possibility of performing data-driven modeling from logged process data without performing new experiments. This is a non-intrusive option that also has economical advantages since it is not necessary to stop an operating process to perform tests for collecting data. However, it is firstly necessary to consider several aspects before using these data for data-driven modeling. Since processes in continuously operated plants are only occasionally excited, logged data exhibit seldom changes which can affect parameter estimation.

In parameter estimation methods, some operations such as matrix inversions can experience numerical problems when using data sets as previously described. These numerical problems can be avoided by estimating parameters from sequences that were recorded during dynamic changes in the process. Useful *data intervals* for system identification may be located around *change times* where a process leaves an operating point. The limits of an informative interval will be referred to as lower and upper interval bounds. These will be further explained in [section 5.3](#) and [section 5.4](#). Data selection for system identification consists of three main tasks: locating change times, interval bounding and data merging. Lower and upper interval bounds are defined by tests performed in the first two stages. A model is estimated by merging informative data intervals in a further stage.

[Figure 1.1](#) positions the research area of the present thesis. Development of data selection methods for identification requires to adapt different tools from statistical signal processing and from data-driven modeling into this new field. Data selection methods can be divided mainly into three groups: signal-, model- or expert-based as shown in [Figure 1.1](#). In the first approach, useful intervals are selected by performing tests separately on input and output signals. In the second approach, interval bounds are defined according to the conditioning of the information matrix associated to a chosen estimation method. Expert-based methods are performed manually which is too expensive and too time-demanding for most applications.



**Figure 1.1:** Data selection as a research field related to system identification

In order to propose a data selection method, several aspects should be addressed. System identification using predominantly stationary data sets is explored and major challenges are discussed. Available approaches for data selection are investigated and their advantages and drawbacks are described. Then, suitable methods are explored and adapted for selection of informative data. A novel data selection method is finally proposed which improves the current state of the art in several aspects.

Available methods are limited to single-input single-output (SISO) systems which is a significant drawback as a large number of multivariate processes are found in different applications. Changes are detected by scanning input and output signals separately. Moreover, additional tests are required to confirm informative intervals which can be computationally demanding.

In real applications, measurement noise can be colored due to filters implemented in sensors and logging devices. Approaches used in current data selection methods assume signals with added WGN which may differ from real situations. The latter can result in wrong interval bounds and thus in wrong models.

Current methods report use of at least four nested tests that are sequentially activated which can be computationally demanding. These approaches can yield a high rate of false alarms and misdetected intervals due to implementation of tests that are not robust to high noise content or colored noise.

The state of the art is explained in the next section. Signal and model-based techniques have dealt with data selection for identification. In the next section, current methods are introduced and their main features are discussed. Moreover, pitfalls are analyzed resulting in proposing a new method, DS4SID, which is considered to address aspects not yet covered by the available methods.

## 1.2 State of the Art

Most of the reported data selection methods for identification can be grouped into signal-based [15, 38, 92] and model-based approaches [5, 12, 17, 73]. Signal-based methods are mainly focused on segmenting data either in steady-state or changing intervals. In these methods, the time series contained in the data sets are analyzed separately. The latter means that no input-output models are considered for identification. Signal-based methods will usually require further tests to confirm preliminary located intervals. Among others, detectors such as the CUSUM test [10, 32, 87] or based on an R-statistic [15] have been tested for location of changes in signals.

One of the first signal-based methods is found in [38]. In this approach, a method as described in [15] is implemented for change detection. The working principle can be summarized as follows: “Start selecting data after each significant set point change and continue until the data has reached steady-state”. False detection rate can be adjusted by setting the filter parameters involved in the method (see [16]).

In [92], a signal-based selection method is presented which follows a similar approach as in [38]. In the first stage, transient changes in the input signals are detected. Then, time instants where the process reaches steady-state are determined. The method is evaluated in SISO processes operating in closed-loop. A review of different approaches found in machine learning and signal processing is presented in [71] where common topics with statistics are highlighted. Data that represent an anomaly are labeled as abnormal, which can be considered as a similar problem as change detection in statistics. In [62], support vector machines are evaluated for anomaly detection in water distribution systems. The proposed approach is tested using historical data of a real system demonstrating its effectiveness in the selected case study.

In model-based methods, input-output models are used for process description. Detection methods, which are usually designed for time series models, need to be extended in the case of input-output models. Potentially informative intervals are related to change times. Upper bounds of these intervals are defined in further tests by evaluating the effect of including more data that does not exhibit so much change in parameter estimation. In these methods, input

signals are assumed to be sufficiently exciting so that generated data sets are informative for data-driven modeling. A detailed analysis on the required degree of excitation for different parametric models using prediction-error identification methods (PEM) is presented in [24].

One of the first model-based methods was reported in [17]. A detailed study on negative effects of using predominantly stationary data for parameter estimation is presented. The proposed method is evaluated in a linear SISO process operating in open-loop with additive colored Gaussian noise. The input transfer function of the model and the system are in the same class. However, the noise transfer function of the system is an autoregressive moving average (ARMA) process while an autoregressive (AR) model is used for identification. The ordinary least squares (OLS) method is implemented for parameter estimation which is performed by singular value decomposition (SVD) of the regression matrix. The process is seldom excited with an input signal that exhibits long stationary periods with seldom changes. The bias of the estimates is constant when persistently exciting (p.e.) signals are used to excite the process. The least squares method (LSM) cannot guarantee consistent parameters in this case due to the additive colored Gaussian noise.

In contrast, as reported in [17], bias of estimates using LSM increases when predominantly stationary data sets are used for identification. To overcome this limitation, a method is proposed to discard non-informative data *i.e.* observations that negatively impact parameter estimation. Model accuracy can then be improved since parameters are estimated only with informative data. At least three main drawbacks of these methods can be mentioned. Firstly, the proposed approach is limited to SISO systems. However, a large number of applications are multivariate processes. Secondly, the identification method used is non-recursive which is computationally demanding in the case of large data sets. Thirdly, no change detection algorithm is used which results in an approach where the parameter estimation method is always active while scanning the data.

A second model-based approach was reported in [12, 70]. This method differs in several aspects with the previous approach. A change detection algorithm was implemented to determine potentially informative intervals. Input and output signals are scaled between 0 and 1 before being processed for change detection. However, detection tests are applied separately to the input and output signals which leads to additional tests to confirm the validity of identified intervals. Interval bounds are defined based on the conditioning of the information matrix associated with the parameter estimation method used. This scalar measure is compared with a user-defined parameter that is initially set by trial and error.

Several aspects such as computational precision are considered for setting this user parameter.

In this model-based method, ARX-Laguerre filters are used for the process description. A recursive least squares (RLS) method is used for identification instead of OLS. Causal relation between changes in the output generated by an excitation in the input signal are confirmed in a last stage by the Granger causality test. Four tests are proposed in this approach which can be computationally demanding. Moreover, informative intervals may be misdetected due to wrong choice of user-defined parameters and false rejections in selection tests. This method was applied to logged data from SISO systems of a pulp and paper plant and it was concluded that only around 1.5% of the process data should be used for identification.

Several drawbacks of the method in [12] can be discussed. Firstly, the method is oriented to SISO systems which represents a relevant pitfall as previously described. Secondly, the implemented detection tests will usually yield a high false alarm rate. Detection tests are based on the change in the sample mean and in the variance assuming signals as DC levels embedded in WGN. Those tests based on threshold of the sample variance are sensitive to high noise content which usually result in false change detections. The proposed approach for upper interval bounding may yield high misdetection rates *i.e.* a large number of informative sequences that are discarded despite being informative. Misdetection may occur due to wrong choices of the model order and the threshold for upper interval bounding. A data pretreatment for scaling between 0 and 1 increases processing times. Thirdly, the QR-RLS method, which is used for parameter estimation, is not robust to colored Gaussian noise which may yield wrong interval bounds. Although consistent parameter estimation is not the main goal in [12], this property is typically desired.

In [74], an extended version of the former method is presented and the effect of the design parameters in the approach is analyzed. The performance of the method is mainly affected by the parameters associated with the Laguerre model and the threshold for upper interval bounding. A high number of ARX-Laguerre models can be obtained depending on the quantity of retrieved intervals. Thus, a merging procedure is proposed to reduce the number of models. A measure based on entropy difference between input and output signals is used for data merging. This modification improves the last stages in the data selection procedure. However, this modified version still suffers from most of the drawbacks discussed for the approach in [12].

In [73], the condition number of the sample information matrix is proposed as criterion to evaluate data quality for identification in routine operating

records. Different thresholds for the condition number are proposed based on computationally numerical precision. The presented quality index is analyzed in simulation case studies using autoregressive with external input (ARX) models. In [66], a slightly modified version of the method in [12] is introduced. The first stages of the method, namely change detection and changes in the variance, are identical to the procedure explained in [12]. Data selection is performed using a two-stage procedure where noise and input transfer functions are estimated separately. Firstly, a noise model is computed with data that is predominantly at steady-state. Then, the output and input signals are filtered by the inverse of the noise transfer function computed in the first step. The resulting filtered signals are used in a second step to estimate input transfer functions for each informative interval. Since the transient detection follows a similar procedure as in [12], the input-output data need to be normalized for transient detection. Consistency problems can be also expected since LSM is used for identification. Processing times can be large since recursive methods are not used in the proposed technique.

A data selection method is proposed in [55] to extract sequences from process data for identification of a selective catalytic reduction (SCR) system. A model with input and noise transfer functions that are independently parameterized is proposed for process description. Assessment of the quality of the data is based on the evaluation of the associated Fisher information matrix that is computed using the Steiglitz–McBride method. Two main drawbacks can be mentioned related to the proposed approach. Firstly, identification is performed using a batch method which for this scenario is not desirable due to computational costs. Secondly, the beginning of the intervals is selected manually, which can be demanding for very large data sets. Reported results show acceptable performance for the proposed application to a real process.

Data selection was evaluated in [75] for a multivariate process corresponding to a lead zinc concentrator. Several cases are considered: including all available signals or manual selection of most relevant for the process. In the first case, no data could be selected due to correlation between some variables which yields strongly ill-conditioned matrices. Results of this practical example shows that a pre-selection of variables can help for data selection when dealing with multivariate processes.

The formerly described data selection methods are based on linear modeling. Despite the fact that a method described in [81, 83] was not originally proposed for data selection, several common aspects with this field are worth mentioning. In this approach, models are estimated “on demand” of the user or process requirements. Several nonlinear black box models such as nonlinear autore-

gressive exogenous models (NARXs), nonlinear autoregressive moving average models with exogenous inputs (NARMAX) and nonlinear Box-Jenkins (NBJ) models are used for process description. Parameter estimation is performed by local model optimization instead of a global approach. The method is flexible as the user can choose between different models and methods for local parameter estimation. This approach was extended for control of dynamical systems as found in [82]. A model-on-demand (MOD) model predictive controller was compared with a linear model predictive controller (MPC) and the former was proved to lead a better performance at high bandwidths. However, the approach in [81] can be computationally demanding and impractical to process large data bases.

The main limitations of current data selection methods are:

- In data selection methods for multivariate processes as described in [75], detection tests are evaluated sequentially for each output signal to locate changes between operating points. This approach might yield multiple detections of the same change due to possible correlations in the system which results in drawbacks for parameter estimation since similar data are contained in the selected intervals. Although methods for system identification of multivariate system have been widely investigated (see, for instance, [46, 95]) they have not been adapted in approaches for searching for informative data for system identification.
- Current methods experience a high number of false alarms and misdetection of informative intervals. Reported detection tests of such methods are based on evaluation of the mean and variance of the process signals that are treated separately. Those tests can be very sensitive to false alarms in the case of high noise content. Misdetection may occur due to use of thresholds that are applied to the entire data set. However, the value of the threshold can depend on the operating point. Thus, some intervals can be rejected when indeed those data are relevant for identification. Misdetection of intervals yields models with worse performance when compared with estimation using the entire data set.
- Available methods are not robust to colored noise. Evaluation of the conditioning of the information matrix has been proposed for interval bounding in data selection methods. Most of the available methods use LSM whose consistency is guaranteed only in the case of WGN. Interval bounding with this estimation method may yield wrong upper bounds in the case of colored noise. This aspect may be considered for the design of a selection method as signals from real processes may be more likely embedded in colored noise than in WGN.

## 1.3 Contributions

The major contributions of this research work are:

- The data selection method proposed in this work can be applied to multivariate systems operating in open- or closed-loop. Detection tests are extended to input-output models which decreases false alarms and misdetection.
- In the proposed approach for data selection, upper interval bounding is performed considering information retrieved locally from current operating points. Interval misdetection is avoided because of implementation of local thresholding instead of local approaches reported in other methods.
- In the present work, methods robust to colored noise are implemented which cover more general situations regarding additive noise found in real applications. This property can guarantee better interval bounding since thresholding is performed with parameter estimation methods that consider the presence of colored Gaussian noise. Correlated noise is an important aspect that should be considered for the design of data selection methods as data logged from processes often exhibit such behavior.

Besides the former major aspects, other contributions of this research work are presented in the following:

- Data selection is stated as a problem which consists of two main tasks: change detection and interval bounding. A detection method should choose between two hypotheses in the first stage of the problem: either the process is operating at the same point or it was externally excited. In current data selection methods, lower interval bounding is performed by several sequential tests that require data pretreatment including removal of mean and data normalization. In this work, only a detection test is performed on the raw data which simplifies locating of potentially informative intervals and adds robustness to change detection. The proposed detector can also reduce interval misdetection and false alarms.
- Upper interval bounds are determined by a local approach where the conditioning of the sample information matrix is evaluated. Several criteria have been proposed in the literature for upper interval bounding. Normally, a norm retrieved from the information matrix is thresholded with a user-defined parameter that is set without considering the dependency between the value of the norm and the amplitude of the changes in the input signals. Thus, intervals resulting from small changes in the excitation signals may not be detected if the threshold was wrongly chosen. The



proposed criterion for interval bounding can decrease misdetection as local information of each retrieved interval is used instead of using global thresholds.

- The number of user-defined parameters is reduced since only two tests are evaluated. Current approaches require a demanding and time-consuming setting of parameters used in several tests. In the proposed method, user-defined parameters correspond mainly to choosing thresholds for change detection, upper interval bounding and model order. Thus, a more practical selection method is offered when compared with other techniques.
- A model is estimated from retrieved informative intervals using a data merging method robust to colored noise. Current approaches do not estimate a model with the retrieved intervals. Thus, performance of the selection method cannot be evaluated based on the resulting model. In this work, evaluation criteria are proposed to compare parameter estimation with informative sequences and with the entire data set. The performance of the proposed method is evaluated on data sets obtained from a simulated binary distillation column and a lab-scale process unit.
- A recursive instrumental variables (RIV) method is used for upper interval bounding which is robust to colored noise. It is computationally less demanding when compared to other approaches reported for data selection methods. Depending on the choice of instruments, the computed information matrix can have smaller size than the ones found for other techniques. As a consequence, faster computations are expected in comparison with other methods such as QR-RLS. This represents a significant advantage for processing mass data.
- A non-linear data-driven modeling approach in which the model output is computed as the contribution of several linear models associated to each informative interval is proposed. A measure related to the “degree of information” for each interval is determined based on the conditioning of the sample information matrix. The output of the  $i$ -th linear model is weighted by its respective factor  $w_i$  which reflects its contribution in the global output.

## 1.4 Thesis Outline

This thesis is divided into seven chapters where data selection is covered from problem statement to evaluation of the proposed approach. Suitable approaches for the different tasks in data selection are discussed in the first chapters. Then, the proposed selection method, DS4SID, is presented and its working principle

is explained with an example. The last chapter is dedicated to discussion and evaluation of the method.

In [Chapter 2](#), detection methods and time series modeling are introduced. Determining lower interval bounds is treated as a detection problem. Different aspects related to upper interval bounding are introduced in [Chapter 3](#). This chapter also deals with modeling of multivariate processes. Then, parameter estimation methods are evaluated for possible use in data selection. In [Chapter 4](#), data merging for system identification is introduced. Different techniques are explored for model estimation from the retrieved informative intervals.

The proposed method is presented in [Chapter 5](#) and explained using a simulation example. Reduction of misdetection and false alarms are discussed for each of the two tests proposed for data selection. The performance of the method is evaluated in [Chapter 6](#) using multivariate processes. A process unit of the lab-scale factory  $\mu$ Plant, located in the Department of Measurement and Control (MRT) in the University of Kassel, is used as a real case. Conclusions and future research opportunities are formulated in [Chapter 7](#).

## 1.5 Publications

Some results of this thesis were published in peer-reviewed issues. In [\[2\]](#), a detection method with different sliding windows for input and output signals is presented. Benefits of parameter estimation using informative intervals is evaluated on bias and variance analysis of the estimates. An initial approach for multivariate data selection is introduced in [\[1\]](#). Statistical analysis on estimates using informative data is extended to multivariate processes. The sample information matrix is retrieved from the OLS method applied for estimation of ARX models. Benefits in the quality of estimates using informative intervals is discussed for additive noise with different signal-to-noise ratio (SNR). Several possible criteria for upper interval bounding are explored in [\[3\]](#). Among others, trace, smallest eigenvalue and reciprocal of the condition number are proposed for thresholding. Use of the reciprocal of the condition number has some advantages over other scalar measures. This quantity is bound between 0 and 1 which simplifies threshold regardless of the application.

In [\[4\]](#), a robust method to colored noise is used for upper interval bounding. This approach is compared with other methods available in the literature using a multivariate case study. A new data selection method for multivariate processes is presented in [\[5\]](#). Drawbacks of current techniques are discussed and the proposed method is contrasted with current techniques. A case study of a lab-scale process plant is used for evaluation of the developed method under conditions that simulate real scenarios.



# CHAPTER 2

---

## Change Detection

---

Determining lower bounds of potentially informative intervals is a similar task to the detection problem found in statistical hypothesis testing. The detection problem in the context of data selection for identification is to locate changes in the process that result from an external excitation. Two hypotheses are evaluated for locating of such changes that differ based on the model used for process description in each situation. Thus, a detection test should choose between one of both hypotheses. In this chapter, several detection methods are considered for locating informative intervals. Transfer functions for time series are briefly introduced for modeling stochastic processes.

### 2.1 Stochastic Processes

A stochastic or random process is a sequence of random variables ordered in time [45, 47]. A random variable is represented by  $x$  and following the notation proposed in [11], the probabilities of the values that  $x$  can take are defined by its probability density function (PDF) denoted by  $p_x(x)$ . For instance, the probability that  $x$  takes values within the interval  $[a, b]$  is expressed as follows

$$\mathbf{P}(a \leq x \leq b) = \int_a^b p_x(x) dx, \quad x \in \mathbb{R} \quad (2.1)$$

A random variable taking values in an interval  $[a, b]$  is referred to as uniform or uniformly distributed if its PDF is defined as follows

$$p_x(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b, \\ 0, & \text{otherwise} \end{cases} \quad (2.2)$$

A random variable can be described by its first moment and second central moment, namely, expectation and variance, respectively [33]. The expected

value of a random variable  $x$  is the quantity to which the average converges:

$$E(x) = \mu_x = \int_{-\infty}^{\infty} xp_x(x)dx \quad (2.3)$$

The expected value is also known as expectation, mean or average.

The second central moment, also known as *variance* and denoted by  $\text{Var}(x)$ , is given by:

$$\text{Var}(x) = E\left((x - E(x))^2\right) = E(x^2) - (E(x))^2 \quad (2.4)$$

The variance describes the average squared deviation from the mean.

The *Gaussian* distribution also called *normal* distribution and denoted by  $\mathcal{N}(\mu, \sigma^2)$ , is of major interest in signal processing. The PDF of the normal distribution is:

$$p_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) \quad -\infty < x < \infty \quad (2.5)$$

where  $\sigma^2 > 0$  and  $-\infty < \mu < \infty$ . The parameters  $\mu$  and  $\sigma^2$  correspond to the mean and variance of the random variable  $x$ .

Consider that the random variables  $x_1$  and  $x_2$  are related to the same experiment. The probabilities of the values of  $x_1$  and  $x_2$  are described by the joint PDF, denoted by  $p_{x_1, x_2}(x_1, x_2)$ . In particular, the joint PDF that the previous random variables take the values  $(x_1, x_2)$  is described by

$$p_{x_1, x_2}(x_1, x_2) = \mathbf{P}(x_1 = x_1, x_2 = x_2) \quad (2.6)$$

Two random variables  $x_1$  and  $x_2$  are independent if

$$p_{x_1, x_2}(x_1, x_2) = p_{x_1}(x_1)p_{x_2}(x_2) \quad (2.7)$$

holds. Let  $\mathbf{x} = \{x_0, \dots, x_{N-1}\}$  denote a vector of  $N$  random variables. Assume that  $\mathbf{x}$  takes the particular value  $\mathbf{x} = \{x_0, \dots, x_{N-1}\}$ . The joint PDF of the random variables of  $\mathbf{x}$  which is described by an unknown parameter  $\theta$  will be denoted by [11]

$$p_{\mathbf{x}}(\mathbf{x}; \theta) = p_{\mathbf{x}}(x_0, \dots, x_{N-1}; \theta) \quad (2.8)$$

The term  $p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta})$  in (2.8) is called the likelihood function. Considering that the random variables of the sequence  $\mathbf{x}$  are independent, the joint PDF is:

$$p_{\mathbf{x}}(x_0, \dots, x_{N-1}; \boldsymbol{\theta}) = \prod_{i=0}^{N-1} p_{x_i}(x_i; \boldsymbol{\theta}) \quad (2.9)$$

The expected value of a random vector,  $\mathbf{x}$ , is a vector that contains the expectations of each stochastic variable [47, 52]:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}; \quad \boldsymbol{\mu}_{\mathbf{x}} = E(\mathbf{x}) = \begin{pmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_N) \end{pmatrix} \quad (2.10)$$

The covariance matrix  $\mathbf{C}_{\mathbf{x}}$  of a random vector  $\mathbf{x}$  is

$$\begin{aligned} \mathbf{C}_{\mathbf{x}} &= E((\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T) \\ &= \begin{pmatrix} \text{Var}(x_1) & \text{cov}(x_1, x_2) & \dots & \text{cov}(x_1, x_N) \\ \text{cov}(x_2, x_1) & \text{Var}(x_2) & \dots & \text{cov}(x_2, x_N) \\ \vdots & & & \vdots \\ \text{cov}(x_N, x_1) & \text{cov}(x_N, x_2) & \dots & \text{Var}(x_N) \end{pmatrix} \end{aligned} \quad (2.11)$$

where  $\boldsymbol{\mu}_{\mathbf{x}}$  is the mean of  $\mathbf{x}$  as defined in (2.10). The term  $\text{cov}(x_i, x_j)$ ,  $i \neq j$ , is called the covariance between the random variables  $x_i$  and  $x_j$  and is defined as follows:

$$\text{cov}(x_i, x_j) = E((x_i - E(x_i))(x_j - E(x_j))) \quad (2.12)$$

Alternatively, the covariance can be written as

$$\text{cov}(x_i, x_j) = E(x_i x_j) - E(x_i) E(x_j) \quad (2.13)$$

Consider that  $N$  experiments are performed sequentially at different time instants  $k = 0, 1, \dots, N-1$ . The result of each experiment is associated to a random variable  $x[k]$  where the argument in the square brackets is explained further. The random variables  $\{x[0], x[1], \dots, x[N-1]\}$  are independent and each normally distributed with mean values  $\{\mu_0, \mu_1, \dots, \mu_{N-1}\}$  and variances  $\{\sigma_0^2, \sigma_1^2, \dots, \sigma_{N-1}^2\}$ . Each random variable is denoted by  $x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$  and included in the vector  $\mathbf{x} = (x[0] \ x[1] \ \dots \ x[N-1])^T$ . Since the random variables

are independent, the PDF of  $\mathbf{x}$  can be expressed using (2.9) as follows:

$$\begin{aligned} p(x[0], \dots, x[N-1]) &= \prod_{i=0}^{N-1} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \\ &= \frac{1}{\left(\prod_{i=0}^{N-1} \sigma_i\right) (2\pi)^{N/2}} \exp\left(-\frac{1}{2} \sum_{i=0}^{N-1} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right) \end{aligned} \quad (2.14)$$

The PDF of a multivariate Gaussian random vector is expressed by

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} \det^{1/2}(\mathbf{C}_{\mathbf{x}})} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \mathbf{C}_{\mathbf{x}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})\right) \quad (2.15)$$

where  $\boldsymbol{\mu}_{\mathbf{x}}$  and  $\mathbf{C}_{\mathbf{x}}$  are the mean vector and covariance matrix as defined in (2.10) and (2.11), respectively. The covariance matrix is assumed to be positive definite, thus,  $\mathbf{C}_{\mathbf{x}}$  is invertible and  $\det(\mathbf{C}_{\mathbf{x}}) > \mathbf{0}$  holds.

Since sensor measurements are normally affected by noise, signals treated in this work will be mostly assumed to be stochastic. Some examples of signals are the level and temperature of the inventory of a continuously stirred tank reactor (CSTR). Discrete or continuous in time random processes can be modeled by time series models that are also called rational transfer function models [48, 60]. In the present work, conventions for continuous- and discrete-time random processes are taken from [49, 68]. The symbol  $t$  denotes continuous-time processes, whereas  $k$  is used for discrete-time processes. Moreover, these processes are additionally distinguished by enclosing the independent variable  $t$  in parenthesis  $(\cdot)$  whereas the independent variable  $k$  will be enclosed by brackets  $[\cdot]$ . A discrete random process consisting of  $N$  observations is represented by [67]:

$$\{x[k]\} = \{x[0], x[1], \dots, x[N-1]\}, k \in \mathbb{Z}^+ \quad (2.16)$$

The present work focuses on discrete-time systems. Nevertheless, the notation of a continuous random process is briefly introduced and denoted by

$$\{x(t)\} = \{x(t_0), x(t_1), \dots, x(t_{N-1})\}, t \in \mathbb{R} \quad (2.17)$$

The random processes in (2.16) and (2.17) are related by

$$x(t) = x[T_s \cdot k] \quad (2.18)$$

where  $T_s$  is the sampling time used for collection of the measurements. In [18, 41, 61, 65], the selection of  $T_s$  is discussed. For instance, as reported in [41], the sampling time can be chosen as

$$T_s \approx k_s \cdot T_{95} \quad (2.19a)$$

$$k_s = \frac{1}{5}, \dots, \frac{1}{15} \quad (2.19b)$$

where  $T_{95}$  is the 95% settling time of the step response of a proportional acting process.

A sequence of independent and identically distributed (i.i.d.) random variables of zero-mean and variance  $\sigma^2$  is known as white Gaussian noise (WGN) [80]. Figure 2.1 shows different realizations of WGN on the  $x-t$  plane.

Consider a DC level subject to WGN expressed by

$$x[k] = A + w[k] \quad k = 0, 1, \dots, N-1 \quad (2.20)$$

where  $A$  is a DC level to be estimated from data and  $\{w[k]\}$  is WGN with unknown variance. Then, the parameter vector to be determined is  $\Theta^T = (A \ \sigma^2)$ .

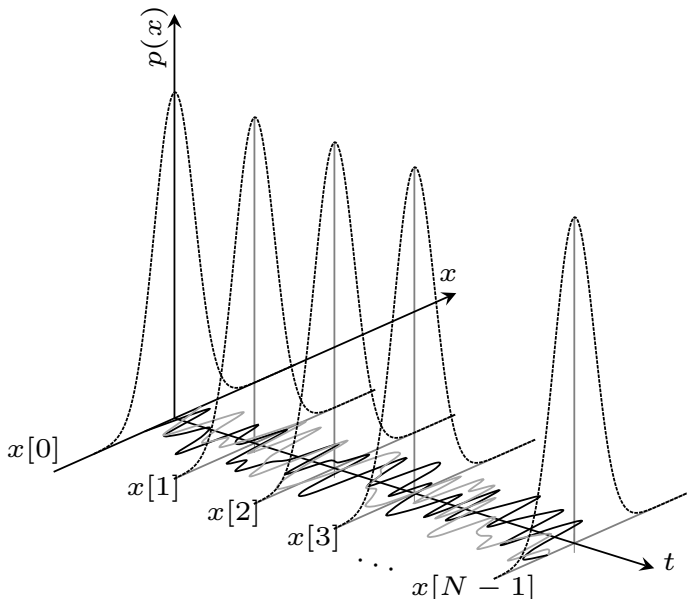


Figure 2.1: Different realizations of WGN



The likelihood function is expressed by

$$p_{\mathbf{x}}(\mathbf{x}; A, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} (x[k] - A)^2\right) \quad (2.21)$$

and the log-likelihood function of (2.21) is

$$\ln p_{\mathbf{x}}(\mathbf{x}; \Theta) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=0}^{N-1} (x[k] - A)^2 \quad (2.22)$$

The estimates of  $\Theta$  are obtained by setting to zero the partial derivatives of (2.22) w.r.t  $\Theta$  (see, for instance, [49] for derivation details).

The estimated parameter vector is

$$\hat{\Theta} = \begin{pmatrix} \hat{A} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{N} \sum_{k=0}^{N-1} x[k] \\ \frac{1}{N} \sum_{k=0}^{N-1} (x[k] - \hat{A})^2 \end{pmatrix} \quad (2.23)$$

$\hat{A}$  and  $\hat{\sigma}^2$  are known as the sample mean and the sample variance, respectively.

Frequently, more complex models as in (2.20) may be required to describe a random process. Some widely known parametric models are addressed in the next section.

## 2.2 Parametric Modeling of Time Series

A large number of discrete-time random processes found in practice can be well described by rational transfer functions, also known as time series models [14, 48]. Consider a time series model, where  $\{x[k]\}$  and  $\{e[k]\}$  are related by

$$x[k] = -\sum_{n=1}^{n_a} a[n] x[k-n] + \sum_{n=0}^{n_c} c[n] e[k-n] \quad (2.24)$$

The model in (2.24) is called ARMA and is used to describe a large number of time series found in practice [48, 78]. This section will focus on the use of ARMA models for description of time series. The terms  $n_a$  and  $n_c$  are the number of parameters of the AR and moving average (MA) polynomials, respectively. The transfer function between  $e[k]$  and  $x[k]$  for an ARMA model is

$$H(q^{-1}, \Theta) = \frac{C(q^{-1}, \theta_c)}{A(q^{-1}, \theta_a)} \quad (2.25a)$$

$$A(q^{-1}, \theta_a) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a} \quad (2.25b)$$

$$C(q^{-1}, \theta_c) = 1 + c_1 q^{-1} + \dots + c_{n_c} q^{-n_c} \quad (2.25c)$$

If  $C(q^{-1}, \theta_c) = 1$  in (2.25a), then  $x[k]$  is an  $\text{AR}(n_a)$  model. In the case of  $A(q^{-1}, \theta_a) = 1$ ,  $x[k]$  results in a  $\text{MA}(n_c)$  model. For ease of use, the parameter vectors of the terms in (2.25b) and (2.25c) will be removed. The latter time series models are summarized as follows:

$\begin{aligned} \text{ARMA : } & A(q^{-1}) y[k] = C(q^{-1}) e[k] \\ \text{AR : } & A(q^{-1}) y[k] = e[k] \\ \text{MA : } & y[k] = C(q^{-1}) e[k] \end{aligned}$	(2.26)
--	--------

Assume that  $x[k]$  can be described by an  $\text{AR}(n_a)$  model. From (2.24), the resulting equation is

$$x[k] = - \sum_{n=1}^{n_a} a[n] x[k-n] + e[k] \quad (2.27)$$

The problem of linear prediction is to predict the value of  $x[k]$  based on given past values  $\{x[k-1], x[k-2], \dots, x[k-n_a]\}$ . A linear predictor, that is function of past samples, is expressed by

$$\hat{x}[k] = - \sum_{l=1}^{n_a} \hat{a}_l x[k-l] \quad (2.28)$$

The coefficients  $\{a_1, a_2, \dots, a_{n_a}\}$  are chosen in order to minimize the power of the prediction error  $\epsilon[k]$  as found in [48]:

$$E(|\epsilon[k]|^2) = E(|x[k] - \hat{x}[k]|^2) \quad (2.29)$$

Note that the prediction error is a variable term. The notation  $\epsilon[k]$  has been used following [80]. If the orders of the process (2.27) and the linear predictor (2.28) are identical, the prediction error,  $\epsilon[k]$ , is equal to  $e[k]$  in (2.27) since

$$\begin{aligned} \epsilon[k] &= x[k] - \hat{x}[k] = x[k] - \left( - \sum_{l=1}^{n_a} \hat{a}_l x[k-l] \right) \\ &= x[k] + \left( \sum_{l=1}^{n_a} \hat{a}_l x[k-l] \right) \\ &= e[k] \end{aligned} \quad (2.30)$$

Let  $\hat{\Theta} = [\hat{a}_1 \hat{a}_2 \dots \hat{a}_{n_a}]^T$  be the parameter vector of (2.28). The linear predictor (2.28) is expressed in a matrix form as follows

$$\hat{\mathbf{x}} = \Phi \hat{\Theta} \quad (2.31a)$$

$$\Phi = \begin{pmatrix} x[n_a] & \dots & x[1] \\ x[n_a + 1] & \dots & x[2] \\ \vdots & & \vdots \\ x[N - 1] & \dots & x[N - n_a] \end{pmatrix} \quad (2.31b)$$

$$\hat{\Theta} = \begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_{n_a} \end{pmatrix} \quad (2.31c)$$

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}[n_a + 1] \\ \hat{x}[n_a + 2] \\ \vdots \\ \hat{x}[N] \end{pmatrix} \quad (2.31d)$$

Using (2.31b) and (2.31c), the resulting likelihood function is

$$p(\mathbf{x}; \Theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{x} - \Phi\Theta)^T (\mathbf{x} - \Phi\Theta)\right) \quad (2.32)$$

The maximum likelihood estimator (MLE) is found by minimizing

$$J(\Theta) = (\mathbf{x} - \Phi\Theta)^T (\mathbf{x} - \Phi\Theta) \quad (2.33)$$

The MLE of (2.32) is (see [49] for derivation details)

$$\hat{\Theta} = (\Phi^T \Phi)^{-1} (\Phi^T \mathbf{x}) \quad (2.34)$$

The estimator of the noise variance is

$$\hat{\sigma}^2 = \frac{1}{N - n_a} (\mathbf{x} - \Phi\hat{\Theta})^T (\mathbf{x} - \Phi\hat{\Theta}) \quad (2.35)$$

The MLE for an AR model corresponds to the least squares estimate (LSE) [49]. The AR model can be used in some situations to describe colored Gaussian noise. For the previous reason, the AR model and its estimator (2.34) and (2.35) is relevant for the next sections where measurement noise of dynamic systems should be considered as colored.

Estimated models should be evaluated and validated to analyze if they can properly describe the process under study. Model validation includes residual analysis to test, for instance, if the resulting residuals can be considered either as WGN or not. The following example shows tests that can be used for residual analysis. Consider a process described by:

$$x[k] = s[k] + n[k], \quad k = 0, 1, \dots, N - 1 \quad (2.36)$$

where  $\{s[k]\}$  and  $\{n[k]\}$  denote deterministic and random sequences, respectively. In (2.36),  $n[k]$  can represent, for instance, either WGN or colored Gaussian noise. The sequence  $n[k]$  is often called measurement noise in signal processing (see *e.g.* [15]). Noise sources are normally related to spurious electronic signals that influence transmitted values.

The noise content in a signal can be defined as the ratio between the variance of the signal and the variance of the noise which is expressed in dB [47]:

$$\text{SNR} = 10 \log_{10} \frac{\sigma_s^2}{\sigma_n^2} \quad (2.37)$$

where  $\sigma_n^2$  and  $\sigma_s^2$  are the variances of the noise and signal, respectively.

Consider a signal that obeys the following relation:

$$x_i[k] = s[k] + n_i[k], \quad k = 0, 1, \dots, N - 1 \quad (2.38a)$$

$$s[k] = \sqrt{2} \sin [0.2\pi k] + \sqrt{2} \sin [0.4\pi k] \quad (2.38b)$$

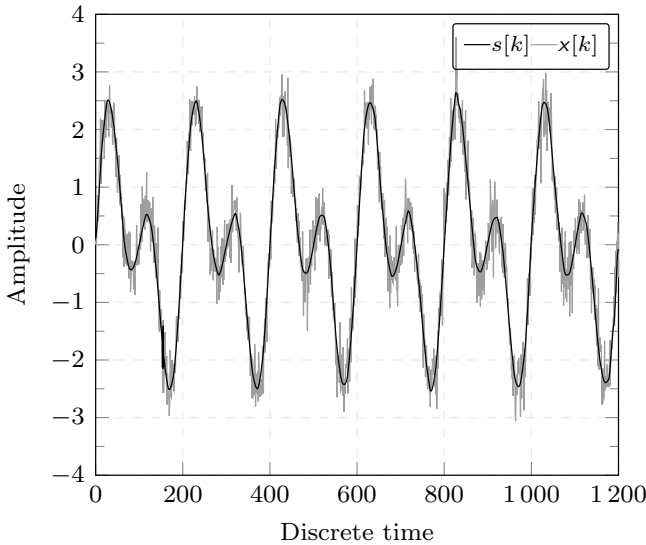
$$n_i[k] = H_i(q^{-1}) e[k] \quad (2.38c)$$

$$H_1(q^{-1}) = 1 \quad (2.38d)$$

$$H_2(q^{-1}) = \frac{1 - 0.9q^{-1}}{1 - 0.25q^{-1}} \quad (2.38e)$$

The term  $e[k]$  is WGN with  $\sigma_e^2 = 0.10$  which yields  $\text{SNR} = 13$  dB. The transfer function  $H_i(q^{-1})$  was chosen differently for  $i = 1, 2$  to analyze the effect of a WGN sequence or colored Gaussian noise. A realization of  $\{x[k]\}$  in (2.38a) for  $H_2(q^{-1})$  is shown in Figure 2.2 for  $H_2(q^{-1})$ .

Since the signal is explicitly defined in (2.38), the stochastic process  $n[k]$  can be computed as the difference between  $x[k]$  and  $s[k]$ . A histogram of  $n[k]$  and its corresponding autocorrelation are shown in Figure 2.3. The sequence  $\{n_2[k]\}$ , which corresponds to colored Gaussian noise, exhibits correlations in the first lags. In contrast, autocorrelations in  $n_1[k]$  are numerically negligible as it corresponds to WGN (see  $H_1(q^{-1})$  in (2.38d)). Sequences as  $\{n_2[k]\}$  are likely to be found in real situations because of filters embedded in sensors



**Figure 2.2:** Measured signal (2.38) with additive white noise

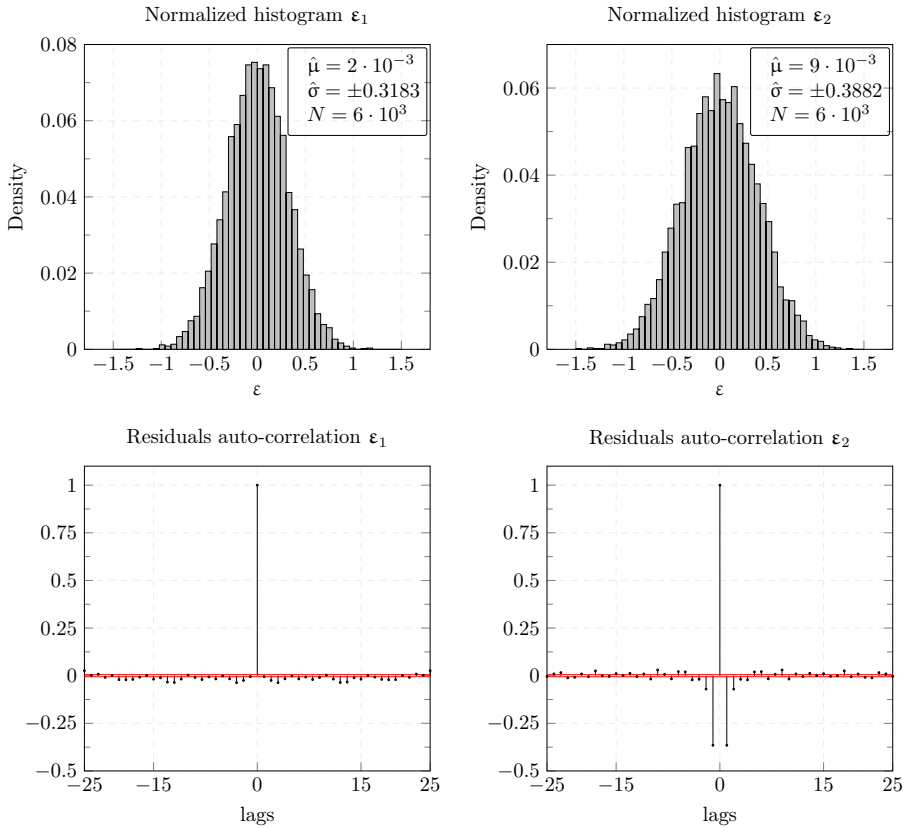
or data logging devices. This situation should be considered when assuming models for process description.

An ARMA process with  $n_a = 1$  and  $n_c = 1$  was chosen to model  $n_2[k]$ . The parameters were computed using the Newton-Raphson algorithm for minimization of the prediction error. Two hypotheses are stated to analyze the model residuals: either  $\{\epsilon[k]\}$  is WGN or not. The first situation is called null-hypothesis  $\mathcal{H}_0$  whereas the second is known as alternative hypothesis  $\mathcal{H}_1$ . The latter hypotheses are evaluated using two normality tests and the results are shown in Table 2.1.

In those tests, the null-hypothesis is confirmed or rejected by comparing a computed measure  $p$  with a threshold  $\alpha$ . Computed  $p$ -values larger than  $\alpha$  in

**Table 2.1:** Normality tests on estimated stochastic sequences in (2.38),  $\alpha = 0.05$ . Left: Shapiro-Wilk (S-W) test. Right: Kolmogorov-Smirnov (K-S) test

	S-W test		K-S test	
	$\hat{v}_1(t)$	$\hat{v}_2(t)$	$\hat{v}_1(t)$	$\hat{v}_2(t)$
$\mathcal{H}$	0	0	0	0
$p$	0.8430	0.8720	0.7586	0.6898



**Figure 2.3:** Histograms and autocorrelation of example signals

Table 2.1 confirm that the estimated ARMA can describe well  $n_2[k]$  since the residuals are normally distributed. Thus, the null-hypothesis is not rejected and the test decides for  $\mathcal{H}_0$ . Results from the Kolmogorov-Smirnov test, that are detailed in Table 2.1, also confirm that the computed residuals are normally distributed. Importance of ARMA models and normality tests will be discussed in further chapters with simulated and real signals. The K-S and S-W tests are two well-known normality tests. These tests were preferred because of their reliability to decide if a sequence is normally distributed as found in [57, 80].

## 2.3 Power Spectral Density

The power spectral density gives information about the power of a signal within a frequency band. The spectral representation of signals plays a relevant role in experiment design for identification (see, for instance, [44] for an optimal

experiment design in closed loop). The power spectral density is defined as the discrete Fourier transform of the covariance function [78, 80, 84]:

$$\phi(\omega) \triangleq \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} r(\tau) e^{-i\tau\omega} \quad (2.39)$$

The expression in (2.39) is derived for discrete time processes following [84]. The square brackets introduced in section 2.1 for the discrete time arguments will not be used in this section aiming at following the literature in spectral estimation [48, 84].

The covariance function of a white noise process  $\{\mathbf{e}(k)\}$  is

$$r_e(\tau) = \sigma_e^2 \delta_{\tau,0} \quad (2.40)$$

where  $\delta_{\tau,0}$  is the Kronecker's delta defined by [80]:

$$\delta_{\tau,0} = \begin{cases} 1; & \tau = 0 \\ 0; & \tau \neq 0 \end{cases} \quad (2.41)$$

From (2.39), the spectrum of a white noise process with variance  $\sigma_e^2$  is:

$$\phi(\omega) = \frac{\sigma_e^2}{2\pi} \quad (2.42)$$

As expressed by (2.42), the spectral density of a white noise process is the same over the entire frequency range. Consider an ARMA process described by

$$y(k) + a_1 y(k-1) + \dots + a_{n_a} y(k-n_a) = \mathbf{e}(k) + c_1 \mathbf{e}(k-1) + \dots + c_{n_c} \mathbf{e}(k-n_c) \quad (2.43)$$

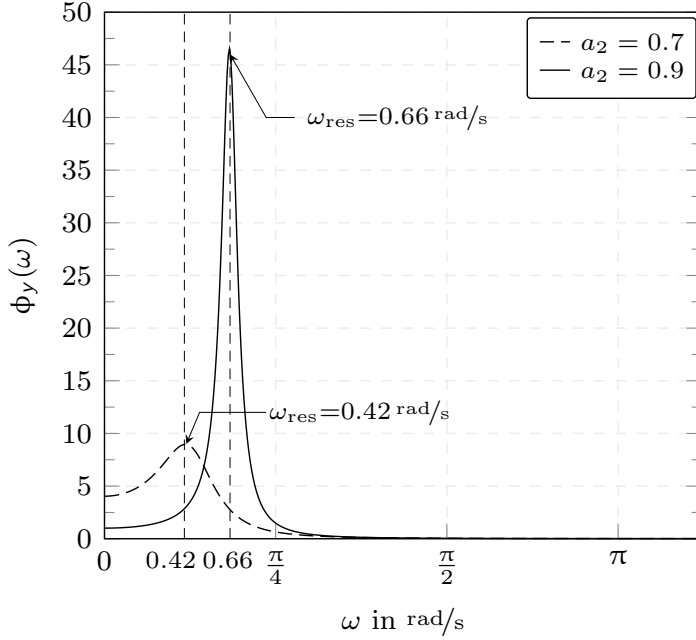
where  $\{\mathbf{e}(k)\}$  is WGN with variance  $\sigma_e^2$ . The power spectral density of (2.43) is:

$$\phi(\omega) = \frac{\sigma_e^2}{2\pi} \frac{A(e^{i\omega})A(e^{-i\omega})}{C(e^{i\omega})C(e^{-i\omega})} = \frac{\sigma_e^2}{2\pi} \left| \frac{A(e^{i\omega})}{C(e^{i\omega})} \right|^2 \quad (2.44)$$

As an example, let  $n_a = n_c = 1$  in (2.43) which yields the following ARMA process

$$y(k) + ay(k-1) = \mathbf{e}(k) + c\mathbf{e}(k-1) \quad |a| < 1, c \neq 0 \quad (2.45a)$$

$$E[\mathbf{e}(t)\mathbf{e}(s)] = \sigma_e^2 \delta_{t,s} \quad (2.45b)$$



**Figure 2.4:** Spectral density of the ARMA process (2.49)

The power spectral density of (2.45) is:

$$\phi(\omega) = \frac{\sigma_e^2}{2\pi} \frac{1 + c^2 + 2c \cos \omega}{1 + a^2 + 2a \cos \omega} \quad (2.46)$$

In general, the spectral density for the ARMA process (2.43) is:

$$\phi(\omega) = \frac{\sigma_e^2}{2\pi} \frac{1 + \phi_{\text{num}}(c, \omega)}{1 + \phi_{\text{den}}(a, \omega)} \quad (2.47)$$

where the numerator and denominator are computed by:

$$\phi_{\text{num}}(c, \omega) = \sum_{n=1}^{n_c} c_n^2 + 2 \sum_{n=1}^{n_c} c_n \cos(n\omega) + 2 \sum_{n=1}^{n_c-1} \sum_{m=n+1}^{n_c} c_n c_m \cos((m-n)\omega) \quad (2.48a)$$

$$\phi_{\text{den}}(a, \omega) = \sum_{n=1}^{n_a} a_n^2 + 2 \sum_{n=1}^{n_a} a_n \cos(n\omega) + 2 \sum_{n=1}^{n_a-1} \sum_{m=n+1}^{n_a} a_n a_m \cos((m-n)\omega) \quad (2.48b)$$



The power spectral density of an ARMA process can be computed by (2.48) and depends on the parameters of the polynomials  $C(q^{-1})$  and  $A(q^{-1})$  and the noise variance  $\sigma_e^2$ .

