Process Analysis and Performance Assessment for Sheet Forming Processes

by

Hui Yuan

A thesis submitted to the Graduate Program in

Department of Chemical Engineering

in conformity with the requirements for the degree of

Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

February 2015

Abstract

Process monitoring and controller performance assessment are essential tools for ensuring that manufacturing processes operate safely, predictably, meet quality targets and operate profitably. Efficient techniques for process analysis and controller assessment facilitates are important for identifying areas for process and control improvement. Sheet- and film-forming processes pose special challenges for these techniques, because they often have periods of non-steady-state operation, and can exhibit both spatial and temporal variations. Existing process evaluation and analysis methodologies are oriented primarily for processes under steady operation, and focus on temporal variations.

New tools are proposed to address limitations on the application of minimum-variance-based controller performance assessment to metal rolling processes. Extensions are proposed which address: 1) non-constant deadtime that arises from changes in rolling speed during startup, steady operation and wind down in the rolling process; 2) constraints on control actions; and 3) different sampling intervals for the manipulated variable input and the process output. The efficacy of the proposed extensions are demonstrated using an aluminum rolling mill and simulation examples.

Singular Spectrum Analysis (SSA) is a promising technique for analyzing time series that decomposes data into a number of interpretable frequency components.

New filtering and spectral interpretations of SSA are proposed in this work, including a modification to the computational procedure that produces a filter with zero-phase lag. Links between SSA are made to other signal processing and time series techniques. The potential for SSA in analyzing chemical manufacturing processes is demonstrated using an extended analysis of a two-tank process under periodic operation. An SSA-based approach is proposed for computing minimum-variance-based controller performance assessment, and as illustrated using an example. This technique has potential for providing more detailed diagnosis of elements causing poor controller performance.

The effectiveness of a recently proposed two-dimensional SSA (2D-SSA) algorithm is investigated for the analysis of two-dimensional problems that frequently arise in sheet forming processes. The use and interpretation of this algorithm is demonstrated using two simple examples of rolling processes with known defects in the roll. 2D spectra are computed using the 2D-SSA algorithm, and are interpreted.

Co-Authorship

The research in Chapter 5 has been published in Physica D: Nonlinear Phenomena. Chapter 3 has been submitted to The Canadian Journal of Chemical Engineering (CJChE). Journal articles based on preliminary research results in Chapter 6 and 7 are in preparation for future submission. I prepared the first and subsequent drafts for all of these coauthored manuscripts, performed all of the calculations and simulations, and generated all of the Figures and Tables. Professors Thomas Harris and James McLellan were co-authors of the journal articles. They helped to formulate research objectives, provided technical advice, and edited the thesis and suggested revisions.

Acknowledgements

A Ph.D. thesis might not be considered as a team effort but it definitely requires the contribution of many individuals. I would like to take this opportunity to express my thanks and appreciations to all those who helped me in this work.

I am greatly indebted to my supervisor Prof. Tom Harris. The thesis could not have not been done without his inspiration as a great teacher and kind guidance throughout this work. I will not be able to thank Prof. James McLellan enough, whom I respect and admire tremendously, for his co-supervision and all his help during my Ph.D. program.

The friendly and cooperative atmosphere in Novelis Global Technology Center is acknowledged. I further appreciate the cooperation and joint work with Matt Fairlie and Florent Bourot.

I would like to thank my parents, Yuan Yiru and Peng Xian'e for their love, support and encouragement throughout my life. And I owe a great deal to my son Yuan Zixuan and my wife Xu Ling. Their support have made everything else possible.

During my program, I benefited a lot from all my labmates and friends at Queen's. I have been so lucky to meet and work with them. And thanks goes to my friends in Kingston community, especially Nancy Smith, Qi Xiaodong, Ian Ritchie, Debra Fieguth, Herb Goold, Stephen Bonnycastle, and many others.

Finally, this work could not have been possible without the financial support from Novelis, Mathematics of Information Technology and Complex Systems (MITACS), Ontario Centres of Excellence (OCE), and Queen's University.

List of Abbreviations

1D-	one-dimensional
2D-	two-dimensional
M-	multi-channel
AGC	Automatic Gauge Control
ARIMA	autoregressive intergrated moving average
CCA	canonical correlation analysis
CD	cross direction
CPA	controller performance assessment
CPA-MVB	controller performance assessment - minimum-variance bound
CPM	control performance monitoring
CUSUM	cumulative sum
CVA	canonical variate analysis
DFT	discrete fourier transform
DPCA	dynamical principal component analysis
EOF	empirical orthogonal functions
EEOF	extended EOF
EWMA	exponentially-weighted moving average
FB	feedback

FF	feedforward
FFT	fast fourier transform
FIR	finite impulse response
FSR	finite step response
GMV	generalised minimum variance
HbH	Hankel-block-Hankel
IIR	infinite impulse response
LQG	linear quadratic gaussian
MD	machine direction
MEM	maximum entropy method
MIMO	multi-input multi-output
MISO	multi-input single-output
MLR	multiple linear regression
MVC	minimum variance control
MPC	mode predictive control
NLD	non-linearity detection
NLPCA	nonlinear principal component analysis
NLSSA	nonlinear singular spectrum analysis
PCA	principal component analysis
PDE	partial differential equation
PID	proportionalintegralderivative
PLS	projection to latent structures OR partial least square
PRBS	pseudo-random binary sequence
RC	reconstructed component

SISO single-input single-output

SPC statistical process control

SPE squared prediction error

SPM statistical process monitoring

SSA singular spectrum analysis

SVD singular value decomposition

Nomenclature

$a_{i,t}$	driving force for the i^{th} disturbance
a_t	random disturbance
$A(q^{-1})$	autoregressive component in a time series model
$\boldsymbol{A}$	Hankel matrix
b	delay for feedback variable
$B(q^{-1})$	exogenous component in a time series model
c	specified constraint
$C(q^{-1})$	moving average component in a time series model
$\boldsymbol{C}$	sample covariance matrix
$ceil(\cdot)$	the least integer greater than or equal of $(\cdot)$
d_i	degree of differencing for i^{th} disturbance model
$D_{i,t}$	i^{th} disturbance at time t
D_t	disturbance
$\hat{D}_{i,t+b\|t}$	b-step ahead minimum variance forecast for i^{th} disturbance
$e_{i,t+b\|t}$	b-step ahead forecast error for i^{th} disturbance
$E(\cdot)$	the operation of taking expectations
f_t	normalized frequencies at temporal direction
f_s	normalized frequencies at spatial direction

$G_C(q^{-1})$	controller transfer function
$G_{d_i}(q^{-1})$	i^{th} disturbance dynamic transfer function
$G_D(q^{-1})$	disturbance transfer function for feedforward variable
$G_{fb}(q^{-1})$	feedback controller transfer function
$G_{ff}(q^{-1})$	feedforward controller transfer function
$G_P(q^{-1})$	discrete-time process transfer function
$G_P(s)$	continuous process transfer function
K	integral controller tuning parameter
P	proportional gain in the PID controller
I	Integral gain in the PID controller
D	Derivative gain in the PID controller
$int(\cdot)$	the integer portion of $(\cdot)$
$\boldsymbol{J}$	exchange matrix
l	delay for feedforward variable
l_j	length-index at time t_j
l_s	distance between the sensor and the roll gap
$\boldsymbol{I}(m)_{s\times s}$	a modified identity matrix
$\tilde{l}_{k+1}$	equally-spaced length index
$\boldsymbol{R}$	sample autocorrelation matrix
q^{-1}	backward shift operator
$round(\cdot)$	the nearest integer to $(\cdot)$
$sign(\cdot)$	the operation of extracting the sign of a real number
$S(f)$	the spectrum of y_t

u_t	manipulated variable at time t (deviation from reference value)
$u_t^\star$	computed control movement by the control law
u_t^{fb}	contribution of feedback controller to manipulated variable
u_t^{ff}	contribution of feedforward controller to manipulated variable
v_t	mill speed
$\boldsymbol{v}_i$	i^{th} eigenvector
$var(\cdot)$	the operation to calculate variance
$\boldsymbol{w}_i$	i^{th} latent variable
$\boldsymbol{W}(k_1, k_2)$	a 2D-window function
$y_{i,t}$	i^{th} output component at time t
y_{sp}	process setpoint
$\tilde{y}_{k+1}$	equally-spaced length-sampled data
$\hat{y}_{t+b\|t}$	b-step ahead forecast
y_t	measured output at time t
y_l	observation at length l
$\boldsymbol{Y}(k_1, k_2)$	a 2D signal

Greek symbols

ϵ_{t+b}	error in a time series model
$\omega(q^{-1})$	numerator polynomial in process dynamics
$\delta(q^{-1})$	denominator polynomial in process dynamics
$\eta(q^{-1})$	numerator polynomial in feedforward dynamics
$\gamma(q^{-1})$	denominator polynomial in feedforward dynamics
$\theta_i(q^{-1})$	moving average component in i^{th} disturbance model
$\phi_i(q^{-1})$	autoregressive component in i^{th} disturbance model
$\boldsymbol{\phi}_i(x)$	i^{th} basis function
$\psi(q^{-1})$	closed loop transfer function
$\psi_i(q^{-1})$	b-step ahead forecast error polynomial for i^{th} disturbance
η	normalized performance index
$\sigma^2_{\hat{a}}$	variance of the random shock estimation $\hat{a}_t$
σ^2_{mv}	variance of theoretical minimum variance controller
σ^2_y	closed loop deviation variance
μ_y	mean process output (deviation from setpoint)
λ	a penalty on the input variance (∇u_t^2)
∇	backwards difference operator
∇e_t	offset deviation
∇u_t	controller output deviation (or implemented control move)
$\nabla \tilde{u}_t$	computed control change by the control law
ζ_i	i^{th} filter coefficient of reconstructed series
$\circ$	convolution computation

Δ integral operator

Δl sampling length interval

Δt sampling time interval

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

In industry, customer/producer demands are ever-increasing, including more critical safety and environmental policies, improved product quality, increased product acceptance, higher production rate, and reduced material and energy assumption. To meet customer/producer demands and eventually fulfill the need of profit maximization, process analysis, control, monitoring, and optimization tools are applied to achieve these goals. Process analysis is often considered as a pre-step for control design, while process monitoring and optimization are post-steps after control implementation. In this work, the research focuses on two areas: process monitoring and analysis.

Typical monitoring techniques are classified into two groups – controller performance assessment (CPA) and statistical process monitoring (SPM). CPA techniques are often identified as diagnosis tools and are used to assess controller performance. When there is a significant performance degradation in a control system, the plant personnel must be notified, and effective monitoring techniques are expected to locate

potential root-causes. SPM techniques are not totally separated from control monitoring activities, meaning that they are a complement to each other as summarized in [1]; they are capable of detecting and diagnosing faults in processes, specifically in process variables, sensors, actuators and equipment. Cinar [1] points out that before implementing control monitoring, the aforementioned process faults can be eliminated by performing process monitoring. Details on two types of monitoring will be provided in Chapter 2.

This thesis is focused on CPA techniques in particular. While the CPA technique has been applied to many different industries, the lack of the applications in sheet forming processes motivates the research in this thesis. Sheet forming processes cover a wide range of industrial processes such as paper-making, metal rolling, and coating. This thesis makes a contribution to address practical issues in CPA applications of metal rolling processes.

While process monitoring helps the investigation of sheet forming processes, there is need for the understanding of complex chemical processes. Analysis techniques are used to summarize data information and understand a process in this thesis. Process modelling is a standard way to do process analysis. The modelling techniques are often classified into two different categories: parametric and nonparametric techniques. In this work, nonparametric techniques are suggested because they are data-driven, and routine process data are always available and not costly. Moreover, there is no need to interrupt a process when preparing the data for analysis, and this is a significant benefit to the industry. Among a large number of nonparametric techniques, a newly developed technique – Singular Spectrum Analysis (SSA) – is proposed. The research described in this thesis makes two significant contributions to the SSA area:

1) a more thorough understanding of this methodology from a filtering and frequency perspective; 2) the promotion of this novel technique to process analysis in the chemical industry.

1.2 Motivations and Contributions

Effective control is the means by which safe operation is ensured, and profitability is maximized. Process monitoring, analysis, and optimization tools assist in control design, and in process operation. Two challenges are regularly encountered in controlling industrial processes: unsuccessful or poorly performing control schemes, and inadequate process models. Questions that arise can include:

- What are the most effective techniques for identifying problems and root causes in the performance of controllers?
- What practical limitations are there on controller performance assessment tools, and can these be alleviated? To what extent do they arise from the type of process being considered (e.g., sheet forming).
- What tools, in addition to those of conventional system identification, can be deployed to provide a better understand of the process behaviour?
- What tools exist to deal with process monitoring and controller performance assessment when there are both temporal and spatial elements to the behaviours that are being assessed? Are there tools that can capture directionality of higher dimensional control problems?

The areas of process monitoring and analysis have been introduced to answer the above questions. Two promising tools – Controller Performance Assessment (CPA)

and Singular Spectrum Analysis (SSA) – are proposed to help realize the goals.