As an example of the effect of the parameters in the power spectral density (PSD), consider the following AR model:

$$(1 - 1.5q^{-1} + a_2q^{-2})y(k) = e(k) \quad (2.49)$$

where  $\{e[k]\}$  is white noise with variance  $\sigma_e^2 = 1$ . The power spectral density of (2.49) for two different values of  $a_2$  are shown in Figure 2.4. The value of  $\phi_y(\omega)$  is concentrated around the resonant frequency. Moreover, the PSD is flatter for smaller values of  $a_2$ .

## 2.4 Detection Problem

Change detection is a well-known problem in statistical hypothesis testing [10, 50, 87] that has been usually applied to time series. In radar applications, the problem is to determine the presence or absence of an approaching aircraft. In speech recognition, the task consists in determining a spoken word from among a set of possible words. More recently, anomaly detection in IP networks for denial of service (DoS) attacks has emerged as an application area in this field. In all these applications, the problem consists in choosing a decision between hypotheses based on observations.

Anomaly detection can be extended to dynamic processes where the task is to find changes from a given operating point. Input-output models are used for description of dynamic processes and a detector should choose between the following hypotheses: either the process is operating at the same point or was moved away from it. Change detection in dynamic processes is relevant for data selection of informative intervals since changes in the process may be associated with excitation of the system. The following  $N$ -point data set is assumed to be available:

$$\mathbf{z}^N = \{y[0], u[0], y[1], u[1], \dots, y[N-1], u[N-1]\} \quad (2.50)$$

where  $\{u[k]\}$  and  $\{y[k]\}$  represent the input and output of process, respectively. In contrast to change detection in time series, in this work, the process input  $\{u[k]\}$  is also analyzed since excitations in the system are associated to changes in its input. This, represents an advantage because a change in the output  $\{y[k]\}$  can be related to the input  $\{u[k]\}$ . Hypotheses are established based on a collected data set. Then, a detection test,  $T$ , is computed using the data

$T(y[0], u[0], \dots, y[N-1], u[N-1])$ . Finally, one of the hypothesis is chosen based on the value of  $T$ .

As an introductory problem, consider the detection of a DC level in WGN. In the null-hypothesis,  $\mathcal{H}_0$ , the observed sequence is WGN and in  $\mathcal{H}_1$  the observed sequence corresponds to a DC level in WGN. The hypotheses are:

$$\begin{aligned} \mathcal{H}_0 : x[k] &= e[k] & k &= 0, 1, \dots, N-1 \\ \mathcal{H}_1 : x[k] &= A + e[k] & k &= 0, 1, \dots, N-1 \end{aligned} \quad (2.51)$$

where  $A$  is unknown with  $-\infty < A < \infty$  and  $e[k]$  is WGN with unknown variance  $\sigma_e^2$ . Thus, the problem is to decide between  $\mathcal{H}_0$  and  $\mathcal{H}_1$  when the PDF for both hypotheses are unknown. The decision problem when the parameters are unknown is called composite hypothesis testing. This problem is addressed in the present work since the model parameters are estimated as part of the detection problem.

The detection problem (2.51) can be also expressed as

$$\begin{aligned} \mathcal{H}_0 : A &= 0, \sigma_e^2 > 0 \\ \mathcal{H}_1 : A &\neq 0, \sigma_e^2 > 0 \end{aligned} \quad (2.52)$$

The likelihood functions for each hypothesis are given as (2.21). Although,  $\sigma_e^2$  is not the central part of the problem, it needs to be estimated since it affects the likelihood functions. The parameter vector is then  $\Theta^T = [A \ \sigma_e^2]$ .

The expressions in (2.51) and (2.52) represent the foundation to formulate the detection problem. Dynamic systems are normally described by other models than used in (2.51). Thus, since the focus of this work is the selection of informative data in dynamic systems, the detection problem will be adjusted using models that can describe such dynamics as it will be presented in section 5.3. As reported in [50], the choice of a detection test depends on the application and the available information. One of the most known detectors is the generalized likelihood ratio test (GLRT) which can be applied to the problem (2.51). In the GLRT, the unknown parameters are replaced by their MLEs. The GLRT chooses  $\mathcal{H}_1$  if

$$T(\mathbf{x}) = L_G(\mathbf{x}) = \frac{p(\mathbf{x}; \hat{A}, \hat{\sigma}_{e_1}^2, \mathcal{H}_1)}{p(\mathbf{x}; \hat{\sigma}_{e_0}^2, \mathcal{H}_0)} > \gamma \quad (2.53)$$

where  $\hat{A}$  and  $\hat{\sigma}_{e_1}^2$  are the MLEs of the parameters under  $\mathcal{H}_1$ . Whereas,  $\hat{\sigma}_{e_0}^2$  is the MLE of  $\sigma_e^2$  under  $\mathcal{H}_0$ .

The PDF under  $\mathcal{H}_1$  is found by maximizing (2.21). The MLEs of  $\Theta$  under  $\mathcal{H}_1$  were found in (2.23). Moreover, under  $\mathcal{H}_0$ , the PDF is obtained by maximizing

$$p(\mathbf{x}; \hat{\sigma}_{e_0}^2, \mathcal{H}_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} x^2[k]\right) \quad (2.54)$$

The MLE of  $\hat{\sigma}_{e_0}^2$  under  $\mathcal{H}_0$  is

$$\hat{\sigma}_{e_0}^2 = \frac{1}{N} \sum_{k=0}^{N-1} x^2[k] \quad (2.55)$$

Using the derived parameters under each hypothesis, the test  $L_G(\mathbf{x})$  results in (see [50] for details on the derivation)

$$L_G(\mathbf{x}) = \left(\frac{\hat{\sigma}_{e_0}^2}{\hat{\sigma}_{e_1}^2}\right)^{N/2} \quad (2.56)$$

The expression in (2.53) is known as a detector. Two main types of errors can occur when implementing detectors. A false alarm or type I error occurs in case (2.53) decides for  $\mathcal{H}_1$  when  $\mathcal{H}_0$  is true. In contrast, a misdetection or type II error occurs in case (2.53) decides for  $\mathcal{H}_0$  when  $\mathcal{H}_1$  is true. A good detector should have a low rate of false alarms as well as low misdetection rate.

Detection of informative intervals in logged data differs from (2.52) in that dynamic models are used in each hypothesis and the problem becomes to decide if the model has changed or not. In selection of informative intervals for identification, the observation data set  $\mathbf{Z}^N$  needs to be evaluated sequentially to search for changes in the process. Moreover, change detection should be multi cyclic which means that the algorithm is reinitialized each time a change is located. These conditions require adaption of the GLRT test to develop multi cyclic detection methods.

The CUSUM test is a procedure for change detection that uses the likelihood ratio of the PDF associated to each hypotheses. Use of the CUSUM test for change detection in time series has been reported in [9, 22, 87] In section 3.5, the CUSUM test is extended to dynamic processes modeled by input-output models. For a further treatment of this topic see [10, 34, 87].

The following derivation of the CUSUM test is based on [30]. Let  $\{x[k]\}$ ,  $k = 0, 1, \dots, N-1$  be a sequence of independent and identically distributed random variables where  $x[0]$  and  $x[N-1]$  represent the first and the last observation,

respectively. Each variable of  $\{x[k]\}$  has the following PDF

$$p_x(x[k]; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x[k] - A)^2\right) \quad (2.57)$$

where  $\theta = [A, \sigma^2]^T$ ,  $\sigma^2 > 0$  and  $-\infty < A < \infty$ .

A change may occur in  $\{x[k]\}$  at the time  $k_{11}$  which is represented by a variation of the parameter  $\theta$ . Thus,  $\theta = \theta_0$  for  $i < k_{11}$  and  $\theta = \theta_1$  for  $i \geq k_{11}$ .

Consider the problem of detecting a change in the parameter  $\theta$  at some unknown time  $k_{11}$  in the sequence  $\{x[k]\}$ . The corresponding hypotheses are:

$$\begin{aligned} \mathcal{H}_0 : x[k] &= A_0 + e[k], & e[k] &\sim \mathcal{N}(0, \sigma_0^2), & k &= 0, 1, \dots, k_{11} - 1 \\ \mathcal{H}_1 : x[k] &= A_1 + e[k], & e[k] &\sim \mathcal{N}(0, \sigma_1^2), & k &= k_{11}, k_{11} + 1, \dots, N - 1 \end{aligned} \quad (2.58)$$

The log-likelihood ratio of  $x[k]$  considering the hypotheses in (2.58) is derived based on (2.57) as follows

$$s[k] = \ln \left( \frac{p_x(x[k]; \theta_1)}{p_x(x[k]; \theta_0)} \right) = \ln \left( \frac{\frac{1}{(2\pi\sigma_1^2)^{1/2}} \exp\left(-\frac{1}{2\sigma_1^2} (x[k] - A_1)^2\right)}{\frac{1}{(2\pi\sigma_0^2)^{1/2}} \exp\left(-\frac{1}{2\sigma_0^2} (x[k] - A_0)^2\right)} \right) \quad (2.59a)$$

$$= \ln \left( \left( \frac{\sigma_0^2}{\sigma_1^2} \right)^{1/2} \exp \left( \frac{1}{2\sigma_0^2} (x[k] - A_0)^2 - \frac{1}{2\sigma_1^2} (x[k] - A_1)^2 \right) \right) \quad (2.59b)$$

$$= \frac{1}{2} \ln \frac{\sigma_0^2}{\sigma_1^2} + \frac{1}{2\sigma_0^2} (x[k] - A_0)^2 - \frac{1}{2\sigma_1^2} (x[k] - A_1)^2 \quad (2.59c)$$

$$= \frac{1}{2} \ln \frac{\sigma_0^2}{\sigma_1^2} + \frac{\varepsilon_0^2[k]}{2\sigma_0^2} - \frac{\varepsilon_1^2[k]}{2\sigma_1^2} \quad (2.59d)$$

where  $\varepsilon_i[k] = x[k] - A_i$ ,  $i = 0, 1$ . Following [10, 30], the CUSUM algorithm is summarized:

$$g[k] = (g[k-1] + s[k])^+ \quad (2.60a)$$

$$(g[k-1] + s[k])^+ = \max(0, g[k-1] + s[k]) \quad (2.60b)$$

$$\text{If } g[k] > \gamma_c \text{ then } k_{11} = k \text{ (change detected)} \quad (2.60c)$$

The algorithm is initialized with  $s[0] = g[0] = 0$ .

In general, the value of  $\theta_0$  and  $\theta_1$  are unknown and they are replaced by their maximum likelihood estimates using the sequence  $\{x[0], x[1], \dots, x[k-1]\}$ . The parameters under each hypotheses are computed using different data sets determined by data windows as proposed in [10] and discussed in section 5.3. The

beginning of potentially informative intervals will be associated with change times determined by the detectors such as the CUSUM. Some parametric dynamic models will be introduced in [Chapter 3](#) and an extension of detectors to these models will be treated in [Chapter 5](#).

Data selection methods as described in [12] implement detection tests to time series even though the studied process is a dynamic system. As a consequence, additional tests are required to determine if a detected change was caused by an excitation of the process. In those detectors, the sample mean and variance are compared with thresholds which can result in false alarms for signals with high noise content. Misdetection can mainly occur in tests used for upper interval bounding. These drawbacks will be further discussed in [Chapter 5](#).

## 2.5 Discussion

In this chapter, a brief introduction on stochastic processes, time series models and detection problems is presented. A signal is a time-indexed collection of process measurements. Measurements collected by sensors contain usually a noise component which results, for instance, from the electronics itself. Thus, signals from dynamical systems may be considered as realizations of stochastic processes. AR models are of special interest since they can be used to describe colored Gaussian noise which is more likely to be found in real applications. The MLE of AR models can be computed using the LSM which is simple to be implemented.

The GLRT was introduced as a detection method. This approach can be extended to dynamical systems. In data selection for system identification, detection of such instants is of interest because they represent possibly informative data sequences as a result of excitation in the process. The GLRT is a well-known test which is computed as the ratio between the MLEs under each hypothesis. Alternatively, the CUSUM test can be used for change detection based on the prediction errors.

A detection problem is stated as a decision between different hypotheses. The problem of detecting a change in a Gaussian process was presented and detectors were derived. In the null-hypothesis,  $\mathcal{H}_0$ , the mean value is assumed to be close to zero. Whereas in the alternative hypothesis,  $\mathcal{H}_1$ , the mean value is assumed to be apart from zero. Lower bounds of informative intervals can be determined using detectors as presented in this section. However, they should be adapted for dynamic systems described by input-output models *i.e.* the data set consists of an input and an output sequence.

# CHAPTER 3

---

## System Identification of Parametric Models

---

This chapter introduces mathematical modeling of dynamical systems. Some of the most known batch and recursive parameter estimation methods as well as their computational implementation are introduced. The content of this chapter is based on [13, 19, 57, 80].

### 3.1 Discrete-Time Systems

Most of the systems are time-continuous but process variables are collected by digital logging systems. A paired input-output observation in continuous-time is represented by  $\{z(t)\} = \{y(t), u(t)\}$ . Assuming that signals are logged with the same sampling period  $T_s$ , the following definitions are stated:

$$y[k] := y(kT_s) = y(t) \quad (3.1a)$$

$$u[k] := u(kT_s) = u(t) \quad (3.1b)$$

with  $k \in \mathbb{Z}$ ,  $t \in \mathbb{R}$  and  $\{y(t)\}, \{u(t)\} \in \mathbb{R}^{N \times 1}$  where  $N$  is the number of observations.

A difference equation expresses the output at instant  $k$  in terms of earlier values of the input and output. For a linear digital system, the latter is expressed as follows

$$y[k] + a_1 y[k-1] + \dots + a_{n_a} y[k-n_a] = b_1 u[k-1] + \dots + b_{n_b} u[k-n_b] \quad (3.2a)$$

$$y[k] + \sum_{i=1}^{n_a} a_i y[k-i] = \sum_{i=1}^{n_b} b_i u[k-i] \quad (3.2b)$$

Applying the  $z$ -transform to (3.2b) yields (see [67])

$$\mathcal{Z}\{y[k]\} + \sum_{i=1}^{n_a} a_i z^{-i} \mathcal{Z}\{y[k]\} = \sum_{i=1}^{n_b} b_i z^{-i} \mathcal{Z}\{u[k]\} \quad (3.3a)$$

$$\mathcal{Z}\{y[k]\} = \frac{\sum_{i=1}^{n_b} b_i z^{-i}}{1 + \sum_{i=1}^{n_a} a_i z^{-i}} \mathcal{Z}\{u[k]\} \quad (3.3b)$$

The transfer function from  $U(z)$  to  $Y(z)$  is defined by

$$G(z) = \frac{Y(z)}{U(z)} = \frac{\sum_{i=1}^{n_b} b_i z^{-i}}{1 + \sum_{i=1}^{n_a} a_i z^{-i}} \quad (3.4)$$

Let  $q^{-1}$  be the backward shift operator:

$$q^{-1}u[k] = u[k-1] \quad (3.5)$$

Using (3.5), the linear difference equation (3.2) can be expressed as

$$y[k] + a_1 q^{-1}y[k] + \dots + a_{n_a} q^{-n_a}y[k] = b_1 q^{-1}u[k] + \dots + b_{n_b} q^{-n_b}u[k] \quad (3.6a)$$

$$y[k] + \sum_{i=1}^{n_a} a_i q^{-i}y[k] = \sum_{i=1}^{n_b} b_i q^{-i}u[k] \quad (3.6b)$$

Analogously to (3.4), the transfer operator in (3.6) is

$$G(q^{-1}) = \frac{\sum_{i=1}^{n_b} b_i q^{-i}}{1 + \sum_{i=1}^{n_a} a_i q^{-i}} \quad (3.7)$$

The term transfer function should be reserved for the  $z$ -transform as in (3.4). However, it will also be used for expressions as in (3.7) to agree with most of the literature in system identification [29, 57, 80].

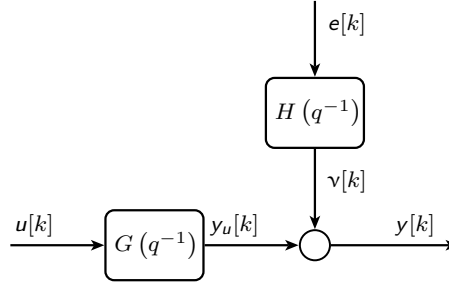
## 3.2 Parametric Models

A system transforms a set of input signals into set of output signals. Systems addressed in this work are dynamic which implies that their current states depends on past state values. A set of  $N$  collected measurements from a system is represented by

$$\mathbf{z}^N = \{u[0], y[0], u[1], y[1], \dots, u[N-1], y[N-1]\} \quad (3.8)$$

It is assumed that the system dynamics can be modeled as follows [81]

$$y[k] = f(\boldsymbol{\varphi}[k-1]) + \mathbf{v}[k], \quad k = 0, 1, \dots, N-1 \quad (3.9)$$



**Figure 3.1:** A linear dynamic system with additive disturbance

where  $\boldsymbol{\varphi}[k-1]$  is called regression vector which is constructed from past elements of  $\mathbf{Z}^k$  defined as follows:

$$\mathbf{Z}^k = \{u[k-1], u[k-2], \dots, u[k-n_b], y[k-1], y[k-2], \dots, y[k-n_a]\} \quad (3.10)$$

where  $\mathbf{n} = [n_a \ n_b]$  the model order and  $v[k]$  is a stochastic term that is usually assumed as a sequence of i.i.d. random variables *i.e.* white noise. Dynamic systems can be approximately described by mathematical models.

Processes may exhibit nonlinearities that should be accounted for in the chosen model. However, a linear time-invariant model can be used to describe a nonlinear process in the vicinity of a certain operating point [35, 91].

Figure 3.1 shows a system with  $G(q^{-1})$  and  $H(q^{-1})$  as the input and noise transfer function, respectively. A digital system is used for data logging and the resulting data set is represented by (3.8). The output sequence  $\{y[k]\}$  is assumed to be stochastic. Moreover,  $\{e[k]\}$  is WGN with  $\mathcal{N}(0, \sigma_e^2)$ . A linear time-invariant (LTI) model is used for process description. The input transfer function,  $G(q^{-1})$ , is defined in (3.7). A time series model from (2.26) is used to describe the stochastic sequence  $\{v[k]\}$ .

The one-step-ahead prediction of  $y[k]$  for a linear model is [57]

$$\hat{y}[k|k-1] = H^{-1}(q^{-1}, \hat{\boldsymbol{\theta}}) G(q^{-1}, \hat{\boldsymbol{\theta}}) u[k] + (1 - H^{-1}(q^{-1}, \hat{\boldsymbol{\theta}})) y[k] \quad (3.11)$$

where  $u[k]$  and  $y[k]$  are the process input and output, respectively. The sequence  $\{u[k]\}$  is considered stochastic as it represents a more general case. In the case of open-loop processes,  $\{u[k]\}$  can be considered as a deterministic quantity. Terms denoted by  $(\hat{\cdot})$  represent estimated values .



Thus,  $\hat{\theta}$  is an estimation of the parameter vector from data. From (3.11), the model residuals or prediction error is

$$\epsilon[k] = y[k] - \hat{y}[k|k-1] = -H^{-1}(q^{-1}, \hat{\theta}) G(q^{-1}, \hat{\theta}) u[k] + H^{-1}(q^{-1}, \hat{\theta}) y[k] \quad (3.12)$$

The term  $\epsilon[k]$  represents the part of  $y[k]$  that cannot be predicted from past data. For ease of use, the model residuals are represented in a non-cursive notation following [80]. In a good prediction model,  $\epsilon[k]$  should be as small as possible and represent white noise. Moreover, the estimated input transfer function  $G(q^{-1}, \hat{\theta})$  may be as close as possible to the true transfer function  $G(q^{-1}, \theta_0)$ . The filters  $G(q^{-1}, \hat{\theta})$  and  $H(q^{-1}, \hat{\theta})$  in (3.11) represent the input and noise transfer functions of the model, respectively, that depend on the estimated parameter vector  $\hat{\theta}$ .

If  $H(q^{-1}) = 1$  in Figure 3.1 the stochastic sequence  $\{v[k]\}$  is white noise. However, in the case of  $H(q^{-1}) \neq 1$ , the noise transfer function is an AR process and  $\{v[k]\}$  is colored noise. In real situations, a sequence corresponding to colored noise is more likely to be expected than white noise because data logging may include filtering or because the process noise is more complicated to be modeled only as WGN.

In what follows, the parameter vector  $\hat{\theta}$  will be omitted from the input and noise transfer function for short notation. Thus,  $G(q^{-1})$  and  $H(q^{-1})$ , will be used instead. Several models can be defined depending on the parameterization of the transfer functions in (3.11). Consider the following linear model called ARX:

$$A(q^{-1})y[k] = B(q^{-1})u[k] + e[k] \quad (3.13a)$$

where  $e[k]$  is white Gaussian noise and the transfer functions are:

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} \quad H(q^{-1}) = \frac{1}{A(q^{-1})} \quad (3.13b)$$

and the polynomials are defined by:

$$A(q^{-1}) = 1 + a_1q^{-1} + a_2q^{-2} + \dots + a_{n_a}q^{-n_a} \quad (3.13c)$$

$$B(q^{-1}) = b_1q^{-1} + b_2q^{-2} + \dots + b_{n_b}q^{-n_b} \quad (3.13d)$$

If  $u[k] = 0$  in (3.13) the model reduces to an AR process. As mentioned earlier, in real situations, the additive noise may be represented by colored Gaussian noise. Thus, a more general model is proposed as follows:

$$A(q^{-1})y[k] = \frac{B(q^{-1})}{F(q^{-1})}u[k] + \frac{C(q^{-1})}{D(q^{-1})}e[k] \quad (3.14)$$

with the following polynomial definitions

$$C(q^{-1}) = 1 + c_1 q^{-1} + c_2 q^{-2} + \dots + c_{n_c} q^{-n_c} \quad (3.15a)$$

$$D(q^{-1}) = 1 + d_1 q^{-1} + d_2 q^{-2} + \dots + d_{n_d} q^{-n_d} \quad (3.15b)$$

$$F(q^{-1}) = 1 + f_1 q^{-1} + f_2 q^{-2} + \dots + f_{n_f} q^{-n_f} \quad (3.15c)$$

In some processes, the system dynamics from  $u[k]$  to  $y[k]$  contain a delay of  $n_k$  samples. Thus, it is convenient to rewrite (3.14) as:

$$A(q^{-1})y[k] = \frac{B(q^{-1})}{F(q^{-1})}\bar{u}[k] + \frac{C(q^{-1})}{D(q^{-1})}e[k] \quad (3.16a)$$

$$\bar{u}[k] = \left\{ q^{-(n_k-1)}u[k] \right\} \quad (3.16b)$$

In the case of a process with a single time delay *i.e.*  $n_k = 1$ , the delay operator in (3.16b) vanishes ( $n_k - 1 = 0$ ).

Several parametric models are listed in Table 3.1. These models can be divided into two main categories: equation and output error models. Input and noise transfer functions have common poles in equation error models. Note that the  $A(q^{-1})$  polynomial appears in the input and noise transfer function of these models. Equation error models are a suitable choice for closed-loop identification as their predictor is stable. In contrast to equation error models, the transfer

**Table 3.1:** Common black-box model structures

Model structure	Model name	$G(q^{-1})$	$H(q^{-1})$
Equation error	ARX	$\frac{B(q^{-1})}{A(q^{-1})}$	$\frac{1}{A(q^{-1})}$
	ARMAX	$\frac{B(q^{-1})}{A(q^{-1})}$	$\frac{C(q^{-1})}{A(q^{-1})}$
	ARARX	$\frac{B(q^{-1})}{A(q^{-1})}$	$\frac{1}{A(q^{-1})D(q^{-1})}$
	ARARMAX	$\frac{B(q^{-1})}{A(q^{-1})}$	$\frac{C(q^{-1})}{A(q^{-1})D(q^{-1})}$
Output error	FIR	$B(q^{-1})$	1
	OE	$\frac{B(q^{-1})}{F(q^{-1})}$	1
	BJ	$\frac{B(q^{-1})}{F(q^{-1})}$	$\frac{C(q^{-1})}{D(q^{-1})}$

functions are independently parametrized in output error models. These models should be preferred for system identification in open-loop, since consistent estimates can be obtained regardless of the presence of colored Gaussian noise.

Two models in Table 3.1 are of a particular interest: finite impulse response (FIR) and ARX. These models are linear-in-the-parameters (LiP) and simple computational methods for parameter estimation can be implemented. However, FIR models may require a large number of parameters to describe the system dynamics adequately. The residuals of ARX models are assumed to be white noise which may be rarely fulfilled in real systems. Moreover, a large number of terms in the  $B(q^{-1})$  polynomial can be required for processes with a large input-output delay.

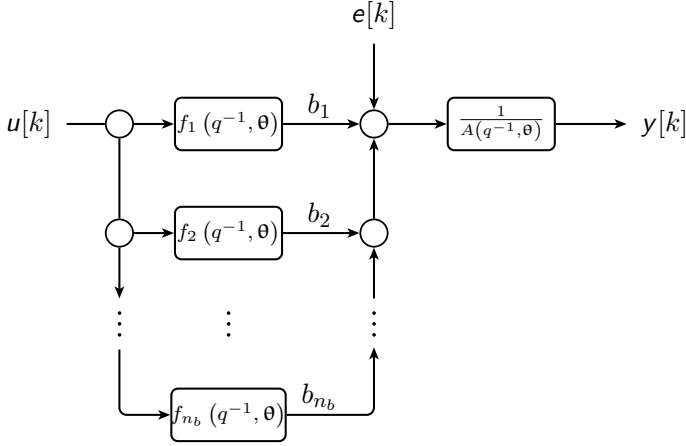
More flexible models can be obtained by replacing the  $B(q^{-1})$  polynomial in (3.16a) by filters constructed with orthonormal basis functions (OBF). Two transfer functions are called orthonormal if their vectorial inner product equals zero and their norms are 1, *i.e.*

$$\begin{aligned} \langle f_1(q^{-1}), f_2(q^{-1}) \rangle &= 0 \\ \|f_1(q^{-1})\| &= \|f_2(q^{-1})\| = 1 \end{aligned}$$

Laguerre filters are well-known OBF and represent a good choice for processes that do not exhibit resonating dynamics. These filters can be used in ARX models resulting in an alternative model called ARX-Laguerre model. Figure 3.2 shows a schematic representation of an ARX-OBF model. In contrast to “classical” ARX models, the input  $u[k]$  is filtered using OBF. The main properties of this model are:

- ARX-Laguerre models are LiP which results in a convex optimization problem for parameter estimation.
- In contrast to FIR models, less parameters are required for acceptable process description.
- Some prior knowledge of the system dynamics is required to set the poles of the Laguerre polynomials.
- The model can be used for closed-loop identification as its predictor is stable.

The transfer functions  $f_i(q^{-1}, \theta)$ ,  $i = 1, 2, \dots, n_b$  in Figure 3.2 are constructed using orthonormal basis filters whose parameters are set according to the dynamic of the system. A first choice is to use Laguerre polynomials but other functions can be used to describe processes with resonant poles as reported in [89]. First-order lag filters with real poles are obtained with Laguerre polynomials



**Figure 3.2:** Laguerre representation

that are suitable to describe well-damped systems. The resulting filters are described by

$$f_i(q^{-1}, \xi) = \frac{\sqrt{1-\xi^2}}{q^{-1}-\xi} \left[ \frac{1-\xi q^{-1}}{q^{-1}-\xi} \right]^{i-1} \quad i = 1, 2, \dots, n_b \quad (3.17)$$

where  $\xi$  is the dominating system pole which is approximated by:

$$\xi = e^{-\frac{T_s}{\tau}} \quad (3.18)$$

where  $T_s$  is the sampling time and  $\tau$  the dominating time constant of the system. Laguerre filters can also be expressed as state space models as described in [37]. Robust parameter estimation methods can be implemented using this latter representation.

Next sections address modeling of single-input single-output (SISO) and multi-input multi-output (MIMO) models.

### 3.2.1 Single-Input Single-Output models

Consider the general model (3.16) with  $\dim(y) = \dim(u) = 1$ . The transfer functions are defined using the polynomials  $A(q^{-1})$ ,  $B(q^{-1})$  and  $C(q^{-1})$ . A block diagram of the SISO model is depicted in Figure 3.1. The additive noise is represented as an MA process:

$$v[k] = C(q^{-1}) e[k] \quad (3.19)$$

where  $\{\mathbf{e}[k]\}$  is white Gaussian noise with variance  $\sigma_e^2$ . The input and noise transfer functions are:

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} \quad (3.20a)$$

$$H(q^{-1}) = \frac{C(q^{-1})}{A(q^{-1})} \quad (3.20b)$$

And the difference equation is

$$y[k] + a_1 y[k-1] + \dots + a_{n_a} y[k-n_a] = b_1 u[k-1] + \dots + b_{n_b} u[k-n_b] + \mathbf{e}[k] + c_1 \mathbf{e}[k-1] + \dots + c_{n_c} \mathbf{e}[k-n_c] \quad (3.21)$$

The model described by (3.21) is called autoregressive moving average with external input (ARMAX) (see Table 3.1) and is commonly used for data-driven modeling as the additive stochastic sequence is represented by colored Gaussian noise. Parameter estimation of ARMAX models should be performed with numerical methods such as the Newton-Raphson algorithm.

System identification of other models such as ARX or ARX-Laguerre is carried out using the LSE method. These models are also an attractive choice since satisfactory process description is reached and parameter estimation is computationally simple. Consider an ARX-Laguerre model with the following input transfer function

$$G(q^{-1}, \boldsymbol{\theta}) = \frac{G_f(q^{-1}, \boldsymbol{\theta}_f)}{A(q^{-1})} \quad (3.22a)$$

$$G_f(q^{-1}, \boldsymbol{\theta}_f) = \sum_{i=1}^{n_b} b_i f_i(q^{-1}) \quad (3.22b)$$

where  $\{b_i\}_{i=1,2,\dots,n_b}$  are the model parameters and  $f_i(q^{-1})$  are the orthonormal basis filters as in (3.17). Note that (3.20a) and (3.22a) differs on the numerator. In the second case, the  $B(q^{-1})$  polynomial is replaced by a Laguerre filter.

Consider the state-space representation of (3.22b) as presented in [37]

$$\mathbf{x}[k+1] = A_f \mathbf{x}[k] + B_f u[k] \quad (3.23a)$$

$$y[k] = C_f \mathbf{x}[k] \quad (3.23b)$$

where  $\mathbf{x}[k] \in \mathbb{R}^{n_b \times n_b}$ . A balanced realization using Laguerre polynomials is obtained as follows (see [36, 37] for details of the derivation). Let  $G(q^{-1})$  be a first order stable SISO-system with a real pole in  $\xi$  and let  $[\xi, b, c, d]$  be a

realization of  $G(q^{-1})$ , then  $\left[\xi, \eta^{\frac{1}{2}}, \eta^{-\frac{1}{2}}bc, d\right]$  is an input balanced realization of  $G(q^{-1})$ . Let  $\xi \in \mathbb{R}$ , with  $|\xi| < 1$  and let  $\eta := 1 - \xi^2$ . A balanced state-space representation of an OBF using Laguerre filters is:

$$A_f = \begin{pmatrix} \xi & 0 & & & & \\ \eta & \xi & 0 & & & \\ -\xi\eta & \eta & \xi & 0 & & \\ \xi^2\eta & -\eta\xi & \eta & \xi & 0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \quad B_f = \begin{pmatrix} \sqrt{\eta} \\ -\xi\sqrt{\eta} \\ \xi^2\sqrt{\eta} \\ \vdots \end{pmatrix} \quad (3.24a)$$

$$C_f = (b_1 \ b_2 \ \dots \ b_{n_b}) \quad D_f = 0 \quad (3.24b)$$

where  $C_f$  contains the parameters  $\{b_i\}_{i=1,2,\dots,n_b}$  of the ARX-Laguerre model.

The realization (3.24) of  $G_f(q^{-1}, \theta_f)$  in (3.22b) is practical for implementation purposes. Note that  $G_f(q^{-1}, \theta_f)$  represents the numerator of the transfer function in (3.22a) whereas the denominator is defined by the polynomial  $A(q^{-1})$ . Once the Laguerre pole  $\xi$  and the order of  $G_f(q^{-1}, \theta_f)$  are defined, a state space representation can be constructed using (3.24). The output of an ARX-Laguerre model can be expressed in vectorial form as

$$y[k] = \varphi^T[k] \theta \quad (3.25a)$$

$$\varphi[k] = (-y[k-1] \ -y[k-2] \ \dots \ -y[k-n_a] \ u_f[k-1] \ u_f[k-2] \ \dots \ u_f[k-n_b])^T \quad (3.25b)$$

$$\theta = (a_1 \ a_2 \ \dots \ a_{n_a} \ b_1 \ b_2 \ \dots \ b_{n_b})^T \quad (3.25c)$$

In contrast to an ARX model, the regression vector contains filtered versions of the input signal by the OBF. Parameter estimation of ARX-Laguerre models results in a linear regression problem that can be solved using the LSM.

### 3.2.2 Multi-Input Multi-Output models

In this section, the parametric model introduced in (3.16) is extended for description of multivariate processes. For a more detailed treatment on multivariate modeling and estimation see [63, 95].

Consider the model (3.16) with  $\dim(\mathbf{y}) = n_y$ ,  $\dim(\mathbf{u}) = n_u$ . The noise sequences for each output  $\{\mathbf{e}_i[k]\}_{i=1,2,\dots,n_y}$  are WGN with variance  $\{\sigma_{\mathbf{e}_i}^2\}_{i=1,2,\dots,n_y}$ . The resulting covariance matrix is  $\Lambda_{\mathbf{e}} = \text{diag}(\sigma_{\mathbf{e}_1}^2, \sigma_{\mathbf{e}_2}^2, \dots, \sigma_{\mathbf{e}_{n_y}}^2)$ .

The general parametric MIMO model is described by

$$\mathbf{A}(q^{-1}) \mathbf{y}[k] = \mathbf{B}(q^{-1}) \bar{\mathbf{u}}[k] + \boldsymbol{\varepsilon}[k] \quad (3.26a)$$

$$\bar{\mathbf{u}}[k] = \left\{ q^{-(n_{k_{ij}}-1)} \mathbf{u}[k] \right\} \quad (3.26b)$$

$$\mathbf{A}(q^{-1}) = \mathbf{I} + \mathbf{A}_1 q^{-1} + \dots + \mathbf{A}_{n_a} q^{-n_a} \quad (3.26c)$$

$$\mathbf{B}(q^{-1}) = \mathbf{B}_1 q^{-1} + \dots + \mathbf{B}_{n_b} q^{-n_b} \quad (3.26d)$$

where  $\mathbf{A}(q^{-1})$  and  $\mathbf{B}(q^{-1})$  are matrix polynomials with size  $(n_y|n_y)$  and  $(n_y|n_u)$ , respectively. The term  $n_{k_{ij}}$ ,  $i = 1, 2, \dots, n_u$ ,  $j = 1, 2, \dots, n_y$  in (3.26b) represent the input-output delay from the  $i$ -th input to the  $j$ -th output. The MIMO model can be expressed in compact form by:

$$\mathbf{y}[k] = \Phi^T[k] \Theta + \varepsilon[k] \quad (3.27a)$$

$$\Phi^T[k] = \begin{pmatrix} \varphi^T[k] & \mathbf{0} \\ & \ddots \\ \mathbf{0} & \varphi^T[k] \end{pmatrix} \quad (3.27b)$$

$$\varphi^T[k] = (-\mathbf{y}^T[k-1] \dots -\mathbf{y}^T[k-n_a] \mathbf{u}^T[k-1] \dots \mathbf{u}^T[k-n_b]) \quad (3.27c)$$

$$\Theta^T = (\theta^1 \dots \theta^{n_y}) \quad (3.27d)$$

$$\begin{pmatrix} (\theta^1)^T \\ \vdots \\ (\theta^{n_y})^T \end{pmatrix} = (\mathbf{A}_1 \dots \mathbf{A}_{n_a} \mathbf{B}_1 \dots \mathbf{B}_{n_b}) \quad (3.27e)$$

The model (3.26) is called matrix fraction description (MFD) which, in contrast to the SISO case, (3.26) is not a unique representation for a given process. Canonical forms have been proposed in order to have unique models for a given system. One of the simplest canonical forms is the diagonal MFD that has the following advantages:

- The model order selection is done for each multi-input single-output (MISO) submodel which reduces the complexity of the problem.
- The identification algorithms for SISO systems can be extended straightforward when using the diagonal MFD form.
- The input-output delay term can be adjusted independently for each submodel.

The advantages of the diagonal MFD form is analyzed with an example. Consider a multivariate process with  $n_y = 3$ ,  $n_u = 4$ , with orders  $n_a = 1$ ,  $n_b = 2$ . The regression and parameter vectors are described by:

$$\Phi^T[k] = (-y_1[k-1] \quad -y_2[k-1] \quad -y_3[k-1])$$

$$\begin{aligned} & u_1[k-1] u_2[k-1] u_3[k-1] u_4[k-1] \\ & u_1[k-2] u_2[k-2] u_3[k-2] u_4[k-2] \end{aligned} \quad (3.28a)$$

$$\boldsymbol{\theta}^1 = (a^{11} \ a^{12} \ a^{13} \ b_1^{11} \ b_1^{12} \ b_1^{13} \ b_1^{14} \ b_2^{11} \ b_2^{12} \ b_2^{13} \ b_2^{14})^T \quad (3.28b)$$

$$\boldsymbol{\theta}^2 = (a^{21} \ a^{22} \ a^{23} \ b_1^{21} \ b_1^{22} \ b_1^{23} \ b_1^{24} \ b_2^{21} \ b_2^{22} \ b_2^{23} \ b_2^{24})^T \quad (3.28c)$$

with the following parameter matrices:

$$\mathbf{A}_1 = \begin{pmatrix} a^{11} & a^{12} & a^{13} \\ a^{21} & a^{22} & a^{23} \end{pmatrix} \quad (3.29a)$$

$$\mathbf{B}_1 = \begin{pmatrix} b_1^{11} & b_1^{12} & b_1^{13} & b_1^{14} \\ b_1^{21} & b_1^{22} & b_1^{23} & b_1^{24} \end{pmatrix} \quad (3.29b)$$

$$\mathbf{B}_2 = \begin{pmatrix} b_2^{11} & b_2^{12} & b_2^{13} & b_2^{14} \\ b_2^{21} & b_2^{22} & b_2^{23} & b_2^{24} \end{pmatrix} \quad (3.29c)$$

Using the diagonal form, the multivariable process (3.26) can be decoupled into  $n_y$  MISO subprocesses. A MISO model is described by:

$$A_j(q^{-1}) y_j[k] = \mathbf{B}(q^{-1}) \bar{\mathbf{u}}[k] + \boldsymbol{\varepsilon}[k] \quad (3.30a)$$

$$A_j(q^{-1}) y_j[k] = \sum_{i=1}^{n_u} B_i(q^{-1}) \bar{u}_i[k] + \boldsymbol{\varepsilon}[k] \quad (3.30b)$$

with the following definitions

$$A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a} \quad (3.30c)$$

$$B_i(q^{-1}) = b_{1,i} q^{-1} + \dots + b_{n_b,i} q^{-n_b,i} \quad (3.30d)$$

$$\bar{\mathbf{u}}[k] = \left( q^{-(n_{k_1}-1)} u_1[k] \ \dots \ q^{-(n_{k_{n_u}}-1)} u_{n_u}[k] \right)^T \quad (3.30e)$$

In order to express the model output in a compact form, the following vectors are introduced:

$$\begin{aligned} \boldsymbol{\varphi}^T[k] = & (-y[k-1] \ \dots \ -y[k-n_a] \quad u_1[k-n_{k_1}] \ \dots \ u_1[k-n_{k_1}-n_{b,1}+1] \\ & u_2[k-n_{k_2}] \ \dots \ u_2[k-n_{k_2}-n_{b,2}+1] \ \dots \\ & u_{n_u}[k-n_{k_{n_u}}] \ \dots \ u_{n_u}[k-n_{k_{n_u}}-n_{b,n_u}+1]) \end{aligned} \quad (3.31a)$$

$$\boldsymbol{\Theta} = (a_1 \ \dots \ a_{n_a} \ b_{1,1} \ \dots \ b_{n_b,1} \ \dots \ b_{1,n_u} \ \dots \ b_{n_b,n_u})^T \quad (3.31b)$$

Then, the model in (3.30a) is expressed in vector notation as

$$y[k] = \boldsymbol{\varphi}^T[k] \boldsymbol{\Theta} + \boldsymbol{\varepsilon}[k] \quad (3.32)$$



where  $\boldsymbol{\varphi}[k]$  and  $\Theta$  are column vectors of dimension  $\dim(\Theta) = \dim(\boldsymbol{\varphi}) = n_a + \sum_{i=1}^{n_u} n_{b_i}$ .

The ARX-Laguerre model presented in (3.22) can be extended to the multivariate case as follows. Consider a MISO model as in (3.30). The input transfer function for the  $j$ -th output is

$$\mathbf{G}(q^{-1}, \boldsymbol{\theta}) = \frac{\mathbf{G}_f(q^{-1}, \boldsymbol{\theta}_f)}{A_j(q^{-1})} \quad (3.33a)$$

$$\mathbf{G}_f(q^{-1}, \boldsymbol{\theta}_f) = \sum_{i=1}^{n_u} \sum_{l=1}^{n_{b_i}} b_{il} f_{il}(q^{-1}) \quad (3.33b)$$

where  $A_j(q^{-1})$  is defined as in (3.30c). A state-space model of (3.33b) is

$$\mathbf{X}^{(i)}[k+1] = \mathbf{A}_f^{(i)} \mathbf{X}^{(i)}[k] + \mathbf{B}_f^{(i)} \mathbf{u}[k] \quad (3.34a)$$

$$\mathbf{y}^{(i)}[k] = \mathbf{C}_f^T \mathbf{X}^{(i)}[k] \quad (3.34b)$$

where the state space matrices are described by

$$\mathbf{A}_f^i = \begin{pmatrix} \mathbf{A}_{f_1}^{(i)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{A}_{f_{n_u}}^{(i)} \end{pmatrix} \quad \mathbf{B}_f^i = \begin{pmatrix} \mathbf{B}_{f_1}^{(i)} & 0 & \cdots & 0 \\ 0 & \mathbf{B}_{f_2}^{(i)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{B}_{f_{n_u}}^{(i)} \end{pmatrix} \quad (3.35a)$$

$$\mathbf{C}_f^{(i)} = (b_{11} \ b_{12} \ \cdots \ b_{1n_1} \ b_{21} \ b_{22} \ \cdots \ b_{2n_2} \ b_{1n_u} \ b_{2n_u} \ \cdots \ b_{n_y n_u})^T$$

where  $\mathbf{C}_f^{(i)}$  is the parameter vector and the block-diagonal matrices  $\mathbf{A}_{f_j}^{(i)}$  and  $\mathbf{B}_{f_j}^{(i)}$  are defined similarly to (3.24).

The models (3.30b) and (3.33a) will be further used in this work for modeling of multivariate process.

### 3.3 Parameter Estimation

Parameter estimation from input-output data is introduced in this section. One of the first publications related to the parameter estimation problem is found in [7]. For a practical and brief treatment of this topic, the reader is referred to [51]. Three parameter estimation methods are introduced in this section. They can be applied for identification of SISO and MIMO systems in open or closed-loop. Firstly, the LSM as well as its advantages and drawbacks are

described. Secondly, the instrumental variables method (IVM) is introduced and its robustness to colored noise is discussed. Lastly, the PEM are introduced as a more general procedure that can also deal with colored noise.