1.2.1 Why Controller Performance Assessment (CPA)?

Bialkowski [2] has claimed that 60% of industrial controllers are non-optimal. Controller performance assessment performance assessment (CPA) techniques are proposed for fault detection and diagnosis of control schemes. In a recent summary of industrial implementations of CPA methods, Jelali [3] states that controller performance assessment (CPA) — benchmarked by Minimum Variance Control (MVC) [4] — is implemented in approximately 60% of the case studies. Note that not much work has been done in the applications of process monitoring in sheet forming processes, especially the metal rolling industry. This shortage motivates CPA applications in batch rolling processes. Three industrial concerns for CPA implementation in rolling processes have been discussed in detail in Chapter 3. Modified CPA algorithms have been proposed for each issue and validated by case studies. The intention of this work is to draw more attention to the need of process monitoring in sheet forming processes. Although proposed CPA algorithms in this thesis have been limited to a simple system — a feedback control system with or without a feedforward control — it is possible to extend these ideas to a system having similar issues with more advanced control strategies, e.g., cascade control and model-predictive control.

1.2.2 Why Singular Spectrum Analysis (SSA)?

Process analysis can play a key role in the root cause analysis and performance improvement of control systems. The more characteristics known about a process, the better control can be achieved. Nonparametric methods for analysis offer the

advantage that they can be executed in a more economical way. It must be stressed though that the results are based on the data considered, and that extrapolation outside the range in which the data are collected must be done carefully, and may lead to inconsistent results. In this work an established time-series analysis method in signal processing, singular spectrum analysis (SSA), has been chosen. Diverse SSA applications have proven a promising future of SSA in data analysis. However, SSA techniques have not been applied to manufacturing plant data [5, 6, 7]. The aim of this research is to introduce this innovative signal processing technique to process analysis and monitoring in chemical and sheet forming industries. The SSA approach complements existing techniques from time-series analysis and multivariate statistics. One characteristic that arises commonly in the sheet forming industry is behaviour in both temporal and spatial directions. In this thesis, an extension of SSA to two-dimensional analysis and monitoring is proposed. The SSA technique is mainly motivated by the featured decomposed components in analyzed data. The results in this thesis emphasize understanding of the algorithm itself, features of decomposed components, and interpretation in industrial applications.

1.3 Thesis Overview

This thesis is organized as follows:

- Chapter 2 reviews two categories of process monitoring techniques: controller performance assessment (CPA) and statistical process monitoring (SPM). Techniques in these two categories are complementary to each other and can work together in order to provide comprehensive information about a process.

- Chapter 3 examines applications of the controller performance assessment (CPA) technique in metal-rolling processes, and provides solutions for three practical issues for CPA implementation in metal rolling: 1) changes in time delay from changing rolling speed during startup, steady operation, and wind down of the roll; 2) constraints on control actions; and 3) different output sampling intervals and control intervals. Because a conventional CPA cannot be implemented directly, modified CPA algorithms are proposed to address these issues. Furthermore, industrial aluminum-rolling examples are provided to prove the effectiveness and efficiency of the proposed extensions.

- Chapter 4 briefly reviews process analysis techniques in two groups — parametric and nonparametric methods. A nonparametric method, singular spectrum analysis (SSA), is chosen in this work. The history and advantages of the SSA technique are presented. Applications of SSA in different areas motivate this research. The aim of the research is to promote this novel tool in the process analysis of manufacturing processes.

- Chapter 5 starts with an introduction of a basic SSA algorithm with some variations, followed by new filtering and spectral interpretations of SSA. It is shown that the variables reconstructed from diagonal averaging of reduced rank approximations to a trajectory matrix can be obtained from a non-causal convolution filter with zero-phase characteristics. The reconstructed variables are readily made using a two-pass filtering algorithm that is well known in the signal processing literature. When the number of rows in the trajectory matrix is much larger than the number of columns, many results reported in the literature can be used to derive properties of the resulting filters and their spectra. New features of the reconstructed series

are revealed using these results. Two examples illustrate the derived results.

- Chapter 6 promotes the SSA application to the chemical process analysis. One simulation example is investigated for demonstration. Moreover, a new controller performance monitoring extension based on SSA is derived. Linear regression theory and properties of symmetric Toeplitz matrices form the foundation of the new developed performance assessment measure. Together with structural insights into reconstructed components, Chapter 6 evaluates the feasibility of the proposed technique in both performance assessment and process analysis.

- Chapter 7 extends the use of 1D-SSA in one-dimensional analysis to two-dimensional analysis in sheet forming processes. Before introducing the selected two-dimensional analysis tool, 2D-SSA, conventional multiple regression methods, multivariate time series methods, and reduced-rank techniques are reviewed. In comparison with other techniques, the ones based on data-driven basis functions are the focus of this thesis. Among data-driven methods, only 2D-SSA provides 2D insights from a windowing perspective. Together with 2D periodogram, 2D-SSA is proposed to implement 2D process analysis. Preliminary results on filtering interpretations of 2D-SSA and meaningful physical interpretations of industrial processes are demonstrated via case studies.

- Chapter 8 concludes this work with the research contributions. Potential research directions are recommended for future study.

Chapter 2

Literature Review on Process Monitoring Techniques

Process monitoring encompasses both monitoring of process behaviour - in either open-loop or closed-loop form - and monitoring of controller performance. The range of process monitoring techniques is quite broad, and approaches can be found that use mechanistic models, or empirical models in both parametric and non-parametric forms. This review will focus on two classes of monitoring techniques: control and statistical monitoring techniques. Controller performance monitoring has grown as a new field in process control community over the last two decades. Control monitoring techniques are often identified as diagnosis tools and are used to assess controller performance. When there is a significant performance degradation in a control system, the plant personnel are notified, and various techniques are used to locate potential root-causes. Statistical monitoring techniques complement control monitoring activities [1], and are capable of detecting and diagnosing faults in processes, specifically in process variables, sensors, actuators, and equipments. Cinar [1] recommends that

before implementing control monitoring, process faults can be eliminated by performing process monitoring. More broadly, statistical process monitoring and control monitoring schemes should be implemented together. In the event that a controller performance problem is detected, the statistical process monitoring system can be used as part of the root-cause detection step. In addition, the statistical process monitoring system should be implemented so that it can alarm independently on process faults, on a continuing online basis.

A literature review on control monitoring and statistical process monitoring is now provided.

2.1 Controller Performance Assessment (CPA)

Bialkowski [2] reports that up to 60% of industrial controllers often do not perform well. To diagnose root causes and improve performance of degraded control systems, over the past two decades, the area of controller performance assessment (CPA) in both industry and academia has attracted significant attention. CPA covers techniques available for monitoring control loop performance. In a recent summary of industrial applications of CPA methods, Jelali [3] draws the following observations:

1) approximately 60% of case studies apply minimum variance control (MVC) benchmarking [8];

2) about 25% of case studies use oscillation detection methods [9];

3) roughly 10% of case studies implement advance benchmarking methods [10];

4) the remaining 5% employ other techniques [11].

The most widely used measures implement an optimal performance benchmark associated with unconstrained minimum variance control (MVC). Note that MVC-based CPA techniques are often referred to as control performance monitoring (CPM) approaches in the literature. The terms 'CPM' and 'CPA' are often used interchangeably. In this thesis, the term 'CPA' is used primarily.

2.1.1 MVC-benchmarking CPA

Minimum variance controllers for linear discrete time processes were first proposed by Åström [12]. MVC controllers provide the best control in the mean square sense for all linear feedback controllers for a process [4, 13]. Further reduction in the process output variance is not achievable via re-tuning the existing control strategy or applying a more sophisticated linear feedback controller; instead, this is only possible via process modifications, e.g., the reduction of the inherent variability or deadtime, as well the implementation of a new feedforward controller or another manipulated variable [4]. Harris [4] proposed using MVC as a basis for CPA because of this optimality property. Note that in industry MVC is not implemented that frequently because the resulting control action is often too aggressive. However, MVC represents the best possible control when there are no constraints on the manipulated variable action, and provides a useful best case benchmark for determining whether current controllers are working well or not. In the conventional CPA approach, a normalized performance index is typically computed to assess the performance of the overall control scheme. One appealing aspect of the conventional CPA approach is that only routine operating data and a process delay are required; additionally, process perturbations are not required.

Desborough and Harris [14] defined the standard MVC-benchmarking CPA for single-input single-output (SISO) feedback control systems. Different MVC-benchmarking CPA techniques have been developed for industrial use, and are often implemented as a screening tool to identify loops requiring attention [3, 8]. Various extensions of standard MVC-benchmarking CPA have been reported dealing with different process systems: 1) multi-input single-output (MISO) systems [15] - The MVC-benchmarking is extended to feedback control systems with feedforward controllers. An analysis of variance method is proposed to take into consideration the contribution to the minimum variance term from different resources, i.e., the feedback controller or feedforward controllers. Priori system knowledge required in MISO applications are minimal, i.e., only the process time delay and individual feedforward delays. 2) multi-input multi-output (MIMO) systems [16, 17, 18, 19, 20, 21] - The framework of MVC-benchmarking CPA is extended to multivariate systems. The primary issue in this extension is that it requires the time delay structure of the system, i.e., an interactor matrix. The construction of an interactor matrix demands essentially the knowledge of the process impulse response. Various methods have been proposed to construct the interactor matrix in different forms. The computation and interpretation of MVC-based CPA is much more difficult for MIMO problems. 3) time-varying systems [22, 23, 24, 25, 26, 27, 28] - Developments have enabled MVC-benchmarking CPA to be extended to monitor time-varying systems. The problem is motivated by the observation of time-varying disturbance dynamics. Note that in this application the type of disturbance is often required to be specified. 4) nonlinear systems [29, 30, 31, 32] - Harris and Yu and his co-workers develop the MVC-benchmarking CPA for a class of nonlinear systems. Non-linear autoregressive moving average with exogenous inputs

(NARMAX) models are used to fit models for closed-loop routine operating data and then calculate the MVC benchmark. Harris and Yu state in [29] that challenges in this research come from: (a) the complexity of non-linear behaviour; (b) system and disturbance representation; (c) efficiency of the model fitting.

Additional details on recent CPA developments can be found in a number of excellent review papers [3, 8, 10, 33, 34, 35].

2.1.2 Other CPA Techniques

Besides the conventional MVC-benchmarking CPA technique, a number of other CPA techniques have been proposed in the last few decades from both academic and industry researchers. A number of these techniques are summarized below, and additional details can be found in the review papers of Jelali [3, 36] and the citations contained therein.

1) Modified versions of MVC benchmarking

Instead of MVC, advanced controllers, such as GMV (Generalised Minimum Variance), MPC (Mode Predictive Control), and LQG (Linear Quadratic Gaussian) controllers, are considered to be optimal linear controllers in terms of both input and output variance. Straightforward extensions of MVC benchmarking have been provided using GMV [37], MPC [38, 39], and LQG [40] benchmarking. Similar to MVC benchmarking, these techniques do not require the implementation of the optimal controllers. Rather, performance bounds and more realistic performance indices are introduced.

Optimal PID benchmarking for performance assessment has been proposed [41, 42], motivated by the widespread use of PID controllers in industry. The key feature

of this benchmarking is to provide a performance bound in terms of variance, which is possibly achievable by a PID controller. Another benefit from this approach is that optimal PID parameters can be obtained as a by-product of the computation.

2) Fault detection methods

Poor control performance may be arise from different sources, such as system non-linearity, disturbances, and system changes. Detection measures have been proposed for each of these areas:

a. Non-linearity detection (NLD);

   A linearity assumption is used in most CPA methods; however, non-linearity does exist in processes and may induce severe performance problems, e.g., oscillations. Measures of nonlinearity have been proposed by a number of authors, including Stack and Doyle [43], Haber and Keviczky [44], Guay and co-workers [45, 46], and Liu et al. [47].

b. Oscillation detection;

   In industry, closed-loop oscillations can occur fairly often. The main causes are typically aggressive control, disturbances, and non-linearity. A key feature of oscillation is that it can be propagated from one control loop to others. A number of oscillation detection methods have been presented for different purposes. Hägglund [48] has introduced a detection method to identify oscillation due to high friction in control valves. Horch [49] has proposed two approaches to verify oscillation due to valve stiction, while Thornhill and her co-workers [50, 51, 52] have published a series of papers from an experimental and practical perspective addressing problems of oscillation assessment and root-cause diagnosis.

c. Sluggish control;

When controllers are retuned after operating conditions change, the issue of sluggish control often occurs. Moreover, very often no further retuning will be implemented again even if controllers are tuned conservatively. This oversight can result in large deviations from set point, and can potentially lead to decreased product quality. Hagglund [53, 54] finds that control performance and product quality can be crucially improved by locating and retuning sluggish control loops; further, an idle index, which expresses the relation between the times of positive and negative correlation between the process output and controller output increments, is proposed to define the sluggishness of control loops.

3) User-defined metrics

User-defined metrics have been proposed for different goals. For instance, a control performance benchmark using historical data from a well-tuned control system is defined by control and maintenance engineers [55, 56]. Higher-level metrics from a business/operational perspective on a plant-wide basis can be established to assess overall control performance of a plant (see, for example, [57, 58]).

2.2 Statistical Process Monitoring (SPM)

In the 1930s Shewhart [59] proposed the concept of being in a state of statistical control. Statistical control refers to a process with only common-cause variation, while out-of statistical control indicates special-cause variation. Statistical process monitoring (SPM) or statistical process control (SPC) refers to statistical methods in process

monitoring. Recent overview papers [1, 60, 61, 62] cover SPM research and applications in the areas of fault detection, diagnosis, estimation, and reconstruction. SPM detects abnormal events and diagnoses root causes. To improve a process, these root causes are removed instead of compensated. Note that SPM is essentially different from automatic feedback control, as stated by MacGregor and Kourti [63]. Feedback controllers adjust manipulated variables to compensate for disturbance effects in key process variables [64]. Accordingly, disturbance variability is only transferred to less important variables, but not essentially removed. In the next section univariate and multivariate SPM are discussed; additionally more attention is paid to multivariate SPM techniques due to the multivariate nature of most processes.

2.2.1 Univariate SPM

Univariate SPM techniques monitor a single quality variable at a time. Control charts are used to provide a visual means of monitoring process variables. The most common control charts are Shewhart [59], CUSUM (Cumulative Sum) [65] and EWMA (Exponentially-Weighted Moving Average) [66, 67] charts.

These SPM techniques are well established, and discussions can be found in many engineering statistics texts (see for example, [68]). Cinar [1] and MacGregor [63] provide discussions of univariate control charts from a control context. Shewhart charts can be used to monitor attributes such as the mean, range, and standard deviation of a process over given time. CUSUM charts monitor accumulated differences from target in order to detect small sustained process shifts. EWMA charts represent a weighted average of several consecutive observations, and are designed to account for serial time correlation in the measurements used to compute the charted statistics.

To identify process shifts, univariate control charts typically keep watch on a very small group of key product/quality variables.

Martin et al. [69] summarize major drawbacks of univariate charts: 1) lack of full product/quality information: a small number of key variables may not be able to fully define quality information, and quality variables tend to not be as available on a frequent basis; 2) failure to account for quality variable interaction: quality variables are often dependent, which makes process monitoring and diagnosis more difficult and may lead to false results; 3) incomplete use of process data: any special event in a process is likely to have a footprint in both process and quality variables, while charts based in particular on single quality variables will not be able to assess this footprint. In our current data-rich society, massive amounts of process data are routinely collected and available for use. Consequently, univariate charts are inadequate for processes with multivariate and highly correlated nature. Multivariate SPM techniques address many of the limitations of univariate techniques. The key feature in multivariate techniques is to take into account all variables of interest simultaneously; hence, distinct behaviour of each variable together with interaction behaviour between variables is accounted for, resulting in more comprehensive and accurate process information. [60] provides a thorough discussion on univariate and multivariate SPM.

2.2.2 Multivariate SPM

Multivariate SPM provides an effective means for making use of the large number of measurements available for most manufacturing processes.

2.2.2.1 Multivariate Control Charts Based on Original Variables

Natural multivariate extensions can be developed for univariate control charts by taking into account the multivariate distribution of the monitored variables 'within sample'. As reviewed in [70, 71, 72], Hotelling T^2, multivariate CUSUM, and multivariate EWMA control charts are direct extensions of univariate control charts, respectively. Multivariate charts monitor several quality characteristics simultaneously, condensed to a single charted statistic. Note that key assumptions of using all these charts are that observations are multivariate normally distributed with a within-sample covariance matrix at a given point in time, and are independent between samples in time. The covariance matrix can be used to check these assumptions and to calculate control limits (see for example, [71]).

Multivariate control charts have been proposed to address the issues in a univariate context, however a number of challenges can remain: 1) even though several key quality variables are summarized in a multivariate form, there may be instances where certain quality variables are infrequently collected; 2) the assumptions of multivariate normality and temporal independence only hold for certain processes. Strictly speaking, these charts cannot be applied to temporally correlated processes, however correlation naturally occurs in many processes. The multivariate EWMA chart accounts for temporal correlation, while the CUSUM chart will account for non-stationary behaviour.

2.2.2.2 Multivariate Control Charts Based on Latent Variables

As an alternative to multivariate control charts based on original variables, Jackson [72] proposed forming multivariate charts based on latent variables. The original

quality or process variables are transformed into new orthogonal latent variables, i.e., principal components. The resulting transformed charted variables are independent because of the orthogonality between the principal components.

The latent variable approach provides the means of combining different types of data - e.g., quality variables measured less frequently and online variables measured more frequently - and of finding significant relationships buried within large numbers of monitored variables that have a strong association with the quality variables of interest. In addition, latent variable techniques provide an effective approach for dealing with missing data, and for collapsing high-dimensional data down to manageable dimensions [60].

Latent variable methods rely on multivariate projection to lower dimensional subspaces. The most widely used approaches are Principal Component Analysis (PCA) [72], Canonical Correlation Analysis (CCA) [73], Reduced Rank Regression [74], and Projection to Latent Structures or Partial Least Squares (PLS) [75]. Dimension reduction is the foundation of these projection methods. Note that the discussion below focuses on PCA and PLS, because they dominate the application literature in process monitoring.

PCA and PLS are based on eigen-decomposition of $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$ or Singular Value Decomposition (SVD) of the data matrix $\mathbf{X}$, and extract significant principal components in terms of variance explained. For process data matrix $\boldsymbol{X}$ and quality data matrix $\boldsymbol{Y}$, where each column in two matrices represents a process or quality variable, and each row represents an observation at a given time, PCA concentrates on a decomposition of $\boldsymbol{X}$, while PLS conducts a decomposition of both $\boldsymbol{X}$ and $\boldsymbol{Y}$. After decomposition, the original variables are projected into lower dimension spaces formed by orthogonal

latent variables. PCA and PLS models for future monitoring are built via latent variables for existing datasets. Reference models can be established using process data with or without quality data from in-control processes, i.e., well-operated and high-quality product making processes. Consequently, multivariate control charts can be constructed for the nominal operating region and applied to on-line monitoring and fault diagnosis.

Typical monitoring charts for latent variable approaches include T-score plots, P-loading plots, Hotelling's T^2 charts, and Squared Prediction Error (SPE) charts. These charts can detect faulty processes and poor-quality products at an early stage through on-line monitoring. Additionally, aforementioned work has been extended to areas such as fault diagnosis [76, 77] and fault reconstruction [78, 79]. For instance, when an out-of-control situation is identified, a contribution plot is often used to diagnose root causes. Note that many successful applications of these control charts can be found in continuous processes [76, 80]. Furthermore, Nomikos and MacGregor [81, 82, 83] extend the idea of multivariate process monitoring to batch and semi-batch processes, i.e., multiway PCA and PLS. Additionally, Kramer [84] has presented a concept of nonlinear principal component analysis. Successful industrial applications can be found in overview papers [1, 61, 63] and references therein.

Multivariate control charts based on latent variables have a better chance of providing an earlier fault warning than univariate ones, because co-dependencies in the multivariate data have been taken into account. Mass balance, energy balance, and other process constraints introduce within sample dependencies, which can be inferred through variable correlations.

Latent variable charts have some limitations as well: 1) Interpreting multivariate

charts is not as straightforward as univariate ones, because latent variables may not have any physical meaning; and 2) there is some information loss due to orthogonal projection and dimension reduction, particularly if the number of latent variables retained is much smaller than the number of original variables.

2.3 Summary

Statistical Process Monitoring (SPM) complements Controller Performance Assessment (CPA) by focusing on the behaviour of the process. SPM can identify the occurrence of faults, such as failure of actuators, or excessive fouling in a heat exchanger. Later in this thesis, Singular Spectral Analysis will be proposed both for SPM, as well as for CPA. Practical applications of CPA in metal rolling processes are discussed in the next chapter.

Chapter 3

Applications and Extensions of Controller Performance Assessment (CPA) To Metal-Sheet-Forming Processes[2]

3.1 Introduction

Initial work on data-driven controller performance assessment (CPA) can be traced to Åström ([12]) and Devries and Wu ([85]). Åström [12] first proposed the Minimum-Variance Control (MVC) strategy for systems described by the superposition of a

[2]The work summarized in this Chapter has been submitted to *The Canadian Journal of Chemical Engineering (CJChE)*. Drs. James McLellan and Thomas Harris were co-authors of this research work. Please refer to the preface for the nomenclature used in this Chapter.

discrete-linear dynamic model and an additive unmeasured disturbance that is modelled by a linear transfer function driven by white noise. Åström showed that a necessary and sufficient condition for the process to be achieving its minimum-variance bound is that the closed-loop output is a moving average process of finite order b. b is directly related to the process delay and control interval. Åström noted that the autocorrelation function of the closed-loop output could be used to test for minimum variance performance. Devries and Wu [85] defined an effectiveness index as the ratio of the variance of the one-step-ahead prediction error to the variance of the measured output.

The minimum-variance bound can be estimated from routine operating data by fitting a linear time-series model to the output [4]. The minimum-variance bound can also be interpreted as the variance of the b-step ahead minimum-variance prediction error. Since then, CPA has received considerable academic and industrial attention, with the development of enhancements to the variance-based methods, and the development of alternate data-driven techniques that address for example deterministic performance monitoring, the response speed, 'sluggishness', settling time, overshoot and damping ratio, stability margin, and the detection of oscillations. To distinguish variance-based methods from many other techniques, we will use the acronym CPA-MVB for Controller Performance Assessment - Minimum-Variance Bound.

The main attraction of most CPA-MVB performance measures is that only routine closed-loop data and knowledge of the process delay are required for the performance assessment. Controller performance can be assessed without disturbing the process. The CPA-MVB does not account for controller structure limitations, e.g., a feedback system with or without a feedforward controller. The excessive control action

associated with minimum-variance control is typically not desirable in practice, and so realizing the minimum-variance bound may not be a desirable control objective. This does not invalidate its utility as a performance monitoring technique [14]. Simple modifications have been proposed to overcome the controller structure limitation [86]. An important point that is often overlooked is that the estimated lower bound may not be sufficient to meet economic, safety or environmental requirements. In these cases process modifications, such as reducing the dead-time, reconfiguring the controller to use a different manipulated variable, or attenuating the disturbances through process improvements must be contemplated. When one is able to identify a model for the dynamics and disturbances, more CPA tools can be implemented [19].

CPA-MVB applications have been reported for a range of process control problems for single-input single-output (SISO) systems [14, 27], multi-input single-output (MISO) systems [15, 87], and multi-input multi-output (MIMO) systems with linear, nonlinear, feedback, feedforward, model-based and MPC controllers, and both constant and time-varying disturbances [16, 28, 88, 89]. Recent developments have enabled variance-based methods to be extended to a class of nonlinear systems [29, 30, 31, 32]. Developments in CPA-MVB have been covered in detail in a number of review papers [3, 8, 10, 33, 34, 35, 90].

Different CPA indices have been developed for industrial uses [18]. MV-based CPA typically serves as a first-pass monitoring layer. One single CPA technique cannot provide sufficient information for monitoring purposes, which is why different techniques are investigated and typically combined for practical use. Jelali [3] reported that CPA-MVB benchmarking had been applied to about 60% of the case studies in industrial applications. Mature application areas include oil refineries

[38, 87], petrochemical and chemical sectors [9, 91], and the pulp and paper industry [41, 48, 49, 54, 92, 93, 94, 95, 96].

Relatively little work has been reported in the metal-processing industry with the exception of [36, 97]. A number of issues can arise that require modifications to conventional CPA-MVB methods. First, rolling mill speeds may be variable, moving from startup to steady operation to roll wind down. In addition, constraints on control action (e.g., the hydraulic actuators of the roll stand) are often imposed to avoid excessive wear and resonance. Finally, the rapid sampling rates sometimes require that outputs and inputs are sampled and/or stored at different intervals because of control system limitations. CPA-MVB methods are valuable in analyzing rolling systems, as the dead time is often a significant portion of the process settling time and are therefore an important limitation on achieving desired performance.

In this thesis we develop extensions to traditional CPA-MVB methods to address these issues. The application and efficacy of the solutions are illustrated using actual data and simulated data from an industrial aluminum-rolling process. A number of key results make use of developments in econometrics.

The work is organized as follows. First, metal sheet-forming processes are briefly described, with an emphasis on Automatic Gauge Control (AGC) systems in aluminum cold rolling mills. This is followed by a brief review of the CPA-MVB method. The three challenges to industrial implementation in metal rolling are discussed, and modified CPA-MVB techniques for process monitoring are proposed. Finally, case studies are presented for each challenge to demonstrate the utility of the proposed techniques.

3.1.1 Sheet-Forming Processes

Sheet- and film-forming processes, also called web-forming processes, cover a wide class of industrial processes, including paper-making, polymer film extrusion, steel/metal rolling, coating, laminating, and plate glass manufacturing [98]. There is significant industrial interest in developing identification, parameter estimation, control, and monitoring techniques for these processes, which exhibit both time-varying and spatially-varying behaviour. Effective control can mean improved product quality, including spatial uniformity, increased product acceptance, higher production rates, and reduced material and energy consumption. The control objective is always to maintain uniform flat profiles in both the Machine Direction (MD) and Cross Direction (CD). The machine direction refers to the direction in which the sheet is moving, while cross direction is the direction across the sheet, perpendicular to the machine direction. In this thesis, the focus will be on gauge control of metal-rolling processes.