### 3.3.1 Least Squares Method

The one-step-ahead predictor of  $y$  for the linear models of [Table 3.1](#) is

$$\hat{y}[k|k-1; \hat{\Theta}] = H^{-1}(q^{-1}; \hat{\Theta})G(q^{-1}; \hat{\Theta})u[k] + (1 - H^{-1}(q^{-1}; \hat{\Theta}))y[k] \quad (3.36)$$

where  $u[k]$  and  $y[k]$  are the process input and output, respectively. Replacing the input and noise transfer functions of the ARX-Laguerre model ([3.13c](#)) and ([3.24](#)) yields:

$$\hat{y}[k|k-1; \hat{\Theta}] = \boldsymbol{\varphi}^T[k] \hat{\boldsymbol{\theta}} \quad (3.37)$$

where  $\boldsymbol{\varphi}$  and  $\boldsymbol{\theta}$  are defined as in ([3.25b](#)) and ([3.25c](#)) and the prediction error is:

$$\boldsymbol{\varepsilon}[k, \hat{\boldsymbol{\theta}}] = y[k] - \boldsymbol{\varphi}^T[k] \hat{\boldsymbol{\theta}} \quad (3.38)$$

The sum of the squared errors is:

$$V_N(\boldsymbol{\Theta}) = \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varepsilon}^2[k] \quad (3.39)$$

The parameter vector  $\boldsymbol{\Theta}$  which minimizes ([3.39](#)) is (see [[49](#)] for derivation details):

$$\hat{\boldsymbol{\theta}}_N = \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}[k] \boldsymbol{\varphi}^T[k] \right)^{-1} \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}[k] y[k] \right) \quad (3.40)$$

The method in ([3.40](#)) is called the LSM. It is computationally simple but has some consistency issues in the case of colored noise as following discussed.

Consistent estimates is a desired property for a chosen identification method. A parameter estimate  $\hat{\boldsymbol{\theta}}$  is consistent if it converges to the true value  $\boldsymbol{\Theta}_0$  as the number of data,  $N$ , tends to infinity. Assume that the data is generated by the following system:

$$A_0(q^{-1})y[k] = \mathbf{B}_0(q^{-1})\mathbf{u}[k] + v[k] \quad (3.41a)$$

$$y[k] = \boldsymbol{\varphi}^T[k] \boldsymbol{\Theta}_0 + v[k] \quad (3.41b)$$

where  $\Theta_0$  is the “true” parameter vector. Replacing (3.41b) in (3.40) yields:

$$\hat{\Theta}_N = \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}[k] \boldsymbol{\varphi}^T[k] \right)^{-1} \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}[k] (\boldsymbol{\varphi}^T[k] \Theta_0 + \mathbf{v}[k]) \right) \quad (3.42a)$$

$$\hat{\Theta}_N = \Theta_0 + \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}[k] \boldsymbol{\varphi}^T[k] \right)^{-1} \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}[k] \mathbf{v}[k] \right) \quad (3.42b)$$

$$\hat{\Theta}_N - \Theta_0 = \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}[k] \boldsymbol{\varphi}^T[k] \right)^{-1} \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\varphi}[k] \mathbf{v}[k] \right) \quad (3.42c)$$

The estimate  $\hat{\Theta}_N$  tends to  $\Theta_0$  if

$$E(\boldsymbol{\varphi}[k] \boldsymbol{\varphi}^T[k]) \text{ is non-singular and} \quad (3.43a)$$

$$E(\boldsymbol{\varphi}[k] \mathbf{v}[k]) = 0 \quad (3.43b)$$

The condition (3.43a) is satisfied assuming that the input signal is (*p.e.*) of sufficient order. In contrast, (3.43b) is satisfied only if  $\{\mathbf{v}[k]\}$  is white noise (a sequence of uncorrelated random variables). In that particular case, (3.43b) holds since  $\{\mathbf{v}[k]\}$  will be uncorrelated with all past data and in particular with  $\boldsymbol{\varphi}[k]$ . Measured signals in real applications are rarely embedded in white noise. Colored noise is a more realistic situation since signals are frequently filtered in the logging process. Alternative identification methods should be considered if consistent estimates are pursued. The IVM and PEM, as discussed in next sections, can overcome the drawbacks related to LSM.

### 3.3.2 Instrumental Variable Methods

Instrumental variables methods are well known approaches for parameter estimation (see, for instance, [94] for a detailed treatment on this topic). Parameter estimation of models for multivariate systems using instrumental variables can be found in [42, 85]. In contrast to LSM, IVM are robust to colored Gaussian noise. In these methods, the identification problem is formulated so that input and noise transfer functions are independently parametrized. The measurement noise can be colored and described, for instance, using an ARMA process.

Instrumental variables (IV) methods are mainly focused on estimating the input transfer function (3.20a). However, estimation of the noise transfer function is also possible when an optimal instrumental variables (IV) approach is used since this filter is estimated within the method. An extensive comparison between optimal IV methods is presented in [79].

Assume that the data is described by (3.41). Consider an instrumental vector  $\zeta[k]$  of size  $(n_z|1)$  whose elements are uncorrelated with  $\mathbf{v}[k]$ . The elements of  $\zeta[k]$  are called instruments or instrumental variables. Pre-multiplying  $\zeta[k]$  by the equation errors yields

$$\frac{1}{N} \sum_{k=1}^N \zeta[k] \varepsilon[k] = \frac{1}{N} \sum_{k=1}^N \zeta[k] (y[k] - \boldsymbol{\varphi}^T[k] \boldsymbol{\Theta}) = 0 \quad (3.44)$$

If  $\zeta[k]$  is chosen so that  $(n_\zeta = n_\Theta)$ , the basic IV estimate is computed by

$$\hat{\boldsymbol{\Theta}}_N = \left( \sum_{k=1}^N \zeta[k] \boldsymbol{\varphi}^T[k] \right)^{-1} \left( \sum_{k=1}^N \zeta[k] y[k] \right) \quad (3.45)$$

To analyze the consistency of IVM, replace (3.41b) in (3.45)

$$\hat{\boldsymbol{\Theta}}_N = \boldsymbol{\Theta}_0 + \left( \frac{1}{N} \sum_{k=1}^N \zeta[k] \boldsymbol{\varphi}^T[k] \right)^{-1} \left( \frac{1}{N} \sum_{k=1}^N \zeta[k] \mathbf{v}[k] \right) \quad (3.46a)$$

$$\hat{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}_0 = \left( \frac{1}{N} \sum_{k=1}^N \zeta[k] \boldsymbol{\varphi}^T[k] \right)^{-1} \left( \frac{1}{N} \sum_{k=1}^N \zeta[k] \mathbf{v}[k] \right) \quad (3.46b)$$

From (3.46b), the following is required for consistent estimates using IVM:

$$E(\zeta[k] \boldsymbol{\varphi}^T[k]) \text{ is non-singular and} \quad (3.47a)$$

$$E(\zeta[k] \mathbf{v}[k]) = 0 \quad (3.47b)$$

Conditions in (3.47) are satisfied if the instruments are correlated with the regression vector  $\boldsymbol{\varphi}[k]$  and uncorrelated with the measurement noise  $\mathbf{v}[k]$ . Instruments can be chosen in different ways as described in the next section.

### Choice of instruments

The choice of the instrumental variables can be influenced by the operating mode of a system. In the case of open-loop operation, the input signal  $u[k]$  is uncorrelated with the measurement noise  $\mathbf{v}[k]$ . A suitable choice for instruments can be, for instance, a vector of delayed input terms and a filtered sequence of the input signal as follows

$$\zeta[k] = (-\boldsymbol{\eta}[k-1] \ \dots \ -\boldsymbol{\eta}[k-n_a] \ u[k-1] \ \dots \ u[k-n_b])^T \quad (3.48a)$$

$$\text{where, } C(q^{-1}) \boldsymbol{\eta}[k] = D(q^{-1}) u[k] \quad (3.48b)$$

The sequence  $\eta[k]$  is obtained by filtering the input signal  $u[k]$ . The filters in (3.48b) can be initial estimates of  $A(q^{-1})$  and  $B(q^{-1})$ , respectively. Alternatively, let  $C(q^{-1}) = 1$  and  $D(q^{-1}) = q^{-n_a}$  which results in the following instruments

$$\zeta[k] = (u[k-1] \ \dots \ u[k-n_a-n_b])^T \quad (3.49)$$

In the case of closed loop identification, delayed values of the reference signal  $r[k]$  are a suitable choice as instruments. Assuming  $C(q^{-1}) = 1$  and  $D(q^{-1}) = q^{-n_a}$  in (3.48b), the resulting instrumental vector is

$$\zeta[k] = (r[k-1] \ \dots \ r[k-n_a-n_b])^T \quad (3.50)$$

An instrumental vector, that can be used for open or closed-loop identification, is obtained by choosing its elements as delayed inputs and outputs starting from certain time  $k_\zeta$ . Consider a MISO model as in (3.30) and assume that the noise is described by an ARMA process with  $C(q^{-1})$  and  $D(q^{-1})$  as monic polynomials described by (3.15a) and (3.15b), respectively. The model output at time  $k$  can be expressed as:

$$\begin{aligned} y[k] = & f(\Theta, y[k-1], \dots, y[k-n_a], \\ & u_1[k-n_{k_1}-1], \dots, u_1[k-n_{k_1}-n_{b,1}+1], \dots, \\ & u_{n_u}[k-n_{k_{n_u}}-1], \dots, u_{n_u}[k-n_{k_{n_u}}-n_{b,n_u}+1], \\ & e[k], \dots, e[k-n_c]) \end{aligned} \quad (3.51)$$

Delayed inputs and outputs from  $k = k_\zeta - 1$ , where  $k_\zeta \geq n_c$ , are no more correlated with the additive disturbance. Thus, a suitable choice for instruments that can be used in open and closed loop identification is

$$\begin{aligned} \zeta[k] = & (-y[k-k_\zeta-1] \ \dots \ -y[k-k_\zeta-\tilde{n}_a] \\ & u_1[k-k_\zeta-n_{k_1}-1] \ \dots \ u_1[k-k_\zeta-n_{k_1}-\tilde{n}_{b,1}+1] \ \dots \\ & u_{n_u}[k-k_\zeta-n_{k_{n_u}}-1] \ \dots \ u_{n_u}[k-k_\zeta-n_{k_{n_u}}-\tilde{n}_{b,n_u}+1])^T \end{aligned} \quad (3.52)$$

where  $\tilde{n}_a > n_a$  and  $\tilde{n}_{b_i} > n_{b_i}$ .

Optimal IVMs yield consistent estimates for a correct choice of the instruments. Implementation of optimal IVMs for (3.45) requires an iterative approach where the instruments are obtained by filtering the noise-free regression vector as described in [57, 80]. In [43], optimal instrumental variable methods are extended to multivariate processes. In the present work, IVM are preferred over LSM because of their robustness to colored Gaussian noise. Moreover, since processes

addressed in this work can be operating in open or closed-loop,  $\zeta[k]$  will be chosen as in (3.52).

### 3.3.3 Prediction Error Methods

In many applications, models are used for prediction. For instance, in control applications, it is necessary to know what the model output is likely to be at time  $k$  considering past information up to  $k - 1$ . In order to obtain good prediction models, the parameter vector is estimated by minimizing the prediction errors:

$$\varepsilon[k, \Theta] = y[k] - \hat{y}[k|k-1; \hat{\Theta}] \quad (3.53)$$

The term  $\hat{y}[k|k-1; \hat{\Theta}]$  denotes the model prediction of  $y[k]$  computed with data up to and including time  $k - 1$  and using  $\hat{\Theta}$ . This can be expressed by:

$$\hat{y}[k|k-1; \hat{\Theta}] = f(\hat{\Theta}, y[k-1], y[k-2], \dots, u[k-1], u[k-2], \dots) \quad (3.54)$$

Consider the following general predictor of (3.11)

$$\hat{y}[k|k-1; \hat{\Theta}] = L_1(q^{-1}; \hat{\Theta})y[k] + L_2(q^{-1}; \hat{\Theta})u[k] \quad (3.55)$$

which is a function of past data if the following holds for the filters  $L_1(q^{-1}; \hat{\Theta})$  and  $L_2(q^{-1}; \hat{\Theta})$ :

$$L_1(0; \hat{\Theta}) = 0 \text{ and } L_2(0; \hat{\Theta}) = 0 \quad (3.56)$$

The assumption (3.56) implies that  $\hat{y}[k|k-1; \hat{\Theta}]$  depends only on previous inputs and outputs and not on  $y[k]$  and  $u[k]$ .

The predictor (3.55) can be defined in different ways for any of the models presented in Table 3.1. Commonly, the predictor is chosen so that the filters  $L_1(q^{-1}; \hat{\Theta})$  and  $L_2(q^{-1}; \hat{\Theta})$  make the sample variance as small as possible [57, 80]. The optimal linear predictor and the prediction errors are obtained by:

$$\hat{y}[k|k-1; \hat{\Theta}] = H^{-1}(q^{-1}; \hat{\Theta})G(q^{-1}; \hat{\Theta})u[k] + (1 - H^{-1}(q^{-1}; \hat{\Theta}))y[k] \quad (3.57a)$$

$$\varepsilon[k; \hat{\Theta}] = H^{-1}(q^{-1}; \hat{\Theta})(y[k] - G(q^{-1}; \hat{\Theta})u[k]) \quad (3.57b)$$

The predictor (3.57) for each model in Table 3.1 is obtained by replacing the corresponding input and noise transfer functions. In addition to the optimal predictor, a minimizing criterion for the prediction errors in (3.57b) should be chosen.

A class of criteria can be proposed based on the sample covariance matrix  $\hat{R}_N$ :

$$\hat{R}_N(\hat{\Theta}) = \frac{1}{N} \sum_{k=0}^{N-1} \boldsymbol{\varepsilon}[k, \hat{\Theta}] \boldsymbol{\varepsilon}^T[k, \hat{\Theta}] \quad (3.58)$$

where  $N$  is the number of measurements. In the case of SISO systems,  $\hat{R}_N(\hat{\Theta})$  is a scalar that can be used directly as minimizing criterion. For multivariable processes,  $\hat{R}_N(\hat{\Theta})$  is a positive definite matrix and the following function is proposed:

$$V_N(\hat{\Theta}) = h(\hat{R}_N(\hat{\Theta})) \quad (3.59)$$

where  $h(\hat{R}_N(\hat{\Theta}))$  is a scalar-valued function defined upon the set of positive definite matrices. Some suitable choices for  $h(\hat{R}_N(\hat{\Theta}))$  are:

$$h(\hat{R}_N(\hat{\Theta})) = \text{tr } S \hat{R}_N(\hat{\Theta}) \quad (3.60a)$$

$$h(\hat{R}_N(\hat{\Theta})) = \det \hat{R}_N(\hat{\Theta}) \quad (3.60b)$$

where  $S$  in (3.60a) is a symmetric positive definite weighting matrix.

Once the optimal predictor and the minimizing cost function have been selected, parameter estimation using the PEM is performed as follows:

- Choose a model structure of the form (3.11) and determine its predictor as in (3.57).
- Select a scalar-valued function  $h(\hat{R}_N(\hat{\Theta}))$ .
- Compute the parameter estimate  $\hat{\Theta}$  as the (global) minimum of the loss function  $V_N(\hat{\Theta})$

$$\hat{\Theta} = \underset{\Theta}{\text{arg min}} h(\hat{R}_N(\hat{\Theta})) \quad (3.61)$$

To evaluate  $V_N(\hat{\Theta})$  at any  $\hat{\Theta}$ , the prediction errors  $\{\boldsymbol{\varepsilon}[k, \hat{\Theta}]\}_{k=0}^{N-1}$  are computed by (3.53) using the optimal model predictor. Then, the sample covariance matrix is evaluated using (3.58).

The minimum of  $V_N(\hat{\Theta})$  in (3.59) can be found analytically for the ARX or ARX-Laguerre model. In that case, the solution is the least squares estimate (3.40). Thus, the LSM is a special case of the PEM. For other models, the solution cannot be found analytically and numerical methods such as the Newton-Raphson algorithm (see [80]) should be implemented. In this numerical

method, the parameter vector is computed by:

$$\hat{\Theta}^{(m+1)} = \hat{\Theta}^{(m)} - \alpha_m \left( V_N''(\hat{\Theta}^{(m)}) \right)^{-1} V_N'(\hat{\Theta}^{(m)})^T \quad (3.62)$$

where  $\hat{\Theta}^{(m)}$  denotes the estimate at the  $m$ -th iteration. The scalar quantity  $\alpha_m$  is used to control the step length and is a user-defined parameter. The following choice can, for instance, improve the convergence of (3.62):

$$\alpha_m = \arg \min_{\alpha} V_N \left( \hat{\Theta}^{(m)} - \alpha \left( V_N''(\hat{\Theta}^{(m)}) \right)^{-1} V_N'(\hat{\Theta}^{(m)})^T \right) \quad (3.63)$$

To find the derivatives of  $V_N(\hat{\Theta})$ , let  $\psi[k, \Theta]$  be:

$$\psi[k, \Theta] = - \left( \frac{\partial \varepsilon[k, \Theta]}{\partial \Theta} \right)^T \quad (3.64)$$

Then, by straightforward differentiation and for  $n_y = 1$ :

$$V_N'(\hat{\Theta}) = - \frac{2}{N} \sum_{k=0}^{N-1} \varepsilon[k, \Theta] \psi^T[k, \Theta] \quad (3.65a)$$

$$V_N''(\hat{\Theta}) = - \frac{2}{N} \sum_{k=0}^{N-1} \psi[k, \Theta] \psi^T[k, \Theta] + \frac{2}{N} \sum_{k=0}^{N-1} \varepsilon[k, \Theta] \frac{\partial^2}{\partial \Theta^2} \varepsilon[k, \Theta] \quad (3.65b)$$

The resulting method, called Gauss-Newton algorithm, is

$$\hat{\Theta}^{(m+1)} = \hat{\Theta}^{(m)} - \alpha_m \left( \sum_{k=0}^{N-1} \psi[k, \hat{\Theta}^{(m)}] H \psi^T[k, \hat{\Theta}^{(m)}] \right)^{-1} \cdot \left( \sum_{k=0}^{N-1} \psi[k, \hat{\Theta}^{(m)}] H \varepsilon[k, \hat{\Theta}^{(m)}] \right) \quad (3.66)$$

where  $H$  is the noise covariance matrix.

In [88], a PEM method for OBF-ARMAX models is introduced and demonstrated for open and closed loop identification. This practical implementation can be useful in real applications when prior knowledge about input-output delay is limited.

Despite the fact that the predictor was introduced in this section for parametric models, it also plays an important role in subspace identification [21]. Additional aspects regarding consistency, however, should be considered for



subspace methods as explained in [20]. In [90], a subspace method for closed loop identification is explained and evaluated with a simulation case study.

In multivariable processes, simultaneous or independent excitation should also be examined as it can affect parameter estimation. In [26] a detailed variance analysis for the multivariate models in Table 3.1 is performed. In the case of equation error models, the covariance of all parameter estimates will decrease if the system is simultaneously excited in all the inputs.. The authors conclude that it is always recommended to excite all inputs simultaneously in multivariable processes.

An interesting alternative to (3.66) is to apply a recursive algorithm to the data several times. The final values of a run are used as initial ones for the next run. This approach will often result in faster convergence and will save computational time.

## 3.4 Recursive Parameter Estimation

In the identification methods presented in section 3.3, called off-line or batch methods, the entire identification data set is used at once for parameter estimation. In contrast, recursive methods are computationally less demanding as the estimate  $\hat{\Theta}[k]$  is computed recursively by modifying  $\hat{\Theta}[k - 1]$ . For a detailed discussion on these approaches, the reader is referred to [6, 28, 59]. Recursive identification methods are less demanding than batch methods because estimates are updated recursively based on past information. Recursive methods can be easily adapted for real time applications, for instance, tracking time-varying parameters. These methods represent a first choice for fault detection as changes in the parameters may be related to anomalies in the process. Recursive identification methods are suitable choices for upper interval bounding for data selection methods. The next sections describe the recursive versions of the least squares and the instrumental variables method.

### 3.4.1 Recursive QR Least Squares Method

In order to explain this method, the QR solution to the batch least squares problem is firstly introduced. Given an identification data set  $\mathbf{Z}^N$  the weighted least squares cost function is formulated as (cf. (3.39))

$$V(\Theta) = \left\| \mathbf{W}_N^{1/2} (\mathbf{y} - \Phi\Theta) \right\|^2 = \sum_{k=0}^{N-1} \lambda[k] \varepsilon^2[k] \quad (3.67a)$$

$$\varepsilon[k] = y[k] - \boldsymbol{\varphi}^T[k]\Theta \quad (3.67b)$$

with

$$\mathbf{y} = (y[0] \ \dots \ y[N-1])^T \quad (3.67c)$$

$$\Phi = (\boldsymbol{\varphi}^T[0] \ \dots \ \boldsymbol{\varphi}^T[N-1])^T \quad (3.67d)$$

$$\mathbf{W} = \text{diag}(\lambda[0] \ \dots \ \lambda[N-1]) \quad (3.67e)$$

where  $\boldsymbol{\varphi}$  and  $\Theta$  are vectors defined by (3.31a) and (3.31b), respectively. In (3.67e),  $\mathbf{W}$  is a diagonal positive definite weighting matrix with  $[\mathbf{W}]_{ii} = \lambda^{k-i}$  and  $\lambda$  is the forgetting factor.

Consider that instead of  $\mathbf{Z}^N$ , a data set  $\mathbf{Z}^m$ ,  $m < N$  is given. The norm in (3.67a) is not affected if it is pre-multiplied by an orthonormal matrix  $\mathbf{Q} \in \mathbb{R}^{m \times m}$ :

$$V(\Theta_m, \mathbf{Z}^m) = \left\| \mathbf{Q} \left( \mathbf{W}_N^{1/2} (\mathbf{y} - \Phi \Theta) \right) \right\|^2 \quad (3.68)$$

Applying QR-factorization to (3.68) results in

$$\left( \mathbf{W}_N^{1/2} \Phi \ \mathbf{W}_N^{1/2} \mathbf{y} \right) = \mathbf{Q} \mathbf{R}, \quad \mathbf{R} = \begin{pmatrix} \mathbf{R}_0 \\ \dots \\ \mathbf{0} \end{pmatrix} \quad (3.69)$$

where  $\mathbf{R}_0 \in \mathbb{R}^{(n_\theta+1) \times (n_\theta+1)}$  is an upper triangular and  $n_\theta = \dim(\theta)$ . The matrix  $\mathbf{R}_0$  can be decomposed as

$$\mathbf{R}_0 = \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ 0 & \mathbf{R}_3 \end{pmatrix}, \quad \mathbf{R}_1 \in \mathbb{R}^{n_\theta \times n_\theta}, \mathbf{R}_2 \in \mathbb{R}^{n_\theta \times 1}, \mathbf{R}_3 \in \mathbb{R} \quad (3.70)$$

Using QR-factorization, the cost function (3.68) can be reformulated as follows

$$\begin{aligned} V(\hat{\Theta}_m, \mathbf{Z}^m) &= \left\| \mathbf{Q}^T \mathbf{W}_N^{1/2} (\mathbf{y} - \Phi \hat{\Theta}) \right\|^2 = \left\| \begin{pmatrix} \mathbf{R}_2 \\ \mathbf{R}_3 \end{pmatrix} - \begin{pmatrix} \mathbf{R}_1 \hat{\Theta} \\ 0 \end{pmatrix} \right\|^2 \\ &= \left\| \mathbf{R}_2 - \mathbf{R}_1 \hat{\Theta} \right\|^2 + \left\| \mathbf{R}_3 \right\|^2 \end{aligned} \quad (3.71)$$

which is minimized for

$$\hat{\Theta}_m = (\mathbf{R}_1)^{-1} \mathbf{R}_2 \quad \text{and the cost function yields} \quad V(\hat{\Theta}_m, \mathbf{Z}^m) = \left\| \mathbf{R}_3 \right\|^2 \quad (3.72)$$

The solution (3.72) is a batch method since the entire data set is used simultaneously. A sequential solution can be proposed as follows [86]. Assume that the

weighting matrix at time  $k$  satisfies the following partitioning scheme

$$\mathbf{W}[k] = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \lambda & & \\ & & \lambda^2 & \\ \vdots & & & \vdots \\ 0 & \dots & 0 & \lambda^{k-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & \lambda \mathbf{W}[k-1] & & \\ 0 & & & \end{pmatrix} \quad (3.73)$$

then, the cost function  $V$  at the instant  $k$  can be expressed as

$$V_k(\hat{\Theta}) = \left\| \begin{pmatrix} \lambda^{1/2} \mathbf{W}^{1/2}[k-1] \mathbf{y}[k-1] \\ \mathbf{y}[k] \end{pmatrix} - \begin{pmatrix} \lambda^{1/2} \mathbf{W}^{1/2}[k-1] \Phi[k-1] \\ \boldsymbol{\varphi}[k] \end{pmatrix} \hat{\Theta} \right\|_2^2 \quad (3.74)$$

where  $\lambda$  is a forgetting factor, which ranges between 0 and 1, representing the relevance of past measurements in the computations at  $k$ . A typical value for the forgetting factor is  $\lambda = 0.975$  [80]. Similarly to (3.71), the cost function (3.74) can be expressed as

$$V_k(\hat{\Theta}) = \left\| \begin{pmatrix} \lambda^{1/2} \mathbf{R}_2[k-1] \\ \lambda^{1/2} \mathbf{R}_3[k-1] \\ \mathbf{y}(t) \end{pmatrix} - \begin{pmatrix} \lambda^{1/2} \mathbf{R}_1[k-1] \\ 0 \\ \boldsymbol{\varphi}(t) \end{pmatrix} \hat{\Theta} \right\|_2^2 \quad (3.75)$$

Applying the QR-factorization to (3.75) yields

$$\begin{pmatrix} \lambda^{1/2} \mathbf{R}_1[k-1] & \lambda^{1/2} \mathbf{R}_2[k-1] \\ 0 & \lambda^{1/2} \mathbf{R}_3[k-1] \\ \boldsymbol{\varphi}(t) & \mathbf{y}[k] \end{pmatrix} = \begin{pmatrix} \lambda^{1/2} \mathbf{R}_0[k-1] \\ \boldsymbol{\varphi}[k] & \mathbf{y}[k] \end{pmatrix} = \mathbf{Q}[k] \mathbf{R}[k] \quad (3.76)$$

The solution to (3.76) is obtained by:

$$\hat{\Theta}[k] = \mathbf{R}_1^{-1}[k] \mathbf{R}_2[k] \quad (3.77)$$

In (3.77),  $\hat{\Theta}[k]$  is computed recursively at each  $k$ . The matrix  $\mathbf{R}_0$  in (3.76) is initialized as  $\mathbf{R}_0 = 1/\beta \cdot \mathbf{I}_{(n_\theta+1) \times (n_\theta+1)}$  where  $\beta$  is a user-defined parameter with  $0 < \beta < 0.01$ .

The recursive QR least squares method has been already used for data selection as reported in [12]. Upper interval bounds are defined according to the conditioning of the matrix  $\mathbf{R}_1[k]$ . This recursive method has similar

properties to its batch version. Thus, estimates will be biased in the case of colored Gaussian noise. This drawback can result in wrong interval bounds since the information matrix is also affected by correlated noise. Alternatively, identification methods robust to colored noise are required to overcome these limitations. In the next section, a recursive approach based on instrumental variables is introduced as an alternative method for data selection.

### 3.4.2 Overdetermined Recursive Instrumental Variables Method

Optimal instrumental variables methods are computationally simpler than PEM and can guarantee consistent estimates in the case of correlated noise [80]. The parameters at time  $k$  are estimated in an iterative procedure where the instruments are obtained by filtering the regressor vector at  $k - 1$ . Implementation of optimal IV methods in closed loop can be more demanding than direct PEM as the controller may be required for computations.

Extended instrumental variables (EIV) methods can be used for open or closed-loop identification. In this method, the number of instruments is larger than the number of parameters. The instrumental variables can be chosen as delayed inputs and outputs as presented in (3.52). This choice of instruments yields an overdetermined system which is solved, as found in [80], by:

$$\hat{\theta}_N^{\text{EIV}} = (\mathbf{R}_N^T \mathbf{Q} \mathbf{R}_N)^{-1} (\mathbf{R}_N^T \mathbf{Q} \mathbf{r}_N) \quad (3.78a)$$

$$\mathbf{R}_N = \frac{1}{N} \sum_{k=1}^N \zeta[k] F(q) \boldsymbol{\varphi}^T[k] \quad (3.78b)$$

$$\mathbf{r}_N = \frac{1}{N} \sum_{k=1}^N \zeta[k] F(q) y[k] \quad (3.78c)$$

where  $F(q)$  is an asymptotically stable pre-filter and  $Q$  is a positive definite weighting matrix.

The algorithm (3.78) is called extended instrumental variables (EIV). This method is not suitable for practical applications as it can be computationally unstable and it is time demanding due to the large size of the matrices to be processed [80]. Instead, a recursive implementation, proposed in [23, 80] and called ORIV, can be used. Two user-defined parameters should be set to define instruments as in (3.52). The delay time is often selected with  $k_\zeta > 5$  which indicates that from that instant possible correlation between the input and noise practically vanishes. The number of instruments,  $\tilde{n}_a$  and  $\tilde{n}_b$ , are normally chosen larger than the model order. A detailed derivation of the ORIV is found

in [23]. The algorithm is summarized as follows

$$(\mathbf{L}[k] \mathbf{S}^T[k]) = \lambda (\mathbf{L}[k-1] \mathbf{S}[k-1]^T) + \zeta[k] (y[k] \boldsymbol{\phi}[k]) \quad (3.79a)$$

$$\hat{\boldsymbol{\theta}} = (\mathbf{S}^T[k])^\dagger \mathbf{L}[k] \quad (3.79b)$$

where  $\mathbf{L}[k] \in \mathbb{R}^{n_\zeta \times 1}$  and  $\mathbf{S}[k] \in \mathbb{R}^{n_\theta \times n_\zeta}$ . The symbol  $(\cdot)^\dagger$  corresponds to the pseudoinverse of  $(\cdot)$ . The forgetting factor  $\lambda$  is normally set as  $0.98 \leq \lambda \leq 0.995$ . The algorithm can be initialized as follows:

$$\mathbf{S}[0] = \beta [\mathbf{I} \mid \mathbf{0}] \quad \mathbf{L}[0] = \mathbf{0} \quad (3.80)$$

where  $\beta$  is a scalar parameter in the range  $\beta \in [0.001, 0.01]$ . The choice of  $\lambda$  and  $\beta$  is discussed in [Chapter 5](#).

In the ORIV, the conditioning of  $\mathbf{S}^T[k]$  can be evaluated to define upper interval bounds. The ORIV can then be proposed as an alternative for data selection to the method in [section 3.4.1](#). The use of this method can yield better interval bounds particularly in real applications and it is a distinguishing feature when compared with available approaches.

### 3.5 The CUSUM Test for Input-Output Models

In the case of an ARX model as in (3.13), the log-likelihood ratio increment is:

$$s[k] = \frac{1}{2} \ln \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} + \frac{\varepsilon_0^2[k]}{2\hat{\sigma}_0^2} - \frac{\varepsilon_1^2[k]}{2\hat{\sigma}_1^2} \quad (3.81)$$

where  $\varepsilon_i[k] = y[k] - \hat{y}_i[k]$  is the residual at time  $k$  for the  $i$ -th hypothesis.

The test statistic  $\mathbf{g}[k]$  of the CUSUM algorithm is:

$$\mathbf{g}[k] = (\mathbf{g}[k-1] + s[k])^+ \quad (3.82)$$

where  $(\mathbf{g}[k-1] + s[k])^+ = \max(0, \mathbf{g}[k-1] + s[k])$ . A change is detected if

$$\mathbf{g}[k] > \gamma_c \quad (3.83)$$

where  $\gamma_c$  is a user-defined parameter corresponding to the threshold for detection. The algorithm is initialized with  $\mathbf{g}[k] = 0$  and this term is reset each time (3.83) holds. The CUSUM test can be extended to different transfer functions as in [Table 3.1](#) by adapting (3.81) to the corresponding model residuals. The models under each hypothesis  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are estimated using sliding windows that are further explained in [section 5.6](#).

### 3.6 Persistence of Excitation

Persistently exciting signals are required for system identification when using methods such as instrumental variables or PEM. The degree of excitation depends among others on the model order, the operating mode *i.e.* if the data is collected from a process in open or closed-loop and the order of the controller in the latter case. A detailed analysis on the required conditions for system identification with the models listed in Table 3.1 and using the PEM is presented in [24]. The following definition of persistence of excitation is based on [80]. Consider a truncated weighting function model defined by:

$$y[k] = \sum_{l=0}^{n-1} h[k]u[k-l] + w[k], \quad k = 0, 1, \dots, N-1 \quad (3.84)$$

For the asymptotic case ( $N \rightarrow \infty$ ), the parameters  $\{h[k]\}_{k=0}^{n-1}$  are obtained by solving

$$\begin{pmatrix} r_u[0] & \dots & r_u[n-1] \\ \vdots & \ddots & \vdots \\ r_u[n-1] & \dots & r_u[0] \end{pmatrix} \begin{pmatrix} h[0] \\ \vdots \\ h[n-1] \end{pmatrix} = \begin{pmatrix} r_{yu}[0] \\ \vdots \\ r_{yu}[n-1] \end{pmatrix} \quad (3.85)$$

To solve the system of equations in (3.85), the matrix on the left side needs to be inverted *i.e.* it must be non-singular. The latter condition yields the concept of persistent excitation which is introduced as presented in [80]. A signal  $\{u[k]\}$  is persistently exciting (*p.e.*) of order  $n$  if:

(1) the following limit exists:

$$r_u(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} u[k+\tau]u^T[k], \quad \text{and} \quad (3.86a)$$

(2) the matrix:

$$R_u[p] = \begin{pmatrix} r_u[0] & r_u[1] & \dots & r_u[n-1] \\ r_u[-1] & r_u[0] & & \vdots \\ \vdots & \ddots & & \vdots \\ r_u[1-n] & \dots & & r_u[0] \end{pmatrix} \quad (3.86b)$$

is positive definite.

If the stochastic process is ergodic, the term  $\lim_{N \rightarrow \infty} \frac{1}{N}$  can be replaced by the expectation operator  $E(\cdot)$ . Then,  $R_u[p]$  is the usual covariance matrix. The

ergodicity theorem states that, for a stationary i.i.d. random signal  $\mathbf{x}$  with mean  $E(\mathbf{x})$ , the sample mean converges with probability one to  $\mu_{\mathbf{x}}$  provided that the number of observations  $N$  tends to infinity. The definition for persistence of excitation previously presented applies for univariate signals and asymptotic cases *i.e.*  $N \rightarrow \infty$ . In contrast to univariate signals, additional aspects such as orthogonality need to be considered for the design of persistently exciting signals for multivariate processes as reported in [41]. The following definition can be stated for persistently exciting multivariate signals.

**Definition 1** *Persistent excitation – finite-length multivariate signal [46]. A multivariate deterministic sequence of length  $N$  and size  $p$ ,  $\mathbf{r} \in \mathbb{R}^{N \times p}$ , is said to be persistently exciting (p.e.) of order  $m$  if the matrix:*

$$\mathbf{R}_{N-1} = \begin{pmatrix} \mathbf{r}[0] & \mathbf{r}[1] & \dots & \mathbf{r}[N-m] \\ \mathbf{r}[1] & \mathbf{r}[2] & \dots & \mathbf{r}[N-m+1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{r}[m-1] & \mathbf{r}[m] & \dots & \mathbf{r}[N-1] \end{pmatrix} \quad (3.87)$$

where  $\mathbf{r}[k] = (r_1[k] \ r_2[k] \ \dots \ r_p[k])^T$ , has rank  $p \cdot m$ .

White noise is persistently exciting of all orders because the resulting covariance matrix,  $R_u(n) = \sigma^2 \mathbf{I}_n$ , is always positive definite. In contrast, a step function of magnitude  $L$  is persistently exciting of order 1. Consider an input-output model with input  $u[k]$  and additive disturbance  $\varepsilon[k]$  as in (3.13). The regression and parameter vector are defined by (3.25b) and (3.25c), respectively. The model is expressed in vector notation as in (3.25a). The existence of an estimate for  $\hat{\Theta}$  using, for instance, the LSM is subject to the nonsingularity (positive definiteness) of the covariance matrix:

$$\mathbf{C}(\Theta) = E(\Phi^T \Phi) \quad (3.88)$$

The input requires some *degree of excitation* to fulfill *i.e.*  $\mathbf{C}(\Theta) > \mathbf{0}$ . Consider  $\varepsilon = 0$ :

$$\Phi = \begin{pmatrix} -b_1 & \dots & -b_{n_b} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & -b_1 & \dots & -b_{n_b} \\ a_1 & \dots & a_{n_a} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & a_1 & \dots & a_{n_a} \end{pmatrix} \begin{pmatrix} \frac{1}{A(q^{-1})} u[k-1] \\ \vdots \\ \vdots \\ \frac{1}{A(q^{-1})} u[k-n_a-n_b] \end{pmatrix} \quad (3.89a)$$

with the definition  $\Phi \triangleq \mathcal{I}(-B, A)\tilde{\Phi}$ , where the model parameters are grouped in  $\mathcal{I}(-B, A)$ . Then, replacing  $\Phi$  from (3.89a) in (3.88):

$$\mathbf{R} = \mathcal{I}(-B, A)E\tilde{\Phi}[k]\tilde{\Phi}^T[k]\mathcal{I}^T(-B, A) \triangleq \mathcal{I}(-B, A)\tilde{\mathbf{R}}\mathcal{I}^T(-B, A) \quad (3.90)$$

it can be concluded from (3.90) that

$$\{\mathbf{R} > 0\} \iff \left\{ \mathcal{I} \text{ nonsingular and } \tilde{\mathbf{R}} > 0 \right\}$$

The term  $\mathcal{I}(-B, A)$  is nonsingular if the polynomials  $(-B, A)$  are coprime. Thus,  $\mathbf{R}$  is positive definite if and only if  $u[k]$  is (*p.e.*) of order  $n_a + n_b$ .

Next, consider  $\varepsilon \neq 0$  in (3.25a):

$$\bar{\varepsilon}[k] = -\frac{1}{A(q^{-1})}\varepsilon[k] \quad (3.91a)$$

$$x[k] = \frac{B(q^{-1})}{A(q^{-1})}u[k] \quad (3.91b)$$

In the case of open-loop *i.e.*  $E[u[k]\varepsilon[s]] = 0$  for all  $k$  and  $s$ , then it follows that:

$$R = E \begin{pmatrix} x[k-1] \\ \vdots \\ x[t-n_a] \\ u[k-1] \\ \vdots \\ u[k-n_b] \end{pmatrix} (x[k-1] \ \dots \ x[k-n_a] \ u[k-1] \ \dots \ u[k-n_b]) +$$

$$E \begin{pmatrix} \bar{\varepsilon}[k-1] \\ \vdots \\ \bar{\varepsilon}[k-n_a] \\ 0 \\ \vdots \\ 0 \end{pmatrix} (\bar{\varepsilon}[k-1] \ \dots \ \bar{\varepsilon}[k-n_a] \ 0 \ \dots \ 0) \quad (3.92a)$$

$$\triangleq \begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} + \begin{pmatrix} \bar{\mathbf{A}} & 0 \\ 0 & 0 \end{pmatrix} \quad (3.92b)$$

As explained in [80],  $\mathbf{C} > 0$  is necessary for  $R > 0$ . If this holds, then

$$\tilde{\mathbf{A}} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T \geq 0 \quad (3.93)$$



and the following equation can be stated

$$\text{rank } \mathbf{R} = n_b + \text{rank} (\tilde{\mathbf{A}} + \bar{\mathbf{A}} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T) \quad (3.94)$$

Assuming  $\bar{\mathbf{A}} > 0$ ,

$$\text{rank } \mathbf{R} = n_a + n_b \quad (3.95)$$

Finally, under weak conditions on the noise *i.e.*  $\tilde{\mathbf{A}} > 0$ , the required order of excitation of the input signal is

$$\mathbf{R} > 0 \iff u[k] \text{ is } (p.e.) \text{ of order } (n_b) \quad (3.96)$$

Identifiability conditions can be derived for different models depending on the operating mode (open or closed loop) and the model to be computed. A detailed analysis on identifiability conditions for ARMAX systems in closed-loop is presented in [76]. In multivariate processes, excitation signals should be (*p.e.*) and uncorrelated. An analysis for the MISO case is presented in [25]. Although simultaneous excitation is preferred as the variance of the parameters can be decreased, this scenario is not commonly found in processes addressed in this work. In such processes, it is more frequent that inputs are excited only one at a time. In this situation, parameter estimation can then suffer numerical problems due to ill-conditioning in matrix operations. Data merging, as it will be shown in Chapter 4, can alleviate this problem.

### 3.7 Discussion

Algorithms for parameter estimation can suffer numerical problems when considering data sets that exhibit few changes. This concept will be used further to propose a selection method for informative data. Matrix conditioning for parameter estimation will be evaluated to determine useful sequences for identification. LTI models and different methods for parameter estimation were introduced in this chapter. The one-step-ahead predictor computes the output of a process at time  $k$  from past input-output data up to  $k - 1$ . Parametric models are classified into two groups: equation and output error models. The input and noise transfer functions are independently parameterized in output error models whereas, in the equation error class the transfer functions share common parameters.

Three identification methods were introduced: LSM, IVM and PEM. The LSM can yield consistent estimates only in case of WGN which is a strong requirement as, in real situations, noise is more likely to be colored. To

overcome this drawback, IVM and PEM can be used as they are robust to colored noise. In the present work, implementation of IVM is preferred as they are less computationally demanding than PEM and have satisfactory accuracy. In batch identification methods, the entire data set is used at once which can be computationally demanding in case of large data sets. Instead, recursive methods can be considered. Although the convergence rate of recursive estimation methods depends on the initialization of the algorithm and user-defined parameters, these methods are a practical solution for interval bounding. This idea will be used as part of the data selection method in [Chapter 5](#).

Persistently exciting signals are required to generate informative data. This relevant concept associated to input signals was presented at the end of this section. The concept was introduced for FIR models and further extended to input-output models. Under weak conditions on the measurement noise, a signal needs to be (*p.e.*) of order  $n_b$  to permit to identify an input-output model. Identification methods are formulated assuming persistently exciting signals [[57](#), [80](#)]. This condition is also considered in DS4SID since estimation methods were adapted based on the state of the art. However, identifiability conditions as in [\(3.96\)](#) can be used in a data selection method to conclude if the data is informative for the given conditions.



# CHAPTER 4

---

## Combination of Data Sets for System Identification

---

In some applications, parameter estimation needs to be performed with data sets collected from different experiments. This context has some similarities with data selection for identification. Models should be computed with selected sequences that correspond to informative intervals. Informative intervals cannot be simply concatenated for model estimation. Suitable data merging techniques for identification are introduced in this chapter. They will be used in a later stage of the proposed data selection method.

### 4.1 Data Merging Problem

The data merging problem [57], applied to data selection, is to estimate a model from separate informative intervals. This problem is called combination of data sets in [54]. Data intervals cannot be simply concatenated as the connecting points may generate transients that will affect the parameter estimation. Thus, identification methods should be modified to estimate a model with informative intervals. These useful sequences can be interpreted as data obtained from separate experiments. The topic treated in this chapter is a practical aspect of system identification. Other issues such as possibly non-zero mean data (see, for instance, [40]) need also to be considered. In fact, the mean values of  $\{y[k]\}$  and  $\{u[k]\}$  corresponding to an operating point should be estimated. (see [41, 80]) Next, the data merging problem and suitable solutions are presented.

The combination of  $q$  informative intervals yields a new subset  $\mathbf{Z}^D$ :

$$\mathbf{Z}^D = \bigcup_{i=1}^q \mathbf{Z}^{n_i} \quad (4.1a)$$

where  $i = 1, 2, \dots, q$  represents the  $i$ -th informative interval of length  $n_i$

$$\mathbf{Z}^i = \{y[k_{i1}], u[k_{i1}], y[k_{i1} + 1], u[k_{i1} + 1], \dots, y[n_i], u[n_i]\} \quad (4.1b)$$

the lower and upper bounds of the  $i$ -th interval are represented by

$k_{i1}$  and  $k_{i2}$ , respectively, where  $k_{i2} = k_{i1} + n_i - 1$

Informative intervals as presented in (4.1b) will be expressed in compact notation by

$$\{\mathbf{Z}[k]\}_{k=k_{i1}}^{k_{i2}} \quad (4.2)$$

A model  $\mathcal{M}(\hat{\Theta}_i)$  can be estimated for each informative interval. Assume that all intervals have models of the same class. Then, the resulting estimate considering the subset  $\mathbf{Z}^D$  is [57]:

$$\hat{\Theta} = \hat{P} \sum_{i=1}^q \left( \hat{P}^{(i)} \right)^{-1} \hat{\Theta}^{(i)} \quad (4.3a)$$

$$\hat{P} = \left( \sum_{i=1}^q \left( \hat{P}^{(i)} \right)^{-1} \right)^{-1} \quad (4.3b)$$

where the least squares estimate  $\hat{\Theta}^{(i)}$  is computed by:

$$\hat{\Theta}^{(i)} = \left( \sum_{k=k_{i1}}^{k_{i2}} \varphi[k] \varphi^T[k] \right)^{-1} \left( \sum_{k=k_{i1}}^{k_{i2}} \varphi[k] y[k] \right) \quad (4.4)$$

The computation of  $\hat{\Theta}$  in (4.3a) also requires the estimated covariance matrix of the parameters obtained for the  $i$ -th interval. The covariance matrix  $P^{(i)}$  is estimated by:

$$P^{(i)} = \hat{\sigma}_{e_i}^2 \left( \sum_{k=k_{i1}}^{k_{i2}} \varphi[k] \varphi^T[k] \right)^{-1} \quad (4.5)$$

The noise variance  $\hat{\sigma}_{e_i}^2$  is estimated by:

$$\hat{\sigma}_{e_i}^2 = \frac{1}{n_i} \sum_{k=k_{i1}}^{k_{i2}} \left( y[k] - \varphi^T[k] \hat{\Theta}^{(i)} \right)^2 \quad (4.6)$$

Parameter estimation for each informative interval involves matrix inversion that can suffer numerical problems due to poorly excited processes. The value of a scalar measure can also be used to evaluate the contribution of each new observation in the numerical conditioning of the corresponding information matrix. A weighting matrix can then be included in the parameter estimation method whose weights are determined based on a scalar measure of the information matrix. The next section introduces weighted identification methods that will be used together with data merging in the last stage of data selection.