3.1.2 Aluminum-Rolling Process

Gauge is critically important to customers of rolled metal products [99]. In this thesis, gauge is defined as the centreline thickness in the MD. Automatic Gauge Control (AGC) systems are implemented to minimize an exit sheet thickness deviation. As it is difficult to remove disturbances once they are imprinted into a metal, AGC systems are used to eliminate disturbances caused by roll eccentricities in present or previous passes, tension variations, hardness variations, or vibrations. Different control schemes can be implemented to achieve this goal. The AGC system considered here comprises a feedback and a feedforward controller. Other control strategies, such as tension control, mass flow control, and roll eccentricity compensation, are beyond

the scope of this thesis. Additional details can be found in [36].

Figure 3.1 illustrates an AGC control system that incorporates both feedback and feedforward control. An exit thickness measurement is compared to a target thickness, and a tracking error is used to adjust a hydraulic gap controller, i.e., an actuator, usually in a cascade-controller arrangement. The AGC sends a set-point target to the gap position controller (not shown), which adjusts the hydraulic jacks. Clearly, a time delay is present since there is a distance between the output (gauge measurement) and the actuator (hydraulic gap), and this time delay has an impact on a feedback controller performance. To make a feedback AGC more effective, it is better to place the gauge sensor closer to the roll stand. Feedforward control is also used to provide a faster response relative to feedback control. An entry thickness gauge is tracked, and a controller is tuned to make necessary corrective actions to offset thickness variations entering the mill.

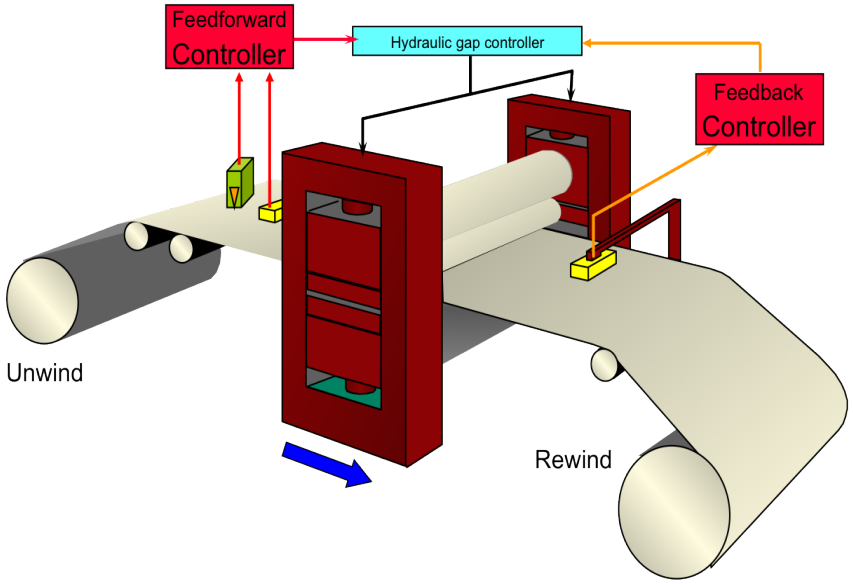


Figure 3.1: **Automatic gauge control system (after [99])**

The block diagram for the AGC system is shown in Figure 3.2, assuming linear models for the process and disturbances. y_t and u_t are the process output and input at time t, respectively. The process is subject to both unmeasured (D_t) and measured ($D_{1,t}$) disturbances. These disturbances in turn are represented as time-series models driven by random shocks a_t and $a_{1,t}$, passing through dynamic elements $G_d(q^{-1})$ and $G_{d_1}(q^{-1})$, respectively. $G_D(q^{-1})$ is a disturbance transfer function representing the impact of $D_{1,t}$ on the output, while $G_P(q^{-1})$ is the process transfer function. The control input is computed by feedback and feedforward controllers $G_{fb}(q^{-1})$ and $G_{ff}(q^{-1})$ respectively. The symbol q^{-1} denotes a backwards shift operator, i.e., $q^{-1}y_t = y_{t-1}$.

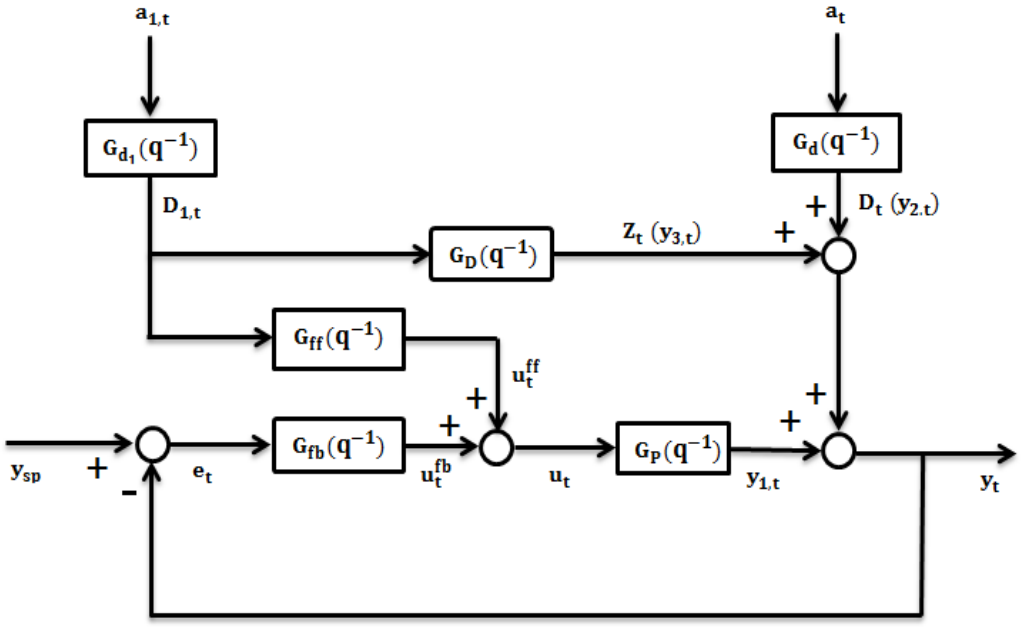


Figure 3.2: **AGC system block diagram**

3.2 Review of Controller Performance Assessment

3.2.1 CPA-MVB Benchmarking

Many industrial processes can be described by the superposition of discrete-time transfer function models of the form:

$$\begin{aligned} y_t &= y_{1,t} + y_{2,t} + y_{3,t} \\ &= G_P(q^{-1})u_t + D_t + G_D(q^{-1})D_{1,t} \\ &= G_P(q^{-1})(u_t^{fb} + u_t^{ff}) + D_t + Z_t \\ &= [G_P(q^{-1})u_t^{fb} + D_t] + [G_P(q^{-1})u_t^{ff} + Z_t] \end{aligned}$$

where:

$$\begin{aligned} G_P(q^{-1}) &= \frac{\omega(q^{-1})q^{-b}}{\delta(q^{-1})} \quad and \quad G_D(q^{-1}) = \frac{\eta(q^{-1})q^{-b_1}}{\gamma(q^{-1})} \\ u_t &= u_t^{fb} + u_t^{ff} \\ D_t &= G_d(q^{-1})a_t = \frac{\theta(q^{-1})}{\phi(q^{-1})\nabla^d}a_t \\ D_{1,t} &= G_{d_1}(q^{-1})a_{1,t} = \frac{\theta_1(q^{-1})}{\phi_1(q^{-1})\nabla^{d_1}}a_{1,t} \quad and \quad Z_t = G_D(q^{-1})D_{1,t} \end{aligned}$$

Note that

- a_t and $a_{1,t}$ are a sequence of independent and identically distributed (i.i.d) random variables with zero mean and variances $\sigma_a^2, \sigma_{a1}^2$, respectively.

- b and b_1 are delays for the process and measured disturbance transfer functions, respectively (by definition $b \geq 1$ and $b_1 \geq 1$). Delay b influences the feedback (FB) controller performance, while delay b_1 influences the feedforward (FF) controller performance.

- $\{\omega(q^{-1}), \eta(q^{-1})\}$ are polynomials in the backshift operator q^{-1}, $\{\delta(q^{-1}), \gamma(q^{-1}), \theta(q^{-1}),$

$\phi(q^{-1}), \theta_1(q^{-1}), \phi_1(q^{-1})\}$ are monic polynomials in q^{-1}, ∇ is the backward difference operator $(1 - q^{-1})$, and $\{d, d_1\}$ are non-negative integers (typically 0, 1, or 2).

As shown in [15], the closed-loop output can be written as:

$$\begin{aligned} y_{t+b} &= [\tfrac{\omega(q^{-1})}{\delta(q^{-1})} u_t^{fb} + \hat{D}_{t+b|t} + e_{t+b|t}] + [\tfrac{\omega(q^{-1})}{\delta(q^{-1})} u_t^{ff} + \hat{Z}_{t+b|t} + e_{1,t+b|t}] \\ &= [e_{t+b|t} + e_{1,t+b|t}] + [\tfrac{\omega(q^{-1})}{\delta(q^{-1})} u_t^{fb} + \hat{D}_{t+b|t} + \tfrac{\omega(q^{-1})}{\delta(q^{-1})} u_t^{ff} + \hat{Z}_{t+b|t}] \end{aligned} \tag{3.1}$$

where $\hat{D}_{t+b|t}$ is the b-step ahead minimum-variance forecast for the unmeasured disturbance, and $e_{t+b|t}$ is the associated b-step ahead forecast error. $\hat{Z}_{t+b|t}$ is the b-step ahead minimum-variance forecast for the effect of the measured disturbance on the process output, and $e_{1,t+b|t}$ is the associated b-step ahead forecast error. The development of the minimum-variance forecasts is standard, which uses well-known methods from time-series analysis (see [15] for more details).

The key points that follow from Equation (3.1) are:

- The forecast errors, i.e., $\{e_{t+b|t},\, e_{1,t+b|t}\}$, are contemporaneously and serially uncorrelated with each other. As well, $\{e_{t+b|t} + e_{1,t+b|t}\}$ are contemporaneously and serially uncorrelated with the right-bracketed expression in Equation(3.1). This latter term can be interpreted as the minimum-variance forecast for y_{t+b} given information available up to and including time t, i.e., $\hat{y}_{t+b|t}$.

- The minimum-variance forecast errors are of the form:

$$\begin{aligned} e_{t+b|t} &= \psi^{mv}(q^{-1}) a_{t+b} = [1 + \psi_1 q^{-1} + \cdots + \psi_{b-1} q^{-(b-1)}] a_{t+b} \\ e_{1,t+b|t} &= \psi_1^{mv}(q^{-1}) q^{-m} a_{1,t+b} \\ &= [\psi_{1,0} q^{-m} + \psi_{1,1} q^{-(m+1)} + \cdots + \psi_{1,b-m-1} q^{-(b-1)}] a_{1,t+b} \end{aligned}$$

where $m = \min[b, b_1]$, $\psi^{mv}(q^{-1})$ is a monic polynomial of order $(b-1)$, and $\psi_1^{mv}(q^{-1})$ is

a polynomial of order $(b-b_1-1)$ when $b > b_1$. When $b \leq b_1$, $e_{1,t+b|t} = 0$, i.e. it is theoretically possible to eliminate the effect of the feedforward disturbance before it appears in the feedback loop. The coefficients of the polynomials $\{\psi^{mv}(q^{-1}), \psi_1^{mv}(q^{-1})\}$ are obtained as the coefficients of q^{-k}, $k = 0, \cdots, b-1$, in the series expansion of the polynomials $\frac{\theta(q^{-1})}{\phi(q^{-1})\nabla^d}$ and $\frac{\eta(q^{-1})q^{-b_1}}{\gamma(q^{-1})} \cdot \frac{\theta_1(q^{-1})}{\phi_1(q^{-1})\nabla^{d_1}}$, respectively. The coefficients of these polynomials are identical with the impulse response coefficients of the respective transfer functions.

- Under minimum-variance control the feedback and feedforward control actions are chosen to set the forecast $\hat{y}_{t+b|t}$ to zero, i.e.,

$$u_t^{fb} = -\frac{\delta(q^{-1})}{\omega(q^{-1})}\hat{D}_{t+b|t} \quad \text{and} \quad u_t^{ff} = -\frac{\delta(q^{-1})}{\omega(q^{-1})}\hat{Z}_{t+b|t}$$

  resulting in the closed-loop output:

$$y_{t+b}^{mv} = e_{t+b|t} + e_{1,t+b|t}$$

- Under minimum-variance control, the process variance is given by:

$$\sigma_{mv}^2 = \text{var}(e_{t+b|t}) + \text{var}(e_{1,t+b|t})$$

  This variance serves as a lower bound on the theoretically achievable performance that can be obtained. All other controllers must result in a variance-based performance that exceeds this bound. The minimum-variance controller, and some trivial modifications, are also optimal for a wide class of symmetric and nonsymmetric cost functions [100].

  σ_{mv}^2 serves as a minimum-variance benchmark for this feedforward/feedback control system and can be used to compute the minimum-variance-based controller performance index between 0 and 1 [15]. Note that when computing this index in practice

it is more useful to apply a mean square error (mse) instead of an output variance:

$$\eta = \frac{\sigma_{mv}^2}{\text{mse}(y_t)} = \frac{\sigma_{mv}^2}{\sigma_y^2 + \mu_y^2}$$

where σ_y^2 denotes an output variance, μ_y denotes an output mean, and a mean square error is a summation of σ_y^2 and μ_y^2. Values of η close to 1 indicate that opportunities for variance reductions are limited.

- In many cases, the implemented feedforward/feedback controller will not be of minimum-variance design, and the closed-loop can be written as:

$$\begin{aligned} y_{t+b} = [e_{t+b} + e_{1,t+b}] \quad &+ \quad [\frac{B_0(q^{-1})}{F_0(q^{-1})} y_t + \frac{B_1(q^{-1})}{F_1(q^{-1})} D_{1,t} \\ &+ \quad \frac{B_2(q^{-1})}{F_2(q^{-1})} u_t^{fb} + \frac{B_2(q^{-1})}{F_2(q^{-1})} u_t^{ff}] \end{aligned} \tag{3.2}$$

  The transfer functions $\frac{B_i(q^{-1})}{F_i(q^{-1})}$, $i = 0, 1, 2$, are obtained from algebraic manipulation of the system transfer functions and transfer functions obtained in constructing the various forecast errors. If linear controllers are used, the right-bracketed term in Equation (3.2) can be written solely in terms of transfer functions involving the output y_t and the feedforward variable $D_{1,t}$.

The key result from [4, 14, 15] was to show that the minimum-variance performance bound could be estimated from routine operating data by using standard time-series methods with a knowledge of $\{b, b_1\}$. In these papers, it was assumed that linear controllers were used. Consequently only records of the process output and measured disturbance were needed to construct the variance bound.

3.3 Industrial Considerations for Rolling Mills

A number of practical issues arise when implementing the CPA-MVB analysis in rolling mills. In the discussion that follows, three potentially limiting issues are described in detail, and solutions are proposed to overcome these challenges. The solutions for the second and third issues are applicable for other methods of data-driven analysis for industrial processes.

3.3.1 Non-constant Time Delay Horizon Due to Changing Speed

Metal-coil rolling is a batch process, and a typical batch consists of a startup phase where a sheet is fed to the bite of the roll stand and the speed accelerates to a steady operating point. As the batch nears the end of the coil, the roll speed decreases. Since the thickness gauge is a fixed distance from a roll gap, the time delay between the point of actuation (the roll gap) and the point of measurement changes with roll speed. Uniform time steps no longer correspond to uniform increments in sample-strip length as the speed changes. These changes in the time delay occur primarily at the start and end of a rolling batch, and violate the usual constant time delay assumption for many controller performance and assessment methods.

One solution to this problem is to compute a CPA-MVB index exclusively using data from the period of steady operation in the middle of a batch. This will not provide any information about how well a controller is performing during acceleration and deceleration phases. An alternative approach is to convert data from a non-uniform time-based setting to a constant length-based setting, motivating the development of

a length-based CPA-MVB index. A length-based CPA-MVB was first proposed by [97] for metal processing, however very little information is provided about how to implement such a measure. In this thesis the derivation of a length-based CPA-MVB is provided in Section 3.4, and the approach is demonstrated using mill data.

3.3.2 Constraints on Controller Outputs

The second issue is how best to deal with constraints implemented on controller actions. Several types of constraints can be encountered. Maximum-change constraints typically limit the change in a controller output to a prescribed fraction of the total controller range (e.g., from zero position to saturation or full actuation). If the constraint is active, the controller output is assigned a fixed value. The second type of constraint is a thresholding or dead-band constraint, in which the calculated control action is not sent to the actuator unless it exceeds a minimum threshold. If this threshold is not exceeded, the control action sent to the plant is held at its previous value. Such dead-band constraints are often implemented in order to avoid excessive actuator manipulation that can cause premature wear, or can lead to resonance behaviour. Another type of threshold constraint occurs when the manipulated variable is only adjusted when the process output exceeds a pre-defined value. In all cases, the controller is also prevented from attempting to implement control moves that are outside of the physical range for the output.

In these cases, a standard MVC controller will not necessarily be able to cancel the effect of the predicted disturbance at the time delay horizon because constraints limit control actions that would otherwise cancel this portion of the disturbance. Fundamentally however, the overall best achievable control performance is still the

variance associated with unconstrained minimum-variance controller, which is again the prediction error over the time delay horizon. Constraints pose potential challenges for conventional CPA-MVB calculations because they introduce a nonlinear element into the control loop. Consequently, the standard AR or ARMA linear time-series transfer function model for the closed-loop tracking error cannot be used to compute the minimum-variance bound. A modified CPA-MVB algorithm is presented to address this issue which sidesteps the problem of determining a time-series model for the tracking error by using the manipulated or controller variables to construct the minimum-variance bound. By directly accounting for the manipulated variable actions as implemented in the plant, it is possible to avoid problems associated with introducing the thresholding element in the closed-loop expression.

3.3.3 Unequal Output Sampling Intervals and Control Intervals

Metal-rolling processes are inherently high-speed processes requiring rapid sampling. Consequently, a large amount of data is recorded during the rolling of a coil. To increase data storage efficiency, process outputs are sometimes down-sampled, leading to differences between process output sampling intervals and control intervals. Often, output measurements are available more frequently than control movements. Finally, control intervals will sometimes be longer than output sampling intervals to enable the control system to operate in real time. The challenge from a CPA standpoint is how to best accommodate the different sampling intervals while making the most effective use of the information available.

The key concept for CPA is to focus on what a controller sees and does, i.e., using a

sampling interval for an output measurement that corresponds to the control interval. Two scenarios arise: i) the output sampling interval exceeds the control interval, and ii) the process output is measured more frequently than the control interval. The latter situation is easily remedied. To address the theoretical and practical issues associated with the first situation, we utitlize results from the econometrics literature.

3.4 Length-based CPA-MVB

In this section, detailed steps for implementing a new length-based CPA-MVB algorithm are described. Note that in the discussion that follows, the subscript t is used to denote a point in the time domain, and the subscript l is used to denote a point in the length domain. For the presentation below, we consider that we have a sample of time data $\{t, y_t\},\ t = 1, 2, \cdots, N$. This sample would typically be the start-up or slow-down phase of the batch. We could also take the entire time series across the batch, comprising both transient and steady operation, and convert to a length basis to calculate an overall CPA-MVB for the batch.

3.4.1 Proposed Algorithm

- Step 1: Conversion of Time-Based Observations to Length-Based Observations $\{t, y_t\} \longrightarrow \{l, y_l\}$

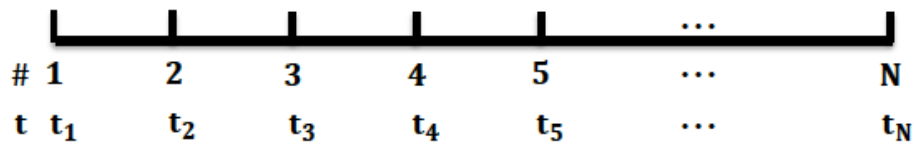


Figure 3.3: **Equally-spaced time-based data**

Given measurements of a mill speed v_t, time-based output measurements y_t can be converted to observations at equivalent length-based intervals. Assuming that the sampling interval is Δt, and the speed is constant over a given sampling time interval, we obtain:

$$l_j = l_{j-1} + \Delta t \cdot v_{j-1}, \qquad j = 2, 3, \cdots, N, \quad l_1 \equiv 0$$

The length l_j is used to define the equivalent length index for the observation at the time t_j.

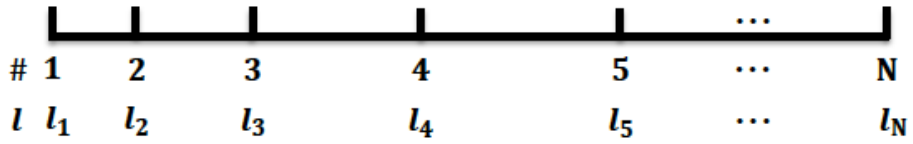


Figure 3.4: **Unequally-spaced length-based data**

The observation y_l corresponds to y_t. Note that in Figure 3.4, the observations are not equally-spaced since the speed is changing in the transient parts.

- Step 2: Conversion to Equally-Spaced Length-Based Observations $\{l, y_l\} \longrightarrow \{\tilde{l}_k, \tilde{y}_k\}$

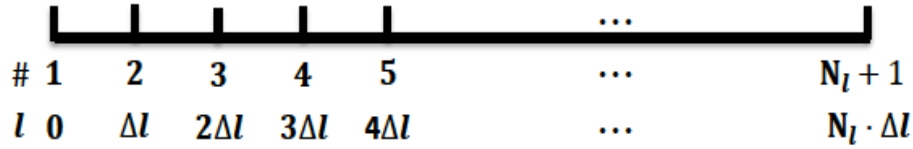


Figure 3.5: **Equally-spaced length-based data**

In this step, the N unequally-spaced length-based data are converted to N_l+1 equally-spaced points using linear interpolation in the length index. (More sophisticated

interpolation methods, such as splines could be used.) The length interval Δl must be specified. If fewer interpolated points are required to reduce the computational load for performance monitoring and analysis, then a larger length sampling interval Δl can be specified. Note that sampling is typically in the millisecond range, and large amounts of data can accumulate quickly. However, aliasing problems can occur if Δl is too large. Thus, there is a tradeoff in specifying Δl. Here we use the maximum length interval from the conversion of the time sampling to length sampling:

$$\Delta l = \max\{l_j - l_{j-1},\ j = 2, 3, \cdots, N\}$$

$(N_l + 1)$ data points will be calculated where $N_l = \text{int}(\frac{l_N}{\Delta l})$, and int$(\cdot)$ is the integer part of $(\cdot)$ with $(N_l + 1) \leq N$.

When linear interpolation is used, the equally-spaced length-sampled data are calculated recursively as:

$$\begin{aligned} \tilde{l}_{k+1} &= k\Delta l \\ \tilde{y}_{k+1} &= y_j + \tfrac{\tilde{l}_{k+1} - l_j}{l_{j+1} - l_j}(y_{j+1} - y_j) \end{aligned} \qquad k = 1, 2, \cdots, N_l$$

with initial conditions determined by

$$\tilde{y}_1 = y_1$$

$$\tilde{l}_1 = l_1 = 0$$

where the index j is such that $l_j \leq \tilde{l}_{k+1} \leq l_{j+1}$ In the general case, it is possible for the sampling intervals of mill speed and output to be different, i.e., $\Delta t_v \neq \Delta t_y\ (= \Delta t)$ and often $\Delta t_v > \Delta t_y$. The calculations for l_j, l_{j+1} are modified as follows:

$$l_j = \sum_{i=1}^{k} \Delta t_v \cdot v_i,\ k = \text{round}(\frac{(j-1) \cdot \Delta t_y}{\Delta t_v}) + 1$$

$$l_{j+1} = \sum_{i=1}^{n} \Delta t_v \cdot v_i,\ n = \text{round}(\frac{j \cdot \Delta t_y}{\Delta t_v}) + 1$$

where round$(\cdot)$ returns the nearest integer to $(\cdot)$.

- Step 3: CPA-MVB Analysis

  The conventional CPA-MVB analysis algorithm for time-based data can be applied directly to the equally-spaced length-based data, using a 'time' delay expressed in terms of the number of length sample intervals (yielding what is really a 'length delay'). If the sensor is located at a distance l_s from the roll gap, the delay required for the CPA-MVB calculation is $\text{int}(l_s/\Delta l) + 1$. Interpretation remains the same as in the time-based case.

3.4.2 Case Study

To illustrate the effectiveness of the proposed length-based CPA-MVB algorithm, industrial data from a rolling mill under the control of an AGC system of the type shown in Figure 3.1 are considered. The data have been scaled to maintain confidentiality.

The first panel in Figure 3.6 provides a mill speed plot, including a steady-state segment in the centre, and acceleration and deceleration transient segments during the startup and wind-down phases of the batch respectively.