## 4.2 Weighted Identification Methods

The identification methods presented in [section 3.3.1](#) and [section 3.3.2](#) can be modified by including a weighting matrix  $\mathbf{W}$  that is associated, for instance, to the relevance of each regressor. In data selection methods, large weights will be associated to data located close to changes in the process. The weighted least squares cost function for the MISO model (3.30) is

$$V(\hat{\Theta}) = (\mathbf{y} - \Phi\hat{\Theta})^T \mathbf{W} (\mathbf{y} - \Phi\hat{\Theta}) \quad (4.7a)$$

$$\mathbf{y} = (y[0] \ y[1] \ \dots \ y[N-1])^T \quad (4.7b)$$

$$\Phi = \begin{pmatrix} \boldsymbol{\varphi}^T[0] \\ \vdots \\ \boldsymbol{\varphi}^T[N-1] \end{pmatrix} \quad (4.7c)$$

where  $\boldsymbol{\varphi}$  and  $\Theta$  are represented by (3.31a) and (3.31b), respectively. The term  $\mathbf{W}$  denotes a positive definite weighting matrix. In the case  $\mathbf{W}$  is a diagonal matrix with  $w_i > 0$ , the LS error is:

$$V(\hat{\Theta}) = \frac{1}{N} \sum_{k=0}^{N-1} w[k] \varepsilon^2[k] \quad (4.8)$$

The solution to the weighted least squares problem (4.7a) is:

$$\hat{\Theta} = (\Phi^T \mathbf{W} \Phi)^{-1} (\Phi^T \mathbf{W} \mathbf{y}) \quad (4.9)$$

In the case  $\mathbf{W} = \mathbf{I}$ , the solution is the ordinary least squares estimate as presented in (3.40).

Weighted IVM are analogous to the EIV method (3.78). The resulting cost function of the MISO model (3.30) including instrumental variables and a weighting matrix is:

$$V(\hat{\Theta}) = \frac{1}{2} \boldsymbol{\varepsilon}^T \mathbf{W}^T \zeta \zeta^T \mathbf{W} \boldsymbol{\varepsilon} \quad (4.10a)$$

$$\zeta = \begin{pmatrix} \zeta^T[0] \\ \vdots \\ \zeta^T[N-1] \end{pmatrix} \quad (4.10b)$$

where  $\zeta$  are instrumental variables as in (3.52).  $V(\hat{\Theta})$  is minimized by:

$$\hat{\Theta} = (\zeta^T \mathbf{W} \Phi)^{-1} (\zeta^T \mathbf{W} \mathbf{y}) \quad (4.11)$$

where  $\mathbf{y}$  and  $\Phi$  are defined by (4.7b) and (4.7c), respectively.

Further, it will be shown that information retrieved from the data selection method can be used to propose a weighting matrix. The resulting matrix  $\mathbf{W}$  can be used in the last step of the data selection method aiming at estimating better models than when using “standard” identification methods.

### 4.3 Least Squares Approach for Data Merging

Consider informative intervals  $\mathbf{Z}^i$ ,  $i = 1, 2, \dots, q$  associated to  $q$  informative intervals each of length of lengths  $n_i$ . The resulting data set  $\mathcal{D}$  is obtained as the union of the informative sequences:

$$\mathcal{D} : \{ \{ \mathbf{Y}_1, \mathbf{U}_1 \}, \{ \mathbf{Y}_2, \mathbf{U}_2 \}, \dots, \{ \mathbf{Y}_{n_q}, \mathbf{U}_{n_q} \} \} \quad (4.12)$$

with,

$$\mathbf{Y}_i^T = (\mathbf{y}[0]^T \dots \mathbf{y}[n_i - 1]^T) \in \mathbb{R}^{n_i \times n_y} \quad (4.13)$$

$$\mathbf{U}_i^T = (\mathbf{u}[0]^T \dots \mathbf{u}[n_i - 1]^T) \in \mathbb{R}^{n_i \times n_u} \quad (4.14)$$

and  $n_u$  and  $n_y$  correspond to the number of inputs and outputs, respectively.

Let  $V_i$  be a cost function applied to the  $i$ -th informative interval. The total cost function is expressed by:

$$V_{\text{total}}(\hat{\Theta}) = \sum_{i=1}^q V_i(\hat{\Theta}_i) \quad (4.15)$$

The parameter vector  $\hat{\Theta}$  is then estimated by minimizing the total cost function (4.15).

Assume that two informative intervals are available ( $q = 2$ ), the least squares error is used as a minimizing criterion in the following cost functions  $V_i$ .

$$V_i(\hat{\Theta}_i, \mathbf{Z}^i) = \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i \quad (4.16)$$

Replacing (4.16) in (4.15) and taking the derivative for  $i = 1, 2$  yields

$$\frac{\partial V_{\text{total}}}{\partial \Theta} = \Phi_1^T \Phi_1 \hat{\Theta} - \Phi_1^T \mathbf{Y}_1 + \Phi_2^T \Phi_2 \hat{\Theta} - \Phi_2^T \mathbf{Y}_2 \quad (4.17)$$

The gradient (4.17) is zero if and only if the following holds:

$$(\Phi_1^T \Phi_1 + \Phi_2^T \Phi_2) \hat{\Theta} = \Phi_1^T \mathbf{Y}_1 + \Phi_2^T \mathbf{Y}_2 \quad (4.18)$$

The least squares estimate is then computed as

$$\hat{\Theta} = \left( \sum_{i=1}^q H_i \right)^{-1} \sum_{i=1}^q g_i \quad (4.19)$$

$$\text{with, } H_i = \Phi_i^T \Phi_i, g_i = \Phi_i^T \mathbf{Y}_i \text{ and } q = 2 \quad (4.20)$$

The method (4.19) will yield biased estimates in the case of colored noise. Thus, alternative identification methods are necessary to deal, for instance, with correlated additive noise. data merging by the LSM should be performed using (4.19) since better conditioned matrices are obtained instead of solving (4.3).

## 4.4 Instrumental Variables Approach for Data Merging

Consider  $q$  informative intervals each of size  $n_i = k_{i2} - k_{i1}$ ,  $i = 1, 2, \dots, q$ . The resulting data set is

$$\mathcal{D} : \{ \{\mathbf{y}_1, \mathbf{U}_1\}, \{\mathbf{y}_2, \mathbf{U}_2\}, \dots, \{\mathbf{y}_q, \mathbf{U}_q\} \} \quad (4.21a)$$

$$\mathbf{y}_i = (y[k_{i1}] \dots y[k_{i2}])^T \in \mathbb{R}^{(k_{i2}-k_{i1}+1) \times 1} \quad (4.21b)$$

$$\mathbf{U}_i = (\mathbf{u}[k_{i1}] \dots \mathbf{u}[k_{i2}])^T \in \mathbb{R}^{(k_{i2}-k_{i1}+1) \times n_u} \quad (4.21c)$$

where  $k_{i1}$  and  $k_{i2}$  are the lower and upper interval bounds, respectively. The equation (3.44) needs to be solved for each of the retrieved intervals as follows

$$\frac{1}{k_{12} - k_{11} + 1} \sum_{k=k_{11}}^{k_{12}} \zeta_1[k] \boldsymbol{\varepsilon}_1[k] + \dots + \frac{1}{k_{q2} - k_{q1} + 1} \sum_{k=k_{q1}}^{k_{q2}} \zeta_q[k] \boldsymbol{\varepsilon}_q[k] = 0 \quad (4.22a)$$

$$\boldsymbol{\varepsilon}_i[k] = y[k] - \boldsymbol{\varphi}^T[k] \hat{\Theta} \quad k = k_{i1}, k_{i1} + 1, \dots, k_{i2} \quad (4.22b)$$

The parameter vector is computed by substituting (4.22b) in (4.22a):

$$\hat{\Theta} = \left( \sum_{i=1}^q H_i \right)^{-1} \left( \sum_{i=1}^q g_i^{-1} \right) \quad (4.23a)$$

$$H_i = \sum_{k=k_{i1}}^{k_{i2}} \zeta[k] \boldsymbol{\varphi}^T[k] \quad g_i = \sum_{k=k_{i1}}^{k_{i2}} \zeta[k] y[k] \quad (4.23b)$$



Parameter estimation using (4.23) yields better conditioned matrices than when considering the entire data set. Moreover, the computational time decreases since the resulting matrices have a smaller size than presented in (3.78).

## 4.5 Multiple-cost Approach for Data Merging

The multiple-cost approach as presented in [54] aims at finding a minimum of a cost function  $V_{\text{total}}$  w.r.t the parameter vector  $\Theta$ . The total cost function is:

$$V_{\text{total}} = \sum_{i=1}^q V_i \quad (4.24)$$

where  $V_i$  is the cost function for the  $i$ -th informative interval. Cost functions as in (3.39) or (3.44) can be used for  $V_i$ . The estimation problem using the multiple-cost approach is formulated as:

$$\hat{\Theta} = \hat{\Theta}(\mathbf{Z}^D) = \arg \min_{\Theta \in D_{\mathcal{M}}} V_{\text{total}} \quad (4.25)$$

where  $D_{\mathcal{M}}$  is the set of values over which  $\Theta$  ranges. Consider that the parameter vector associated with each interval  $\hat{\Theta}_i$  and  $\hat{\Theta}$  are related as follows

$$\hat{\Theta} = \sum_{i=1}^q \hat{\Theta}_i + \Delta \hat{\Theta}_i \quad (4.26)$$

where  $\Delta \hat{\Theta}_i$  represent increments required to equal the individual-run estimates to the multiple-cost one. Minimization of (4.24) yields

$$\left. \frac{\partial V_{\text{total}}}{\partial \Theta} \right|_{\Theta} = \sum_{i=1}^q \left. \frac{\partial V_i}{\partial \Theta_i} \right|_{\hat{\Theta}_i + \Delta \hat{\Theta}_i} \quad (4.27)$$

For a minimum, the following equality must hold

$$\sum_{i=1}^q \left. \frac{\partial V_i}{\partial \Theta} \right|_{\Theta_i + \Delta \hat{\Theta}_i} = 0 \quad (4.28)$$

Expanding the left-hand side in the vicinity of the estimates and neglecting higher-order terms results in

$$\sum_{i=1}^q \left. \frac{\partial V_i}{\partial \Theta} \right|_{\hat{\Theta}_i} + \sum_{i=1}^q \left. \frac{\partial^2 V_i}{\partial \Theta^2} \right|_{\hat{\Theta}_i} \Delta \Theta_i \approx 0 \quad (4.29)$$

By definition, the first term in (4.29) is zero. Then, replacing (4.26) in (4.29) yields

$$\sum_{i=1}^q \frac{\partial^2 V_i}{\partial \Theta^2} \Big|_{\Theta_i} (\Theta - \Theta_i) \approx 0 \quad (4.30)$$

The parameter vector  $\Theta$  is then estimated by:

$$\hat{\Theta} = \left( \sum_{i=1}^q \frac{\partial^2 V_i}{\partial \Theta^2} \Big|_{\hat{\Theta}_i} \right)^{-1} \left( \sum_{i=1}^q \frac{\partial^2 V_i}{\partial \Theta^2} \Big|_{\hat{\Theta}_i} \hat{\Theta}_i \right) \quad (4.31)$$

In short notation, (4.31) can be expressed by

$$\hat{\Theta} = \left( \sum_{i=1}^q \mathbf{M}_i \right)^{-1} \left( \sum_{i=1}^q \mathbf{M}_i \hat{\Theta}_i \right) \quad (4.32)$$

where  $\mathbf{M}_i$  is the information matrix associated to the  $i$ -th informative interval.

The solutions in (4.19) and (4.23a) yield similar results as when using the multiple-cost approach. However, an additional operation is required when using multiple-cost as local estimates need to be computed for each informative interval. The information matrix  $\mathbf{M}_i$  can be replaced by one of the obtained in section 3.3. Models estimated with entire data sets and with informative intervals will be further discussed in Chapter 6.

## 4.6 Discussion

The data merging problem refers to estimating a model with data retrieved from different informative intervals. The data cannot be simply joined as each informative interval has a different non-zero mean value and, consequently, the total estimate will be affected. Models associated to informative sequences should have the same structure and order for data merging. The total estimate is obtained by minimizing a total cost function which is a sum of cost functions for each interval.

Three solutions to this problem were presented in this chapter: least squares, instrumental variables and multiple-cost. The first two approaches have similar features as their counterpart methods used for batch off-line identification. The least squares approach cannot guarantee consistent estimates in case of correlated additive noise. As previously discussed, this is a significant drawback as correlated noise is likely to be expected in real applications.

The IV method for data merging is robust against correlated noise. It is based on the EIV and represents an alternative to the least squares approach which is commonly found in the literature. A multiple-cost approach is a more general technique where the estimate is computed minimizing the sum of individual cost functions. Although it represents a general method, it is computationally more demanding than the other two approaches as estimates are computed for each interval. An approach for nonlinear system identification using different data sets is presented in [56]. Results from simulation and real case studies show that suitable models are obtained using the method in [56] with different data sets. The procedures presented in this chapter will be compared and discussed using simulation and real case studies.

# CHAPTER 5

---

## Data Selection Method for System Identification (DS4SID)

---

In the present chapter, a new and modular method, DS4SID, is introduced as data selection approach which addresses limitations found in the available methodologies. The problem of selecting informative intervals is divided into three main tasks: change detection, interval bounding and data merging. DS4SID differs in several aspects from current methods. Firstly, it can be applied to multivariate processes which is a relevant extension to the current state of the art. Additional tests to evaluate if a change in the process was generated by an excitation in the input, are not required. This aspect is implicitly evaluated in the detection test. Secondly, upper interval bounds are defined by thresholding the gradient of the reciprocal of the condition number computed for the information matrix obtained from a chosen parameter estimation method. This feature can decrease misdetection of informative intervals and enhances the performance when compared to current approaches. Additionally, data normalization is not required which saves processing time. Thirdly, the proposed detection method is robust to colored noise that is most likely to be found in real applications.

### 5.1 Identifying Data Sets Predominantly at Steady-state

Parameter computation with predominantly stationary data sets can have negative effects on the estimates. To analyze this situation, consider a process described by:

$$(1 - 0.3q^{-1} + 0.5q^{-2})y[k] = 1.5q^{-1}u[k] + (1 + 0.8q^{-1} + 0.3q^{-2})e[k] \quad (5.1)$$

where  $\{e[k]\}$  is an i.i.d. normally distributed random sequence with  $\mathcal{N}(0, \sigma_e^2)$ . From (5.1), the input transfer function is

$$G_0(q^{-1}) = \frac{1.5q^{-1}}{1 - 0.3q^{-1} + 0.5q^{-2}} \quad (5.2)$$

The “true” parameter vector of the input transfer function is

$$\Theta_0 = (-0.3 \ 0.5 \ 1.5)^T \quad (5.3)$$

The order for the parameters of  $A(q^{-1})$  and  $B(q^{-1})$  are  $n_a = 2$  and  $n_b = 1$ , respectively.

The following model is used to identify (5.1):

$$(1 + a_1q^{-1} + a_2q^{-2})y[k] = b_1q^{-1}u[k] + \epsilon[k] \quad (5.4)$$

The input transfer function of the model (5.4) is:

$$G(q^{-1}; \hat{\Theta}) = \frac{b_1q^{-1}}{1 + a_1q^{-1} + a_2q^{-2}} \quad (5.5)$$

The process (5.1) can be identified using an input signal with order of excitation  $n_b$  (see section 3.6). The following signal is used for experiments:

$$u[k] = \begin{cases} 0; & 0 \leq k < 25 \\ 1; & 25 \leq k \leq 1000 \end{cases} \quad (5.6)$$

The use of step signals as (5.6) represents a typical case found in continuously operated plants. Processes are moved from a current to a new operating point as a result of excitation signals as previously described.

The stochastic process in (5.1) is colored Gaussian noise which represents a situation that is likely to be found in logged data. The variance of  $\{e[k]\}$  was adjusted for several SNR (cf. (2.37)). Two identification methods are used for parameter estimation: the LSM and optimal IVM (see section 3.3.1 and section 3.3.2). Estimated input transfer functions and resulting information matrices are analyzed. The following criterion is used to evaluate the resulting input transfer function:

$$Norm = \frac{1}{l} \sum_{j=1}^l \int \left| G_0(e^{i\omega}) - \hat{G}_j(e^{i\omega}) \right|^2 d\omega \quad (5.7)$$

where  $\hat{G}_j(e^{i\omega})$  represents the estimated input transfer function for the  $j$ -th experiment,  $l$  is the total number of experiments and  $Norm$  evaluates the difference in the magnitude of the frequency response between the real and the estimated input transfer functions. Even though repetition of experiments is constrained in real applications, the proposed example helps to give an

overview of the effect of poorly excited data sets on parameter estimation. The conditioning of the auxiliary information matrix at  $k$ ,  $\tilde{\mathbf{I}}(k, \hat{\Theta})$  is evaluated with the reciprocal of the condition number  $\tilde{\kappa}[k]$

$$\tilde{\kappa}[k] = \frac{\lambda_{\min}(\tilde{\mathbf{I}}(k; \hat{\Theta}))}{\lambda_{\max}(\tilde{\mathbf{I}}(k; \hat{\Theta}))} \quad (5.8)$$

Since the additive disturbance in (5.1) is colored noise, the least squares estimate will be biased. The estimated parameter vector will not converge asymptotically to the true value but will exhibit a difference called bias. In the case of using input signals with high order of excitation, sample information matrices are well-conditioned. However, a different situation is observed if signals with low order of excitation as in (5.6) are used.

The following notation is adopted for the estimated input transfer functions, the sample information matrices and the reciprocal of the condition number:

$\hat{G}_1(k; q^{-1}), \mathbf{I}(k; \hat{\Theta}_1), \tilde{\kappa}(\mathbf{I}(k; \hat{\Theta}_1))$  : estimated input transfer function,  
information matrix and reciprocal of the  
condition number using least squares

$\hat{G}_2(k; q^{-1}), \mathbf{I}(k; \hat{\Theta}_2), \tilde{\kappa}(\mathbf{I}(k; \hat{\Theta}_2))$  : estimated input transfer function,  
information matrix and reciprocal of the  
condition number using instrumental variables

The input transfer functions as well as the information matrices were computed for the following identification data sets:

$$\mathbf{Z}^0 = \{y[0], u[0], y[1], u[1], \dots, y[k_0 - 1], u[k_0 - 1]\} \quad (5.9a)$$

$$\mathbf{Z}^1 = \{y[0], u[0], y[1], u[1], \dots, y[k_0], u[k_0]\} \quad (5.9b)$$

$\vdots$

$$\mathbf{Z}^{k_2} = \{y[0], u[0], y[1], u[1], \dots, y[k_2 - 1], u[k_2 - 1]\} \quad (5.9c)$$

$\vdots$

$$\mathbf{Z}^N = \{y[0], u[0], y[1], u[1], \dots, y[N - 1], u[N - 1]\} \quad (5.9d)$$

where  $\mathbf{Z}^N$  corresponds to the entire data set and  $N$  is the number of observations. The initial data set  $\mathbf{Z}^0$  is chosen so that  $k_0 \gg n_p$ , where  $n_p$  is the model order. This choice will avoid numerical problems. In  $\mathbf{Z}^{k_2}$ ,  $k_2 = k_{\text{step}} + n$  where  $k_{\text{step}}$  is the instant when the system is excited with the step signal and  $n > n_p$ .

The term  $\mathbf{Z}^{k_2}$  will represent the ideal data set for parameter estimation since, for  $k > k_2$ , the conditioning of the sample information matrix will start to exhibit ill-conditioning. This means, that, in case the model order is correctly chosen, the entire data set should not be preferred for parameter estimation. Instead, the estimates should be computed with selected sequences as otherwise numerical problems can appear when performing computations.

A total of 100 experiments were performed using different realizations of  $\{\mathbf{e}[k]\}$  in (5.1) and the same input signal (5.6). The norm (5.7) of the estimated transfer functions using the LSM is shown in the top left plot of Figure 5.1.

In the case of input signals with high order of excitation, for instance a pseudorandom binary sequence (PRBS), the bias of the estimated input transfer function is approximately constant [17]. However, (5.7) increases for  $k > k_2$  with the input signal in (5.6). Thus, parameters should be estimated with the data set  $\{\mathbf{Z}\}_{k=k_1}^{k_2}$  so that the parameter bias does not increase as more data are used for identification. In real applications, the time  $k_2$  cannot be retrieved from a criterion such as (5.7) since the “true” transfer function  $G_0(q^{-1}, \Theta_0)$  is unknown.

The reciprocal of the condition number is a criterion that has been shown to work well in practice. It is scaled between 0 and 1, which facilitates the choice of thresholds that can be used in different applications. The reciprocal of the condition number  $\tilde{\kappa}$  for the information matrix using the LSM is shown in the bottom left plot of Figure 5.1. As a general guideline, large values of  $\tilde{\kappa}$  indicate well conditioned matrices whereas ill-conditioning is associated to small values of  $\tilde{\kappa}$ .

The information matrix  $\mathbf{I}(k; \hat{\Theta}_1)$  is better conditioned with larger data sets as shown in the bottom left of Figure 5.1. This is observed for different SNRs used in simulations. The conditioning of  $\mathbf{I}(k; \hat{\Theta}_1)$  using the LSM starts to improve after a certain instant. However, worse estimates of the input transfer function are obtained as shown in the top left of Figure 5.1 because identification with largely stationary data sets increases the parameter bias. A criterion for selecting informative data based on the conditioning of the information matrix resulting from the LSM is unreliable in the considered example since the matrix conditioning improves as more data are used for estimation but, simultaneously, parameter bias increases. A discarding criterion based on  $\tilde{\kappa}(\mathbf{I}(k; \hat{\Theta}_1))$  will indicate that more stationary data should be used for identification since the conditioning is improving but, indeed, this will increase the parameter bias.

When dealing with very large data sets, practical implementations show that model quality evaluated for instance using normalized root mean square error (NRMSE) will just improve slightly if more observations are used for estimation.

[5, 12]. Thus, only a selection of the identification data set should be used in the case of processes that are externally seldom excited. Alternative identification methods other than LSM need to be considered for selection of informative intervals. Among others, methods based on instrumental variables (IV) and the PEM can be used as part of a procedure for selection of informative data based on the evaluation of the conditioning of the corresponding information matrix.

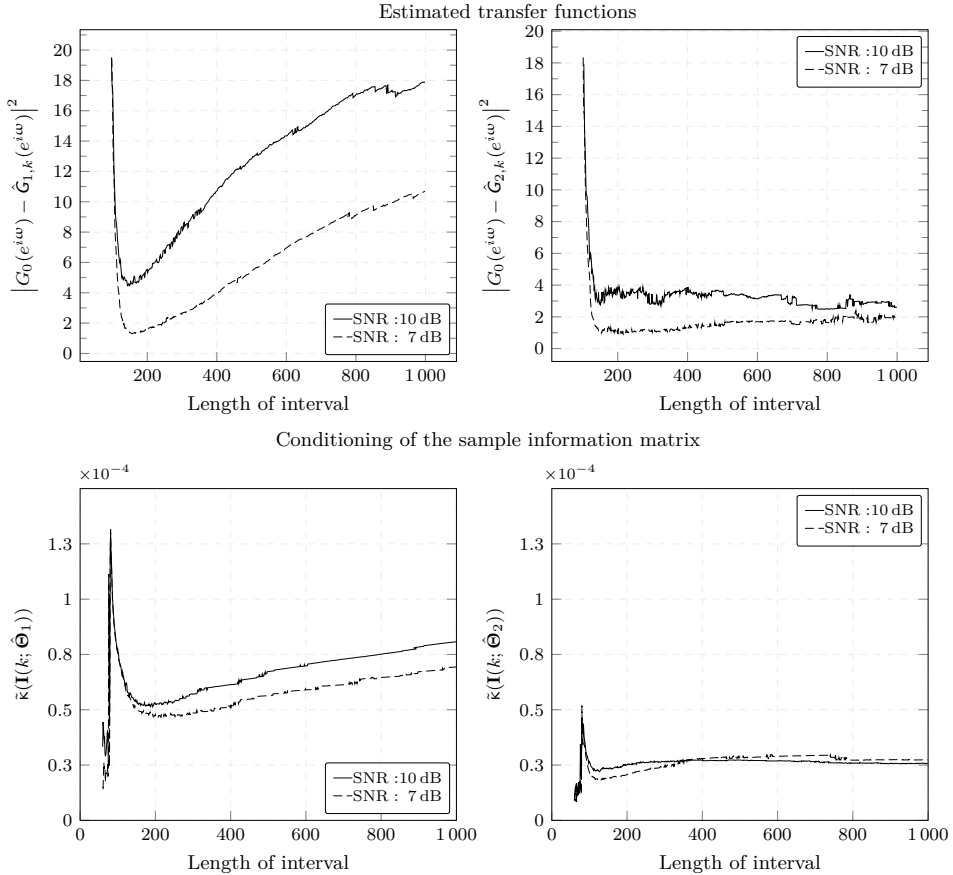
The top right plot of Figure 5.1 shows the norm (5.7) of input transfer functions for different SNRs estimated using an optimal IV method. In contrast to the LSM, the bias of the input transfer function vanishes as the length of the interval increases for 10 dB. However, IV estimates also show an increasing bias in case of 7 dB. This can be explained by the “short” length of the data sets used for estimation.

As shown in Figure 5.1, the optimal IV method is robust to colored noise since the bias of  $\hat{\mathbf{G}}_2(k; q^{-1})$  asymptotically tends to zero. The bias of the estimates will slightly change with the length of intervals used for computations. Thus, parameter estimation with  $\{\mathbf{Z}\}_{k=k_1}^{k_{21}}$  or  $\{\mathbf{Z}\}_{k=k_1}^{k_{22}}$  where  $k_2 < k_{21} < k_{22}$  will yield slightly improved estimates but computations using those data sets will be more demanding. Consequently, a small data set instead of the entire data set should be used for identification. In practical situations, the sample information matrix  $\mathbf{I}(k; \hat{\Theta}_2)$  should be evaluated to define informative intervals for identification.

The conditioning of  $\mathbf{I}(k; \hat{\Theta}_2)$  is shown in the bottom right of Figure 5.1. The condition number worsens when using an optimal IV method and identification data sets that are predominantly at steady-state. The condition number reaches a maximum around  $k = k_2$  which corresponds to an instant some samples after the process is excited with the step signal. From that instant, the larger the identification data set becomes the smaller is  $\tilde{\kappa}(\mathbf{I}(k; \hat{\Theta}_2))$  that shows a degradation in the conditioning of the corresponding information matrix. Thus, as suggested by the bottom right plot of Figure 5.1, parameters should be estimated with a selected data set which will avoid numerical problems. If the subset  $\{\mathbf{Z}\}_{k=k_1}^{k_2}$  was used for identification, it will represent around 10 % of the entire data set  $\mathbf{Z}^N$ .

The information matrix associated with the LSM is better conditioned than its counterpart obtained by the IV method as can be concluded from the values of  $\tilde{\kappa}(\tilde{\mathbf{I}}(k; \hat{\Theta}_1))$  and  $\tilde{\kappa}(\tilde{\mathbf{I}}(k; \hat{\Theta}_2))$ , respectively. In case of using the optimal IV method, additional parameters are estimated since the input and noise transfer functions are independently parametrized. Thus, the information matrices are  $\tilde{\mathbf{I}}(k, \hat{\Theta}_1) \in \mathbb{R}^{n_1 \times n_1}$  when using LSM and  $\tilde{\mathbf{I}}(k, \hat{\Theta}_2) \in \mathbb{R}^{n_2 \times n_2}$  when using the IV

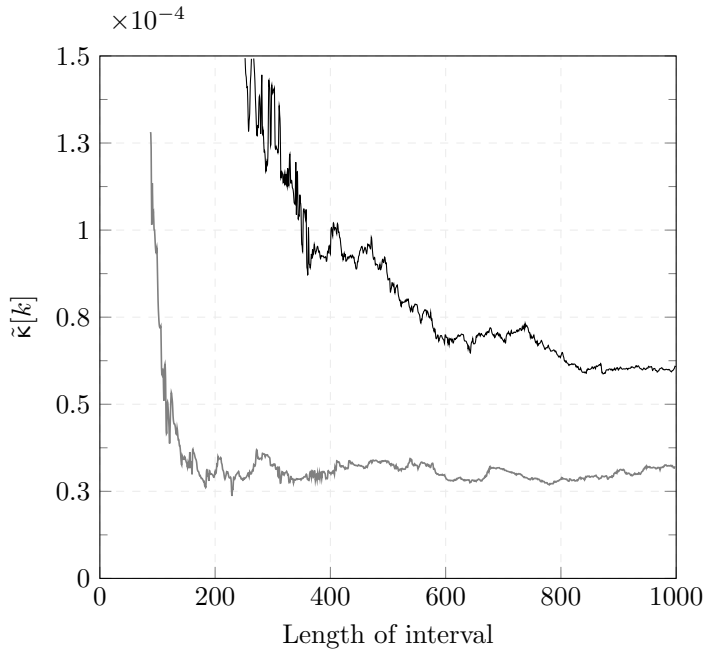




**Figure 5.1:** Estimated transfer functions and conditioning of the sample information matrices dependent on the data interval length using the LSM (left) and an optimal IV method (right)

method where  $n_1 < n_2$ . Consequently, computation of additional parameters in case of using the IV method yields lower values of  $\tilde{\kappa}$  than when using the LSM.

The effect of identification data sets predominantly at steady-state on parameter estimation was examined with the system (5.1). The condition number  $\tilde{\kappa}(\tilde{\mathbf{I}}(k; \hat{\Theta}_2))$  degrades as shown in the bottom right of Figure 5.1. It was shown that the corresponding information matrix, called instrumental product moment matrix (IPMM), degrades as larger identification data sets are used for parameter estimation. Former analysis suggests that the “most” informative data for system identification represents only a small part of predominantly stationary data sets.



**Figure 5.2:** Reciprocal of the condition number for different models for the system (5.1) excited with (5.6) (Solid black line:  $n_a = 3$ . Solid gray line:  $n_k = 2$ )

The behavior of  $\tilde{\kappa}(\tilde{\mathbf{I}}(k; \hat{\Theta}))$  when the model and the process are not in the same class *i.e.*  $\mathcal{G} \notin \mathcal{M}$  is analyzed for two different models described as follows:

$$\mathcal{M}_1 : (1 + a_1q^{-1} + a_2q^{-2} + a_3q^{-3}) y[k] = b_1q^{-1}u[k] + \epsilon[k] \quad (5.10)$$

$$\mathcal{M}_2 : (1 + a_1q^{-1} + a_2q^{-2}) y[k] = b_1q^{-2}u[k] + \epsilon[k] \quad (5.11)$$

The order of the “true” model (5.1) is  $n_a = 2$ ,  $n_b = 1$ ,  $n_k = 1$ , where  $n_k$  denotes the dead time. Note that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  differ wrt.  $\mathcal{G}$  in  $n_a$  and  $n_a$ , respectively. The process in (5.1) was excited with the input signal in (5.6), the model (5.4) was used for identification and the noise variance  $\sigma_\epsilon^2$  was chosen such that SNR = 10 dB. The models and corresponding information matrices were computed using identification data sets as described in (5.9). A total of 100 experiments were performed to analyze the behavior of the information matrix for  $\mathcal{G} \notin \mathcal{M}$ . The results of the conditioning of the information matrix are shown in in Figure 5.2. The solid black line represents a wrong choice in the model order for the  $A(q^{-1})$  polynomial, *i.e.*  $n_a = 3$ , whereas the gray solid line shows the behavior for  $n_k = 2$  *i.e.* the time delay is wrongly chosen. For both proposed

cases, the information matrix is worse conditioned than when the model order is correctly chosen. The conditioning of the information matrix for a wrong choice of  $n_a$  shows a similar progression as when the model and the process are in the same class. Low values of the reciprocal of the condition number as depicted in Figure 5.2 are explained by the computation of an additional parameter with a data set generated by an input signal of low order of excitation (cf. (5.6)). A criterion for selection of informative data as reported in [12, 73] might not be robust for a situation as the one previously described. Consider that the threshold to define the bound for discarding uninformative data for estimation is set to  $\tilde{\kappa}_{\min} = 0.8 \cdot 10^{-4}$  since considering more elements will yield computational problems. The interval bound is defined at  $k \approx 600$  for  $n_a = 3$  and  $k \approx 100$  for  $n_k = 2$ . The choice of the model order affects the retrieval of informative intervals. Since in real cases the true process is not known, the condition  $\mathcal{G} \in \mathcal{M}$  cannot be accessed. Thus, a model that represents the process “well” should be chosen and used for a data selection method.

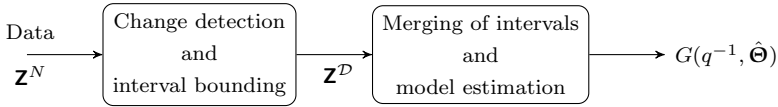
Selection of informative data can support system identification when dealing with data sets predominantly at steady-state. The IV method shows the negative effect on the corresponding information matrix as depicted in the bottom right of Figure 5.1. In the next sections, a novel method, DS4SID, is introduced as a suitable solution for the problem of selecting informative sequences for identification.

## 5.2 General Overview of DS4SID

The main goal of DS4SID is to retrieve informative data intervals from logged process records to support system identification. The input transfer function,  $G(q^{-1}, \hat{\Theta})$ , is also estimated using retrieved intervals. A simplified block diagram of DS4SID is shown in Figure 5.3. The input of DS4SID is logged data represented by  $\mathbf{Z}^N$ . The output of DS4SID are informative data intervals and a model estimated with the retrieved data.

Two main tasks are performed in the block “change detection and interval bounding”. Changes caused by external excitations are located by a detection test and lower bounds of potentially informative intervals are set to these change times. The notation  $\mathbf{Z}^i = \{\mathbf{Z}\}_{k=k_{i1}}^{k_{i2}}$ ,  $i = 1, 2, \dots, q$  will be used to represent informative intervals. The lower and upper bounds of the  $i$ -th interval are represented by  $k_{i1}$  and  $k_{i2}$ , respectively. The subset containing informative intervals is expressed as the following union (cf. Figure 5.3):

$$\mathbf{Z}^D = \bigcup_{i=1}^q \mathbf{Z}^i \quad (5.12)$$



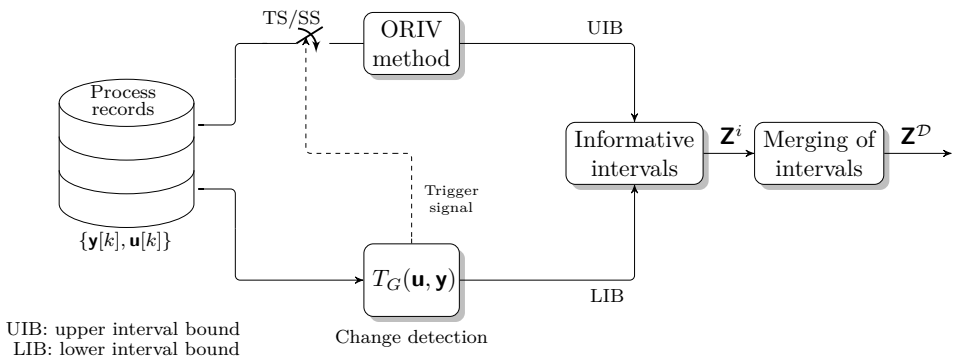
**Figure 5.3:** Simplified block diagram of the DS4SID

where  $q$  is the number of potentially informative intervals. A model  $G(q^{-1}, \hat{\theta})$  is estimated with  $\mathbf{Z}^D$  by merging data of informative intervals.

A more detailed block diagram of DS4SID is shown in Figure 5.4. The detection test  $T_G(\mathbf{u}, \mathbf{y})$  is computed in a first step represented by the lower branch of Figure 5.4. In the null-hypothesis  $\mathcal{H}_0$  of the detection test, the parameter vector is defined by  $\hat{\theta}_0$  whereas the parameter for the alternative hypothesis  $\mathcal{H}_1$  is  $\hat{\theta}_1$ . The detection test chooses  $\mathcal{H}_0$  if the process is not moved from the current operating point. The hypothesis  $\mathcal{H}_1$  is accepted if the process is externally excited which causes a change in the model.

If  $T_G(\mathbf{u}, \mathbf{y})$  chooses  $\mathcal{H}_1$ , the lower bound of a potentially informative interval is set to the instant when a change was located which is denoted by  $k_{i1}$ . Then, the ORIV is used to evaluate the conditioning of the information matrix  $\tilde{\kappa}(\mathbf{I}(\hat{\theta}_1))$  with a data set  $\{\mathbf{Z}\}_{k=k_{i1}}^{k_{i2}}$  where  $k_{i2} = k_{i1} + m \cdot n_p$ . In the second test of DS4SID,  $\tilde{\kappa}(\mathbf{I}(\hat{\theta}_1))$  is updated recursively using the ORIV until the sample information matrix is ill-conditioned due to data that exhibit poor excitation.

The data selection method proposed in this thesis differs from current approaches mainly in the following aspects. Firstly, DS4SID can be applied to multivariate processes in open or closed-loop which is an advantage over current approaches that are limited to SISO systems. Secondly, misdetection of informative intervals is reduced because of utilizing robust detection methods



**Figure 5.4:** Data selection using DS4SID

and interval bounding. Thirdly, DS4SID is robust to colored noise which is a situation likely to be found in real applications. As an added feature, a final model is computed with the retrieved informative intervals. For first analysis of DS4SID, consider the process (5.1) operating in closed loop with the following PI controller:

$$C(q^{-1}) = 0.112 \frac{1 + q^{-1}}{1 - q^{-1}} \quad (5.13)$$

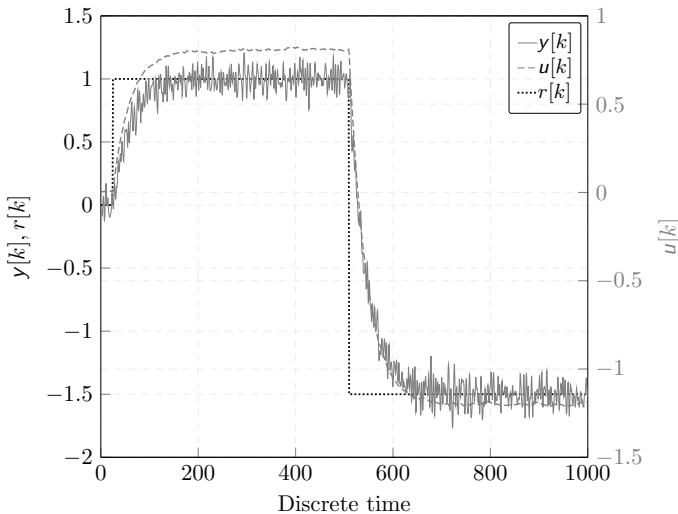
The input signal is defined by

$$u[k] = C(q^{-1})(r[k] - y[k]) \quad (5.14)$$

where  $r[k]$  is the external reference signal described by

$$r[k] = \begin{cases} 0; & 0 \leq k < 25 \\ 1; & 25 \leq k < 485 \\ -1.5; & 485 \leq k \leq 1000 \end{cases} \quad (5.15)$$

Figure 5.5 shows the system response to  $r[k]$  according to (5.15). The noise variance was adjusted so that  $\text{SNR} = 7$  dB. Two steps of different sign and



**Figure 5.5:** Closed-loop response of the system (5.1)

amplitude were entered at  $k = 25$  and  $k = 485$  to evaluate DS4SID under different process conditions.

The input-output data set used for searching of informative intervals is denoted by  $\mathbf{Z}^N = \{u[0], y[0], \dots, u[N-1], y[N-1]\}$ . The first task is to locate change times in  $\mathbf{Z}^N$  using a detection test. These instants are related to lower bounds of informative intervals. Next, these data sequences are evaluated in a further stage and bounds of informative intervals are set by evaluating the conditioning of the information matrix resulting from ORIV.

### 5.3 Determining Lower Interval Bounds

Lower bounds of potentially informative intervals are change times located by a detection test. Dynamic models are used to describe the process under each hypothesis of the detection test. Consider that the process can be modeled around an operating point by:

$$y[k] = y_{00} + s[k] + v[k], \quad k_0 \leq k \leq k_2 \quad (5.16)$$

where the data set for the operating point is defined between  $k_0$  and  $k_2$  and given by  $\mathbf{Z} = \{y[k_0], u[k_0], y[k_0+1], u[k_0+1], \dots, y[k_2], u[k_2]\}$ . The term  $y_{00}$  represents a possible non-zero mean and  $v[k] \sim \mathcal{N}(0, \mathbf{C}(\theta_v))$  is colored Gaussian noise. The sequence  $\{s[k]\}$  is described by a linear model so that  $\mathbf{s} = \Phi\Theta$  where  $\Phi$  and  $\Theta$  are defined as in (3.25b) and (3.25c), respectively.

In vector notation, (5.16) can be described by:

$$\mathbf{y} = \mathbf{y}_{00} + \Phi\Theta + \mathbf{e}, \quad k_0 \leq k \leq k_2 \quad (5.17)$$

A data selection method should choose between the following two hypotheses in order to locate changes due to external excitations:

$$\begin{aligned} \mathcal{H}_0 : \mathbf{y} &= \mathbf{y}_{00} + \Phi\Theta_0 + \mathbf{e}, \quad k_0 \leq k \leq k_1 - 1 \\ \mathcal{H}_1 : \mathbf{y} &= \mathbf{y}_{01} + \Phi\Theta_1 + \mathbf{e}, \quad k_1 \leq k \leq k_2 \end{aligned} \quad (5.18)$$

where  $k_2 - k_0 \gg n_p$ .

Assume that the parameter vector is  $\Theta = \Theta_0$  for  $k < k_0$ . Then, in  $\mathcal{H}_0$ , the process is assumed to be operated at the same operating point *i.e.* the process is not excited by changes in the input. In contrast,  $\mathcal{H}_1$  indicates that the process was moved from that operating point at  $k = k_1$  which is denoted by the change in the parameter vector  $\Theta = \Theta_1$ .

A detection test chooses between the latter hypotheses based on the data set  $\{\mathbf{Z}\}_{k=k_0}^{k_2}$ . The PDFs under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  differ in the unknown parameter

vector. The CUSUM algorithm as presented in [section 3.5](#) can be used for change detection in dynamic systems. This test decides  $\mathcal{H}_1$  if

$$g[k | \mathbf{y}, \mathbf{u}] > \gamma_c \quad (5.19)$$

where  $g[k | \mathbf{y}, \mathbf{u}]$  denotes that the test statistic is computed based on input and output sequences. As presented in (3.81), the log-likelihood ratio increment  $s[k]$  is computed based on the PDF under each hypothesis.

Since the change time  $k_1$  in (5.18) is unknown, (5.19) must be computed for different possible change times. The detector in (5.19) can be sequentially evaluated using two windows for computations (see [10, 87]). The PDF under  $\mathcal{H}_0$  is estimated using a growing time window with size  $M_0 = k_2 - k_0$ . Whereas, PDF under  $\mathcal{H}_1$  is computed using a sliding fixed-size time window of size  $M_1$ .

Consider part of the identification data set described by:

$$\mathbf{Z}^M = \{y[0], u[0], y[1], u[1], \dots, y[M-1], u[M-1]\} \quad (5.20)$$

where  $M < N$  and a change occurs at  $k = k_1$ . The first window has size  $M_0$  and corresponds to the data set  $\{\mathbf{Z}\}_{k=k_0}^{k_2}$ . At each iteration,  $M_0$  is increased by one so that a new data set is generated and the PDF under  $\mathcal{H}_0$  is updated. The resulting data sets for computation of the PDFs under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  after  $n$  iterations, where  $l$  is the iteration counter, are

$$\begin{aligned} l = 0 \quad \mathcal{H}_0 : \{\mathbf{Z}\}_{k=k_0}^{k_2} \quad \mathcal{H}_1 : \{\mathbf{Z}\}_{k=k_2-M_1}^{k_2} \\ l = 1 \quad \mathcal{H}_0 : \{\mathbf{Z}\}_{k=k_0}^{k_2+1} \quad \mathcal{H}_1 : \{\mathbf{Z}\}_{k=k_2-M_1+1}^{k_2+1} \\ \vdots \\ l = n \quad \mathcal{H}_0 : \{\mathbf{Z}\}_{k=k_0}^{k_2+n} \quad \mathcal{H}_1 : \{\mathbf{Z}\}_{k=k_2-M_1+n}^{k_2+n} \end{aligned} \quad (5.21)$$

As expressed by (5.21), the data sets used to compute the PDF under  $\mathcal{H}_0$  correspond to growing time windows that are increased at each iteration. Whereas, the PDF under  $\mathcal{H}_1$  is computed with a sliding fixed-size time window of length  $M_1$ . The test statistic  $g[k]$  is then evaluated and compared at each iteration with the threshold  $\gamma_c$ . The procedure to locate lower interval bounds is shown in Algorithm 1.

Lower bounds of informative intervals are denoted by:

$$k_{i1}, i = 1, 2, \dots, q \quad (5.22)$$

---

**Algorithm 1:** Detection of lower interval bounds using the CUSUM test

---

**Input:** Threshold  $\gamma_c$ , data window length  $M_1$  and model order  $n = [n_a \ n_b \ n_k]$ . The data set for computation of  $\hat{\Theta}_0$  is defined by setting  $k = m \cdot n_p$ . Whereas,  $M_1 = m/2 \cdot n_p$  where  $n_p$  is the number of parameters and  $4 \leq m \leq 10$  as discussed in [section 5.7](#). The value of  $M_1$  determines the size of the data set for computation of  $\hat{\Theta}_1$ .

**Output:** Lower bounds of informative intervals  $k_{i1}$ ,  $i = 1, 2, \dots, q$

**Data:**  $\mathbf{Z}^N = \{y[0], \mathbf{u}[0], y[1], \mathbf{u}[1], \dots, y[N-1], \mathbf{u}[N-1]\}$

**Initialization:**  $g[i] = s[i] = 0$ ,  $i = 0, 1, \dots, k-1$

**while**  $k \leq N-1$  **do**

Compute the current estimates of  $\hat{\Theta}_i[k]$ ,  $i = 0, 1$  for each hypotheses with the following data sets:

$$\mathcal{H}_0 : \mathbf{Z}_0 = \{y[0], \mathbf{u}[0], y[1], \mathbf{u}[1], \dots, y[k-1], \mathbf{u}[k-1]\} \quad (5.23a)$$

$$\mathcal{H}_1 : \mathbf{Z}_1 = \{y[k-M_1-1], \mathbf{u}[k-M_1-1], y[k-M_1], \mathbf{u}[k-M_1], \dots, y[k-1], \mathbf{u}[k-1]\} \quad (5.23b)$$

Compute  $\boldsymbol{\varepsilon}_i[k]$  and  $\hat{\boldsymbol{\sigma}}_i^2[k]$  for  $i = 0, 1$  as follows

$$\boldsymbol{\varepsilon}_0[k] = y[k] - \boldsymbol{\varphi}^T[k] \hat{\Theta}_0[k] \quad (5.24a)$$

$$\boldsymbol{\varepsilon}_1[k] = y[k] - \boldsymbol{\varphi}^T[k] \hat{\Theta}_1[k] \quad (5.24b)$$

$$\hat{\boldsymbol{\sigma}}_0^2[k] = \frac{1}{k - n_p} \sum_{i=0}^{k-1} \boldsymbol{\varepsilon}_0^2[i] \quad (5.24c)$$

$$\hat{\boldsymbol{\sigma}}_1^2[k] = \frac{1}{M_1 - n_p} \sum_{i=k-M_1-1}^{k-1} \boldsymbol{\varepsilon}_1^2[i] \quad (5.24d)$$

Compute the log-likelihood ratio as follows

$$s[k] = \frac{1}{2} \ln \frac{\hat{\boldsymbol{\sigma}}_0^2}{\hat{\boldsymbol{\sigma}}_1^2} + \frac{\boldsymbol{\varepsilon}_0^2[k]}{2\hat{\boldsymbol{\sigma}}_0^2[k]} - \frac{\boldsymbol{\varepsilon}_1^2[k]}{2\hat{\boldsymbol{\sigma}}_1^2[k]} \quad (5.25)$$

Compute  $g[k]$  as follows

$$g[k] = (g[k-1] + s[k])^+ = \max(0, g[k-1] + s[k]) \quad (5.26)$$

**if**  $g[k] > \gamma_c$  **then**

$k_{i1} \leftarrow k$

Evaluate conditioning of the information matrix to define upper interval bounds (refer to [section 5.4](#))

**else**

$k = k + 1$

---



Initial upper bounds of informative intervals are denoted by  $k_{i1} + n_p$ . Thus, an informative interval is initially denoted by  $\{\mathbf{Z}\}_{k=k_{i1}}^{k_{i1}+n_p}$  but the upper bound should be defined in a further step. Once (5.19) holds, the lower bound of a potentially informative interval is set and a further procedure is started to evaluate the conditioning of the information matrix. Informative intervals with defined upper bounds are denoted by:

$$\{\mathbf{Z}\}_{k=k_{i1}}^{k_{i2}}, \quad i = 1, 2, \dots, q \quad (5.27)$$

## 5.4 Determining Upper Interval Bounds

Parameter estimation can be negatively affected when data sets that exhibit poor excitation are used for computations. Data-driven modeling usually involves matrix inversion which can yield numerical problems since matrix columns may become nearly linearly dependent. In the case of the LSM, the information matrix is:

$$\mathbf{I}(\hat{\Theta}) = \sigma_e^{-2}(\Phi^T \Phi) \quad (5.28)$$

where  $\sigma_e^2$  is the noise variance and  $\Phi$  is the regression matrix as defined in (3.25b). The information matrix in (5.28) can also be expressed as

$$\mathbf{I}(\hat{\Theta}) = \sigma_e^{-2}\tilde{\mathbf{I}}(\hat{\Theta}) \quad (5.29)$$

where  $\tilde{\mathbf{I}}(\hat{\Theta}) = \Phi^T \Phi$ ,  $\tilde{\mathbf{I}}(\hat{\Theta}) \in \mathbb{R}^{n_p \times n_p}$  will be called auxiliary information matrix. In the LSM,  $\tilde{\mathbf{I}}(\hat{\Theta})$  should be full-rank and well-conditioned since it is inverted for parameter estimation as expressed in (3.42). To evaluate the conditioning of information matrices, the use of  $\tilde{\mathbf{I}}(\hat{\Theta})$  is preferred over  $\mathbf{I}(\hat{\Theta})$  because computation of parameters involves matrix operations on  $\tilde{\mathbf{I}}(\hat{\Theta})$ . The conditioning of  $\tilde{\mathbf{I}}(\hat{\Theta})$  can degrade with data sets that exhibit poor excitation.

Upper bounds of informative intervals can be defined based on a test that evaluates the conditioning of the auxiliary information matrix. Lower bounds of useful intervals,  $k_{i1}$ ,  $i = 1, 2, \dots, q$ , are located by detectors as explained in section 5.3.

An initial data interval used to compute  $\tilde{\mathbf{I}}(\hat{\Theta})$  is:

$$\begin{aligned} \mathbf{Z}^{(n_{i0})} &= \{\mathbf{Z}\}_{k=k_{i1}-n_l-1}^{k_{i1}} \\ &= \{y[k_{i1}-n_l-1], u[k_{i1}-n_l-1], y[k_{i1}-n_l], u[k_{i1}-n_l], y[k_{i1}], u[k_{i1}]\} \end{aligned} \quad (5.30)$$

where  $n_l \gg n_p$  should be chosen so that the size of the resulting data set guarantee a reliable computation of  $\tilde{\mathbf{I}}(\hat{\Theta})$  since identification methods can be affected when using small data sets. The data set in (5.30) is used only for initial computations of the condition number of  $\tilde{\mathbf{I}}(\hat{\Theta})$ . This is a reference value that is required in further steps for determining upper interval bounds. The data set in (5.30) is used at the iteration  $l = 0$ . The upper bound of potentially informative intervals is increased by 1 at each iteration and the conditioning of the auxiliary information matrix is evaluated. The resulting data sets after  $m$  iterations are:

$$\begin{aligned}
 l = 1 : \mathbf{Z}^{(n_{i1})} &= \{\mathbf{Z}\}_{k=k_{i1}-n_l-1}^{k_{i1}+1} \\
 l = 2 : \mathbf{Z}^{(n_{i2})} &= \{\mathbf{Z}\}_{k=k_{i1}-n_l-1}^{k_{i1}+2} \\
 &\vdots \\
 l = m : \mathbf{Z}^{(n_{im})} &= \{\mathbf{Z}\}_{k=k_{i1}-n_l-1}^{k_{i1}+m}
 \end{aligned} \tag{5.31}$$

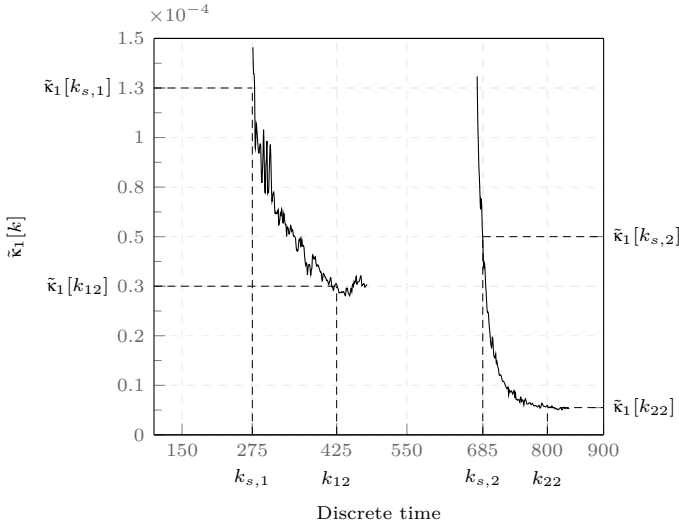
The matrix  $\tilde{\mathbf{I}}(\hat{\Theta})$  is computed until the  $m$ -th iteration where the conditioning starts to deteriorate and that computation of  $\hat{\Theta}$  can be affected by numerical problems. This procedure is repeated for each change time located by the detection test. Consider the following short notation for  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  computed using data sets as defined by (5.30) and (5.31):

$$\begin{aligned}
 \tilde{\kappa}^{(n_{i0})} &: \text{computed with } \mathbf{Z}^{(n_{i0})} \\
 \tilde{\kappa}^{(n_{i1})} &: \text{computed with } \mathbf{Z}^{(n_{i1})} \\
 &\vdots \\
 \tilde{\kappa}^{(n_{im})} &: \text{computed with } \mathbf{Z}^{(n_{im})}
 \end{aligned} \tag{5.32}$$

The conditioning of  $\tilde{\mathbf{I}}(\hat{\Theta})$  with  $\mathbf{Z}^{(n_{i0})}$  yields low values since the data exhibit seldom changes. The value of  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  becomes larger until some instant  $k = m$ . Thus, the following can be stated:

$$\tilde{\kappa}^{(n_{i0})} < \tilde{\kappa}^{(n_{i1})} < \dots < \tilde{\kappa}^{(n_{im})} < \tilde{\kappa}^{(n_{im}+1)} < \tilde{\kappa}^{(n_{im}+2)} < \dots < \tilde{\kappa}^{(n_{im}+n)} < \tilde{\kappa}^{(m)} \tag{5.33}$$

Locating upper bounds for informative intervals consists of determining the instant  $k$  where  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  changes from high to low values which means a worsening of the matrix conditioning. Figure 5.6 shows  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  for the two useful sequences generated by the excitation signal (5.15). The value of  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  exhibit an abrupt change at  $k = 100$  which represents that the conditioning



**Figure 5.6:** Upper interval bounding using the ORIV method

of  $\tilde{\mathbf{I}}(\hat{\Theta})$  improved significantly. Low values of  $\tilde{\mathbf{I}}(\hat{\Theta})$ , as shown before the abrupt change, describe poor conditioning of the auxiliary information matrix since data used for estimation exhibit little excitation.

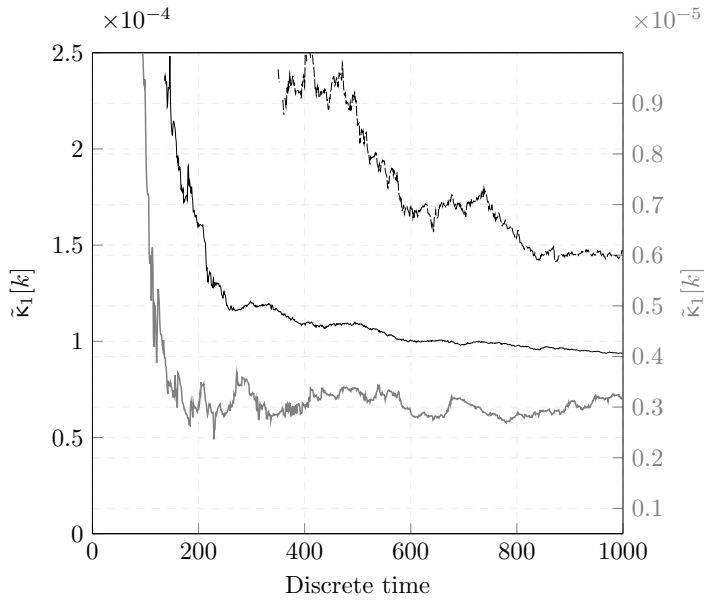
As shown in Figure 5.6, conditioning of  $\tilde{\mathbf{I}}(\hat{\Theta})$  starts to degrade after the change time as more data that exhibit poor excitation are used for estimation. This indicates that system identification should be performed with a selected data set to avoid numerical problems.

Upper interval bounds should be defined based on a trade-off between the quality of the resulting model and the conditioning of the auxiliary information matrix. Thus, upper interval bounding can be proposed as follows:

$$\boxed{\text{Evaluate recursively } \tilde{\kappa}_1 \text{ and extend the informative interval until } \tilde{\kappa}_1[k] \leq \eta \cdot \tilde{\kappa}_1[k_{s,1}]} \quad (5.34)$$

The user-defined parameter  $\eta$  can be set, for instance to  $0.25 < \eta \leq 0.50$ . This means that an interval is extended until the condition number has decayed below 25% of  $\tilde{\kappa}_1[k_{s,1}]$ . The two informative intervals  $\{\mathbf{Z}\}_{k=k_{11}}^{k_{12}}$  and  $\{\mathbf{Z}\}_{k=k_{21}}^{k_{22}}$  in Figure 5.6 were obtained using this criterion.

In [12], a discarding criterion based on absolute bounds for the reciprocal of the condition number is proposed. The choice of thresholds for that criterion is defined based on computationally numerical precision. The value of  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$



**Figure 5.7:** Reciprocal of the condition number of the information matrix for the model in (5.4) using a ramp signal for excitation. Solid black line: correct model order. Solid gray line:  $n_a = 3$ . Dashed black line:  $n_k = 2$

depends on the data used for computation. Thus, informative sequences can be wrongly discarded if the value of  $\tilde{\kappa}$  is compared with thresholds that are valid for the entire data set as proposed in [12]. In contrast, (5.34) proposes a thresholding based on the value of  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  for each interval. This can decrease misdetection of informative intervals when compared to [12]. Informative intervals can, however, be wrongly discarded using such approach specially in the case of small changes in the amplitude of the input signal. In contrast, interval bounding as in (5.34) can guarantee that upper bounds of informative intervals are located regardless of the magnitude of the change between operating points.

Changes in the set points can be represented by ramp-like signals different to the step representation in (5.15). However, the values follow a similar progression as observed in the case of excitation with a step as shown in Figure 5.7. Input signals with greater order of excitation than a step yield larger interval bounds since the information matrix will be better conditioned than in the first case. However, a criterion as presented in this section can also be proposed for interval bounding in case of excitation with ramp-like signals.

## 5.5 Combination of Informative Intervals

The resulting data set that includes informative intervals is denoted by:

$$\begin{aligned}
 \tilde{\mathbf{Z}} &= \left\{ \{\mathbf{Z}\}_{k=k_{11}}^{k_{12}}, \{\mathbf{Z}\}_{k=k_{21}}^{k_{22}}, \dots, \{\mathbf{Z}\}_{k=k_{q1}}^{k_{q2}} \right\} \\
 \tilde{\mathbf{Z}} &= \{y[k_{11}], u[k_{11}], y[k_{11} + 1], u[k_{11} + 1], \dots, y[k_{12}], u[k_{12}], \\
 &\quad y[k_{21}], u[k_{21}], y[k_{21} + 1], u[k_{21} + 1], \dots, y[k_{22}], u[k_{22}], \\
 &\quad \vdots \\
 &\quad y[k_{q1}], u[k_{q1}], y[k_{q1} + 1], u[k_{q1} + 1], \dots, y[k_{q2}], u[k_{q2}]\}
 \end{aligned} \tag{5.35}$$

where  $q$  is the number of retrieved informative intervals.

Figure 5.8 shows the two informative intervals retrieved by DS4SID for the system (5.1) operating in closed-loop with the controller (5.13) and with reference signal (5.15). Data sets logged in industrial processes of continuously operated plants are normally very large and exhibit seldom changes between operating points. Tests with logged data from real processes show that informative sequences can represent between 5% and 10% of the entire data set (see [5, 12]).

The intervals  $\tilde{\mathbf{Z}}_1$  and  $\tilde{\mathbf{Z}}_2$  are merged using approaches introduced in Chapter 4. Since satisfactory estimates can be computed from the retrieved informative intervals, models estimated with  $\tilde{\mathbf{Z}}$  show similar performance than when using

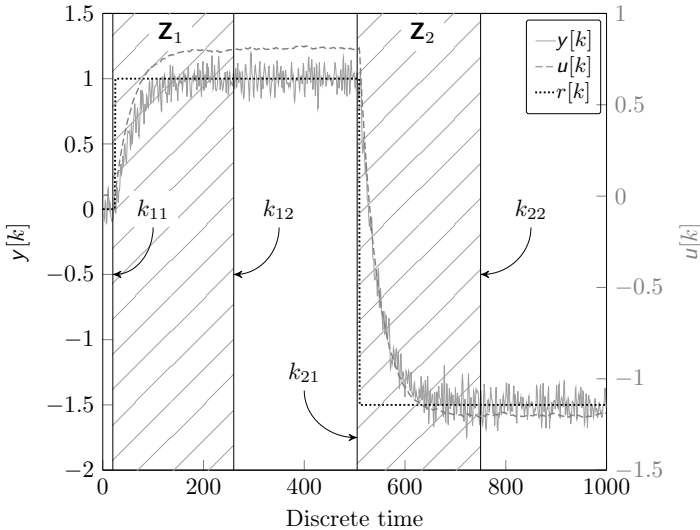


Figure 5.8: Informative intervals selected by DS4SID

the entire data set as will be discussed in [Chapter 6](#). Model estimation with large data sets is, however, computationally demanding. Therefore, selection of informative intervals should be considered as an attractive trade-off between quality of the estimated models and computational cost for estimation.

## 5.6 DS4SID Algorithm

Lower and upper interval bounds are defined by two tests in DS4SID. The number of tests is thus reduced when compared to other approaches that propose at least four sequential evaluation stages [12, 74]. Lower interval bounds are defined using a detection test. Upper interval bounds are determined by evaluating the conditioning of the sample information matrix retrieved from a parameter estimation method. A model is estimated in a final stage by merging the data of the informative intervals.

In the null-hypothesis of the detection test,  $\mathcal{H}_0$ , the model is described using a model parametrized by  $\Theta_0$ . In this application,  $\Theta_0$  denotes the parameter vector of the model for the null-hypothesis. This use should be distinguished from modeling applications where  $\Theta_0$  usually represents the “true” parameter vector which, unless otherwise indicated, is assumed to be unknown in this work. In the alternative hypothesis,  $\mathcal{H}_1$ , the process is described using a model parametrized by  $\Theta_1$ .

Firstly, the detection test is computed with the data set  $\mathbf{Z}^{(1)}$  as in (5.19). The size of this data set is defined by  $M_0 = k_2 - k_0$  with  $M_0 \gg n_p$  where  $n_p$  is the number of parameters of the model and  $k_0$  is the starting discrete-time index of the data set. The value of  $k_2$  is, thus, chosen so that the former condition for  $M_0$  holds. For instance,  $k_2$  can be set as  $k_2 = c \cdot n_p + k_0$  with  $4 \leq c \leq 10$  which guarantees the condition for  $M_0$ . This choice avoids numerical problems in the parameter estimation methods involved in the computation of the detection test. The detection test is iteratively evaluated until a change is detected. Then, the detector is reset and searches for further changes. The detection test is computed with increasing number of data sets as follows,

$$\begin{aligned}
 \mathbf{Z}^{(1)} &= \{y[k_0], \mathbf{u}[k_0], y[k_0 + 1], \mathbf{u}[k_0 + 1], \dots, y[k_2], \mathbf{u}[k_2]\} \\
 \mathbf{Z}^{(2)} &= \{y[k_0], \mathbf{u}[k_0], y[k_0 + 1], \mathbf{u}[k_0 + 1], \dots, y[k_2 + 1], \mathbf{u}[k_2 + 1]\} \\
 &\vdots \\
 \mathbf{Z}^{(m)} &= \{y[k_0], \mathbf{u}[k_0], y[k_0 + 1], \mathbf{u}[k_0 + 1], \dots, y[k_2 + m], \mathbf{u}[k_2 + m]\}
 \end{aligned} \tag{5.36}$$

Change detection by evaluating of the null and the alternative hypothesis with the data windows  $M_0$  and  $M_1$  is shown in [Figure 5.9](#). The bigger the data

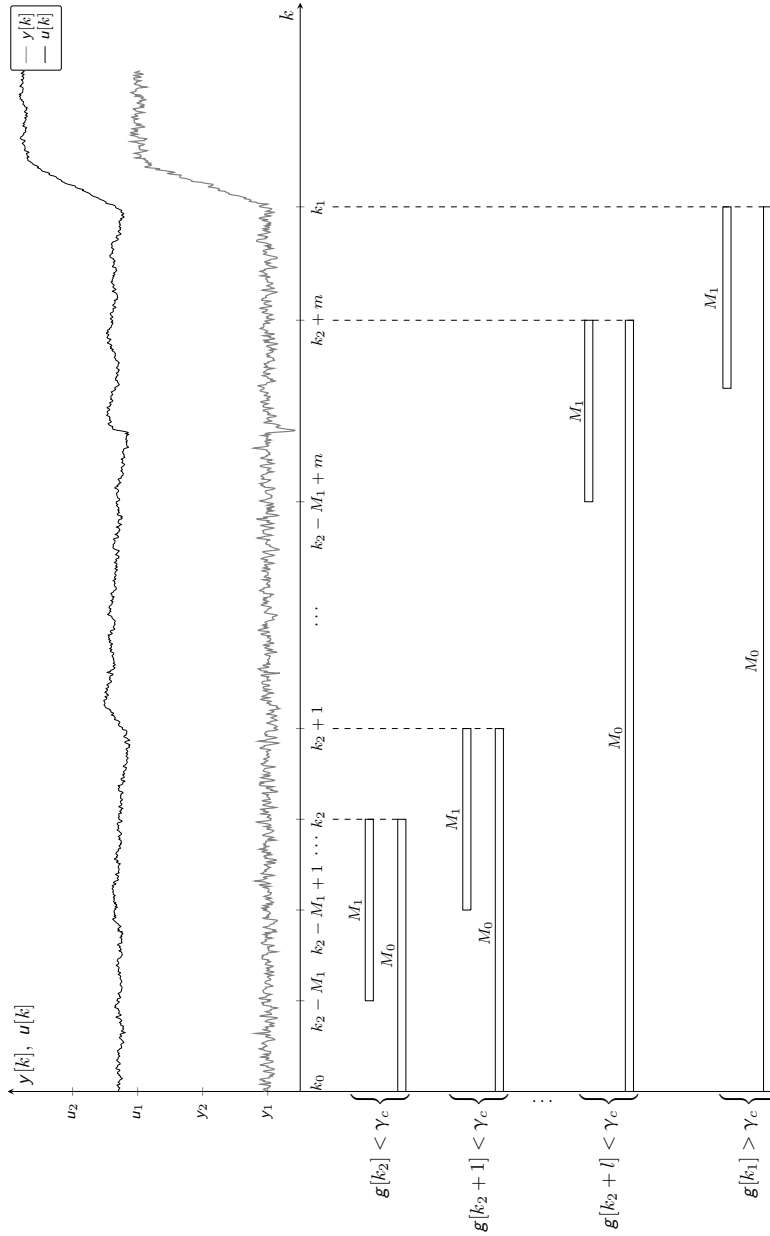
set  $\mathbf{Z}$  becomes the larger the size of the matrices required for computation of the detection test is. If  $\mathcal{H}_0$  is accepted for all iterations (5.36), the detection test can be restarted from  $k_0 = k_2 + m + 1$  as the lower bound of the new data set. Then, the detection test is computed with data sets similarly defined as in (5.36). However, if  $\mathcal{H}_1$  is accepted with the data set  $\mathbf{Z}^{(m)}$ , the lower bound of an informative interval is set and the conditioning of the auxiliary information matrix is evaluated in a second step.

The final informative interval is denoted by  $\{\mathbf{Z}\}_{k=k_{i1}}^{k_{i2}}$ . Then, the detection test is restarted for computations with a data set whose lower bound is  $k_0 = k_{i2} + 1$  and the data sets are similarly defined as in (5.36) but replacing  $k_0$  by  $k_0 = k_{i2} + 1$  and setting  $k_2 = c \cdot n_p + k_{i2} + 1$ . These computations are repeated until the entire data set  $\mathbf{Z}^N$  is processed. Figure 5.9 shows the data windows for evaluating  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . The data window  $M_0$  is increased from  $k_2$  while  $g[k] < \gamma_c$ . The lower bound of an informative interval is set when the detection test chooses  $\mathcal{H}_1$  i.e.  $g[k] > \gamma_c$  which occurs at  $k_1$  as shown in Figure 5.9.

The flow diagram of DS4SID is shown in Figure 5.10. Two user-defined parameters should be chosen for lower interval bounding in the detection test: the model order and the threshold  $\gamma_c$ . Knowledge about the process is required for setting the model order. The threshold  $\eta$  is also required as user-defined parameter for upper interval bounding. Moreover, parameters for the computation of ORIV need to be defined by the user. Less hyper-parameters are required for DS4SID when compared with other approaches which represents a simpler and more practical design. At least six parameters are required, for instance in [12], due to four tests proposed for data selection. Setting of parameters can be more challenging in the previous method. Choice of parameters for DS4SID is discussed in section 5.7.

The initialization block in the flow diagram of Figure 5.10 shows the variables that are set before starting computations. The initial data set  $\{\mathbf{Z}\}_{k=k_0}^{k_2}$  contains observations between  $k_0$  and  $k_2$ . Assuming  $k_0 = 0$ , the value of  $k_2$  is selected so that  $M_0 = k_2 - k_0 \gg n_p$  which avoids numerical problems in the computation of the detection test. At each iteration, the data set used for computation of the detection test is updated by increasing  $k_2$ . Since  $\{\mathbf{Z}\}_{k=k_0}^{k_2}$  can become large after several iterations, the maximum size of this data set is limited to  $n_1$  to avoid large processing times.

The variable  $w$  is used to decide which test is computed: either the detection test or the conditioning of the information matrix. In the case of  $w = 0$ , the detection test is computed with  $\{\mathbf{Z}\}_{k=k_0}^{k_2}$ . If  $g[k] > \gamma_c$  does not hold,  $k_2$  is increased and the detection test is computed with an updated data set. In case a change is detected,  $g[k] > \gamma_c$  holds and the lower interval bound  $k_{i1}$  is



**Figure 5.9:** Data windowing using a sliding fixed-size window  $M_1$  and a growing time window  $M_0$  for change detection in DS4SID. Sample size just for illustrative purposes. (Please refer to Figure 5.10 for definition of data sets used for computation of  $\mathbf{g}[k]$ )



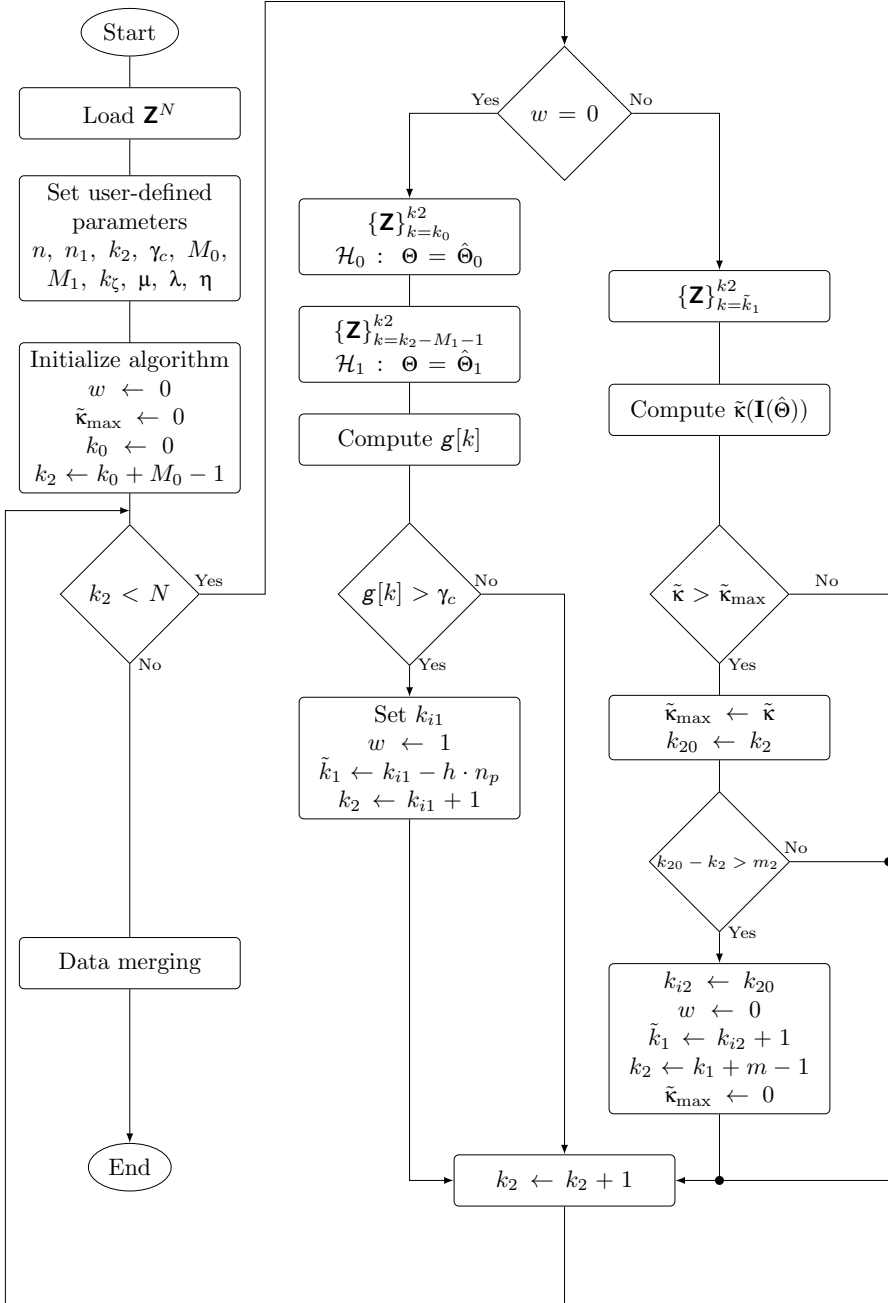


Figure 5.10: Flow diagram of DS4SID

defined. The variable  $w$  is set to 1 which indicates that the lower bound of an informative interval was located and the upper bound should be defined in a second stage. A data set  $\{\mathbf{Z}\}_{k=k_{i1}-n_l-1}^{k_{i1}}$  is defined to start computations using a second test to evaluate the conditioning of the information matrix. The value of  $n_l$  should be chosen so that  $k_{i1} - n_l \gg n_p$ .

The conditioning of the information matrix is evaluated at each iteration by thresholding  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  against the user-defined parameter  $\eta$ . If the information matrix is well-conditioned, the data set used for computations is increased *i. e.*  $\{\mathbf{Z}\}_{k=k_{i1}-n_l-1}^{k_{i1}+1}$ . If the information matrix is ill-conditioned, the upper interval bound  $k_{i2}$  is defined based on the data set  $\mathbf{Z}$  with which this condition is fulfilled. The resulting informative interval is denoted by  $\{\mathbf{Z}\}_{k=k_{i1}}^{k_{i2}}$ . The variable  $w$  is reset so that the detection test is further computed for change detection. Once an informative interval is confirmed, the lower and upper bounds of  $\{\mathbf{Z}\}_{k=k_0}^{k_2}$  are updated, the value of  $k_0$  is set to  $k_{i2} + 1$  and  $k_2 = k_0 + M_0$  which yields a data set of length  $M_0 = k_2 - k_0$ .

The data should be informative for the model chosen in DS4SID since the identification method used as part of the approach considers this assumption. In [24], conditions for informative data are derived for the models introduced in Table 3.1. These requirements are derived for data collected in open and closed-loop. Logged process data can be used for identification when informative conditions are satisfied. These conditions depend, among others, on the model order and the controller when the process is operated in closed-loop. Properties of the estimators obtained by identification methods used in DS4SID are discussed in [57, 80]. The situation when model and system do not belong to the same class is represented by  $\mathcal{S} \notin \mathcal{M}$ . In this situation, computation of the associated information matrix can yield wrong upper interval bounds. To discuss this case, a wrong choice for the process in (5.1) is analyzed.

The detection test in Algorithm 5.10 is based on PDFs resulting from input-output models which agrees with the data collected from a dynamic system. In other approaches such as in [12, 74], detection tests are applied to the time series  $\{\mathbf{u}[k]\}$  and  $\{\mathbf{y}[k]\}$  independently. This may yield errors since changes in the process output cannot be directly linked to an excitation in the input. Thus, additional tests are required to confirm the input-output dependency in the detected change. Only two tests are computed to locate informative intervals instead of four procedures as described in [12]. The CUSUM algorithm and the gradient of the information matrix are used for lower and upper interval bounding, respectively. Use of only two tests can decrease false alarms and misdetection. Other approaches implement detection tests that are applied separately in the input and output signal. Choice of thresholds for these tests

can be challenging and they normally result in large false alarms. Moreover, additional computations such as the Granger causality tests are required to confirm that the dynamic change was caused by an excitation in the input. The detection problem as proposed in DS4SID is stated using two hypothesis that represent a system driven around the same operating point or deviated from it. Input-output models are used in the hypothesis formulation which avoids performing further computations such the Granger causality test.

The value of  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  depends on the auxiliary information matrix which is defined based on observations contained in an informative interval. Thus,  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  can have large values for abrupt changes from an operating point. In contrast, small excitations in the process can yield lower values for the conditioning of the auxiliary information matrix. Next, choice of user-defined parameters is further discussed.

## 5.7 Choice of Design Parameters

General guidelines on the parameters choice for the detection test can be found in [10, 50]. For a more detailed discussion on the parameters for the ORIV, the reader is referred to [23, 80]. The model order and the threshold  $\gamma_c$  are required in the detection test. The choice of an appropriate order of the input-output model requires knowledge about the process. Thus, identification tests with part of the data may be required to determine the model order. For instance, the model order of a MISO ARX model is denoted by  $n = [n_a \ n_b \ n_k]$ . Setting  $n_k$  can be especially challenging for processes with large input-output delay and a wrong choice will affect upper interval bounding. In case an ARX-Laguerre model is used, the pole of the Laguerre filter should also be defined. However, this model can be a suitable choice for processes where setting the input-output delay is a difficult choice due to lack of knowledge about the process. This lack of information can be compensated by a proper selection of the pole of the Laguerre filter. The threshold  $\gamma_c$ , used in the detection test, is set by trial and error as a trade-off between misdetection and false alarms. Depending on the application, the value of  $\gamma_c$  needs to be adjusted as previously described.

For computation of the detection test, data are scanned using two time windows. The PDF for  $\mathcal{H}_0$  is estimated with a growing time window of length  $M_0 = k_2 - k_0$ . The choice of  $M_0$  is set so that  $n_p \ll M_0 \ll N$  where  $n_p$  is the number of parameters and  $N$  the size of the entire data set. Thus,  $M_0$  is selected to be larger than the number of parameters *i.e.*  $M_0 > m \cdot n_p$ ,  $m \in \mathbb{Z}^+ \wedge m > 1$ . Usually, initial choices such as  $4 \leq m \leq 10$  can yield a suitable length for  $M_0$ . However, the value of  $M_0$  requires fine tuning depending on the application. A second data set of length  $M_1$  is called sliding fixed-size time window used for

computation of the PDF under  $\mathcal{H}_1$ . Initial choices for  $M_1$  can be  $M_1 = M_0/2$  but the size of  $M_1$  is also adjusted by trial and error.

Several design parameters are also required in ORIV. The forgetting factor  $\lambda$  used for initialization is normally chosen between 0.95 and 0.985 as found in [57]. The auxiliary information matrix  $\mathbf{S}^T$  is initialized by setting  $0.001 \leq \mu \leq 0.01$ . The parameter  $k_\zeta$ , required for defining the instant from which the instruments are considered, can be set to 1 for processes operating in open loop. In the case of closed loop operation, this value can be chosen in the range  $5 \leq k_\zeta \leq 15$ . The parameters  $\tilde{n}_a$  and  $\tilde{n}_b$  should be larger than the model order. Since the input signals have low order of excitation, a suitable choice is  $1 \leq \tilde{n}_b \leq 2$ . An acceptable value for  $\tilde{n}_a$  may depend on the system dynamics for instance if a process exhibits a well damped or oscillating behavior. As a rule of thumb, this parameter can be adjusted to  $1 \leq \tilde{n}_a \leq 3$ . Finally, the parameter  $\eta$  used for upper interval bounding is chosen between 0.25 and 0.50. According to results obtained in simulation and real case studies treated in [section 6.2](#) and [section 6.3](#) the value of  $\eta$  might be less dependent on the application since a similar degradation of the conditioning of the information matrix was observed for different processes. This behavior was observed in simulation and real case studies that are treated in the further chapter. However, since the size of the information matrix and its conditioning depend on the model order, this parameter should be adjusted according to the application.

## 5.8 Discussion

A new data selection method, DS4SID, was introduced in this chapter. Informative intervals are located only by two tests which is a notable feature when compared to other approaches where at least four tests have been reported. DS4SID differs in the following main aspects from other available methods. Firstly, it can be used for multivariate processes and, therefore, covers a large number of systems that are not addressed by current approaches. Secondly, DS4SID consists of well-defined tests for lower and upper interval bounding that decrease misdetection and false alarms when compared to current methods. Thirdly, the tests used in DS4SID are also robust to colored noise which represents a differencing feature with current methods, particularly, when real applications are addressed. This aspect has not been considered in other approaches which may result in wrong interval bounds.

Some practical problems found in logged process data correspond to compressed data, missing data or outliers. The presence of outliers can trigger the detection test for lower interval bounding. There exist several methods that can be applied in case outliers and missing data occur. Outliers can be either

removed or be included with an associated weight that reduces their influence. Alternatively, they can be replaced by estimated values as described in [69]. Thus, data can be preprocessed before using DS4SID for selection of informative intervals.

Data-driven modeling can be affected by compressed data as the process dynamics cannot be correctly extracted. The previously described situation can yield problems in the upper interval bounding since the model cannot correctly describe the process.

An informative interval is defined by time instants that describe lower and upper bounds of the data sequence. Lower interval bounds are related to change times resulting of moving away a process from a current operating point. Detection tests should be performed with input-output models instead of time series models because dynamic processes are treated in the context of selecting informative intervals. Moreover, we are interested in changes caused by external excitations in the input of a process. Such changes can be observed using input-output models in the detection test. If the change detection is evaluated using only time series models, additional tests such as the Granger causality test [12, 31] are required to confirm that a change observed in the output was generated by an excitation in the input. Performing additional tests is computationally demanding and will increase the false alarms rate. Thus, the approach proposed here can overcome some of the drawbacks of current methods.

Current approaches require a larger number of user-defined parameters than DS4SID since they implement more tests for change detection in time series models and the Granger causality test. This is an advantage of DS4SID since setting suitable parameters is simpler when compared to other methods. The model order is a relevant hyperparameter for current approaches as well as for DS4SID since the results of tests for lower and upper interval bounds depend on this value. Parameters for the detection test are chosen as a trade-off between false alarm and detection rate. The choice of the threshold  $\gamma_c$  can be adjusted with a test data set until satisfactory results are obtained. Choice of parameters for recursive parameter estimation is straightforward and widely discussed in the literature [29, 57, 59, 80].

In the two tests of DS4SID, the parameter estimation method is robust to Gaussian colored noise. In real applications, colored noise is more likely to be found than WGN. Thus, DS4SID has a significant advantage over current approaches whose tests are limited to signals with additive WGN.

The CUSUM algorithm is used for change detection and the PDFs under each hypothesis are computed for dynamic models. These intervals are further

evaluated using an identification method. The ORIV was proposed in [section 5.4](#) as a robust method for parameter estimation where the conditioning of the auxiliary information matrix is evaluated. Upper interval bounds are set to time instants where the conditioning of  $\tilde{\mathbf{I}}(\hat{\Theta})$  decays below values yielding numerical problems for parameter estimation. The ORIV is robust to colored noise and can be applied to systems operating in open- or closed-loop. Instead of using a discarding criterion with absolute values as proposed in available methods, upper interval bounding based on relative values is stated.

The test for lower interval bounding proposed in DS4SID can detect changes from a current operating point regardless of the signal type that caused the excitation. Thus, changes generated by step- or ramp-like signals will be detected with possible adjustment of lower interval bounds.

The size of the data set can be a constraint depending on the processing hardware. This can be overcome by using, for instance, high-performance computing clusters. A second option can be to use a program that reads the logged data and send them sequentially to the selection method. Thus, requirements of the RAM memory are decreased and very large data sets can be processed.

Finally, the proposed data merging procedure in [section 5.5](#) can also yield satisfactory models and can be considered as an alternative to the approaches introduced in [Chapter 4](#). To further develop data merging of informative intervals, a weighted approach can be proposed. Weights that can be defined based on the retrieved sequence  $\{\tilde{\kappa}[k]\}$ . These values are computed in the test to evaluate the conditioning of the auxiliary information matrix. DS4SID is evaluated in simulation and real case studies in the next chapter.

A correct choice of the user defined parameters described in [section 5.7](#) can help to obtain acceptable results using DS4SID. The value of  $\gamma_c$  is set as a trade-off between misdetection and false alarms. Small excitation in the process can result in misdetection of intervals through the computation of  $\mathbf{g}[k]$ . Thus, potentially transient changes in the process might not be located, which results in a small number of intervals and, consequently, degradation of the resulting model. Upper interval bounding can be affected by a wrong choice of the model order as discussed at the end of [section 5.1](#). For practical use, knowledge from an expert can support choosing a suitable model order that yields acceptable data segments.

[Table 5.1](#) compares DS4SID with two related approaches reported in the literature (see [[12](#), [74](#)]). In each of the comparison categories, DS4SID has competitive properties and improved design that results in an improvement of the state of the art of data selection for linear system identification.

Table 5.1: Comparison of DS4SID with available approaches

Method	DS4SID	[12]	[74]
Process addressed	MIMO (Assuming uncorrelated variables. A data pre-selection, as proposed in a practical case study in [75], can support data selection in case of correlation between variables.)	SISO	SISO
Scaling	No scaling required	Scaling of signals between 0 and 1	Scaling as in [10]
Type of detection test	1. Change detection based on input-output models for lower interval bound 2. Evaluation of the conditioning of the information matrix for current operating point to determine upper interval bound	1. Change detection based on time-series models for input signal 2. Change detection based on time-series models for output signal 3. Evaluation of the conditioning of the information matrix based on absolute threshold for upper interval bound 4. Granger causality test to confirm that change in the output was generated by excitation in the input	<ul style="list-style-type: none"> <li>• 1–4 as in [12]</li> <li>• Entropy test to merge neighboring intervals and decrease the number of local models</li> </ul>
User-defined parameters	<ul style="list-style-type: none"> <li>• Detection test: <math>n, n_1, k_2, M_0, M_1, \gamma_c</math></li> <li>• Matrix conditioning (ORIV): <math>k_\zeta, \mu, \lambda, \eta</math> (See section 5.7)</li> </ul>	<ul style="list-style-type: none"> <li>• Detection test: <math>\lambda_w, \lambda_y, \eta_0, \eta_1, \eta_2</math></li> <li>• Matrix conditioning (QR-RLS): <math>n, \alpha, N_g, \mu, \lambda, \eta_3</math></li> <li>• Granger test: <math>\eta_4</math></li> </ul>	<ul style="list-style-type: none"> <li>• Same as in [12]</li> <li>• Entropy test: <math>\eta_5</math></li> </ul>
False alarms	Low, since robust test for change detection	High, due to more false activations of the detection test using time-series models	High, since same approach as in [12]
Detection rate	High	High	High
Robustness to colored noise	Yes due to ORIV for parameter estimation	No due to QR-RLS used for model computation. (WGN assumed)	No, approach similar to [12]

# CHAPTER 6

---

## Case Studies

---

In this chapter, DS4SID is evaluated in simulation and real case studies. Identification data sets are predominantly at steady-state for continuously operated plants since processes are seldom driven away from a given operating point. DS4SID is firstly tested in a simulated process which describes a multivariate binary distillation column operating in closed-loop. A process unit from the lab-scale factory “ $\mu$ Plant” is used as industry-oriented case study. This is a multivariate process that consists of several sub-systems each with different dynamics.

### 6.1 Performance Assessment

The performance of DS4SID is evaluated using two criteria. The UINR describes the ratio between the size of the resulting data set that contains informative intervals and the entire data set. It will be shown that a small part of the data set is sufficient to estimate models that can describe a process well. The model performance is evaluated using the goodness of fit whose computation is based on the NRMSE between the real observations and the model output..

#### 6.1.1 Data Reduction Ratio

The data sets corresponding to the collection of informative intervals and the entire data set itself are represented by  $\mathbf{Z}^n$  and  $\mathbf{Z}^N$ , respectively. The ratio of the size of these data sets is:

$$\text{UINR} = \frac{\text{size}(\mathbf{Z}^n)}{\text{size}(\mathbf{Z}^N)} \cdot 100 \% \quad (6.1)$$

The UINR is a criterion used to assess the down-sizing of the given dataset to what should be used for model estimation.



### 6.1.2 Goodness of Fit

The goodness of fit compares the model output with the real signal using a cost function as follows:

$$\text{fit}(G(q^{-1}, \hat{\Theta})) = (1 - \text{NRMSE}(y(k), \hat{y}(k, \hat{\Theta}))) \cdot 100\% \quad (6.2)$$

where  $\hat{\mathbf{y}}_i$  and  $\mathbf{y}$  are the model output and the output signal, respectively. The NRMSE is used as cost function in (6.2). This scalar measure is described by:

$$\text{NRMSE} = \sqrt{\frac{\sum_{k=0}^{N-1} (y[k] - \hat{y}[k])^2}{\sum_{k=0}^{N-1} (y[k] - \bar{y}[k])^2}} \quad (6.3)$$

The NRMSE ranges between 0 and 1 whereas (6.2) results in 100% for a perfect model and 0% otherwise. Since (6.2) can be better interpreted for model quality, this performance criterion is preferred over the NRMSE. Model quality using (6.2) is evaluated on validation data sets for the simulation case study. For the industry-oriented case study, (6.2) is computed on an identification data set and results are confirmed by cross-validation using additional data sets. In the next section, a simulation case study, with unknown “true” transfer function is presented. Estimated and “real” transfer functions are compared. In the real case study two models are evaluated: one is estimated using the entire data set whereas a second one is computed with informative intervals.