The second panel gives performance index plots for time- and length-based CPA-MVB. The CPA-MVB index values are computed by partitioning the coil data into 30 segments of equal time duration. Some of the segments are contained entirely within the steady operation part of the batch, while others lie in the ramp-up and ramp-down transients. Some intervals also contain observations lying in the transient phase and in the steady operation phase. A number of interesting observations can be made:

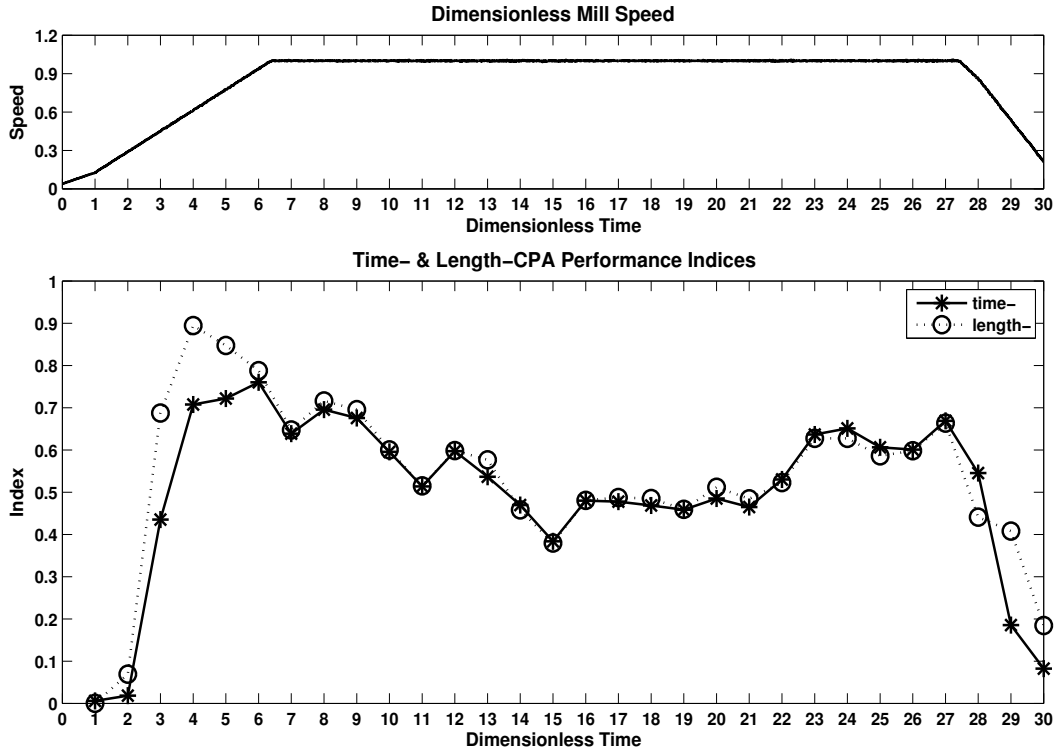


Figure 3.6: **(a) Mill speed plot; (b) Time- and length-based CPA-MVB results**

- In the central steady-state part of the batch, in which the mill speed is approximately constant, the time- and length-based CPA-MVB results are in good agreement. This is expected since the speed is constant and hence the time delay is constant. In addition, the nominal time delay specified in calculating the time-based CPA-MVB is based on the typical mill speed in the steady operating part of the batch. The length delay is computed as the distance between the sensor and the roll gap divided by the length-sample interval, and is inherently fixed by the physical configuration of the mill. The rolling speed appears indirectly in the length-based CPA-MVB during the conversion of the time-based to length-based data.

- In transient parts with changing mill speed, the time- and length-based CPA-MVB

results no longer match, with the time-based CPA-MVB indices generally being lower than the corresponding length-based values. This makes sense from a process standpoint as the mill speed in the transient portions is slower than the nominal speed during the steady operation portion, resulting in a larger time delay. While the time-based CPA-MVB index is being calculated based on a nominal time delay which is smaller than the actual one, the estimated minimum-variance bound will be smaller than the true benchmarking value. Consequently, the time-based CPA-MVB index will be lower, i.e., the length-based CPA-MVB index is larger and more accurate.

- Once the steady-operating region is reached, the time- and length-based indices match very quickly. At the start of the batch, both the time- and length-based CPA-MVB indices are small, suggesting that the controller is not providing very good performance at the very slow speed associated with the start of the batch.

3.5 Constraints on Manipulated Variables

Constraints on manipulated variables are often encountered, and they can be treated as input nonlinearities. Calculation of the CPA-MVB relies on the concept of a feedback invariant, which is not affected by this type of nonlinearity [29]. However, this benchmarking calculation is only possible if a record of the manipulated or control variable is available in addition to the record of the output variable. Calculation of the CPA-MVB is illustrated using maximum-change and dead-band constraints in this thesis.

3.5.1 Maximum-change Constraints

To avoid control saturation or aggressive control moves, maximum-change constraints are often employed:

$$\nabla u_t = \begin{cases} \nabla \tilde{u}_t, & |\nabla \tilde{u}_t| < c \\ \text{sign}(\nabla \tilde{u}_t) \cdot c, & \text{otherwise} \end{cases}$$

where $\nabla \tilde{u}_t$ is the control change computed by the controller, ∇u_t is the control move implemented in the plant, and the constraint c (typically a constant) is specified by the control engineer. Note that sign$(\cdot)$ is the operation of extracting the sign of a real number $(\cdot)$.

3.5.2 Dead-band Constraints

To prevent frequent actuator movements and equipment wear, dead-band constraints are often imposed on the controller outputs in rolling mill applications. The dead-band constraints are the dual problem to the maximum-change constraints discussed above:

$$\nabla u_t = \begin{cases} \nabla \tilde{u}_t, & |\nabla \tilde{u}_t| > c \\ 0, & \text{otherwise} \end{cases}$$

where in this instance c is the dead-band limit.

3.5.3 MV Performance Bounds for Systems with Constraints

Harris and Wu [29] exploited the fact that the unconstrained CPA-MVB is a feedback invariant performance bound, to extend this concept to classes of nonlinear stochastic systems which include the type of input considered in this thesis. A constraint is fundamentally a nonlinear element, and it cannot be expressed in transfer-function

form. Regardless of the type of constraint imposed, it will not be possible to reduce the output variance beyond the CPA-MVB over the time delay horizon, meaning that this CPA-MVB value represents the minimum-variance benchmark for the constrained case as well. Since there is no transfer-function representation for a nonlinear constraint the CPA-MVB cannot be computed by using a transfer-function form that includes output measurements only. The feedback invariant is the prediction error for the b-step ahead prediction. This can be estimated using a model that takes into account the control actions that have actually been implemented. This can be accomplished using a closed-loop predictor of the ARX form:

$$A(q^{-1})y_t = B(q^{-1})u_{t-b} + a_t \tag{3.3}$$

where $\{A(q^{-1}), B(q^{-1})\}$ are polynomials in the back-shift operator q^{-1}, and a_t is a white noise element [14, 29]. The actual values of the manipulated variable are used when constructing the closed-loop predictor. When an ARX model is used, the feedback invariant is obtained by computing the impulse response coefficients from the estimated transfer function $\frac{1}{A(q^{-1})}$ and the estimated noise variance for σ_a^2. When feedforward variables are presented, they are treated as exogenous variables in the time-series model.

3.5.4 Case Study: An AGC Control System

An open-loop step test has been done in an AGC control system of a rolling process, and step response data have been collected. The sampling interval used in the system is 1 ms, and the process delay is 15 ms. The discrete process transfer function is

modelled as:

$$G_P(q^{-1}) = \frac{\omega(q^{-1})}{\delta(q^{-1})} q^{-b} = \frac{0.1945 + 0.1506q^{-1}}{1 - 1.263q^{-1} + 0.4661q^{-2}} q^{-16}$$

assuming a zero-order hold is used for the control input.

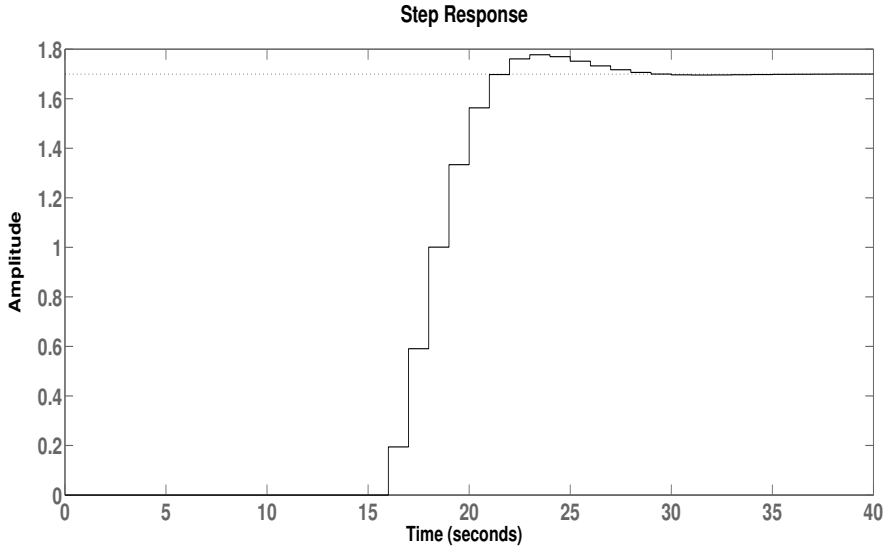


Figure 3.7: **Step response**

The step response for the discrete-time process is shown in Figure 3.7 and displays second-order process characteristics including overshoot. The deadtime is a significant component of the settling time - a common characteristic in many rolling processes.

To evaluate the proposed methodology, the additive disturbance is represented as an autoregressive process with root close to the unit circle. This allows for considerable meandering or nearly nonstationary behaviour in the absence of control:

$$D_t = G_D(q^{-1})a_t = \frac{1}{1 - \phi q^{-1}} a_t = \frac{1}{1 - 0.98q^{-1}} a_t$$

where the driving force a_t is i.i.d. with mean 0 and variance 0.1188, which results in a variance of 3 for the disturbance D_t.

The closed-loop system with no control constraints can be expressed as:

$$y_t = G_P(q^{-1})u_t + G_D(q^{-1})a_t = \frac{0.1945 + 0.1506q^{-1}}{1 - 1.263q^{-1} + 0.4661q^{-2}}u_{t-16} + \frac{1}{1 - 0.98q^{-1}}a_t$$

For this example, the disturbance can be expressed as:

$$\begin{aligned} D_{t+16} = \hat{D}_{t+16|t} + e_{t+16|t} &= \frac{0.98^{16}}{1 - 0.98q^{-1}}a_t + (1 + 0.98q^{-1} + \cdots + 0.98^{15}q^{-15})a_{t+16} \\ &= 0.98^{16}D_t + (1 + 0.98q^{-1} + \cdots + 0.98^{15}q^{-15})a_{t+16} \end{aligned}$$

The resulting CPA-MVB is calculated as:

$$\sigma^2_{mv} = \frac{1 - \phi^{32}}{(1 - \phi^2)}\sigma^2_a = 1.4284$$

The variance ratio of best-achievable control to no control is:

$$\frac{\sigma^2_{mv}}{\mathrm{var}(D_t)} = 1 - \phi^{32} = 0.4761$$

3.5.5 Constrained MVC Controller Design

The unconstrained MVC is given by the expression:

$$\begin{aligned} \tilde{u}_t &= -\frac{\delta(q^{-1})}{\omega(q^{-1})}\hat{D}_{t+16|t} \\ &= -\frac{\delta(q^{-1})}{\omega(q^{-1})}\phi^{16} \cdot D_t = -\frac{\delta(q^{-1})}{\omega(q^{-1})}\phi^{16} \cdot (y_t - \frac{\omega(q^{-1})}{\delta(q^{-1})}\tilde{u}_{t-16}) \\ &= \phi^{16}\tilde{u}_{t-16} - \phi^{16}\frac{1 - 1.263q^{-1} + 0.4661q^{-2}}{0.1945 + 0.1506q^{-1}}y_t \end{aligned} \tag{3.4}$$

The unconstrained MVC has an unacceptably large input variance. To moderate the control action and to enable constraints, we will use the most elementary extension of model predictive controller which was proposed by Clarke and Hastings-James [101]. The controller is designed to minimize $E(y^2_{t+b} + \lambda_t \nabla u^2_t)$, which is equivalent to minimizing the one-step criterion $\hat{y}^2_{t+b|t} + \lambda_t \nabla u^2_t$. The manipulated variable is calculated at each time step by setting the expectation to zero and solving for u_t.

MacGregor and Tidwell [102] noted that a one-stage constrained MVC with fixed λ is not usually optimal from a variance perspective. By contrast, an infinite-stage constrained MVC is guaranteed to provide an input variance that is less than or equal to that obtained from the one-stage optimal controller for the same output variance.

The use of a fixed penalty parameter, however, does not allow for incorporation of hard constraints. To enable enforcement of hard constraints, the penalty weight is chosen to ensure that the control action from the constrained MVC law does not exceed the constraint. If the unconstrained control action $|\nabla\tilde{u}_t| < c$, the penalty weight $\lambda_t = 0$, while if $|\nabla\tilde{u}_t| \geq c$, then the input is fixed at the appropriate constraint. The constrained MVC implementation would use a very large (tending to infinite) value of λ_t when $|\nabla\tilde{u}_t| \leq c$, and a penalty $\lambda_t = 0$ when $|\nabla\tilde{u}_t| > c$. For either type of constraint, when the constraint is active, this corresponds to using a non-zero, finite value for λ_t.