## 6.2 Simulation Case Study: Binary Distillation Column

First results of the DS4SID application were reported in [4]. In this section, a more detailed analysis is presented. The operation of the proposed system yields suitable data sets for the evaluation of DS4SID.

### 6.2.1 Process Description

Consider a multivariate binary distillation column described by the following continuous-time transfer function as described in [93]

$$\begin{pmatrix} Y_1(s) \\ Y_2(s) \end{pmatrix} = \begin{pmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{pmatrix} \begin{pmatrix} U_1(s) \\ U_2(s) \end{pmatrix} \quad (6.4)$$

where  $G_{ij}$  corresponds to the transfer function from the  $j$ -th input to the  $i$ -th output defined by:

$$G_{11}(s) = \frac{12.8e^{-s}}{16.7s + 1} \quad G_{12}(s) = \frac{-18.9e^{-3s}}{21s + 1} \quad (6.5a)$$

$$G_{21}(s) = \frac{6.6e^{-7s}}{10.9s + 1} \quad G_{22}(s) = \frac{-19.4e^{-3s}}{14.4s + 1} \quad (6.5b)$$

Each transfer function has an associated dead-time as expressed in (6.5). The function  $G_{21}$  has the largest dead-time. The input and output signals of the multivariate process represent the following physical variables:

$$\begin{aligned} y_1(t) &: \text{overhead composition in wt. \% methanol} & u_1(t) &: \text{reflux flow in g/s} \\ y_2(t) &: \text{bottom composition in wt. \% methanol} & u_2(t) &: \text{steam flow in g/s} \end{aligned}$$

where wt. refers to weight percent. This process is highly interacting since inputs and outputs are mutually coupled. This interaction is evaluated using the condition number of (6.5). The steady-state relative gain array (RGA) of (6.5) is evaluated at  $\omega = 0$  which yields (dimensionless):

$$\Lambda(0) = \begin{pmatrix} 12.8 & -18.9 \\ 6.6 & -19.4 \end{pmatrix} \quad (6.7)$$

Then, the SVD of (6.7) results in

$$\Sigma = \begin{pmatrix} 30.4048 & 0 \\ 0 & 4.0645 \end{pmatrix} \quad (6.8)$$

The condition number  $\kappa$  is computed from (6.8) as the ratio between eigenvalues

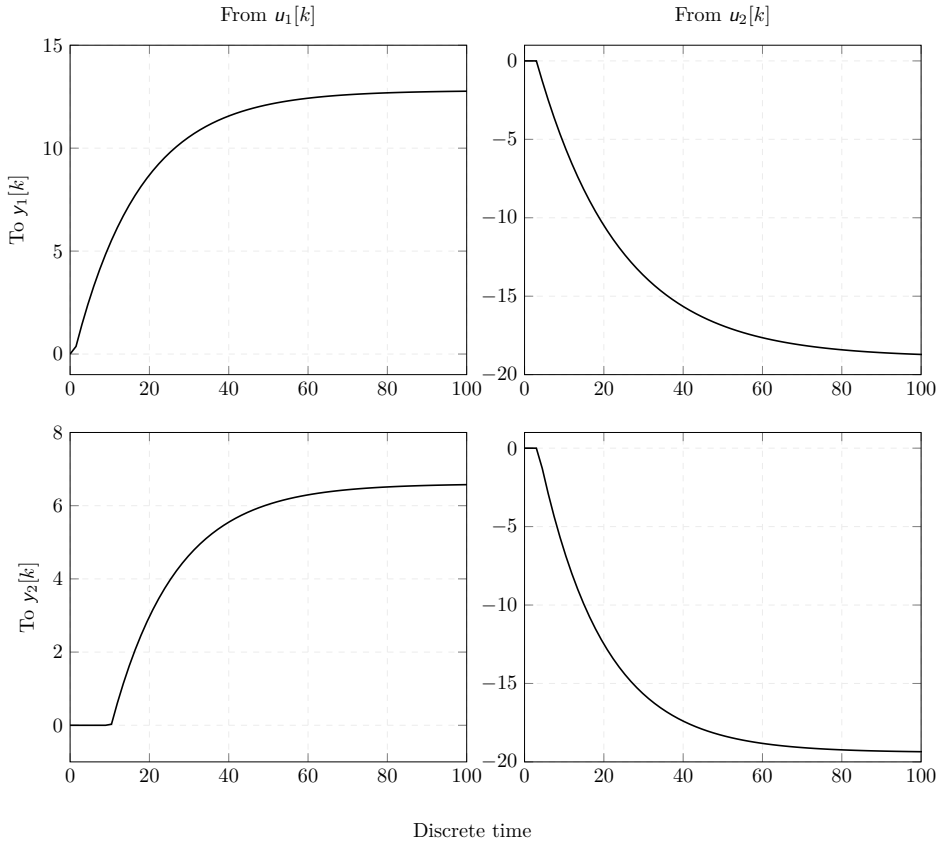
$$\kappa = \frac{30.4048}{4.0645} = 7.48 \quad (6.9)$$

The transfer functions (6.5) were converted to discrete-time using the Zero-Order Hold (ZOH) method with sampling time  $T_s = 1$  s. The resulting discrete-time transfer functions are as follows:

$$G_{11}(q^{-1}) = q^{-1} \frac{0.744q^{-1}}{1 - 0.9419q^{-1}} \quad G_{12}(q^{-1}) = q^{-3} \frac{-0.8789q^{-1}}{1 - 0.9535q^{-1}} \quad (6.10a)$$

$$G_{21}(q^{-1}) = q^{-7} \frac{0.5786q^{-1}}{1 - 0.9123q^{-1}} \quad G_{22}(q^{-1}) = q^{-3} \frac{-1.302q^{-1}}{1 - 0.9329q^{-1}} \quad (6.10b)$$

Figure 6.1 shows the step response of the discrete-time transfer functions.



**Figure 6.1:** Step response of system in (6.4) with transfer functions in (6.5)

Since  $\kappa$  yields a large value, the system is considered to be strongly coupled (see [77]). In the process, the overhead and bottom composition should exhibit large and low values, respectively, which means that most of the raw material is converted to methanol and extracted from the distillation column. This is not the case as shown in Figure 6.1 since a control action in  $u_1(t)$  increases simultaneously  $y_1(t)$  and  $y_2(t)$ .

In this process, the goal is to control simultaneously overhead and bottom composition. This is particularly challenging since the process is highly-interacting. As a result, the two control loops interact which leads to a deterioration in the performance of both composition loops [93]. Thus, a non-interacting controller is required to guarantee a simultaneous control of the overhead and bottom composition. A procedure for controller design of the system (6.4) is described

in [93] and was followed here for implementation in closed-loop. Contrarily to available data selection methods that are limited to SISO processes and cannot be applied to this case study, DS4SID can be used for multivariable processes.

### 6.2.2 Controller Design

Figure 6.2 shows a block diagram of the process (6.4) using the control scheme proposed in [93]. The process outputs are embedded in colored noise obtained by filtering WGN using the transfer functions  $H_1(q^{-1})$  and  $H_2(q^{-1})$ :

$$H_1(q^{-1}) = \frac{1}{1 - 0.3q^{-1} + 0.5q^{-2}} \quad (6.11a)$$

$$H_2(q^{-1}) = \frac{1}{1 - 0.5q^{-1} + 0.8q^{-2}} \quad (6.11b)$$

where  $e_i(t) \sim \mathcal{N}(0, \sigma_{e_i}^2)$ ,  $i = 1, 2$ .

Two noninteracting controllers  $D_{21}(s)$  and  $D_{12}(s)$  are required to eliminate the interacting effect of the control in  $u_1(t)$  and  $u_2(t)$  on  $y_2(t)$  and  $y_1(t)$ , respectively. The desired decoupling is expressed as follows

$$D_{21}(s)P_{22}(s) + P_{21}(s) = 0 \quad (6.12a)$$

$$D_{12}(s)P_{11}(s) + P_{12}(s) = 0 \quad (6.12b)$$

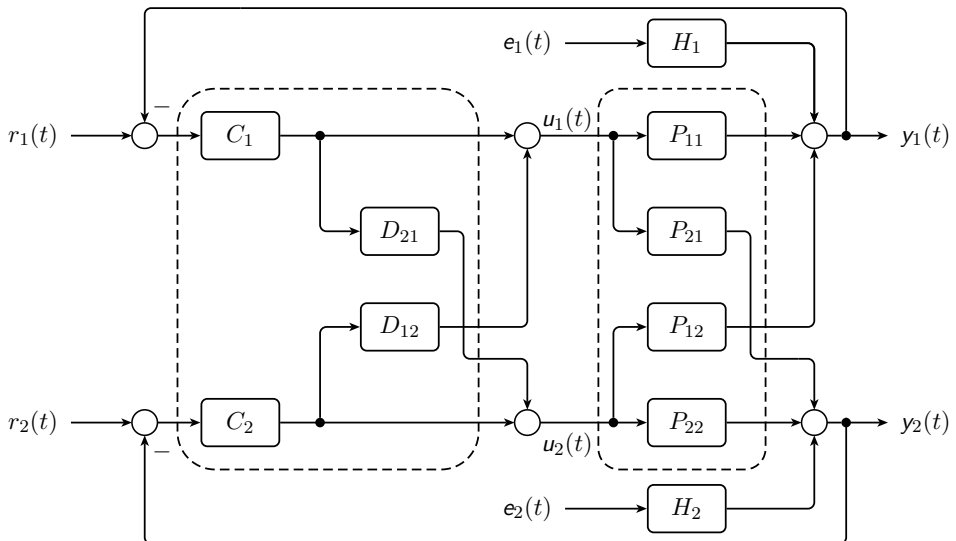


Figure 6.2: Block diagram of the non-interacting control system

where  $D_{ij}(s)$  is the noninteracting compensator to eliminate the effect of the control action associated to the  $i$ -th input on the  $j$ -th output.

In general, the following relation applies to express the desired non-interacting control

$$D_{ij}(s)P_{ii}(s) + P_{ij}(s) = 0 \quad (6.12c)$$

The decoupling controllers are then obtained from (6.12c) by:

$$D_{ij}(s) = -\frac{P_{ij}(s)}{P_{ii}(s)} \quad (6.13)$$

The process transfer functions in (6.5) are of identical order. Then, the decoupling controllers are realizable if the time delay of the transfer function  $P_{ij}(s)$  is larger than the delay of  $P_{ii}(s)$ . This guarantees that the order of the denominator is larger than the numerator and the designed controller is realizable. This condition is satisfied by the transfer functions in (6.5). Thus, (6.12c) is used for designing of the decoupling controllers.

As shown in Figure 6.2, two additional controllers  $C_1$  and  $C_2$  are included in the closed-loop system. Several methods published in the literature such as internal model control (IMC) or symmetric optimum can be used for designing  $C_1$  and  $C_2$ . The IMC showed satisfactory closed-loop system performance and was therefore used in this case study. In order to design the controllers as explained in [93], equations (6.5) were converted to discrete time transfer functions using the Zero-Order Hold (ZOH) method with sampling time  $T_s = 1$  s. The resulting discrete-time controllers are:

$$D_{12}(q^{-1}) = \frac{1.174q^{-3} - 1.106q^{-4}}{1 - 0.9535q^{-1}} \quad (6.14a)$$

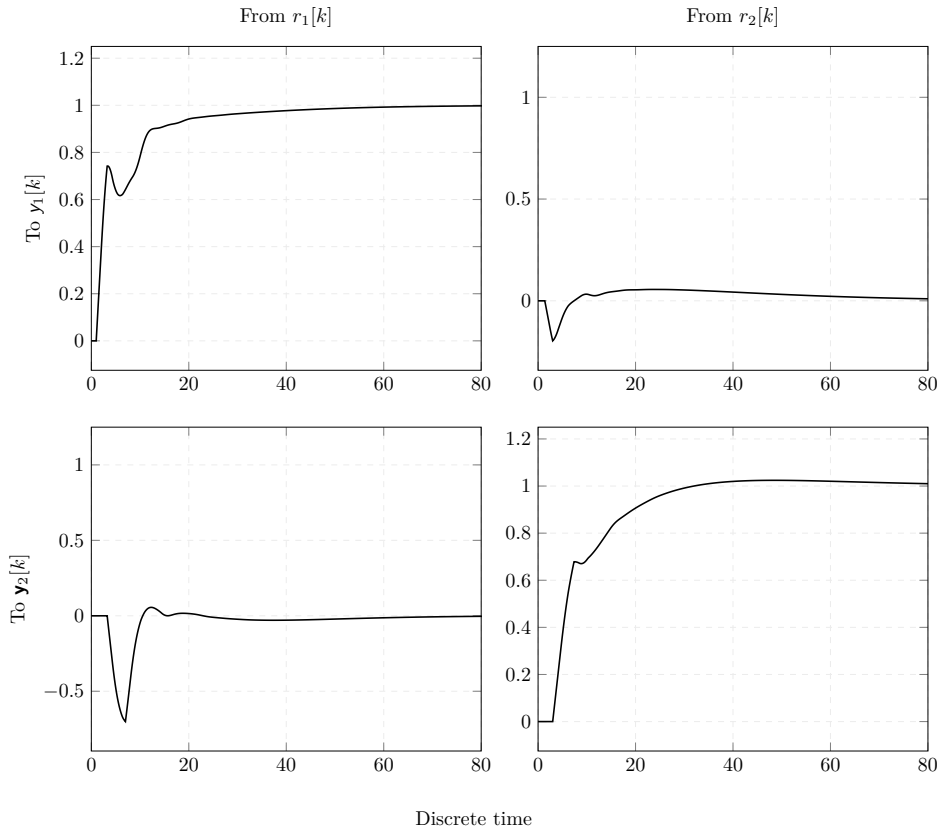
$$D_{21}(q^{-1}) = \frac{0.4494q^{-2} - 0.4196q^{-3}}{1 - 0.9123q^{-1}} \quad (6.14b)$$

$$C_1(q^{-1}) = \frac{0.4886 - 0.4593q^{-1}}{1 - q^{-1}} \quad (6.14c)$$

$$C_2(q^{-1}) = \frac{-0.1375 + 0.1279q^{-1}}{1 - q^{-1}} \quad (6.14d)$$

where  $C_1(q^{-1})$  and  $C_2(q^{-1})$  are PI controllers.

Simulations with discrete time controllers result in a small interaction between the off-diagonal components for  $0 < k < 20$  as shown in Figure 6.3. Note that



**Figure 6.3:** Closed-loop step response of system (6.5) with controllers (6.14)

no effect of the inputs on their non-diagonal output is observed from around  $k = 30$  as shown in the closed-loop response in [Figure 6.3](#).

### 6.2.3 Performed Experiments

Different simulations were performed in an operating range where the model (6.5) applies (see [93]). The continuous-time transfer functions from the process (6.5) were converted to discrete-time and the resulting system is excited using step-like signals that are seldom changed. The external references of the process are described by:

$$r_i[k] = R_{ij}, \quad k_{j1} \leq k \leq k_{j2}, \quad i = 1, 2 \quad j = 1, 2, \dots, n_s \quad (6.15)$$

**Table 6.1:** External reference signals used to generate test data sets

$j$	$r_1[k]$	Disc. time	$r_2[k]$	Disc. time
1	90	$0 \leq k \leq 499$	2	$0 \leq k \leq 499$
2	98	$500 \leq k \leq 1149$	1	$530 \leq k \leq 1189$
3	93	$1150 \leq k \leq 1769$	2	$1190 \leq k \leq 2194$
4	90	$1770 \leq k \leq 2489$	1	$2195 \leq k \leq 2749$
5	95	$2490 \leq k \leq 3500$	3	$2750 \leq k \leq 3500$

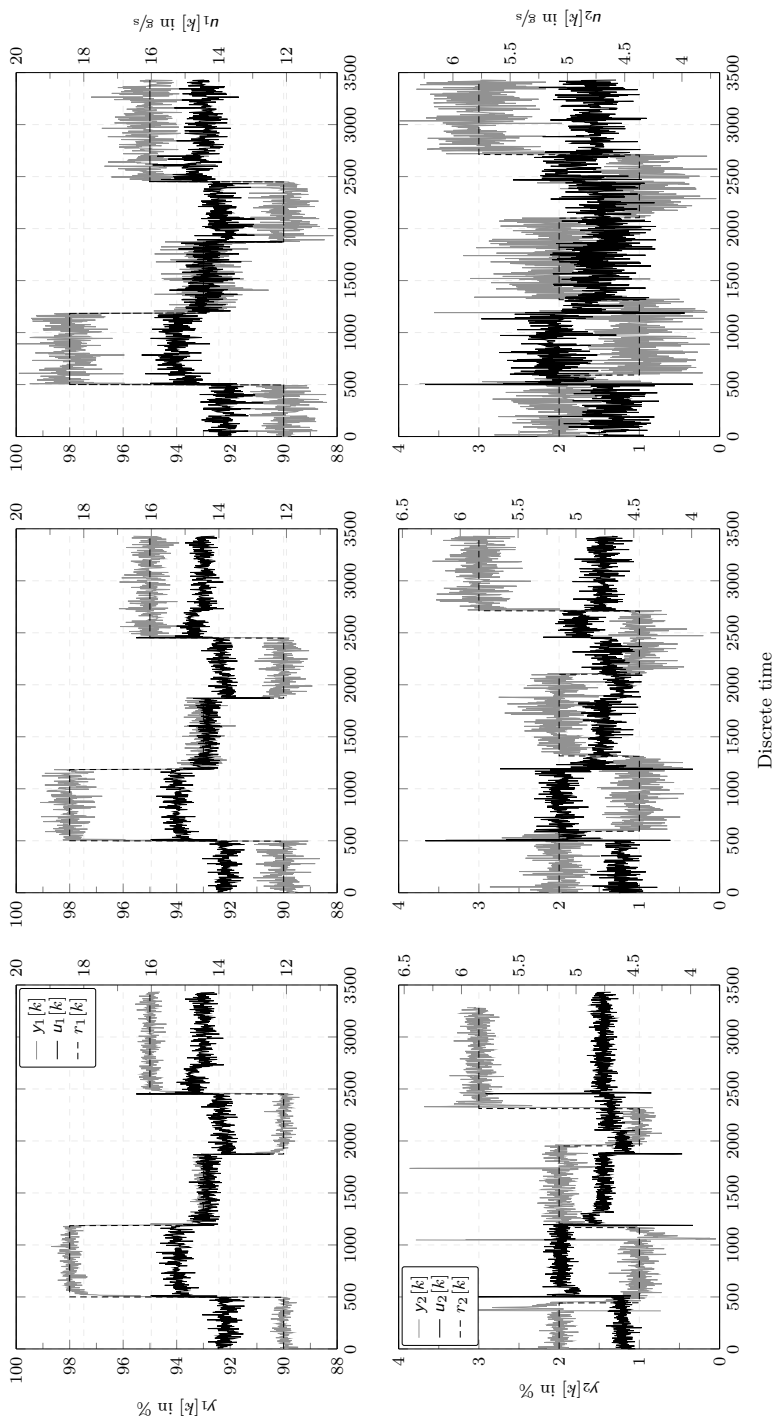
where  $R_{ij}$  is the amplitude of the  $i$ -th reference at the operating point between  $k_{j1} \leq k \leq k_{j2}$  and  $n_s$  is the total number of changes in the reference. The sequences listed in Table 6.1 for  $r_1[k]$  and  $r_2[k]$  were generated considering admissible ranges reported in [93] and used to excite the process (6.5).

The simulations are started with zero initial conditions using the references in Table 6.1. Three values of SNR are used for simulations: 25, 20 and 15 dB that are obtained by adjusting the noise variance  $\sigma_{e_i}^2$ ,  $i = 1, 2$  where  $i$  represents the process output. These SNRs represent values from low to high noise content in the signal and were chosen to evaluate DS4SID under different conditions. The SNR is defined as in (2.37) where  $\sigma_e^2$  and  $\sigma_{y_u}^2$  are the variances of the WGN and the noise-free process output, respectively.

Figure 6.4 shows the process outputs  $y_1[k]$  and  $y_2[k]$  under the formerly described SNRs in one of the data sets used for experiments. The external references  $r_1[k]$  and  $r_2[k]$  are shown as dashed lines for each output. The input signals  $u_1[k]$  and  $u_2[k]$  are shown in the bottom plot of Figure 6.4 and their values can be read on the right y-axis.

The system inputs are also embedded in noise due to the feedback loop. The resulting signals exhibit seldom changes as the system is predominantly at steady-state.

The test data sets are used to evaluate DS4SID. The data selection method reported in [12] was adapted for comparison. For the latter approach, the GLRT was implemented for lower interval bounding and the auxiliary information matrix was computed using the QR-RLS. In DS4SID, the ORIV was implemented for upper interval bounding. For each test data set, the intervals from DS4SID and the method reported in [12] are used for model estimation. Moreover, a third model is computed using the entire data set.



**Figure 6.4:** Selected intervals for the process (6.5) subject to SNR = 25 dB (left), SNR = 20 dB (middle) and SNR = 15 dB (right)



The resulting parameter vectors for each iteration are denoted by:

- $\hat{\Theta}_{1,j}$ : parameter vector estimated using the entire data set and the Gauss-Newton algorithm to minimize the prediction error
- $\hat{\Theta}_{2,j}$ : parameter vector estimated using informative data set obtained using the method adapted from [12] and LSM for data merging
- $\hat{\Theta}_{3,j}$ : parameter vector estimated using informative data set obtained with DS4SID and using an IV method for data merging

where  $j = 1, 2, \dots, n_{\text{exp}}$  denotes the  $m$ -th test data set and  $n_{\text{exp}} = 200$  is the number of test data sets. The goodness of fit (6.2) is computed for each model parametrized by  $\hat{\Theta}_{i,j}$ ,  $i = 1, 2, 3$ ,  $j = 1, 2, \dots, n_{\text{exp}}$ .

The sample mean of this criterion is computed on the total number of test data sets and the following value is used for comparison:

$$\begin{aligned} \text{fit}_{\text{ident}}(G(q^{-1}, \hat{\Theta}_i)) &= \frac{1}{n_{\text{exp}}} \sum_{j=1}^{n_{\text{exp}}} \text{fit}_{\text{ident}}(G(q^{-1}, \hat{\Theta}_{i,j})), \quad i = 1, 2, 3 \\ &= \frac{1}{n_{\text{exp}}} \sum_{j=1}^{n_{\text{exp}}} (1 - \text{NRMSE}(y(k), \hat{y}_{i,j}(k, \hat{\Theta}))) \cdot 100 \% \end{aligned} \quad (6.16)$$

The sample parameter vector over the total number of experiments is:

$$\hat{\hat{\Theta}}_i = \frac{1}{n_{\text{exp}}} \sum_{j=1}^{n_{\text{exp}}} \hat{\Theta}_{ij}, \quad i = 1, 2, 3 \quad (6.17)$$

Three models are parameterized using (6.17) and evaluated in validation data sets. The reference signals also represent step-like signals as in (6.15) but the amplitude and duration of the operating points are generated randomly within admissible ranges. The following criterion for goodness of fit on the validation data sets is also used for comparison

$$\begin{aligned} \text{fit}_{\text{val}}(G(q^{-1}, \hat{\Theta}_i)) &= \frac{1}{n_{\text{val}}} \sum_{j=1}^{n_{\text{val}}} \text{fit}_{\text{val}}(G(q^{-1}, \hat{\Theta}_i)), \quad i = 1, 2, 3 \\ &= \frac{1}{n_{\text{val}}} \sum_{j=1}^{n_{\text{val}}} (1 - \text{NRMSE}(y(k), \hat{y}_{i,j}(k, \hat{\Theta}))) \cdot 100 \% \end{aligned} \quad (6.18)$$

where  $n_{\text{val}} = 50$  is the number of validation data sets.

The length of test and validation data sets is each  $N = 3500$  samples. The starting discrete-time index  $k$  of these data sets correspond to the instants

where the process reaches steady-state for the first operating points defined by  $R_{11}$  and  $R_{21}$  (first row of Table 6.1).

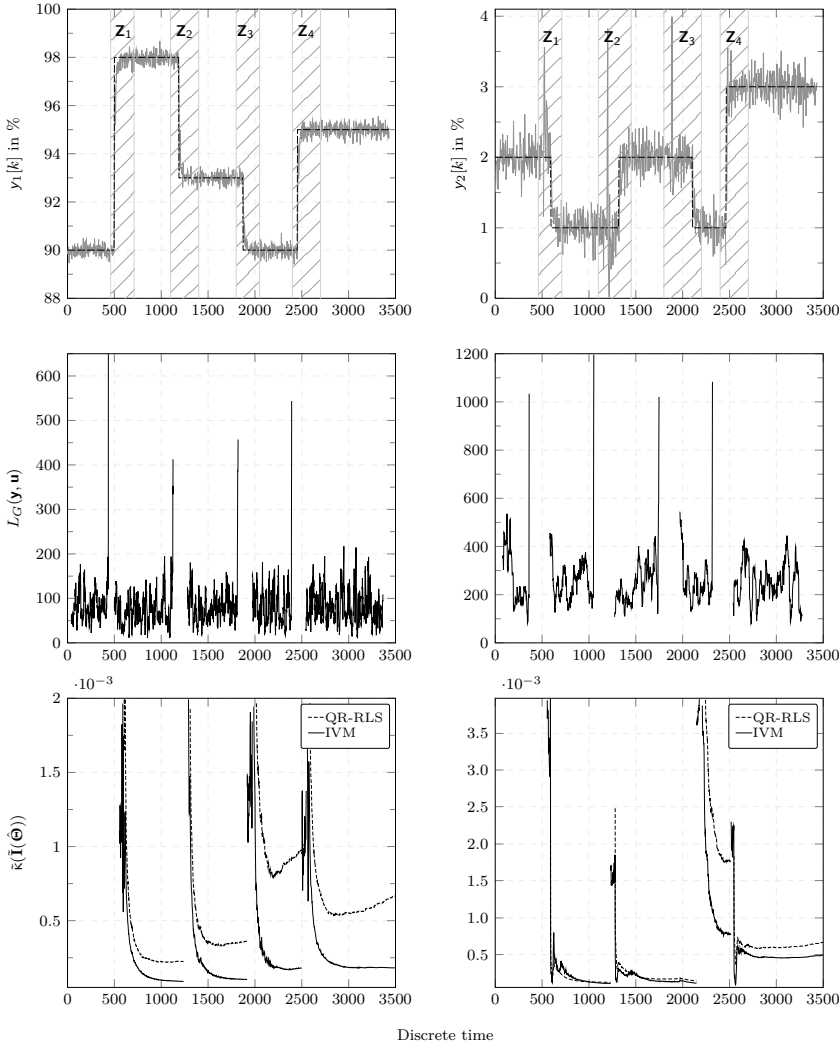
The detection test is used for locating lower interval bounds of possibly informative intervals corresponds to the change times located by the detection test. Then, the conditioning of  $\tilde{\mathbf{I}}(\hat{\Theta})$  is evaluated and upper interval bounds are set to instants where the auxiliary information matrix becomes ill-conditioned.

Current data selection methods, as reported in [12], locate change times based on detection tests whose hypotheses are based on time series models. As a consequence, the relation between an excitation in the input and the change in the output needs to be further evaluated with other tests *e.g.* the Granger causality test. The latter can be avoided by using the detection method proposed in DS4SID since the hypothesis of the detector are stated based on dynamic models. Thus, the causal relation between a dynamic change and excitation in the inputs is implicitly considered in the detection test. Thus, only two tests are required in DS4SID for lower and upper interval bounding which simplifies data selection and choice of user-defined parameters. Moreover, it is further discussed that misdetection and false alarms for change detection are reduced with the methods proposed in DS4SID.

ARX models are used for modeling and the model order,  $\mathbf{n}_j = (n_{a_j} \mathbf{n}_{b_{ij}} \mathbf{n}_{k_{ij}})$ ,  $i, j = 1, 2$ , is assumed to be known and is defined by:

$$\begin{aligned} n_{a_1} &= 1 & \mathbf{n}_{b_{11}} &= (1 \ 1) & \mathbf{n}_{k_{11}} &= (1 \ 3) \\ n_{a_2} &= 1 & \mathbf{n}_{b_{12}} &= (1 \ 1) & \mathbf{n}_{k_{12}} &= (7 \ 3) \end{aligned} \quad (6.19)$$

The parameter vector for the null-hypothesis  $\mathcal{H}_0$  is computed with the data contained in the growing time window  $M_0$ , where  $n_{p_j}$  is the number of parameters for the  $j$ -th MISO model. Thus, the initial size of  $M_0$  should be set so that  $M_0 \gg n_{p_j}$ . For this case study,  $M_0$  was empirically initialized setting  $20n_{p_j} \leq M_0 \leq 50n_{p_j}$ . The estimates for the alternative hypothesis  $\mathcal{H}_1$  are computed with sliding fixed-size time window  $M_1$  whose size is set following a similar approach as formerly discussed for  $M_0$ . However,  $M_1 < M_0$  and by trial and error  $M_1 = M_0/2$  yielded satisfactory results for the estimates under  $\mathcal{H}_1$ . The threshold  $\gamma$  used for the detection test is set as a trade-off between detection rate and false alarms. For this case study,  $250 < \gamma < 400$  yielded acceptable values for the detection test. Upper interval bounds are defined by thresholding  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  with  $\eta = 3 \cdot 10^{-4}$  where the value of  $\eta$  was obtained from [73] which indicates a limit for ill-conditioning of the auxiliary information matrix  $\tilde{\mathbf{I}}(\hat{\Theta})$ .



**Figure 6.5:** Selected intervals for  $\{y_1[k]\}$  and  $\{y_2[k]\}$  from a test data set (SNR = 15 dB). Top: selected intervals using DS4SID. Middle: Detection test for locating lower interval bounds. Bottom: Conditioning of the information matrix for determining upper interval bounds. Continuous line: DS4SID using IVM. Dashed line: method as presented in [12] using QR-RLS.

### 6.2.4 Results and Discussion

A total of 200 test data sets were generated for each SNR. Detectors for data selection methods based on time series models result in a large number of false

alarms, particularly, in the case of high noise content. In contrast, the detection test proposed for DS4SID can successfully locate change times while keeping false alarm rate in acceptable ranges.

Figure 6.5 shows the selected intervals for the process outputs  $y_1[k]$  and  $y_2[k]$  using DS4SID for SNR = 15 dB. The conditioning of the information matrix associated to the method reported in [12] and for DS4SID are shown in the bottom row of Figure 6.5. The detection test detects four intervals that coincide with the number of changes between operating points. Change times are successfully detected by the GLRT since the condition  $L_G(\mathbf{y}, \mathbf{u}) > \gamma$  only holds for the instants where the process is moved out from an operating point. The value of  $L_G(\mathbf{y}, \mathbf{u})$  is lower than the threshold  $\gamma$  regardless the additive noise which results in few false alarms for change detection.

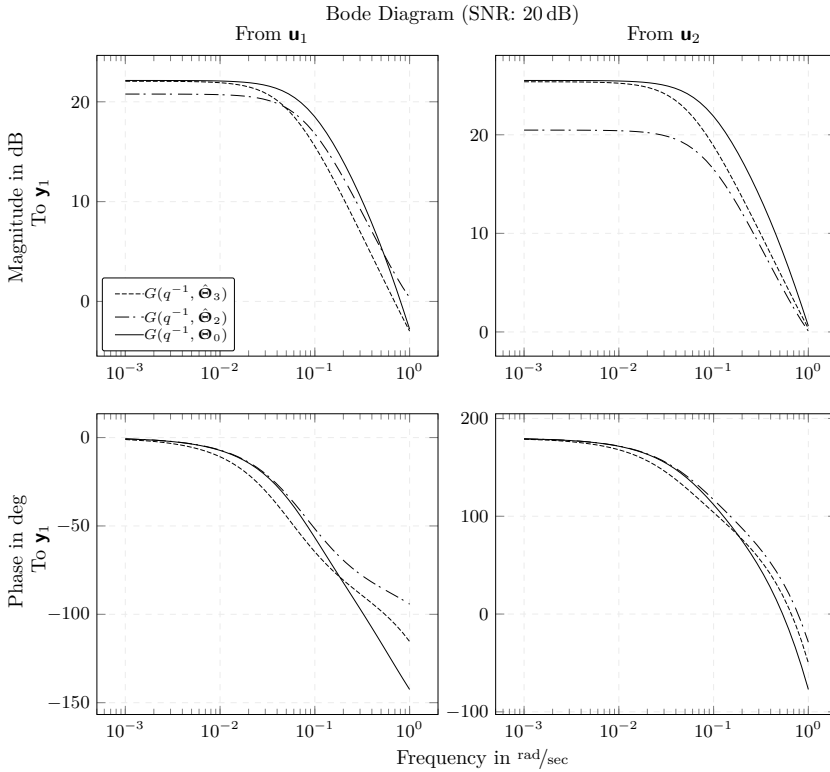
The conditioning of two auxiliary information matrices are computed for each test data set that are denoted by:

$$\begin{aligned} \tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}_2)) \\ \tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}_3)) \end{aligned} \tag{6.20}$$

where  $\hat{\Theta}_2$  and  $\hat{\Theta}_3$  are the estimates with intervals retrieved from the method in [12] and DS4SID, respectively.

The value of (6.20) has a maximum at the instant where the process reaches steady-state conditions for a new operating point. Then, the conditioning of the auxiliary information matrix decays below values that can cause numerical computational problems. Upper interval bounds are defined differently for each method since the noise is colored and this aspect is not addressed by the method in [12]. The auxiliary information matrix from DS4SID becomes ill-conditioned earlier than with the other method. Thus, informative intervals using DS4SID are shorter than the ones obtained from the method adapted from [12]. This results in models with better goodness of fit as further discussed. Consequently, less data are selected and performance of estimated models is satisfactory.

The results of DS4SID and the method adapted from [12] are compared in Table 6.2. In this application, the number of the instruments could be chosen so that computations in ORIV have some computational advantages over QR-RLS. In the ORIV, the left part of (3.79a) has size  $n_\zeta \times (n + 1)$  where  $n$  is the model order and  $n_\zeta$  is the number of instruments. Whereas, an analogous matrix of (3.79a) in the QR-RLS method has size  $(n + 2) \times (n + 1)$ . In the present example,  $n < n_\zeta \leq (n + 2)$  holds and, consequently smaller matrices are obtained for ORIV. This results in a smaller processing time as shown in Table 6.2.



**Figure 6.6:** Bode diagrams of models computed with informative intervals and true process. Dashed line: model estimated using intervals from DS4SID ( $G(q^{-1}, \hat{\Theta}_3)$ ). Dashed-dotted line: model estimated using intervals from [12] ( $G(q^{-1}, \hat{\Theta}_2)$ ). Continuous line: true process ( $G(q^{-1}, \hat{\Theta}_0)$ ).

The multiple-cost approach introduced in [section 4.5](#) was used for model estimation. The information matrices as well as the model parameters for the method in [12] and for DS4SID were obtained using the QR-RLS and the ORIV, respectively. The goodness of fit of  $G(q^{-1}, \hat{\Theta}_3)$  for the test and validation data sets is better than for models obtained using the other approaches which confirms the improved performance of DS4SID. Interestingly, models estimated with informative intervals using DS4SID have a similar goodness of fit as estimates using the entire data set.

[Figure 6.6](#) shows the resulting bode plots of the models computed with informative intervals and listed in [Table 6.2](#). The transfer functions were obtained with parameters averaged over the total number of experiments. Models estimated with the intervals retrieved using DS4SID and using the

**Table 6.2:** DS4SID for  $\{y_1[k]\}$  and  $\{y_2[k]\}$  ( $\hat{\Theta}_1$ : parameter vector using the entire data set;  $\hat{\Theta}_2$ : estimates using intervals retrieved from [12] and  $\hat{\Theta}_3$ : parameter vector using intervals from DS4SID). Test data sets as in Figure 6.4.

		$y_1$			$y_2$		
		SNR in dB			SNR in dB		
		25	20	15	25	20	15
UINR in %	ORIV	27.00	25.61	25.31	18.00	24.10	22.35
	QR-RLS	24.02	22.67	21.87	28.17	24.25	22.15
Computational time in s	ORIV	0.876	0.868	0.889	0.890	0.880	0.885
	QR-RLS	0.821	0.813	0.832	0.842	0.820	0.839
fit( $y_{\text{ident}}, \hat{y}$ ) in %	$G(q^{-1}, \hat{\Theta}_1)$	74.12	67.21	53.47	70.12	62.28	52.10
	$G(q^{-1}, \hat{\Theta}_2)$	65.60	58.81	45.16	62.20	55.25	40.58
	$G(q^{-1}, \hat{\Theta}_3)$	71.77	64.66	48.26	68.54	60.45	45.87
fit( $y_{\text{val}}, \hat{y}$ ) in %	$G(q^{-1}, \hat{\Theta}_1)$	74.01	67.09	53.31	70.01	62.09	50.31
	$G(q^{-1}, \hat{\Theta}_2)$	61.18	54.79	42.26	59.18	52.79	40.26
	$G(q^{-1}, \hat{\Theta}_3)$	71.05	63.44	46.42	70.05	60.44	44.42

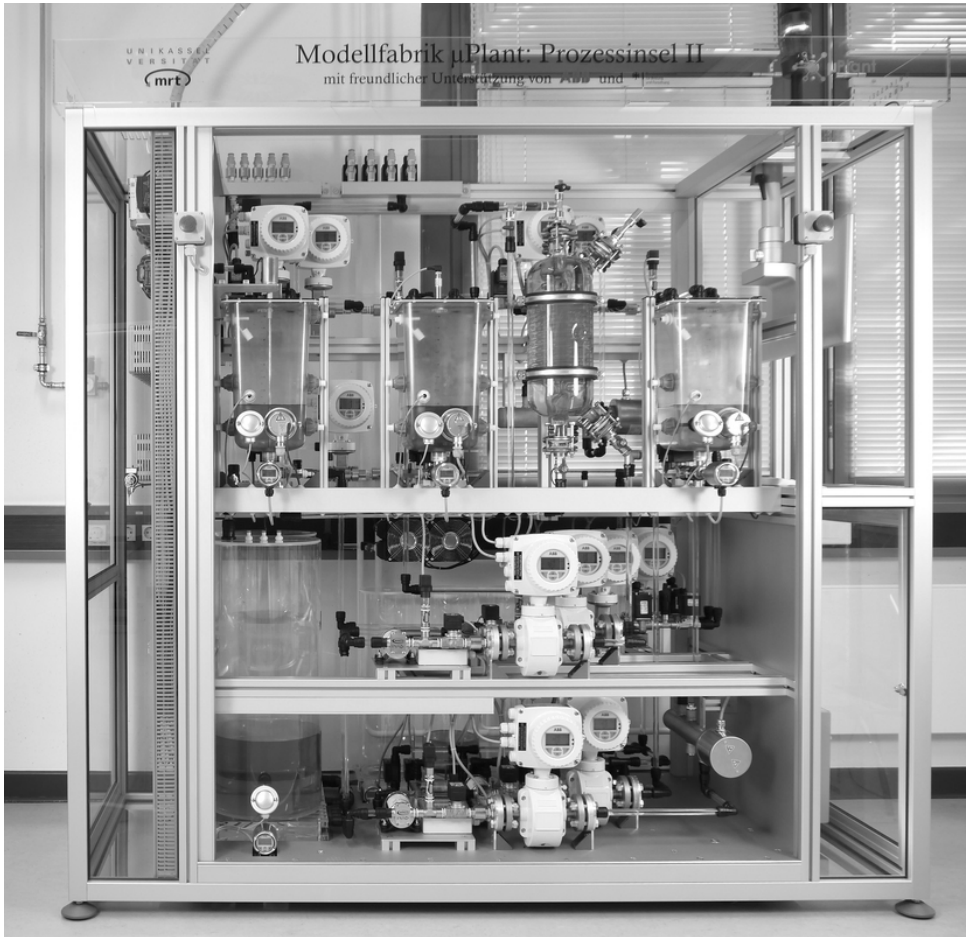
multiple-cost approach for parameter estimation presented in section 4.5 yields better models than when using the intervals obtained with [12]. The benefits of using the proposed method for data merging are even more noticeable as the SNR increases.

## 6.3 Industry-oriented Case Study: The Process Unit II

The process unit II is part of the  $\mu$ Plant [53], which is a lab-scale chemical production plant where continuous and batch industrial operation scenarios can be recreated. The  $\mu$ Plant consists of two process units, two stations for emptying and filling, a product storage area with articulated robot, a finishing unit and four autonomous mobile robots that transport products in the plant. In this section, the process unit II is firstly described and experiments as well as performance of DS4SID are further discussed. In order to avoid operational risks during performing the experiments and to prevent drainage of chemical substances, water was used as fluid in the process unit II.

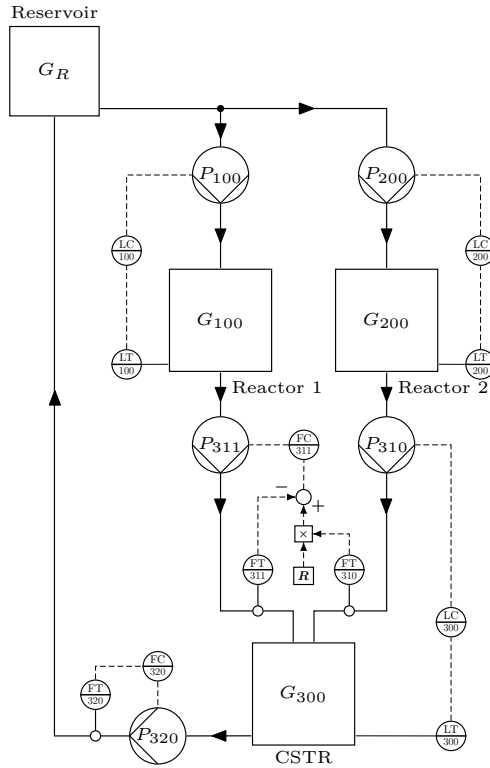
### 6.3.1 Process Description and Control Scheme

The process unit II or *Prozessinsel II* (see Figure 6.7) was designed, built-up and commissioned in the Department of Measurement and Control (MRT) at the University of Kassel. Three main operations are performed in the process unit II: two of them correspond to processing of raw material while a final blending



**Figure 6.7:** Process unit II used for case study

process is carried out in the CSTR. A simplified process and instrumentation diagram (P&ID) of the process unit II is shown in [Figure 6.8](#). The raw material is pumped from the reservoir  $G_R$  to the reactors  $G_{100}$  and  $G_{200}$ . Processing in the former reactors result in intermediate products that further pumped to the CSTR  $G_{300}$ . The final product is obtained by blending the inflow streams in  $G_{300}$  and it can be stored in several storage tanks (not shown). Alternately, it can be pumped back to the reservoir which recreates material recycling. The external references of the closed-loop process are the production rate (outflow from  $G_{300}$ ) and the tank levels.



**Figure 6.8:** Simplified P&ID of the process unit II

The process unit II is equipped with sensors to measure different variables such as pressure, temperature, level and flow. Sensors are wired to input-output modules and measurements are processed in an ABB<sup>®</sup> AC 700F controller. Measurements are logged in a data historian system with a predefined sampling time of  $T_s = 250$  ms. This value is preconfigured by the manufacturer but it can be adjusted by the user in the program of the control system. However, the logged data can be downsampled by setting  $k_{os} = c_{os} \cdot k$  where  $k$  and  $k_{os}$  are the discrete-time index for the original and the modified data sets and  $c_{os} \in \mathbb{Z}$ ,  $c_{os} > 1$ . For the case studies in the process unit II the following holds:

$$\frac{T_s}{T_{95}} \approx \frac{1}{5}, \dots, \frac{1}{15} \quad (6.21)$$

where  $T_{95}$  is the 95% settling time of the step response of a proportional acting process. In the process unit II,  $T_{95} \approx 4.5$  s which yields  $T_s/T_{95} = 0.25/4.5 = 0.056$



that agrees with (6.21). Thus, digital controllers can be successfully designed as described in [41]. The process unit II is supervised and operated using a human machine interface (HMI) that was developed with ABB® Freelance. The logged data represents a process which is operated between 7 h and 18 h with seldom changes between operating points.

Five control loops are implemented in the process unit II. Three controllers regulate the level of the three reactors while the other two control the production rate and the mixing ratio of the reactants, respectively. The control concept is explained from the bottom to the top of Figure 6.8. The signals referred to in the following are defined in Figure 6.9.

The production rate is controlled by the feedback loop FC-320 which regulates the pump  $P_{320}$ . The pump voltage,  $u_{320}(t)$ , is the manipulated variable and the output flow,  $y_{320}(t)$ , measured by the device FT-320, is the controlled variable. The flow streams of the reactants delivered to the CSTR  $G_{300}$  are related by the ratio factor  $R$ . The output flow of the reactant in the first reactor  $G_{100}$  is regulated by the loop FC-220 that controls the pump  $P_{220}$ . The set point of the flow control loop FC-220 is obtained by multiplying the mixing ratio  $R$  with the flow stream of the second reactant from  $G_{200}$  whose control was designed following [72]. The set point of the flow control loop FC-220 is obtained by

$$w_{220}(t) = R \cdot y_{310}(t) \quad (6.22)$$

where  $R$  represents the ratio between the two reactants and is a user-defined parameter. For this case study, the value of  $R$  is set to 1 i.e. it is assumed that the reactant streams are required to be the same.

For performing of experiments, a controller system based on decentralized PID controllers was designed and implemented in the process unit II. The level of the CSTR  $G_{300}$  is regulated by the control loop LC-300. The controlled variable  $y_{300}(t)$  is measured by the instrument LT-300 whereas the manipulated variable is the voltage of the pump  $P_{310}$ ,  $u_{310}(t)$ . In the first reactor, the level is controlled by the feedback loop LC-100. In this loop, the controlled and manipulated variables are  $y_{100}(t)$  and  $u_{100}(t)$ , respectively. The level of the second reactor,  $y_{200}(t)$  is regulated by the control loop LC-200. The voltage of the pump  $P_{200}$ ,  $u_{200}(t)$ , is the manipulated variable. The input flow streams from the reservoir  $G_R$  to the first  $G_{100}$  and second reactor  $G_{200}$  are regulated by  $u_{100}(t)$  and  $u_{200}(t)$ , respectively.

The set points corresponding to the levels of the reactors are  $w_{100}(t)$ ,  $w_{200}(t)$  and  $w_{300}(t)$ . Moreover, the user-defined production rate is represented by  $w_{320}(t)$ . The level controllers are represented by  $R_{P_{100}}$ ,  $R_{P_{200}}$  and  $R_{P_{310}}$ . Whereas, the flow controllers are  $R_{P_{220}}$  and  $R_{P_{320}}$  for the mixing ratio and production rate,

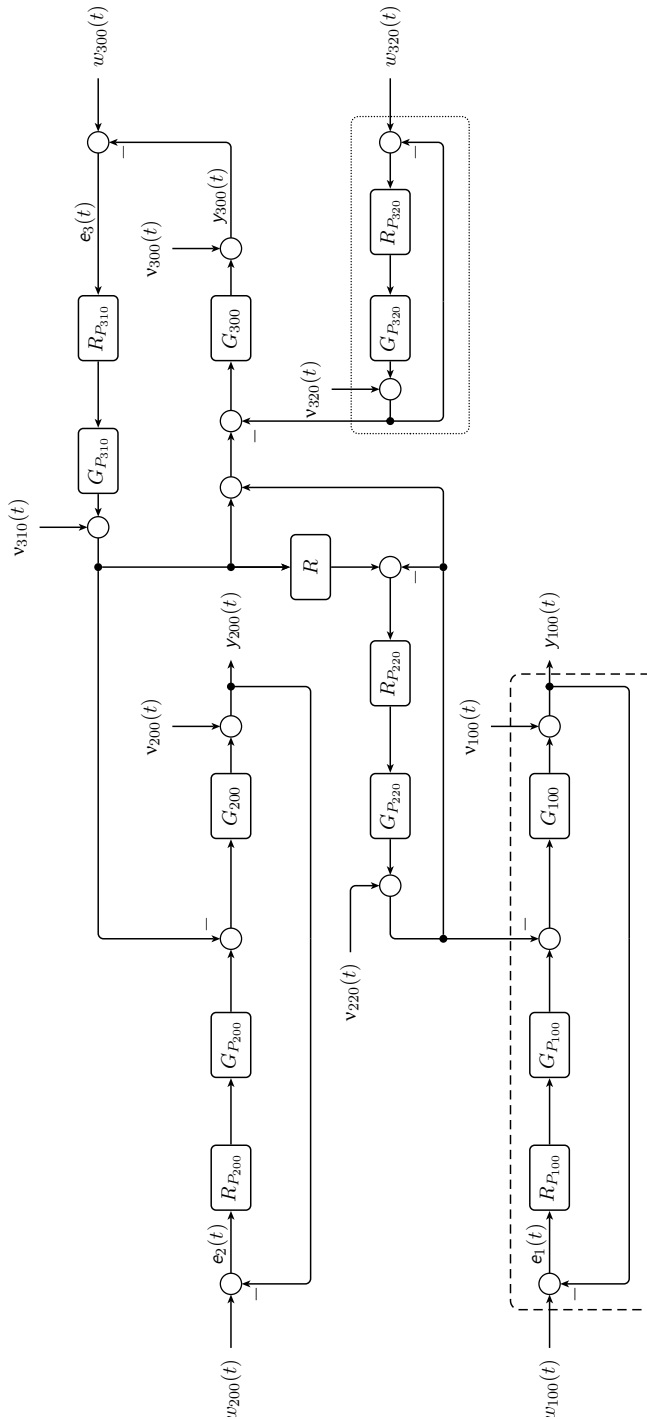


Figure 6.9: Control scheme of the process unit II. Dotted: SISO process. Dashed: MISO process.

respectively. The transfer functions of the reactors are described by  $G_{100}$ ,  $G_{200}$  and  $G_{300}$ . The transfer functions of the pumps used in the level control loops are  $G_{P_{100}}$ ,  $G_{P_{200}}$ ,  $G_{P_{310}}$ . The pumps used in the flow control loops correspond to  $G_{P_{220}}$  and  $G_{P_{320}}$ . All sensors are standard industrial devices and measurements are transmitted to remote I/Os using 4/20 mA standard. Proportional-Integral (PI) controllers were implemented and used for the operation of the process unit II. The designed controllers have an anti-windup feature and the sampling time was chosen as the default value for the process *i.e.*  $T_s = 250$  ms. These regulators were designed within several student projects and are listed in Table 6.3.

### 6.3.2 Noise Analysis

The measurement noise is denoted by  $v(t)$  in Figure 6.9 and it can represent WGN or colored noise. In real applications,  $v(t)$  is more likely to be colored than white noise because measurements are usually filtered before they are logged. This aspect is further discussed for the collected signals from the process unit II in Appendix B. Noise analysis is based on data recorded from open-loop operation. Since the data was collected in a digital control system, the time argument of the variables will be changed to  $k$ , where  $k = 0, 1, \dots, N - 1$  represents the discrete time and  $N$  is the number of observations. Thus,  $y[k]$  represents the value of  $y$  at the discrete time  $k$ . Set points for level and flow were kept at the same value during long periods to collect enough data for analysis. Logged data denoted by the sequence  $x_i[k]$ ,  $i = 1, 2, \dots, n_x$  were firstly modeled as DC level in WGN:

$$x_i[k] = A_i + e_i[k] \quad (6.23)$$

**Table 6.3:** Level and flow controllers used in the process unit II

	Transfer function
Level control	$R_{P_{100}}(q^{-1}) = \frac{10.68 - 8.785q^{-1}}{1 - q^{-1}}$
	$R_{P_{200}}(q^{-1}) = \frac{11.38 - 8.942q^{-1}}{1 - q^{-1}}$
	$R_{P_{300}}(q^{-1}) = \frac{8.923 - 8.456q^{-1}}{1 - q^{-1}}$
Flow control	$R_{P_{220}}(q^{-1}) = \frac{0.18 - 0.1556q^{-1}}{1 - q^{-1}}$
	$R_{P_{320}}(q^{-1}) = \frac{0.4 - 0.3553q^{-1}}{1 - q^{-1}}$

where  $\mathbf{e}[k]$  is WGN and  $A_i$  is the sample mean (DC level),  $A_i = \bar{x}_i[k]$ , computed using (2.23). The following sequence was calculated for analysis of the noise

$$\mathbf{v}[k] = x_i[k] - \bar{x}_i[k] \quad (6.24)$$

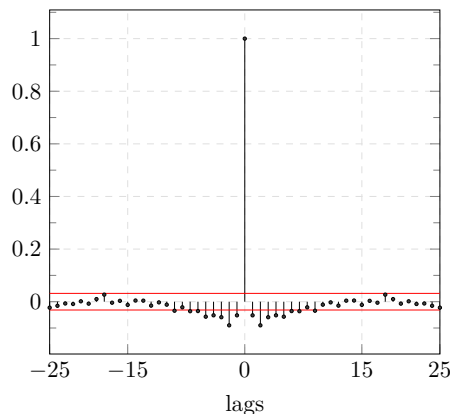
The term  $\mathbf{v}[k]$  in (6.24) should be WGN if the model (6.23) fits to the observations. The autocorrelations and distributions of  $\mathbf{v}[k]$  are shown in Appendix B.

The sequence  $\mathbf{v}[k]$  is considered WGN if the sample autocorrelation yields small values so that  $\mathbf{v}[k]$  is assumed to be a sequence of i.i.d. random variables.

Even though the noise analysis is not the main focus of the present work, the estimation method should be robust to the measurement noise.

As shown in Figure 6.10,  $\{\mathbf{v}[k]\}$  differs from WGN as it exhibits correlation up to approximately  $k = 5$ . The model (6.23) can not properly describe the signal and, consequently, other models should be evaluated. The measurement noise is not white but colored and this important aspect should be considered when proposing methods for data selection. In industrial applications, measurement noise is also more likely to be colored than WGN. Detection and parameter estimation methods should be robust to colored noise so that interval bounds are well defined.

Different time series models, namely AR and ARMA models were chosen to describe the data logged in open-loop. The residuals of the estimated AR models result in a sequence that can be considered WGN. The latter confirms that there exist correlation in the measurement noise and, thus, it should be treated as colored. This conclusion supports choice of robust methods for upper interval bounding such as ORIV instead of LSM.



**Figure 6.10:** Autocorrelation of residuals for model (6.23)

### 6.3.3 Performed Experiments

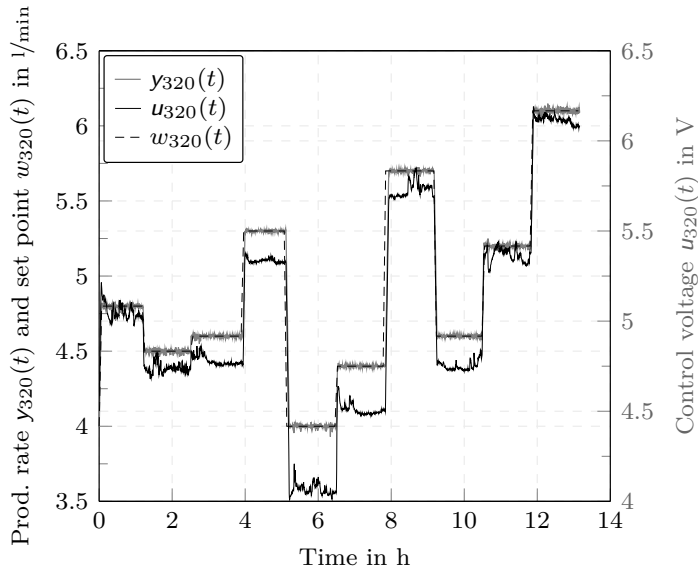
The ratio factor  $R$  in (6.22) was set to 1 *i.e.* the flow rate of the reactants is the same during the experiments. The process unit II is operated in continuous mode and the set points are seldom changed. The purpose is to generate data sets where the process is occasionally excited so that the proposed selection method locates relevant data for identification.

As an example, consider the production rate for one of the experiments. The set point for this variable,  $w_{320}(t)$ , is seldom changed in time periods that range between 50 min and 80 min (see Figure 6.11). The control system of the process unit II was programmed to change automatically the value of the set points. To simplify the implementation of this program, the set points in Figure 6.11 and Figure 6.12 are changed simultaneously. This aspect does not affect the evaluation of DS4SID since, as discussed in section 6.2.4, the proposed method can locate informative intervals if changes occur at different instants. Values for amplitudes are defined based on admissible ranges to avoid saturation of the controllers. Measurements of the production rate for the first experiment are shown in Figure 6.11. DS4SID is evaluated on different test data sets and the goodness of fit of models estimated with retrieved intervals is compared with other models computed using the entire data set.

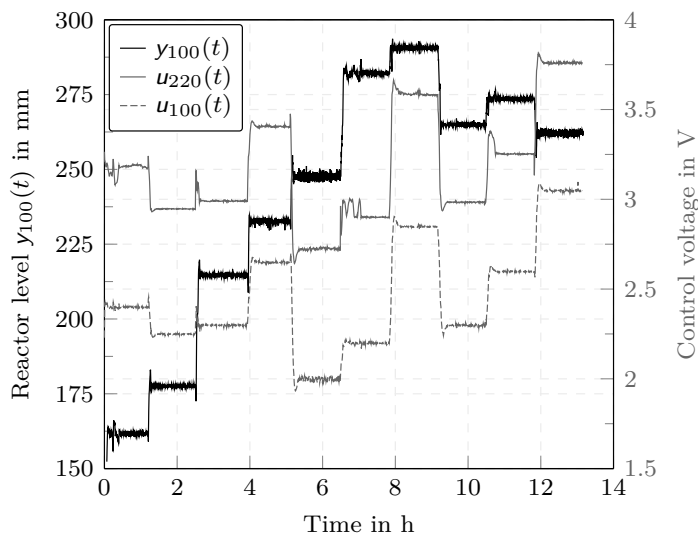
Figure 6.12 shows the level of the first reactor for the experiment previously described. Similarly to the production rate, the references for the level in the first reactor,  $w_{100}(t)$ , is seldom changed (approx. every 60 min). In the experiment of Figure 6.12, the process unit II was operated approximately during 13.5 h and data were logged using the pre-defined sampling time of the historian system,  $T_s = 250$  ms. Since the process dynamics is slower than the sampling time, underlying data-driven modeling in DS4SID is not affected by the predefined value of  $T_s$ .

The performance of DS4SID was assessed on data retrieved from two sub-systems: the production rate  $\{y_{320}[k]\}$  of the CSTR and the level  $\{y_{100}[k]\}$  of the first reactor that correspond to a SISO and a MISO process, respectively. The input and output variables for the first subprocess are  $u_{320}[k]$  and  $y_{320}[k]$ , respectively. The level  $y_{100}[k]$  is regulated by the control signals  $u_{100}[k]$  and  $u_{220}[k]$ .

A total of seven experiments were performed in the process unit II for evaluation of DS4SID. Table 6.4 describes the experiments performed in the process unit II. The duration of the experiments ranges between 6 h and 18 h. Duration of steady-state regions was varied to analyze influence of this condition on the length of the retrieved intervals.



**Figure 6.11:** Production rate and control voltage of the pump  $P_{320}$  for the first experiment (see Table 6.5)



**Figure 6.12:** Output and input signals for the first experiment (see Table 6.5)

**Table 6.4:** Summary of experiments in the process unit II

Exp.	1	2	3	4	5	6	7
Duration in h	13.2	8.0	8.7	11.5	18.0	6.1	6.8
No. of changes	10	6	9	12	21	6	8
Min. change in %	2.5	1.5	4.5	2.5	1.5	18	8
Max. change in %	30	15	35	50	30	40	35

The method proposed in [12] was implemented to compare its performance with DS4SID. The fourth and fifth experiment were selected for tests of the data selection methods. The data sets of the remaining experiments were used for cross-validation to evaluate the models estimated with the retrieved informative intervals. The computed input transfer functions are denoted by  $G(q^{-1}, \hat{\Theta}_i)$  where the parameter vectors are defined as follows:

- $\hat{\Theta}_1$ : parameter vector estimated using the entire data set and the Gauss-Newton algorithm to minimize the prediction error
- $\hat{\Theta}_2$ : parameter vector estimated using informative data set obtained with method adapted from [12] and using LSM for data merging
- $\hat{\Theta}_3$ : parameter vector estimated using informative data set obtained with DS4SID and using an IV method for data merging

The performance of DS4SID was evaluated using the UINR and goodness of fit as introduced in (6.1) and (6.2), respectively, and with (6.3) as cost function. The misdetection rate (MR) is defined as the fraction of non-detected intervals among the total of informative intervals. In a data selection method, this value should be small regardless process conditions such as change in amplitude of the excitation signal or the noise content expressed by the SNR.

For this case study,  $M_0$  was empirically initialized setting  $100 \leq M_0 \leq 200$  samples. Similarly to the simulated case study, the sliding window  $M_1 = \frac{M_0}{2}$  yielded satisfactory results for the estimates under  $\mathcal{H}_1$ . For the industry-oriented case study,  $250 < \gamma < 400$  yielded acceptable values for the detection test. The parameters for the ORIV method were set to  $\mu = 10^{-3}$ ,  $k_\zeta = 10$  and  $\lambda = 0.995$ . The value of  $\tilde{\kappa}(\tilde{\mathbf{I}}(\hat{\Theta}))$  is thresholded with  $\eta = 3 \cdot 10^{-4}$  where the value of  $\eta$  was obtained from [73] which indicates a limit for ill-conditioning of the auxiliary information matrix  $\tilde{\mathbf{I}}(\hat{\Theta})$ .

### 6.3.4 Results and Discussion

Table 6.5 shows the results of the data selection methods applied to the production rate which corresponds to a SISO process.

**Table 6.5:** Subprocess: output flow rate. Goodness of fit (6.2) for selected intervals using DS4SID ( $G(q^{-1}, \hat{\Theta}_3)$ ), using [12] ( $G(q^{-1}, \hat{\Theta}_2)$ ) and for entire data set ( $G(q^{-1}, \hat{\Theta}_1)$ ). Bold font: best results for test experiments.

		$G_{P320}$					
		Experiments used for tests					
Selection Method		Exp. no. 4			Exp. no. 5		
		[12]	DS4SID		[12]	DS4SID	
Missdetection rate		5/12	2/12		12/21	3/21	
UINR in %		7.12	20.28		10.19	18.19	
Exp. no.	Duration in h	$G(q^{-1}, \hat{\Theta}_1)$	$G(q^{-1}, \hat{\Theta}_2)$	$G(q^{-1}, \hat{\Theta}_3)$	$G(q^{-1}, \hat{\Theta}_1)$	$G(q^{-1}, \hat{\Theta}_2)$	$G(q^{-1}, \hat{\Theta}_3)$
1	13.2	85.73	76.64	82.51	85.51	82.15	84.04
2	8.0	74.47	64.18	71.70	73.77	71.48	72.88
3	8.7	90.24	69.83	80.97	87.05	80.46	83.21
<b>4</b>	<b>11.5</b>	<b>95.27</b>	<b>76.36</b>	<b>85.07</b>	91.42	84.52	87.06
<b>5</b>	<b>18.0</b>	90.54	73.85	84.06	<b>88.63</b>	<b>83.63</b>	<b>86.20</b>
6	6.1	83.78	65.66	76.30	81.66	75.83	78.13
7	6.8	79.54	63.88	74.39	78.42	74.03	76.24

The goodness of fit of  $G(q^{-1}, \hat{\Theta}_1)$  is slightly better than for models estimated with informative intervals  $G(q^{-1}, \hat{\Theta}_2)$  and  $G(q^{-1}, \hat{\Theta}_3)$ . However, it is worth noting that considering less than 20% of the entire data sets, the goodness of fit decreases only around 4%. This difference can be acceptable specially when dealing with very large data sets where model estimation can require long computational time.

The method adapted from [12] has a larger misdetection rate than DS4SID which confirms limitations of the first approach where intervals can be wrongly discarded in one of the stages for interval bounding. Misdetection occurs in part due to rejection of the third test reported in [12] since  $\tilde{\kappa}(t) > \eta_3$  does not hold because the threshold is wrongly set. In the first experiment (cf. Figure 6.11), the transition between the second and third operation point is small and the detection test fails detecting this change since  $L_G > \gamma$  does not hold. As shown in Table 6.5, the higher the misdetection rate is, the lower the UINR is. Results obtained from the fifth experiment confirm these observations. The high misdetection rate is explained by the fact that the amplitude of the changes ranges within values that affects largely the test of the condition number.

Results of DS4SID applied to the level in the first reactor are shown in Table 6.6. This system corresponds to a MISO process with two inputs. The misdetection rate of DS4SID is also smaller for this process than the approach in [12] which confirms the results discussed for the SISO case. The increase of



**Table 6.6:** Subprocess: level in the first reactor. Goodness of fit (6.2) for selected intervals using DS4SID ( $G(q^{-1}, \hat{\Theta}_3)$ ), using [12] ( $G(q^{-1}, \hat{\Theta}_2)$ ) and for entire data set ( $G(q^{-1}, \hat{\Theta}_1)$ ). Bold font: results for test experiments.

		$G_{100}$					
		Experiments used for tests					
Selection Method		Exp. no. 4			Exp. no. 5		
Misddetection rate		[12]	DS4SID		[12]	DS4SID	
UINR in %		7/12	3/12		15/21	5/21	
		10.25	24.30		15.35	28.20	
Exp. no.	Duration in h	$G(q^{-1}, \hat{\Theta}_1)$	$G(q^{-1}, \hat{\Theta}_2)$	$G(q^{-1}, \hat{\Theta}_3)$	$G(q^{-1}, \hat{\Theta}_1)$	$G(q^{-1}, \hat{\Theta}_2)$	$G(q^{-1}, \hat{\Theta}_3)$
1	13.2	90.30	80.22	88.37	89.43	88.31	89.00
2	8.0	68.38	59.37	65.15	66.98	65.05	66.30
3	8.7	69.48	56.93	63.49	66.85	63.29	65.41
<b>4</b>	<b>11.5</b>	<b>76.97</b>	<b>67.61</b>	<b>73.15</b>	75.32	73.02	74.43
<b>5</b>	<b>18.0</b>	81.98	74.24	78.81	<b>80.60</b>	<b>78.71</b>	<b>79.89</b>
6	6.1	80.37	70.91	75.40	78.20	75.22	77.01
7	6.8	59.87	45.70	51.12	56.18	50.79	53.94

the misdetection rate is explained by partial excitation in only one of the two inputs. Misddetection in DS4SID occurs when the detection test cannot locate changes in the process particularly in case of small changes in the amplitude. The UINR results in larger values for DS4SID because of its better detection rate when compared to the other selection approach. Larger values of UINR with respect to the SISO case can be explained because the controllers require longer time to regulate the process and, consequently, the MISO system is excited for larger periods when compared to the SISO example.

Models estimated with informative intervals, retrieved using DS4SID, have better goodness of fit than other models estimated used the alternative approach. The results listed in Table 6.6 confirm the satisfactory performance of the DS4SID also for the analyzed multivariate process.  $G(q^{-1}, \hat{\Theta}_3)$  outperforms between 3% and 5%  $G(q^{-1}, \hat{\Theta}_2)$ . Parameter estimation using the entire data set also results in models with better goodness of fit as found for the production rate. However, identification using only selected intervals yields better conditioned information matrices than when using the entire data set. Thus, data selection for identification can offer additional advantages for multivariate processes that are seldom excited.

In the experiments selected for testing DS4SID, relevant data subsets for identification represent between 7% and 30% of the entire data sets. The goodness of fit of the models computed with the informative intervals is satisfactory when

compared with  $G(q^{-1}, \hat{\theta}_1)$ . The computational load can be considerably reduced for model estimation as only a small part of the recorded data is required to estimate satisfactory models.

## 6.4 Discussion

The proposed data selection method, DS4SID, was evaluated on simulation and industry-oriented case studies. A model of a binary distillation column, which represents a  $2 \times 2$  highly interacting multivariate process, was used as simulation example. The controller consists of two non-interacting regulators and two PI controllers. The process was simulated in closed-loop recreating conditions where operating points are seldom changed. Colored Gaussian noise was added to the outputs to recreate real scenarios where this type of noise is more likely to be found than WGN.

A total of 200 experiments were performed for the binary distillation column. Models were estimated using the retrieved intervals from each experiment and the goodness of fit was compared with other estimates obtained using an alternative data selection approach and the entire data set. The UINR was around 20% regardless the SNR which shows the robustness of DS4SID to different levels of noise content. The goodness of fit for models estimated with DS4SID was found to be better than for estimates using the alternative approach which shows a better performance of the proposed method. Moreover, the goodness of fit of models computed with informative intervals is satisfactory when compared with estimates using the entire data set. The goodness of fit degrades between 5% and 8% which represents an acceptable trade-off considering the size of the data subset used for model estimation.

A lab-scale plant operated in continuous mode was used as industry-oriented case study. The process unit II mainly consists of two reactors and a CSTR where batch and continuous processes can be recreated. DS4SID was tested in a SISO and a MISO process. Several experiments with duration between 6.5 h and 13.2 h and seldom changes in the set points were performed.

Location of deviations from current operating points using DS4SID is robust to changes in the amplitude and noise content in the signals. The UINR for the output flow rate changed between 15% and 20% depending on the experiment. This difference is explained by a larger variation in the corresponding set point in the fourth experiment. The goodness of fit for models computed with the entire data set was better by around 4% compared with models estimated with informative intervals using DS4SID. This represents an acceptable trade-off considering that only one fifth of the data was used for model estimation. Models estimated with intervals from DS4SID, have better performance than when

using intervals retrieved from the approach in [12]. This is explained by the fact that DS4SID has better detection rate and can retrieve more informative intervals which results in better data sets for parameter estimation.

The process with output corresponding to the level of the first reactor tank,  $y_{100}[k]$ , and inputs,  $u_{100}[k]$  and  $u_{220}[k]$ , was used as MISO system. The misdetection rate of DS4SID is smaller than when using [12] which confirms results discussed for the SISO case. The goodness of fit of  $G(q^{-1}, \hat{\Theta}_3)$  is acceptable as it represents a difference of around 7% when compared with other models. Models estimated with informative intervals, retrieved using DS4SID, yield better performance than their counterpart computed with data sequences from the method in [12]. The results reported in Table 6.6 confirm the satisfactory performance of the DS4SID also for the multivariate case. The UINR was slightly larger than values obtained for the SISO case which is explained by the control configuration of this sub-process. The process corresponding to the regulation of the level is slightly longer excited and the informative intervals are thus larger. The goodness of fit for models in this sub-process is smaller than for the output flow rate due to additional challenges in modeling of multivariate processes.

# CHAPTER 7

---

## Conclusions and Outlook

---

### 7.1 Conclusions

Selection of informative data for system identification is a promising alternative when performing new experiments is constrained. As mentioned in [section 1.1](#), data-driven modeling involves three main aspects: data, model and identification method. Tools from statistical signal processing and parameter estimation can be adapted to support data selection for system identification. In the present work, available techniques for data selection were investigated and research opportunities were explored. A modular and flexible data selection method, DS4SID, applicable to multivariable systems and robust to colored noise is presented. Three main aspects were investigated in this research: change detection for dynamic systems, interval bounding and merging of informative data. DS4SID addresses limitations and drawbacks found in available approaches and achieves notable performance in terms of goodness of fit for models estimated with retrieved informative intervals.

The DS4SID consists of two main stages to locate and define informative intervals. Firstly, potentially informative intervals are identified using a detection method and lower interval bounds are set where a change is located. Upper interval bounds are defined by evaluating the resulting information matrix using parameter estimation method. Finally, as mentioned before, a model is obtained by merging the retrieved intervals.

Potentially informative intervals are usually associated to changes in the process due to external excitations. In available approaches for selecting informative data in multivariable systems simple tests such as change in the mean and in the variance are performed separately on output signals for detecting changes from an operating point. This methodology might yield multiple activations of the detection test for the same change due to possible correlations present in multivariable processes. In DS4SID, these changes are located using available detection tests that are extended to cover multivariable dynamical models. It is assumed that multivariable systems addressed in this work can

be decomposed using MISO models. The detection test chooses between two hypothesis: either the system is at steady-state or not. In the hypotheses,  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , the considered models are described using parameter vectors  $\Theta_0$  and  $\Theta_1$ , respectively. In the present work, ARX and ARX-Laguerre models are chosen for local modeling in the vicinity of an operating point.

Available data selection methods require a preliminary step to normalize the input and output signals between 0 and 1 for change detection which can be time-demanding in the case of very large data sets. Moreover, in available approaches signals are processed separately for change detection using time series models even though they are collected from a dynamical system. Thus, additional tests are required in a last stage to confirm the causal relation between the detected change and an excitation in the input. These additional steps are avoided in DS4SID as they are implicitly considered within the GLRT since the hypotheses are based on dynamical models.

For change detection in DS4SID, two windows are used for data processing and computations of the GLRT. The PDF under  $\mathcal{H}_0$  is computed using a growing time window whereas for  $\mathcal{H}_1$  a sliding fixed time window is used. The growing time window is increased at each iteration until a change is located by the detector.

Potentially informative intervals are confirmed or rejected in a second test which involves an identification method. Informative intervals are determined by thresholding a scalar measure retrieved from the information matrix associated to the chosen parameter estimation method. Current methods propose global user-defined thresholds for upper interval bounding that are related to computational precision. Despite the pragmatic aspect of this solution, some limitations arise from use of this criterion. Some intervals can be misdetected if the threshold is set to a value that does not consider the matrix conditioning resulting from small amplitude changes in the input. To overcome this limitation, the gradient of the information matrix conditioning is proposed in DS4SID. This criterion considers local information for interval bounding and misdetection can be reduced when compared with global thresholds.

Different models and parameter estimation methods are presented in [Chapter 3](#). An identification method robust to colored noise was chosen for this second stage of DS4SID. Available works propose use of recursive least squares based on QR factorization of the information matrix as identification method. Even though the least squares method is computationally simple, it yields consistent estimates only in the case of WGN. However, in most real applications, signals are embedded in colored noise. Thus, using the previously mentioned methods will result in imprecise interval bounds and biased parameters. This important

aspect is considered in DS4SID since identification is performed using methods robust to colored noise such as instrumental variables and the generalized least squares. Contrarily to QR-RLS, the ORIV can be computationally faster when  $n \leq n_\eta \leq (n + 2)$  holds where  $n$  is the model order and  $n_\eta$  is the number of instruments.

Two categories of user-defined parameters should be set in DS4SID. The first one is the model order for each hypothesis that are required for computation of the detection test. Thresholds for the detection test and upper interval bounding is the second type of user-defined parameters. General guidelines are given in the literature to choose thresholds but a fine tuning is usually necessary in practice.

In the available literature, the problem of data selection is limited to finding informative data sequences. A further aspect, namely merging of useful intervals is also considered in the present work. Suitable techniques found in the literature were presented in [Chapter 4](#). Some of these methods are adapted and implemented as part of DS4SID. As an additional feature, the proposed method can provide a model computed with the retrieved intervals. A weighted least squares data merging method that can improve model quality is introduced in [section 4.2](#). The weights are obtained from the information matrix conditioning. Information retrieved by DS4SID from interval bounding is used in a post-processing stage for model estimation. All the models are assumed to have the same order for data merging. This can be a considerable drawback as the system dynamics can depend on the operating point which may require models of different orders. To overcome this limitation, an alternative local modeling is proposed. In this approach, the output is constructed from the sum of models estimated from each informative interval. The weights of each model are computed based on the information matrix conditioning obtained for each informative interval. This approach is flexible as each local model can be assigned a different order that can be set empirically. The suitability of this method was demonstrated on two case studies.

The validity of the method was tested by comparing the goodness of fit between models estimated with informative intervals and others computed with the entire data set. Tests carried out in simulation and in real case studies show the validity of the IV method for data merging. Models estimated by DS4SID using less than 25% of the data usually yield a goodness of fit comparable to the models computed with the entire data set. Thus, computational time for system identification can be reduced by considering informative intervals instead of the entire data set.

The designed method is endowed with a flexible and modular methodology as explained in [Chapter 5](#). In DS4SID, data selection is stated as a problem consisting of three main tasks: lower and upper interval bounding and data merging. Different detection and parameter estimation methods are investigated and adapted in DS4SID. Additionally, the problem statement proposed in the present work allows the user to explore other techniques and extend the state of the art.

## 7.2 Outlook

The method developed in the present thesis, DS4SID, provides a reliable solution to the problem of data selection for system identification. This is a promising research field where the following directions should be further explored.

**Process description/Modeling:** multivariable processes are described in this work using MISO parametric models. State space models are well-known mathematical descriptions for multivariable systems. DS4SID can be extended with a bank of models. Estimation methods should be adjusted for computation of some specific models. As general features, parameter estimation should be fast, computationally efficient and recursively implementable. The PEM is an attractive approach that covers such requirements. Batch implementations of selected identification methods are also required for data merging.

Decoupling of multivariate processes can be challenging in highly-interacting systems [77]. In such situations, state space modeling is recommended as it offers a general description of the system without requiring decomposition in MISO sub-processes. Detection methods as the one proposed in the present work can be adapted depending on the model selected for each hypothesis. For an extensive analysis on available subspace identification methods the reader might refer to [20]. Alternatively, filters constructed with OBF can be used for modeling [64]. Transfer functions with those filters have been proven to properly describe systems with unknown input-output delays [57].

**Interval bounding:** a criterion based on the gradient of the reciprocal of the condition number is used in the present work to define upper bounds of informative intervals. Alternative criteria such as interval bounding based on statistical hypothesis testing should be further investigated.

Identification methods implemented in the present work for interval bounding can be applied to systems operating in open or closed-loop. Alternative identification methods other than ORIV should be considered for interval bounding in closed-loop systems. Methods for closed-loop identification based on instrumental variables (IV) may require the controller parameters for model estimation

[80]. Implementation of IV methods for closed-loop systems is presented in [27] where measurement noise is assumed to be generated by an ARMA process.

**Data merging:** different estimation methods for data merging can be implemented using the multi-cost approach presented in [section 4.5](#). For instance, the PEM can be evaluated using different models that can better describe the system dynamics. Alternatively, model computation using optimal IV methods can yield better results as parameter estimation is performed by more sophisticated techniques. Further use of estimated models with informative intervals should be considered. Simulation, prediction, controller design or fault detection are possible tasks that can be performed using the resulting models.

**Determination of weights for local models:** a suitable procedure to ponder the relevance of each local model would be required if different identification methods are used for interval bounding. An additional stage previous to data merging would be necessary to determine weights for each local model using a unifying approach. As an alternative, an evaluation criterion such as the NRMSE can be considered. The goodness-of-fit of each local model can be used to ponder individual contributions related to each interval. Use of a common identification method in the model-based stage would simplify this step since the same discarding criterion is used for the entire data set. Computation of weights will be straightforward as relevance of each interval is compared with the others using the same measure.

**Condition monitoring:** satisfactory models can be obtained by merging data of informative intervals. Intervals that reflect changing conditions are included in the model estimation. Using one of the proposed weighting techniques, current system dynamics can be better explained by data merging. Estimated models can be used to track the system state and to detect possible anomalies. This is a relevant application that can be further explored.

**Case studies:** additional tests will improve discussion about the support offered by data selection method to system identification. As the price of storage devices decreases, more process data can be collected which offers new chances for data selection. Further tests on real processes are highly encouraged as challenges encountered in such systems are different from simulation case studies. Measurement noise is more likely colored and differ from WGN. Moreover, real data can exhibit outliers or errors in the data collection that cannot be simulated. Thus, the presented method DS4SID can be improved and further developed accordingly.





---

## List of Figures

---

1.1	Data selection as a research field related to system identification	3
2.1	Different realizations of WGN	17
2.2	Measured signal (2.38) with additive white noise	22
2.3	Histograms and autocorrelation of example signals	23
2.4	Spectral density of the ARMA process (2.49)	25
3.1	A linear dynamic system with additive disturbance	33
3.2	Laguerre representation	37
5.1	Estimated transfer functions and conditioning of the sample information matrices dependent on the data interval length using the LSM (left) and an optimal IV method (right)	74
5.2	Reciprocal of the condition number for different models for the system (5.1) excited with (5.6) (Solid black line: $n_a = 3$ . Solid gray line: $n_k = 2$ )	75
5.3	Simplified block diagram of the DS4SID	77
5.4	Data selection using DS4SID	77
5.5	Closed-loop response of the system (5.1)	78
5.6	Upper interval bounding using the ORIV method	84
5.7	Reciprocal of the condition number of the information matrix for the model in (5.4) using a ramp signal for excitation. Solid black line: correct model order. Solid gray line: $n_a = 3$ . Dashed black line: $n_k = 2$	85
5.8	Informative intervals selected by DS4SID	86
5.9	Data windowing using a sliding fixed-size window $M_1$ and a growing time window $M_0$ for change detection in DS4SID. Sample size just for illustrative purposes. (Please refer to Figure 5.10 for definition of data sets used for computation of $g[k]$ )	89
5.10	Flow diagram of DS4SID	90
6.1	Step response of system in (6.4) with transfer functions in (6.5)	100
6.2	Block diagram of the non-interacting control system	101

6.3	Closed-loop step response of system (6.5) with controllers (6.14)	103
6.4	Selected intervals for the process (6.5) subject to SNR = 25 dB (left), SNR = 20 dB (middle) and SNR = 15 dB (right)	105
6.5	Selected intervals for $\{y_1[k]\}$ and $\{y_2[k]\}$ from a test data set (SNR = 15 dB). Top: selected intervals using DS4SID. Middle: Detection test for locating lower interval bounds. Bottom: Conditioning of the information matrix for determining upper interval bounds. Continuous line: DS4SID using IVM. Dashed line: method as presented in [12] using QR-RLS.	108
6.6	Bode diagrams of models computed with informative intervals and true process. Dashed line: model estimated using intervals from DS4SID ( $G(q^{-1}, \hat{\Theta}_3)$ ). Dashed-dotted line: model estimated using intervals from [12] ( $G(q^{-1}, \hat{\Theta}_2)$ ). Continuous line: true process ( $G(q^{-1}, \hat{\Theta}_0)$ ).	110
6.7	Process unit II used for case study	112
6.8	Simplified P&ID of the process unit II	113
6.9	Control scheme of the process unit II. Dotted: SISO process. Dashed: MISO process.	115
6.10	Autocorrelation of residuals for model (6.23)	117
6.11	Production rate and control voltage of the pump $P_{320}$ for the first experiment (see Table 6.5)	119
6.12	Output and input signals for the first experiment (see Table 6.5)	119
A.1	Experiments 1-3 (see Table 6.4) from left to right	136
A.2	Experiments 4-6 (see Table 6.4) from left to right	137
A.3	Experiment 7 (see Table 6.4) from left to right	138
B.1	Analysis of model residuals for the output flow rate, operating points 1-3 (see Figure A.2 top-left)	141
B.2	Analysis of model residuals for the output flow rate, operating points 4-6 (see Figure A.2 top-left)	142
B.3	Analysis of model residuals for the output flow rate, operating points 7-9 (see Figure A.2 top-left)	143
B.4	Analysis of model residuals for the output flow rate, operating points 10-12 (see Figure A.2 top-left)	144

---

## List of Tables

---

2.1	Normality tests on estimated stochastic sequences in (2.38), $\alpha = 0.05$ . Left: S-W test. Right: K-S test . . . . .	22
3.1	Common black-box model structures . . . . .	35
5.1	Comparison of DS4SID with available approaches . . . . .	96
6.1	External reference signals used to generate test data sets . . . . .	104
6.2	DS4SID for $\{y_1[k]\}$ and $\{y_2[k]\}$ ( $\hat{\Theta}_1$ : parameter vector using the entire data set; $\hat{\Theta}_2$ : estimates using intervals retrieved from [12] and $\hat{\Theta}_3$ : parameter vector using intervals from DS4SID). Test data sets as in Figure 6.4. . . . .	111
6.3	Level and flow controllers used in the process unit II . . . . .	116
6.4	Summary of experiments in the process unit II . . . . .	120
6.5	Subprocess: output flow rate. Goodness of fit (6.2) for selected intervals using DS4SID ( $G(q^{-1}, \hat{\Theta}_3)$ ), using [12] ( $G(q^{-1}, \hat{\Theta}_2)$ ) and for entire data set ( $G(q^{-1}, \hat{\Theta}_1)$ ). Bold font: best results for test experiments. . . . .	121
6.6	Subprocess: level in the first reactor. Goodness of fit (6.2) for selected intervals using DS4SID ( $G(q^{-1}, \hat{\Theta}_3)$ ), using [12] ( $G(q^{-1}, \hat{\Theta}_2)$ ) and for entire data set ( $G(q^{-1}, \hat{\Theta}_1)$ ). Bold font: results for test experiments. . . . .	122
B.1	Normality tests for $\{\epsilon[k]\}$ (cf. (B.2)) for the output rate of the fourth experiment: Shapiro-Wilk (S-W) test: $\alpha = 0.05$ . . . . .	140
B.2	Normality tests for the level of the first reactor in the fourth experiment. Shapiro-Wilk (S-W) test: $\alpha = 0.05$ . . . . .	145



# A Experiments in the Process Unit II

In this appendix, measurements collected during the experiments in the process unit II and discussed in [section 6.3.4](#) are shown. The main features of these experiments are listed in [Table 6.4](#). In order to evaluate the performance of DS4SID using real data, a SISO and a MISO process of the process unit II were used for analysis. The experiments were performed while operating the process unit II in closed-loop as described in [section 6.3.1](#). The detection test of DS4SID was evaluated for small and large changes between operating points. Small changes in the set point are shown in [Figure A.1](#) for the first experiment between the second and third operating point. Another example of small changes in the process is shown in [Figure A.2](#) for the sixth experiment between the third and fourth operating point.

The top rows of [Figure A.1](#) to [Figure A.3](#) show the signals  $y_{320}[k]$ ,  $u_{320}[k]$  as well as  $w_{320}[k]$  that represent the output, input and set point for the production rate, respectively. The output signal  $y_{100}[k]$  and the two input signals  $u_{100}[k]$  and  $u_{220}[k]$  for the multivariable process are shown in the bottom row of [Figure A.1](#) to [Figure A.3](#).

The set points  $w_{320}[k]$  and  $w_{100}[k]$  were firstly generated in a simulation program and further implemented in the control system of the process unit II that was operated mainly automatically. This recreates situations found in continuously-operated plants. Experiments 4 and 5 were used for data selection using DS4SID and an approach based on [\[12\]](#) was also adapted for tests and comparison. Models were estimated with the selected intervals of these experiments and the NRMSE using these data was computed. The models were cross-validated with the remaining data sets.