3.5.6 PI Controller Design

A PI controller is also used for the case study. The PID Tuner GUI in MATLAB^{TM} is used to design a PI controller in velocity form. To optimize the closed-loop controller performance the integral of squared error (ISE) is used as a tuning criterion to calculate PI controller parameters [103]:

$$\nabla\tilde{u}_t = P \cdot \nabla e_t + I \cdot e_t = -(0.27 \cdot (y_t - y_{t-1}) + 0.027 \cdot y_t)$$

Note that the setpoint y_{sp} in this example is zero. The velocity form avoids problems with integral windup and enables easy implementation of input constraints.

3.5.7 Simulation Results for Systems with Maximum-change Constraints

The closed-loop behaviour of the process example under the MVC and PI controllers described above is simulated for the maximum-change constraint scenario. An actuator always has a physical limit, and another hard constraint is imposed to avoid aggressive behaviour in both cases:

$$u_t = \begin{cases} \tilde{u}_t, & |\tilde{u}_t| < 2 \\ \text{sign}(\tilde{u}_t) \cdot 2, & otherwise \end{cases}$$

The process is simulated for $N = 10,000$ observations, which leads to data records typically encountered in rolling-mill operations. A range of constraint values are imposed, and the results are shown in Figures 3.8 and 3.9 for the constrained implementation of MVC and PI cases, respectively. The horizontal axes represents the constraint limit c, while the vertical axes represents the CPA-MVB, the output (y_t) and input (u_t) variances, and performance indices, respectively. The CPA-MVB is calculated two ways: using an ARX model that incorporates the implemented control action, and using an AR model which explicitly ignores the input constraints. The minimum-variance bound (MVB) required for the performance bound is computed in each case from Equation (3.3), i.e., the variance of the b-step ahead prediction error. The dotted lines show the theoretical MVB, which is calculated using the ARX approach.

A number of observations can be made:

- The MVB calculated using an ARX model has almost no dependency on the constraint imposed on the control action and has a value of approximately 1.52, whereas the true value is 1.43. Note that this is not really a bias; rather, the time series is one

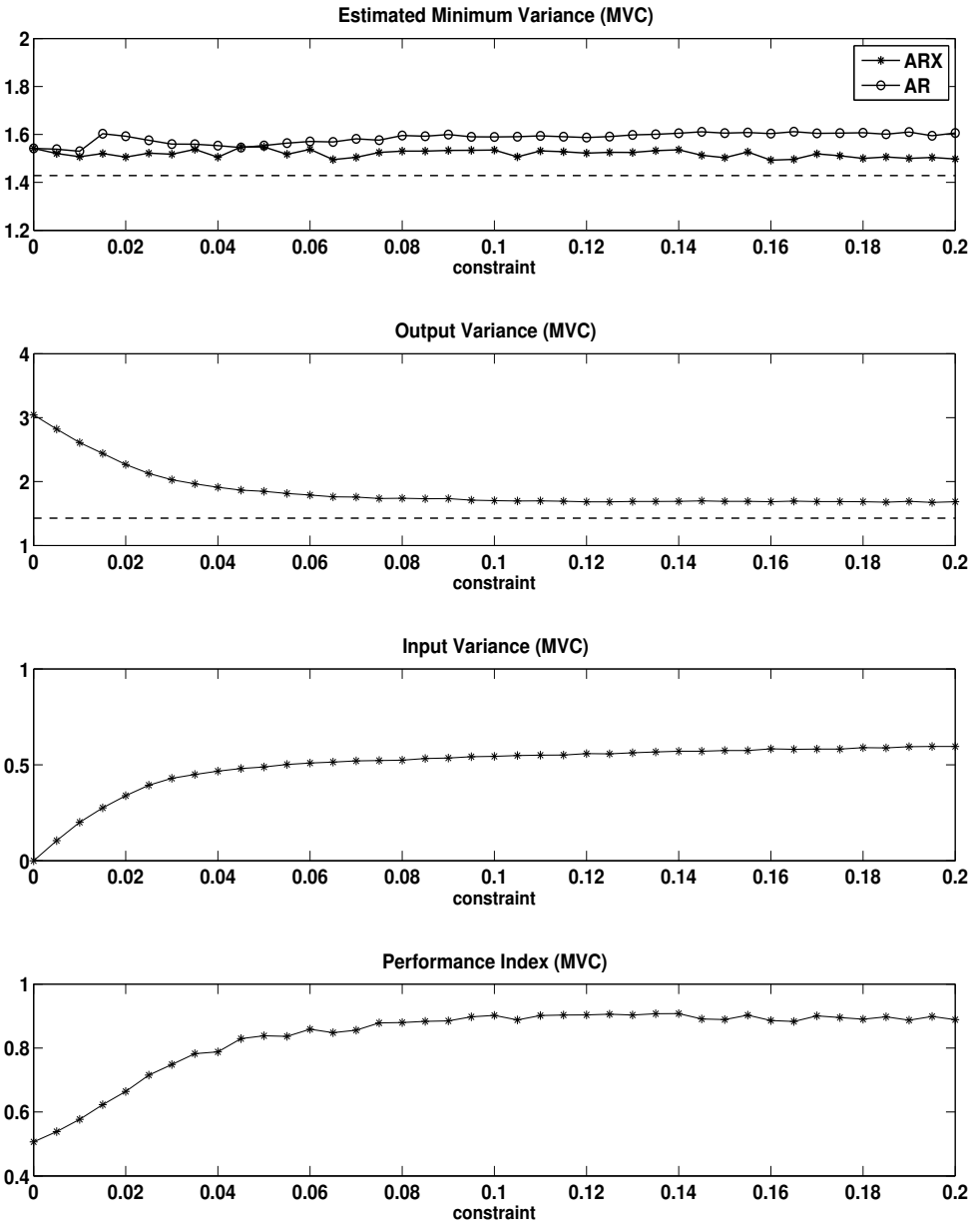


Figure 3.8: **CPA results for MVC controller with maximum-change constraints**

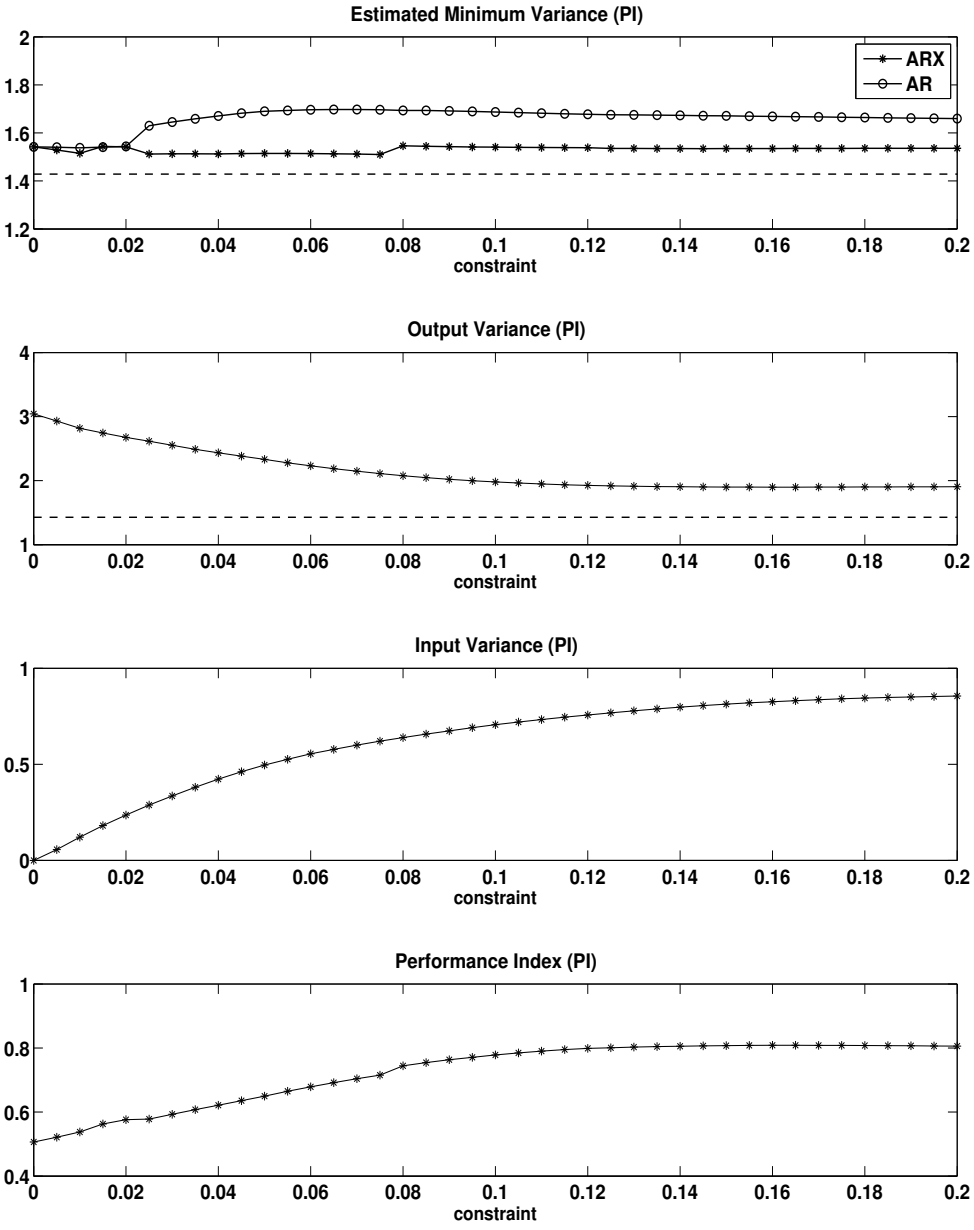


Figure 3.9: **CPA results for PI controller with maximum-change constraints**

realization of the true process, and the MVB is calculated for the actual realization. When the number of samples is increased to 100,000, then the calculated MVB value converges to the true value.

- The top panels in Figures 3.8 and 3.9 show the CPA-MVB values calculated using AR and ARX models. In comparison, inclusion of inputs gives better results in both PI and MVC cases. When inputs are not included, the results are not as good and depend on the controller type, i.e., the top lines in the first panels differ. These plots reveal the effectiveness of including all available information when forming the prediction equation. The calculated MVB is close to its theoretical value when the system is highly constrained, i.e., c is very small. In these constraint regions, the control action has a limited effect on reducing the output variance. The closed-loop dynamics are dominated by the disturbance and the inclusion of the control inputs in the model is of no value. When the constraint is relaxed, i.e., c is increased, the controller dynamics are more important relative to the disturbance dynamics and the inclusion of the constrained manipulated variable in the prediction equation is essential. In either case, the modelling effort is comparable to that of fitting an autoregressive model to the output alone.

- Decreasing trends in output variance and increasing trends in input variance are observed as the constraint is relaxed, as expected. When the maximum-change constraint c is 0, the actuator is held constant; when the maximum change constraint c is large, the performance of the constrained MVC is close to that of an unconstrained MVC system. For both constrained MVC and PI control asymptotes are reached for the output variance. The dotted lines show the theoretical MVB for the unconstrained MVC. In this example the MVC controller gives better performance, i.e.,

smaller output variances, with less control effort than the PI controller for any given constraint limit c over 0.05. Note that this is not generalizable. For this example, re-tuning the PI controller does not improve its performance; there are many 'reasonable' values for the gain and integral time that lead to worse performance than no control at all. Moreover, the variance of the input and output monotonically approach their limits. Again, this is not generalizable. Moden and Soderstrom [104] have established conditions for the one-step constrained MVC with a fixed penalty parameter to have monotonic behaviour.

- The last panels of Figures 3.8 and 3.9 present performance index plots, and the impact of the constraint on closed-loop performance can be clearly seen, especially when the imposed constraint is small. Jelali [97] has commented that the control performance can be claimed to be *practically optimal* when the performance index is great than 0.8. Note that, due to the imposed hard constraint on the actuator, performance index does not approach to 1 even under MVC with no constraints. From a production standpoint, both controllers are optimal when the constraint is relaxed, e.g., any c exceeding 0.12. The controllers are already operating very close to minimum variance, and retuning the controllers will not bring appreciable performance benefits. For this rolling-mill analysis, further potential performance improvement may be achieved via modifying the process itself, for example, 1) reducing the process delay, or 2) reducing the disturbance variability, or 3) implementing a feedforward controller.