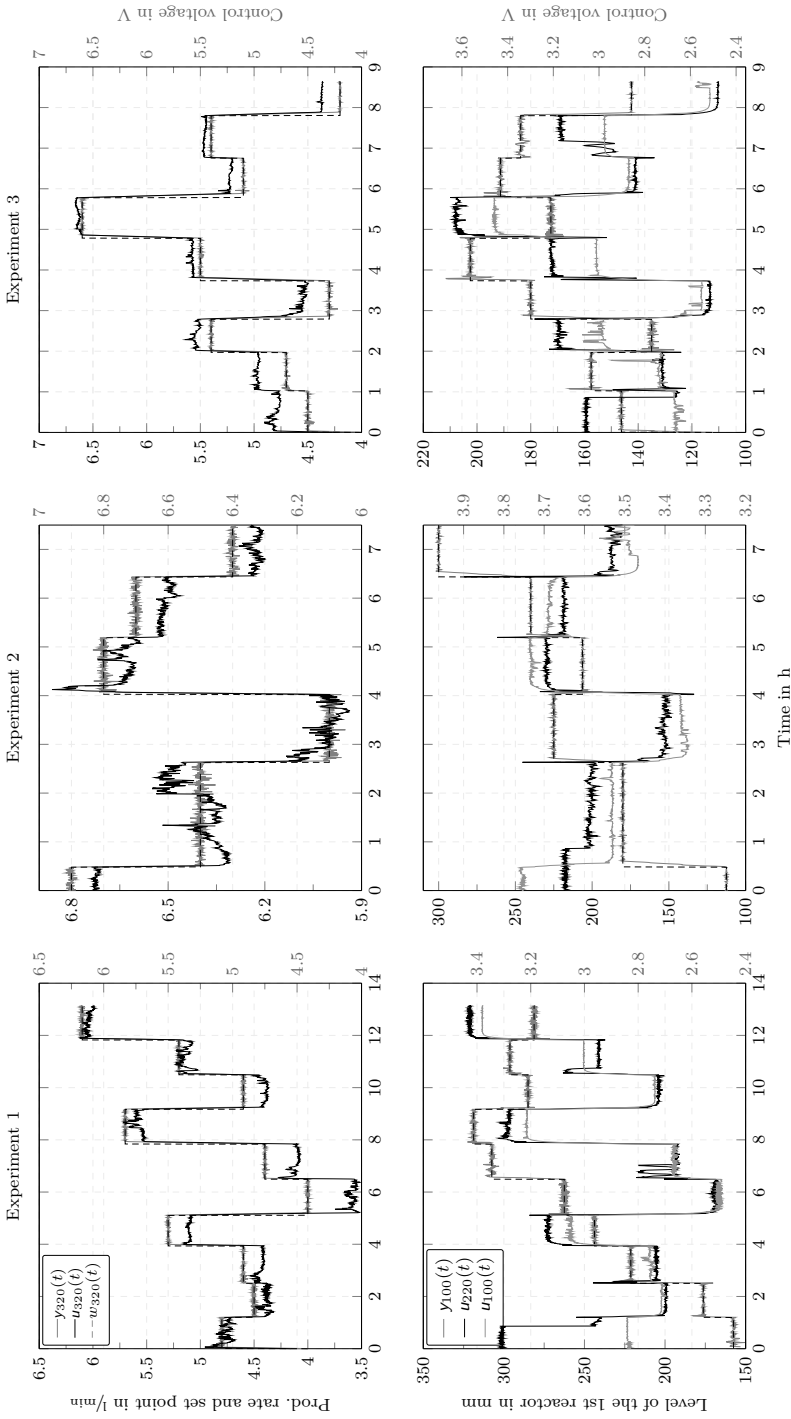
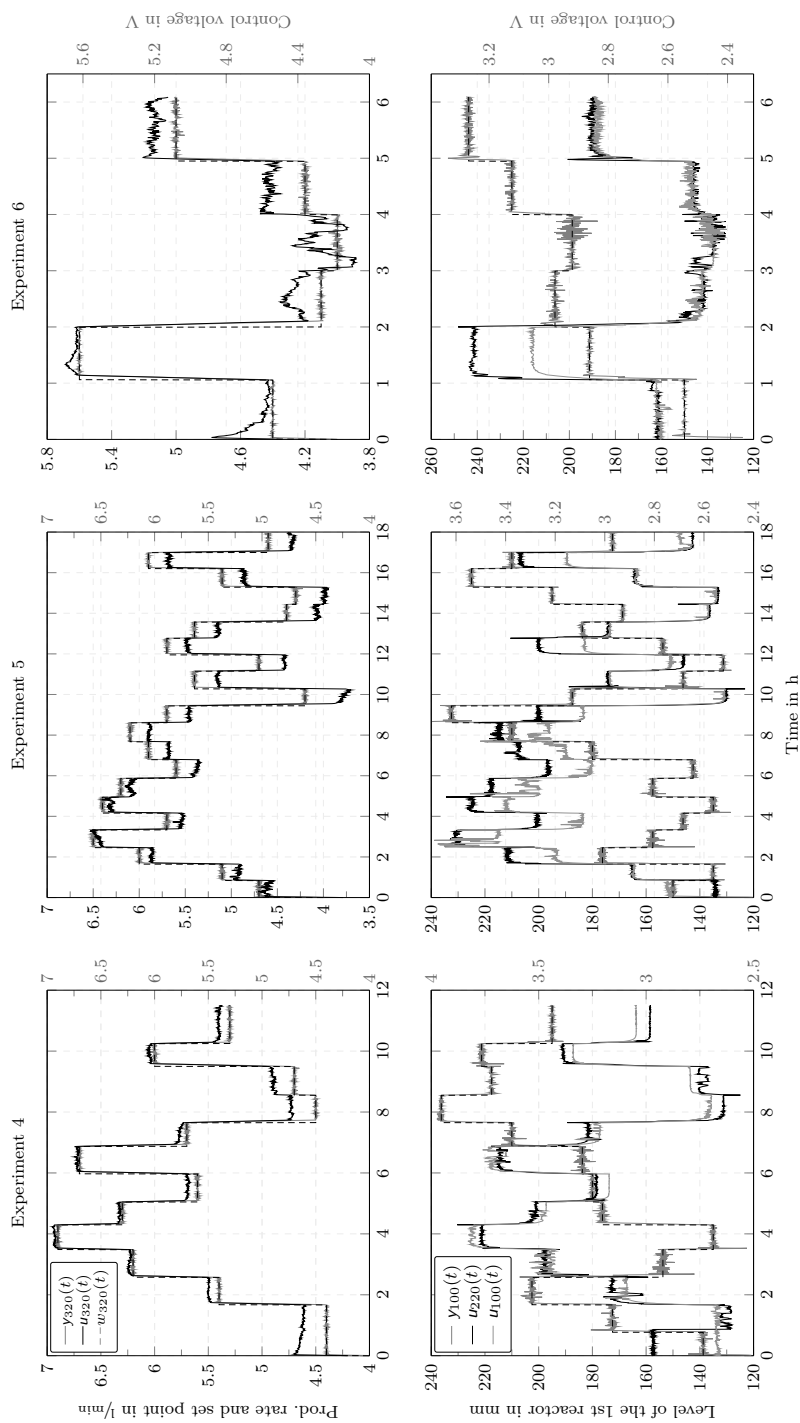
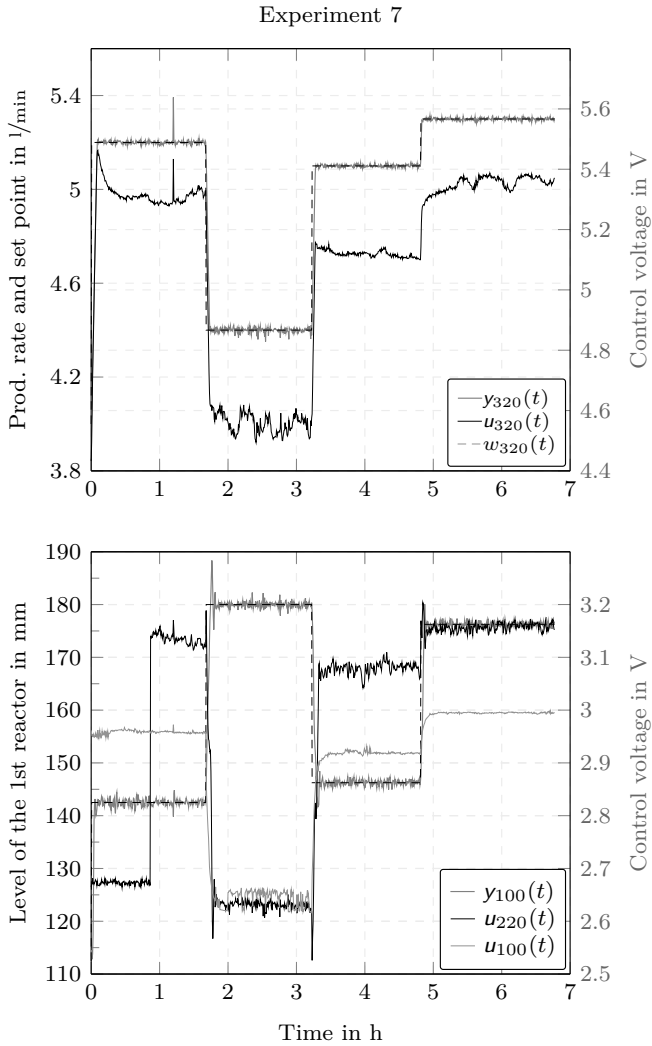


Figure A.1: Experiments 1-3 (see Table 6.4) from left to right



**Figure A.2:** Experiments 4-6 (see Table 6.4) from left to right





**Figure A.3:** Experiment 7 (see Table 6.4) from left to right

## B Evaluation of Model Residuals

In this chapter, model residuals for the two processes considered in the process unit II are evaluated. The model residuals are computed as the difference between the output sequence and the one-step-ahead prediction of the model. The resulting residuals are filtered according to the assumed model and the resulting sequence is evaluated with a normality test. First, results are discussed for the output rate. Then, the model residuals of the level in the first reactor are analyzed.

### B.1 Output Flow Rate

In this section, model residuals for the output flow rate are analyzed. The output signal is denoted by  $y_{320}[k]$  whereas the input signal is the pump control voltage  $u_{320}[k]$  and the process sampling time is  $T_s = 250$  ms. The fourth experiment, shown in the top-left of [Figure A.1](#) was used for evaluation of model residuals. The process is operated between 12 different operating points. An ARMAX model is used for process description with the following model order:

$$\mathbf{n} = [n_a \ n_b \ n_c \ n_k] = [2 \ 1 \ 3 \ 4] \quad (\text{B.1})$$

Model parameters were estimated using the PEM. The model residuals are computed by:

$$\boldsymbol{\varepsilon}[k] = y[k] - \hat{y}[k] \quad (\text{B.2})$$

where  $y[k]$  and  $\hat{y}[k]$  are the observations and the one-step ahead prediction, respectively.

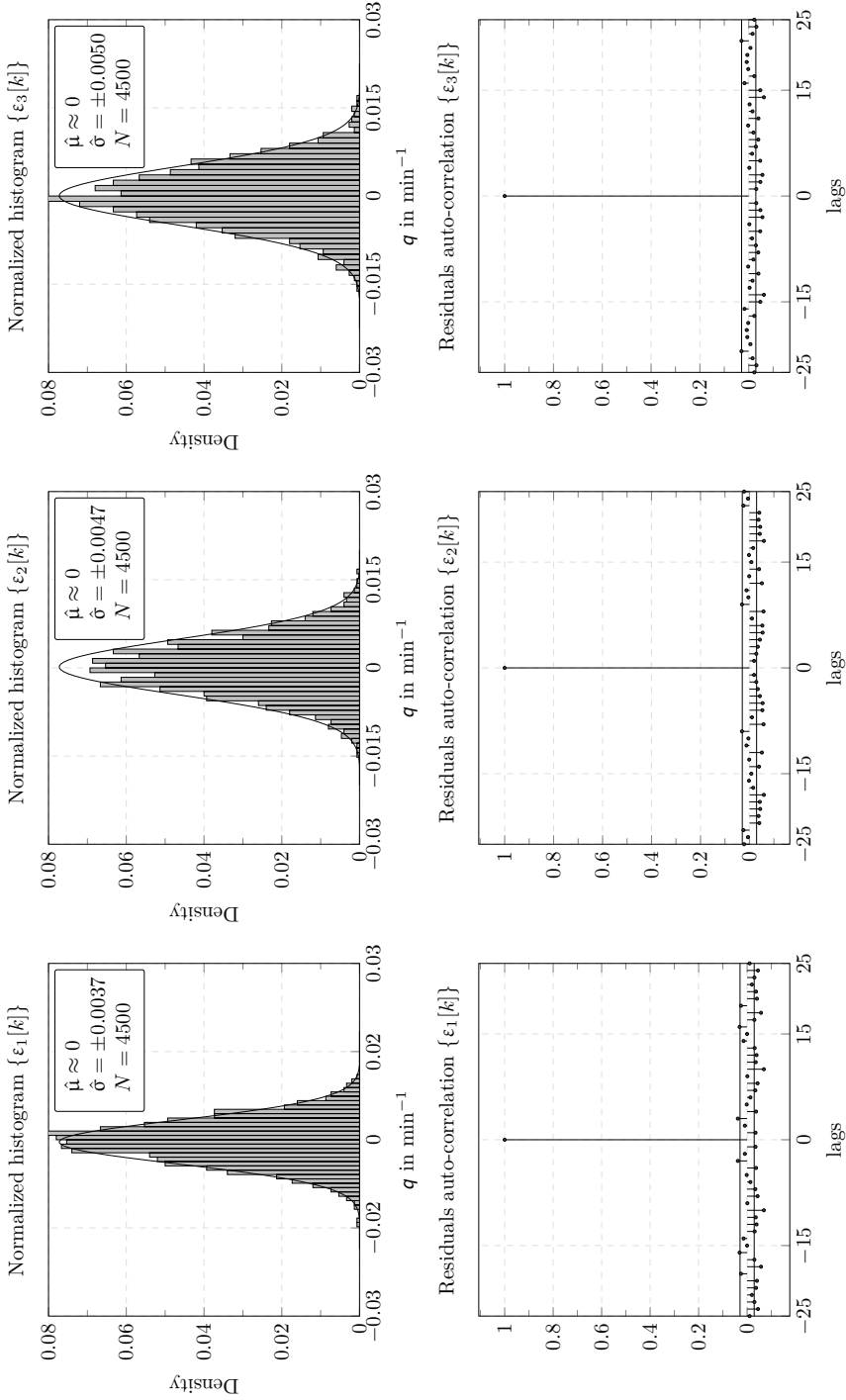
The residuals  $\{\boldsymbol{\varepsilon}[k]\}$  are estimated by the identification method and analyzed using normality tests. As null-hypothesis,  $\mathcal{H}_0$ , it is assumed that  $\{\boldsymbol{\varepsilon}[k]\}$  is considered to be WGN. In the alternative-hypothesis,  $\mathcal{H}_1$ , the sequence  $\{\boldsymbol{\varepsilon}[k]\}$  is considered to be colored noise.

Results of normality tests for the fourth experiment are shown in [Table B.1](#). In most of the operating points, the ARMAX model describes the process well. The operating points, where the null-hypothesis is accepted, are shaded. This indicates that the resulting sequence  $\{\boldsymbol{\varepsilon}[k]\}$  is WGN. Thus, the ARMAX model is suitable to describe this process. In order to confirm these results,

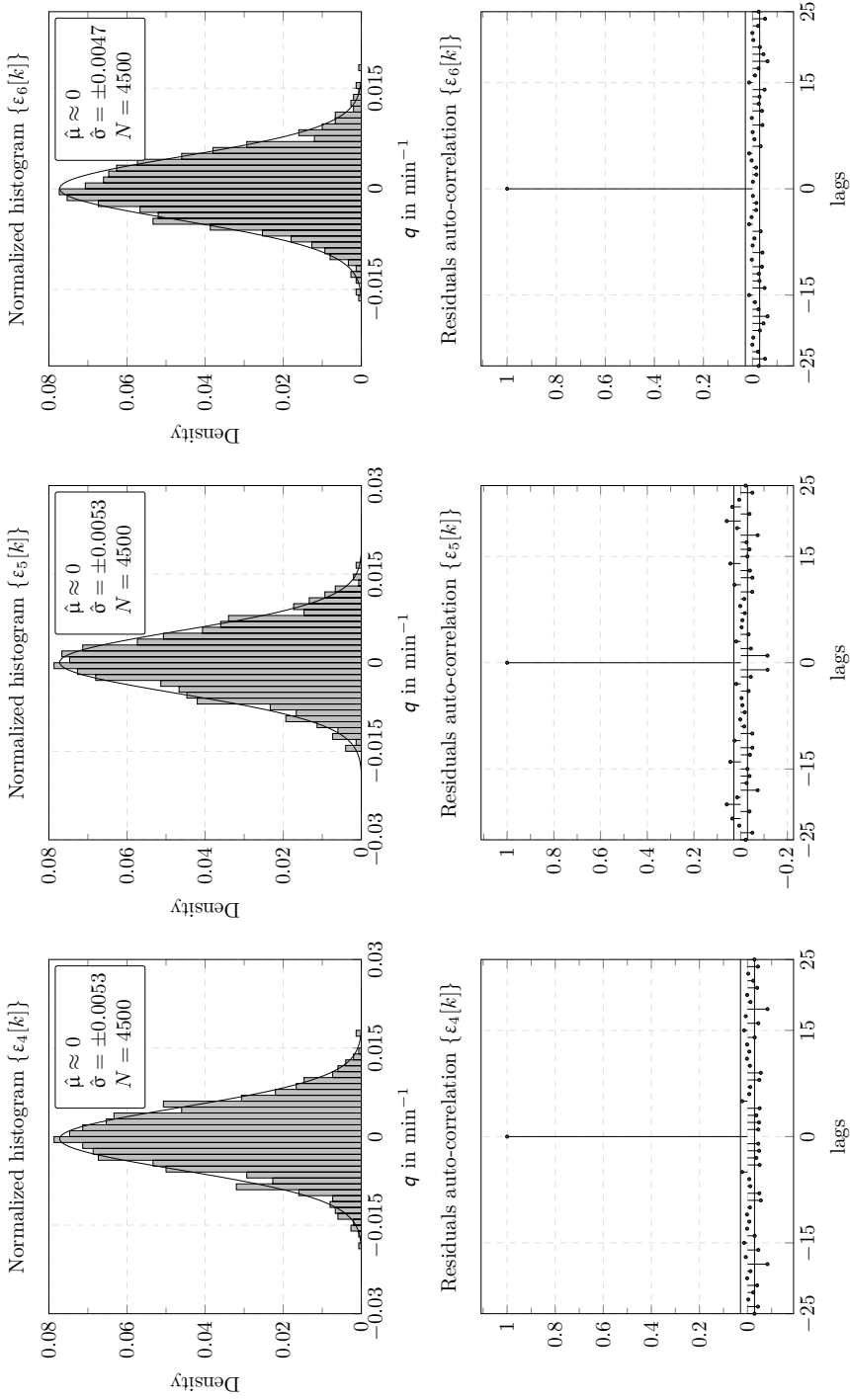
ARX models were also used for modeling. Despite the fact that the model quality was satisfactory, the null-hypothesis was always rejected which indicates that the residuals do not fulfill the assumptions of the ARX model. Histograms and correlations of the computed sequence  $\{\epsilon[k]\}$  are shown in [Figure B.1](#) to [Figure B.4](#).

**Table B.1:** Normality tests for  $\{\epsilon[k]\}$  (cf. (B.2)) for the output rate of the fourth experiment: Shapiro-Wilk (S-W) test:  $\alpha = 0.05$

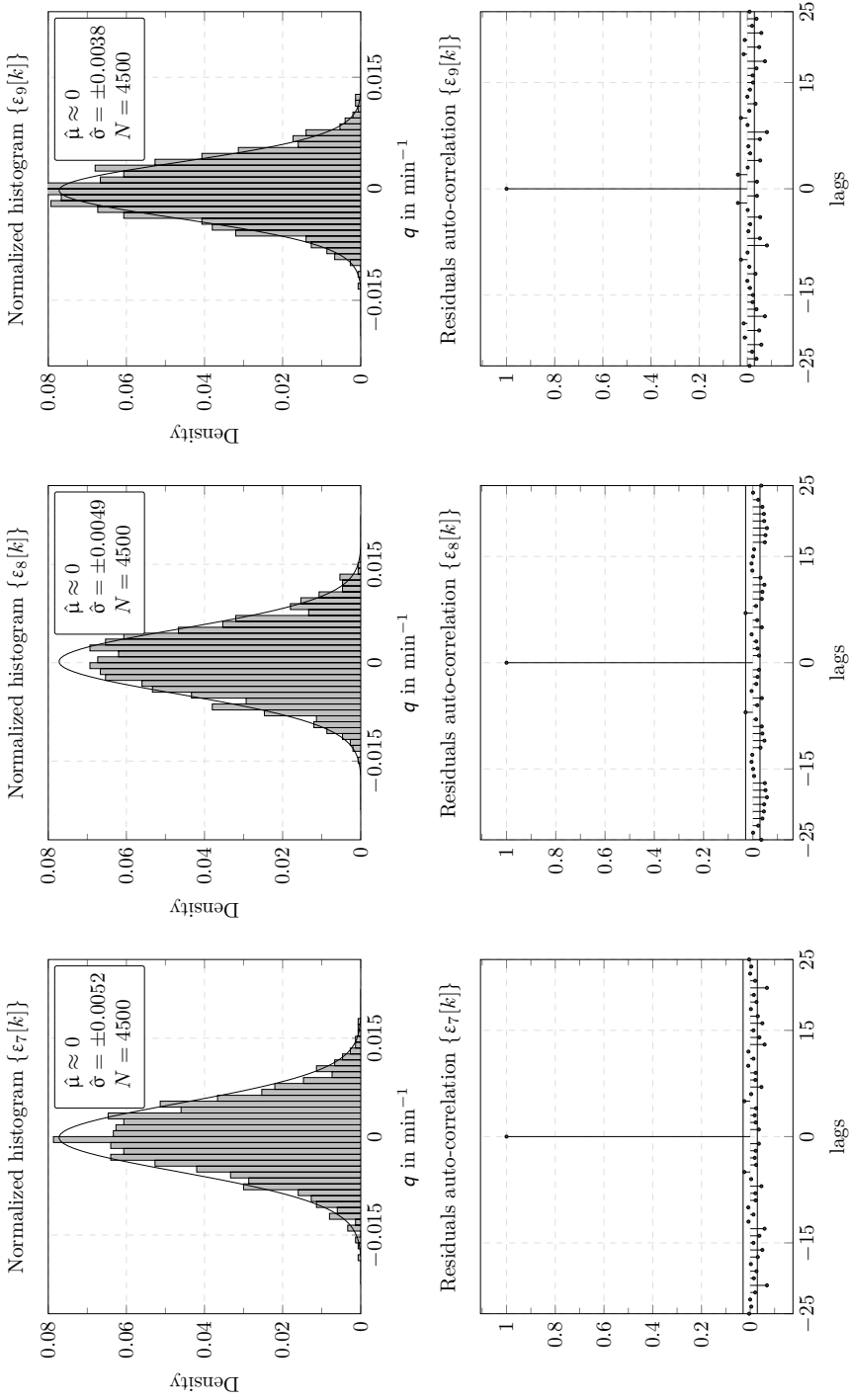
Op. point	$w_{320}$	$\mathcal{H}$	pValue	$\hat{\chi}^2$	dof	$\hat{\sigma}_y$	$\hat{\sigma}_\epsilon$	Meas. error in %	SNR in dB
1	4.4	0	0.1957	8.6259	6	0.0074	0.0037	4.8424	3.1563
2	5.4	1	0.1328	7.0591	4	0.0082	0.0047	8.0690	3.6485
3	6.2	0	0.0020	16.9209	4	0.0075	0.0050	2.1120	3.5857
4	6.9	1	0.0359	13.4889	6	0.0079	0.0053	1.5595	3.5491
5	6.3	0	0.1718	10.3071	7	0.0077	0.0053	1.4086	3.1887
6	5.6	0	0.4019	6.1928	6	0.0076	0.0047	1.9249	4.2071
7	6.7	1	0.0178	15.3387	6	0.0082	0.0052	3.1418	3.9967
8	5.7	0	0.6960	3.0259	5	0.0077	0.0049	3.1461	3.8620
9	4.5	0	0.2198	8.2578	6	0.0061	0.0038	5.1083	4.1483
10	4.7	0	0.4054	7.2288	7	0.0072	0.0041	4.2106	4.7966
11	6.0	0	0.0994	12.0353	7	0.0088	0.0052	4.4777	4.5250
12	5.3	0	0.2661	8.8163	7	0.0075	0.0046	2.0130	4.1725



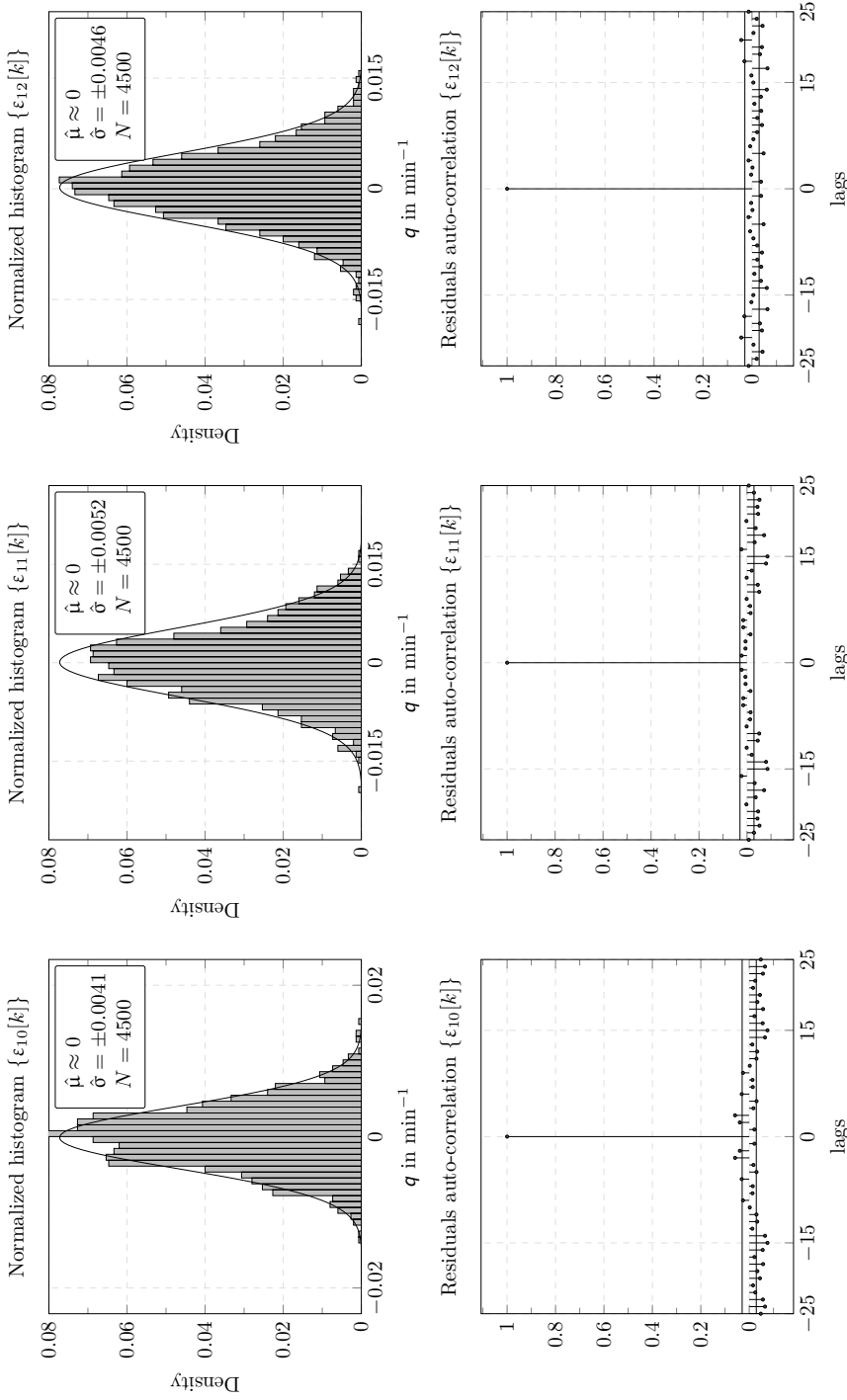
**Figure B.1:** Analysis of model residuals for the output flow rate, operating points 1-3 (see Figure A.2 top-left)



**Figure B.2:** Analysis of model residuals for the output flow rate, operating points 4-6 (see [Figure A.2](#) top-left)



**Figure B.3:** Analysis of model residuals for the output flow rate, operating points 7-9 (see Figure A.2 top-left)



**Figure B.4:** Analysis of model residuals for the output flow rate, operating points 10-12 (see [Figure A.2](#) top-left)

## B.2 Level of the first Reactor

The process is operated in closed-loop using PI controllers that regulate the reactor level. The controlled variable is the reactor level,  $y_{100}(t)$ . The reactor level is operated at the same operating point and it is kept on that set value between 50 to 60 min, approximately. The level is measured using a hydrostatic sensor located at the bottom part of the reactor. The control variables, are  $u_{100}(t)$  and  $u_{220}(t)$ . Similar to the analysis presented in [section B.1](#), the model residuals for the level in the first reactor were evaluated. The level of the reactor is kept at the same value during long periods in order to evaluate characteristics of the measurement noise.

The model residuals are computed by:

$$\epsilon[k] = y[k] - \hat{y}[k], \quad k = 0, 1, \dots, N - 1 \quad (\text{B.3})$$

The residuals  $\epsilon[k]$  were analyzed following a similar procedure as presented in [section B.1](#) and normality tests were applied. Results of the normality tests for the fourth experiment (bottom-left of [Figure A.1](#)) are shown in [Table B.2](#).

**Table B.2:** Normality tests for the level of the first reactor in the fourth experiment. Shapiro-Wilk (S-W) test:  $\alpha = 0.05$

Op. point	$w_{100}$	$\mathcal{H}$	pValue	$\hat{\chi}^2$	dof	$\hat{\sigma}_y$	$\hat{\sigma}_\epsilon$	Meas. error in %	SNR in dB
1	140	0	0.2001	7.3052	6	0.0120	0.0053	5.0120	7.0981
2	172	1	0.1520	6.8251	4	0.0143	0.0062	7.8526	7.2589
3	202	0	0.0014	12.1456	4	0.0134	0.0080	2.2598	4.4803
4	154	1	0.0219	13.9635	6	0.0125	0.0054	1.4926	7.2903
5	135	0	0.2015	9.4895	7	0.0102	0.0053	1.5792	5.6865
6	176	0	0.3584	6.2416	6	0.0148	0.0065	1.7492	7.1470
7	180	1	0.0187	15.4251	6	0.0158	0.0058	3.4852	8.7046
8	185	0	0.4218	3.1423	5	0.0134	0.0064	3.1416	6.4185
9	210	0	0.2036	8.7496	6	0.0192	0.0073	4.9587	8.3996
10	236	0	0.1428	6.5078	7	0.0136	0.0083	4.3216	4.2892
11	215	0	0.1796	12.0125	4	0.0154	0.0087	3.8753	4.9600
12	220	0	0.2248	8.9675	6	0.0110	0.0071	6.4521	3.8027





## Bibliography

- [1] D. Arengas and A. Kroll: ‘A search method for selecting informative data in predominantly stationary historical records for multivariable system identification’. *21st International Conference on System Theory, Control and Computing (ICSTCC)*. IEEE. Sinaia, Romania, Oct. 2017: pp. 100–105.
- [2] D. Arengas and A. Kroll: ‘Searching for informative intervals in predominantly stationary data records to support system identification’. *XXVI International Conference on Information, Communication and Automation Technologies (ICAT)*. IEEE. Sarajevo, Bosnia & Herzegovina, Oct. 2017: pp. 1–6.
- [3] D. Arengas and A. Kroll: ‘Removal of Insufficiently Informative Data to Support System Identification in MISO Processes’. *European Control Conference (ECC)*. IEEE. Limassol, Cyprus, June 2018: pp. 2842–2847.
- [4] D. Arengas and A. Kroll: ‘A Data Selection Method for large Databases for System Identification of MISO Models Based on Recursive Instrumental Variables’. *European Control Conference (ECC)*. IFAC. Naples, Italy, June 2019: pp. 357–362.
- [5] D. Arengas and A. Kroll: ‘Data selection for system identification (DS4SID) from logged process records of continuously operated plants’. *at – Automatisierungstechnik* (2020), vol. 68(5): pp. 347–359.
- [6] K. J. Åström and B. Wittenmark: *Adaptive control*. Courier Corporation, 2013.
- [7] K. J. Åström and P. Eykhoff: ‘System identification a survey’. *Automatica* (1971), vol. 7(2): pp. 123–162.
- [8] M. Atzmueller, B. Klöpper, H. Al Mawla, B. Jäschke, M. Hollender, M. Graube, D. Arnu, A. Schmidt, S. Heinze, L. Schorer, A. Kroll, G. Stumme, and L. Urbas: ‘Big data analytics for proactive industrial decision support’. *atp magazin* (2016), vol. 58(09): pp. 62–74.
- [9] M. Basseville and A. Benveniste: ‘Sequential detection of abrupt changes in spectral characteristics of digital signals’. *IEEE Transactions on Information Theory* (1983), vol. 29(5): pp. 709–724.

- [10] M. Basseville, I. V. Nikiforov, et al.: *Detection of abrupt changes: theory and application*. Vol. 104. prentice Hall Englewood Cliffs, 1993.
- [11] D. P. Bertsekas and J. N. Tsitsiklis: *Introduction to probability*. Athena Scientinis, 2000.
- [12] A. C. Bittencourt, A. J. Isaksson, D. Peretzki, and K. Forsman: ‘An Algorithm for Finding Process Identification Intervals from Normal Operating Data’. *Processes* (2015), vol. 3(3): pp. 357–383.
- [13] C. Bohn and H. Unbehauen: *Identifikation dynamischer Systeme*. Springer, 2013.
- [14] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung: *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [15] S. Cao and R. R. Rhinehart: ‘An Efficient Method for On-line Identification of Steady State’. *Journal of Process Control* (1995), vol. 5(6): pp. 363–374.
- [16] S. Cao and R. R. Rhinehart: ‘Critical values for a steady-state identifier’. *Journal of Process Control* (1997), vol. 7(2): pp. 149–152.
- [17] P. Carrette, G. Bastin, Y. Y. Genin, and M. Gevers: ‘Discarding Data May Help in System Identification’. *IEEE Transactions on Signal Processing* (1996), vol. 44(9): pp. 2300–2310.
- [18] C.-T. Chen: *Analog and digital control system design: transfer-function, state-space, and algebraic methods*. Oxford University Press, Inc., 1995.
- [19] C.-T. Chen: *Linear system theory and design*. Oxford University Press, Inc., 1998.
- [20] A. Chiuso and G. Picci: ‘Consistency analysis of some closed-loop subspace identification methods’. *Automatica* (2005), vol. 41(3): pp. 377–391.
- [21] A. Chiuso and G. Picci: ‘Prediction error vs subspace methods in closed loop identification’. *IFAC Proceedings Volumes* (2005), vol. 38(1): pp. 506–511.
- [22] W. El Falou, M. Khalil, J. Duchêne, and D. Hewson: ‘Automatic threshold determination for a local approach of change detection in long-term signal recordings’. *EURASIP journal on Advances in Signal Processing* (2007), vol. 2007: pp. 1–7.
- [23] B. Friedlander: ‘The Overdetermined Recursive Instrumental Variable Method’. *IEEE Transactions on Automatic Control* (1984), vol. AC-29(4): pp. 353–356.

- [24] M. Gevers, A. S. Bazanella, X. Bombois, and L. Mišković: ‘Identification and the Information Matrix: How to Get Just Sufficiently Rich?’ *IEEE Transactions on Automatic Control* (2009), vol. 54(12): pp. 2828–2840.
- [25] M. Gevers, L. Mišković, D. Bonvin, and A. Karimi: ‘Identification of a two-input system: Variance analysis’. *IFAC Proceedings Volumes* (2005), vol. 38(1): pp. 674–679.
- [26] M. Gevers, L. Mišković, D. Bonvin, and A. Karimi: ‘Identification of multi-input systems: variance analysis and input design issues’. *Automatica* (2006), vol. 42(4): pp. 559–572.
- [27] M. Gilson, H. Garnier, P. C. Young, and P. M. Van den Hof: ‘Refined instrumental variable methods for closed-loop system identification’. *Proceedings of the 15th IFAC Symposium on System Identification*. 2009: pp. 284–289.
- [28] G. C. Goodwin and K. S. Sin: *Adaptive filtering prediction and control*. Courier Corporation, 2014.
- [29] G. C. Goodwin and R. L. Payne: *Dynamic system identification: experiment design and data analysis*. Academic press, 1977.
- [30] P. Granjon: *The CUSUM algorithm a small review*. Available from: <https://hal.archives-ouvertes.fr/hal-00914697/document>. 2014.
- [31] C. Grillenzoni: ‘Testing for causality in real time’. *Journal of econometrics* (1996), vol. 73(2): pp. 355–376.
- [32] F. Gustafsson: *Adaptive Filtering and Change Detection*. John Wiley & Sons, 2006.
- [33] J. D. Hamilton: *Time series analysis*. Vol. 2. Princeton university press Princeton, NJ, 1994.
- [34] C. W. Helstrom: *Elements of signal detection and estimation*. Prentice-Hall, Inc., 1994.
- [35] J. P. Hespanha: *Linear systems theory*. Princeton university press, 2018.
- [36] P. S. Heuberger, P. M. V. den Hof, and B. Wahlberg: *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer, 2005.
- [37] P. S. C. Heuberger: ‘On approximate system identification with system-based orthonormal functions’. Ph.D. Thesis. Delft, The Netherlands: Delft University of Technology, 1992.
- [38] A. Horch: ‘Condition Monitoring of Control Loops’. Ph.D. Thesis. Stockholm: Royal Institute of Technology, 2000.

- [39] A. J. Isaksson: ‘Some aspects of industrial system identification’. *10th IFAC Symposium on Dynamics and Control of Process Systems, DYCOPS 2013; Mumbai; India*. 2013: pp. 153–159.
- [40] R. Isermann: ‘Practical aspects of process identification’. *System Identification*. Elsevier, 1981: pp. 575–587.
- [41] R. Isermann and M. Münchhof: *Identification of dynamic systems: an introduction with applications*. Springer Science & Business Media, 2010.
- [42] A. J. Jakeman, L. P. Steele, and P. C. Young: ‘Instrumental variable algorithms for multiple input systems described by multiple transfer functions’. *IEEE transactions on systems, man, and cybernetics* (1980), vol. 10(10): pp. 593–602.
- [43] A. Jakeman and P. Young: ‘Refined instrumental variable methods of recursive time-series analysis Part II. Multivariable systems’. *International Journal of Control* (1979), vol. 29(4): pp. 621–644.
- [44] H. Jansson and H. Hjalmarsson: ‘Optimal experiment design in closed loop’. *IFAC Proceedings Volumes* (2005), vol. 38(1): pp. 488–493.
- [45] F. K. Jondral and A. Wiesler: *Wahrscheinlichkeitsrechnung und stochastische Prozesse: Grundlagen für Ingenieure und Naturwissenschaftler*. Springer-Verlag, 2002.
- [46] T. Katayama: *Subspace Methods for System Identification*. Springer, 2005.
- [47] S. Kay: *Intuitive probability and random processes using MATLAB®*. Springer Science & Business Media, 2006.
- [48] S. M. Kay: *Modern spectral estimation*. Pearson Education India, 1988.
- [49] S. M. Kay: *Estimation theory, vol. I of Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.
- [50] S. M. Kay: *Detection theory*. Prentice Hall PTR, 1998.
- [51] K. J. Keesman: *System identification: an introduction*. Springer Science & Business Media, 2011.
- [52] J.-P. Kreiss and G. Neuhaus: *Einführung in die Zeitreihenanalyse*. Springer-Verlag, 2006.
- [53] A. Kroll, A. Dürrbaum, D. Arengas, B. Jäschke, H. Al-Mawla, and A. Geiger: ‘ $\mu$ Plant: Model factory for the automatization of networked, heterogeneous and flexibly changeable multi-product plants’. *Proceedings of Automation 2016*. VDI-Berichte 2284. Baden-Baden, 2016.

- [54] D. J. Leith, D. . Murray-Smith, and R. Bradley: ‘Combination of Data Sets for System Identification’. *IEE Proceedings-D (Control Theory and Applications)*. Vol. 140. 1. IET. 1993: pp. 11–18.
- [55] J. Li, R. Shi, C. Xu, and S. Wang: ‘Process identification of the SCR system of coal-fired power plant for de-NO<sub>x</sub> based on historical operation data’. *Environmental technology* (2018), vol.
- [56] P. Li, H.-L. Wei, S. A. Billings, M. A. Balikhin, and R. Boynton: ‘Nonlinear model identification from multiple data sets using an orthogonal forward search algorithm’. *Journal of Computational and Nonlinear Dynamics* (2013), vol. 8(4).
- [57] L. Ljung: *System Identification: Theory for the User*. Prentice Hall information and system sciences series. Prentice Hall PTR, 1999.
- [58] L. Ljung: ‘Perspectives on system identification’. *Annual Reviews in Control* (2010), vol. 34(1): pp. 1–12.
- [59] L. Ljung and T. Söderström: *Theory and practice of recursive identification*. MIT press, 1983.
- [60] H. Madsen: *Time series analysis*. Chapman and Hall/CRC, 2007.
- [61] R. H. Middleton and G. C. Goodwin: *Digital control and estimation: a unified approach*. Prentice Hall Professional Technical Reference, 1990.
- [62] S. R. Mounce, R. B. Mounce, and J. B. Boxall: ‘Novelty detection for time series data analysis in water distribution systems using support vector machines’. *Journal of hydroinformatics* (2011), vol. 13(4): pp. 672–686.
- [63] T.-S. Ng, G. C. Goodwin, and B. D. Anderson: ‘Identifiability of MIMO linear dynamic systems operating in closed loop’. *Automatica* (1977), vol. 13(5): pp. 477–485.
- [64] B. Ninness, J.-C. Gomez, and S. Weller: ‘MIMO system identification using orthonormal basis functions’. *Proceedings of 1995 34th IEEE Conference on Decision and Control*. Vol. 1. IEEE. 1995: pp. 703–708.
- [65] K. Ogata: *Discrete-time control systems*. Prentice-Hall, Inc., 1995.
- [66] J. G. D. Oliveira and C. Garcia: ‘Algorithm-Aided Identification Using Historic Process Data’. *XIII Simpósio Brasileiro de Automação Inteligente*. Porto Alegre – RS, Brazil, Oct. 2017: pp. 1235–1240.
- [67] A. V. Oppenheim and R. W. Schaffer: *Discrete-time signal processing*. Pearson Education, 2014.
- [68] A. Oppenheim, A. Willsky, and S. Nawab: *Signals & Systems*. Prentice-Hall signal processing series. Prentice-Hall International, 1997.

- [69] R. K. Pearson: ‘Outliers in process modeling and identification’. *IEEE Transactions on control systems technology* (2002), vol. 10(1): pp. 55–63.
- [70] D. Peretzki, A. J. Isaksson, A. C. Bittencourt, and K. Forsman: ‘Data Mining of Historic Data for Process Identification’. *Proceedings of the 2011 AIChE Annual Meeting*. Minneapolis, MN, USA, 2011.
- [71] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko: ‘A review of novelty detection’. *Signal Processing* (2014), vol. 99: pp. 215–249.
- [72] D. E. Seborg, D. A. Mellichamp, T. F. Edgar, and F. J. Doyle III: *Process dynamics and control*. John Wiley & Sons, 2010.
- [73] Y. A. W. Shardt and H. Biao: ‘Data quality assessment of routine operating data for process identification’. *Computers and Chemical Engineering* (2013), vol. 55: pp. 19–27.
- [74] Y. A. W. Shardt and L. S. Sirish: ‘Segmentation Methods for Model Identification from Historical Process Data’. *Proceedings of the 14th IFAC World Congress*. Cape Town, South Africa, Aug. 2014: pp. 2836–2841.
- [75] Y. A. Shardt and K. Brooks: ‘Automated system identification in mineral processing industries: a case study using the zinc flotation cell’. *IFAC-PapersOnLine* (2018), vol. 51(18): pp. 132–137.
- [76] Y. A. Shardt and B. Huang: ‘Closed-loop identification condition for ARMAX models using routine operating data’. *Automatica* (2011), vol. 47(7): pp. 1534–1537.
- [77] S. Skogestad and I. Postlethwaite: *Multivariable feedback control: analysis and design*. Vol. 2. Wiley New York, 2007.
- [78] T. Söderström: *Discrete-time stochastic systems: estimation and control*. Springer Science & Business Media, 2012.
- [79] T. Söderström and P. Stoica: ‘Comparison of some instrumental variable methods: consistency and accuracy aspects’. *Automatica* (1981), vol. 17(1): pp. 101–115.
- [80] T. Söderström and P. Stoica: *System Identification*. Prentice Hall International, 1989.
- [81] A. Stenman: ‘Model On Demand: Algorithm, Analysis and Applications’. Ph.D. Thesis. Linköping: Linköping University, 1999.
- [82] A. Stenman, M. Braun, B. McNamara, and D. E. Rivera: ‘Model-on-Demand Identification for Control: An Experimental Study and Feasibility Analysis for MOD-Based Predictive Control’. *12th IFAC Symposium on System Identification, Santa Barbara, CA, USA*. June, 2000: pp. 439–444.

- [83] A. Stenman, F. Gustafsson, and L. Ljung: ‘Just in time models for dynamical systems’. *Proceedings of 35th IEEE Conference on Decision and Control*. Vol. 1. IEEE. 1996: pp. 1115–1120.
- [84] P. Stoica and R. Moses: *Spectral Analysis of Signals*. Pearson Prentice Hall, 2005.
- [85] P. Stoica and T. Söderström: ‘Optimal instrumental-variable methods for identification of multivariable linear systems’. *Automatica* (1983), vol. 19(4): pp. 425–429.
- [86] P. Strobach: *Linear prediction theory: a mathematical basis for adaptive systems*. Vol. 21. Springer Science & Business Media, 2012.
- [87] A. Tartakovsky, I. Nikiforov, and M. Basseville: *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.
- [88] L. D. Tufa, M. Ramasamy, and M. Shuhaimi: ‘Closed-loop system identification using OBF-ARMAX model’. *Journal of Applied Sciences(Faisalabad)* (2010), vol. 10(24): pp. 3175–3182.
- [89] P. M. Van Den Hof, P. S. Heuberger, and J. Bokor: ‘System identification with generalized orthonormal basis functions’. *Automatica* (1995), vol. 31(12). Trends in System Identification: pp. 1821–1834.
- [90] A. van der Klauw, M. Verhaegen, and P. van den Bosch: ‘State space identification of closed loop systems’. *Proceedings of the 30th Conference on Decision and Control*. Jan. 1992: 1327–1332 vol.2.
- [91] M. Verhaegen and V. Verdult: *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.
- [92] J. Wang, S. Jianju, Z. Yan, and Z. Donghua: ‘Searching Historical Data Segments for Process Identification in Feedback Control Loops’. *Computers and Chemical Engineering* (2018), vol.
- [93] R. K. Wood and M. W. Berry: ‘Terminal composition control of a binary distillation column’. *Chemical Engineering Science* (1973), vol. 27: pp. 1707–1717.
- [94] P. C. Young: *Recursive estimation and time-series analysis: an introduction*. Springer Science & Business Media, 2012.
- [95] Y. Zhu: *Multivariable system identification for process control*. Elsevier, 2001.





# Schriftenreihe Mess- und Regelungstechnik der Universität Kassel

Herausgegeben von / Edited by

Univ.-Prof. Dr.-Ing. Andreas Kroll, Universität Kassel

---

- Band 1:** Klassifikationsgestützte on-line Adaption eines robusten beobachterbasierten Fehlerdiagnoseansatzes für nichtlineare Systeme, Kassel 2011  
Patrick Gerland
- Band 2:** Zur Identifikation mechatronischer Stellglieder mit Reibung bei Kraftfahrzeugen, Kassel 2012  
Zhenxing Ren
- Band 3:** Sensordatenfusionsansätze in der Thermografie zur Verbesserung der Messergebnisse, Kassel 2014  
Samuel Soldan
- Band 4:** Multi-Robot Task Allocation for Inspection Problems with Cooperative Tasks Using Hybrid Genetic Algorithms, Kassel 2014  
Chun Liu
- Band 5:** Gasleckortungsmethode für autonome mobile Inspektionsroboter mit optischer Gasfernmesstechnik in industrieller Umgebung, Kassel 2015  
Gero Bonow
- Band 6:** Dynamische Analyse großer verkoppelter Systeme mit Methoden der Komplexen Netzwerke am Beispiel des Inverse-Response-Verhaltens, Kassel 2015  
Andreas Geiger
- Band 7:** Close range 3D thermography: real-time reconstruction of high fidelity 3D thermograms, Kassel 2018  
Antonio Rafael Ordóñez Müller
- Band 8:** Zur regelungsorientierten Identifikation nichtlinearer Systeme mittels lokal affiner Takagi-Sugeno-Fuzzy-Modelle, Kassel 2019  
Alexander Schrod
- Band 9:** System Identification of Stochastic Nonlinear Dynamic Systems using Takagi-Sugeno Fuzzy Models, Kassel 2019  
Salman Zaidi
- Band 10:** Automatic 3D Visualization and Tracking of Gaseous Organic Volatile Compound Emissions by means of Spatial and Temporal Information from an Optical Gas Imaging Stereo System, Kassel 2021  
Johannes Rangel

Performing experiments for system identification of continuously operated plants might be restricted as it can impact negatively normal production. In such cases, using historical logged data can become an attractive alternative for system identification. However, operating points are rarely changed and parameter estimation methods can suffer numerical problems.

Three main drawbacks of current approaches in this research area can be discussed. Firstly, detection tests are not adapted for dynamical systems. Secondly, methods to define upper interval bounds are not robust to colored noise that is more likely to be found in real applications. Thirdly, model estimation with the retrieved data is not supported and the performance of the method cannot be assessed. In the method proposed in this work, called data selection for system identification (DS4SID), previous drawbacks are addressed and robust tests are designed and implemented. The performance of DS4SID is evaluated in a simulated and laboratory multivariate processes. A process unit of the lab-scale factory “ $\mu$ Plant” is used as industry-oriented case study. Models estimated with selected set data are shown to have similar performance than estimates with the entire data set.

ISBN 978-3-7376-1009-4



9 783737 610094 >