- Moden and Soderstrom [104] have introduced a diagram to describe the input and output variances for stochastic systems, which is used for analysis of regulators for stochastic systems from an energy saving or profit perspective. This idea has been employed in drawing Figure 3.10. Two curves for the MVC and PI controllers show

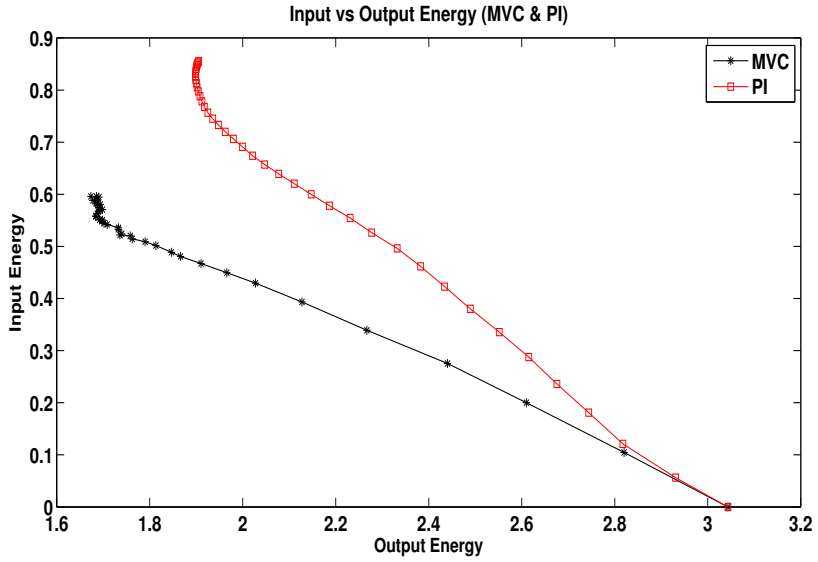


Figure 3.10: **Input-output energy plot**

the variance of the manipulated variable (input energy) as function of the variance of controlled variable (output energy). Note that by letting the value of output energy be the same, the value of input energy associated with the MVC controller is smaller than the one associated with the PI controller, which implies that the MVC controller in this example is more effective than the PI controller from an energy saving standpoint.

Similar results are obtained when dead-band constraints on the control action are applied (not shown here). The calculated MVB indicates that feedback control alone will not enable the process to meet its quality constraints and feedforward control must be implemented, as in standard industrial practice.

3.6 Dealing with Differences in Sampling Rates - Resampling

In rolling mill control systems the sampling and control rates are typically in the millisecond range because of the high-speed nature of the process. As a result, large amounts of data can be obtained. To reduce data storage costs and to free up computing cycles for real-time control systems, process variables are sometimes down-sampled, wherein output measurements may be recorded less frequently than control actions. For example, a mill controller receives observations and computes control actions at 1 ms intervals, and output measurements are only written to storage at 5 ms intervals. It is important to remember that the controller is seeing output measurements every 1 ms and computing actions on this basis, however the output observations are being recorded in the data historian less frequently.

The challenge is how to obtain meaningful estimates of the CPA-MVB in this scenario. There are two general approaches: 1) use only output measurements to construct an appropriate estimate of the CPA-MVB, or 2) utilize controller inputs as well to construct this estimate. In both cases, we cannot theoretically provide as good an estimate as when down-sampling is not used. It is difficult to predict the loss of accuracy a-priori by using down-sampled data. However, the examples shown below demonstrate that very credible estimates can be obtained. The use of down-sampled data has been studied in the econometrics literature to study prediction effectiveness [105, 106, 107, 108]. Again, the research findings support similar conclusions to what are found in this thesis. The use of down-sampled data for determining the optimal control interval is discussed in [109].

We first note that the autocorrelation test for minimum-variance performance can be applied when the output is sampled less frequently than the control interval. Recall that if a process is achieving minimum-variance performance, its autocorrelation function is a moving average process of order $(b-1)$, where b is the discrete-time delay that is given by:

$$b = 1 + \text{int}(\frac{T_d}{T})$$

where $\text{int}(\cdot)$ denotes the integer portion of the requisite term, T_d is the process delay in time, and T is the control interval. If a process is moving average of order $(b-1)$ at control interval T, then the down-sampled process at a sampling interval of KT is moving average of order $\text{int}(\frac{b-1}{K})$, $K \geq 1$ [106, 110].

We will consider the situation where we wish to estimate the CPA-MVB where the controller is implemented at control interval T, the manipulated variables are all recorded at this interval, and the controller variable is sampled at integer multiples of this interval, denoted by KT, $K > 1$. Note that the proposed idea can be extended to a process with both feedforward and feedback control systems. We need to ascertain whether this data will enable the construction of the minimum-variance forecast error at horizon $(b-1)$, where the forecast horizon refers to the control interval. Intuitively the answer to this question is no. One would anticipate a loss in forecast efficiency or accuracy. This problem has been studied extensively in the econometrics literature under the framework of *systematic sampling* and *aggregation of time series*.

To develop these ideas, first consider the open-loop system:

$$\text{Process output} : y_t = G_P(q^{-1})u_t + D_t$$

$$\text{Disturbance} : D_t = G_D(q^{-1})a_t$$

where the random shocks a_t for the disturbance are i.i.d. $(0, \sigma_a^2)$. To justify the approaches that we will take, we need to understand how the system dynamics and disturbances are affected by down-sampling. This is sometimes referred to as decimation in the signal-processing literature. The notation is that the time index, $\{t = 0, 1, 2, \cdots\}$, refers to the input data collected at the control interval, and $\{Kt = 0, K, 2K, \cdots\}$ refers to the sample times at which the output data is recorded.

Consider a disturbance model D_t that can be described as the output of an Autoregressive Integrated Moving Average - ARIMA(p,d,q) process:

$$D_{Kt} = \frac{\theta(q^{-1})}{\phi(q^{-1})(1-q^{-1})^d} a_{Kt} \tag{3.5}$$

It is known that the time-series representation at sampling intervals Kt can be written in terms of p lagged values of $\{D_{Kt}, D_{Kt-K} \cdots\}$ and $q + (K-1) \times (p+d)$ values of past shocks $\{a_{Kt}, a_{Kt-1} \cdots\}$, as [106]:

$$D_{Kt} = \frac{\zeta(q^{-1})\theta(q^{-1})}{\phi^\star(q^{-K})(1-q^{-K})^d} a_{Kt} \tag{3.6}$$

$\zeta(q^{-1})$ and $\phi^\star(q^{-K})$ are obtained from the roots of the polynomial $\phi(q^{-1})$ and ∇. First, factor $\phi(q^{-1})$ as:

$$\phi(q^{-1}) = \prod_{i=1}^{p} (1 - \beta_i q^{-1})$$

then

$$\phi^\star(q^{-K}) = \prod_{i=1}^{p} (1 - \beta_i^K q^{-K})$$

and

$$\zeta(q^{-1}) = \left(\prod_{i=1}^{p} \left(\frac{1 - \beta_i^K q^{-K}}{1 - \beta_i q^{-1}} \right) \right) \left(\frac{1 - q^{-K}}{1 - q^{-1}} \right)^d$$

To verify that $\zeta(q^{-1})$ is a polynomial in q^{-1} of order $(K-1) \times (p+d)$, we note that:

$$\frac{1 - \beta_i^K q^{-K}}{1 - \beta_i q^{-1}} = 1 + \beta_i q^{-1} + \beta_i^2 q^{-2} + \cdots + \beta_i^{K-1} q^{-(K-1)}$$

The model described by Equation (3.6) expresses the output in terms of down-sampled disturbance values $\{D_{Kt}, D_{Kt-K} \cdots\}$, and random shocks that occur at each sampling interval $\{Kt, Kt-1, \cdots\}$ Using fundamental results from time-series analysis, it is known that $\zeta(q^{-1})\theta(q^{-1})a_{Kt}$ can be expressed as:

$$\zeta(q^{-1})\theta(q^{-1})a_{Kt} = \theta^{\star}(q^{-K})a^{\star}_{Kt}$$

where $a^{\star}_{Kt}$ is i.i.d. $(0, \sigma^2_{a\star})$, and is only defined at sampling intervals $\{Kt, Kt-K, \cdots\}$. Note that $a^{\star}_{Kt}$ is not obtained by simply sampling the white noise sequence a_t. The parameters of $\theta^{\star}(q^{-K})$ and $\sigma^2_{a\star}$ are nonlinear functions of the parameters of $\phi(q^{-1})$, ∇^d, $\theta(q^{-1})$ and σ^2_a and are obtained through a straightforward application of a spectral-factorization algorithm. The disturbance is then represented at the down-sampled interval as:

$$D_{Kt} = \frac{\theta^{\star}(q^{-K})}{\phi^{\star}(q^{-K})(1-q^{-K})^d} a^{\star}_{Kt} \tag{3.7}$$

The polynomials $\phi^{\star}(q^{-K})$ and $\theta^{\star}(q^{-K})$ are of order $(p^{\star}, q^{\star})$ in the shift operation q^{-K} [106, 107], where:

$$p^{\star} = p \qquad \text{and} \qquad q^{\star} \leq \text{int}([p+d] + [q-p-d]/K)$$

Furthermore, $\sigma^2_{a^{\star}} > \sigma^2_a$. It is more convenient to write Equation (3.7) in the equivalent form:

$$D^{\star}_{t^{\star}} = \frac{\theta^{\star}(q^{-1})}{\phi^{\star}(q^{-1})(1-q^{-1})^d} a^{\star}_{t^{\star}} \tag{3.8}$$

where $D^{\star}_{t^{\star}}, t^{\star} = \{0, 1, 2, \cdots\}$ represents the disturbance at sampling interval $\{Kt, Kt-K, Kt-2K \cdots\}$, $a^{\star}_{t^{\star}}$ represents the i.i.d random sequence at the same sampling intervals, and the backshift operator q^{-1} is understood to operate as follows:

$$q^{-1}D^{\star}_{t^{\star}} = D^{\star}_{t^{\star}-1} = D_{Kt^{\star}-K} \qquad \text{and} \qquad q^{-1}a^{\star}_{t^{\star}} = a^{\star}_{t^{\star}-1} \neq a_{Kt^{\star}-K}$$

It is known that the effect of less-frequent sampling may make the underlying disturbance structure unobservable due to aliasing effects [108]. This is more pronounced when the more rapidly-sampled series has periodic components. The forecasting efficiency of less-frequently sampled time series was studied in [107]. In the examples studied, there was little loss in forecasting efficiency when the sampled series was "sufficiently long".

To include the effects of down-sampling on the dynamics, assume that the linear dynamics can be written in the form:

$$G_p(q^{-1}) = \frac{\omega(q^{-1})}{\delta(q^{-1})} q^{-b}$$

where the orders of the numerator and denominator are (r, s), respectively. $\delta(q^{-1})$ is a monic polynomial. The overall down-sampled model is of the form:

$$y_{Kt} = \frac{\omega'(q^{-1})}{\delta^\star(q^{-K})}) u_{Kt-b} + D_{Kt}$$

where:

$$\omega'(q^{-1}) = \omega(q^{-1}) \prod_{i=1}^{s} (1 + \alpha_i q^{-1} + \alpha_i^2 q^{-2} + \cdots + \alpha_i^{K-1} q^{-(K-1)})$$

$\delta(q^{-1})$ has been factored as:

$$\delta(q^{-1}) = \prod_{i=1}^{s} (1 - \alpha_i q^{-1}) \qquad \text{and} \qquad \delta^\star(q^{-K}) = \prod_{i=1}^{s} (1 - \alpha_i^K q^{-K})$$

The order of numerator dynamics is inflated from r to $r + (K-1)s$. $\omega'(q^{-1})$ can be written as:

$$\omega'(q^{-1}) q^{-b} = \sum_{i=0}^{K-1} \omega_i'(q^{-K}) q^{-(b-i)}$$

where each of the terms, $\omega_i'(q^{-K})$, is a polynomial at most of $r^\star$, where:

$$r^\star \leq \begin{pmatrix} s - 1 + \text{ceil}(\frac{r-s+1}{K}), & s \geq 1 \\ \text{ceil}(\frac{r+1}{K}) - 1, & s = 0 \end{pmatrix}$$

ceil$(\cdot)$ denotes the ceiling function, or the least integer greater than or equal of the argument $(\cdot)$. The down-sampled output can be written as:

$$y^{\star}_{t^{\star}} = \frac{\sum_{i=0}^{K-1} \omega'_i(q^{-K})}{\delta^{\star}(q^{-K})} u_{Kt^{\star}-b-i} + D^{\star}_{t^{\star}} \tag{3.9}$$

where $q^{-1}y^{\star}_{t^{\star}} = y^{\star}_{t^{\star}-1} = y_{Kt^{\star}-K}$,and $q^{-1}u_t = u_{t-1}$. The down-sampled output is expressed as a function of past down-sampled outputs and disturbances, and inputs that affect the process at the base-sampling interval.

Assume, without loss of generality, that the setpoint y_{sp} is 0. There are several approaches that can be used for CPA-MVB analysis.

- Approach 1: Fit an ARMA or AR model to the sampled outputs alone:

$$A^{\star}(q^{-1})y^{\star}_{t^{\star}} = C^{\star}(q^{-1})a^{\star}_{t^{\star}} \tag{3.10}$$

The notation is that the values at the more slowly-sampled intervals are denoted by the superscript $\star$, and that the backshift operator q^{-1} is understood to provide the appropriate time shift to the more slowly-sampled value. This approach assumes that the manipulated variable is generated by a linear feedback controller of the form $u_t = -G_C(q^{-1})y_t$ and is no constraints on the manipulated variable are imposed. In these circumstances, the closed-loop is described by:

$$y_t = \frac{D_t}{1 + G_P(q^{-1})G_C(q^{-1})} = \psi(q^{-1})a_t$$

where $\psi(q^{-1})$ is ratio of polynomials in the back-shift operator. The down-sampled system will be represented by the theory for time-series only models presented earlier, and a regression model of the form Equation (3.10) is the appropriate model for estimation.

- Approach 2: Fit an ARMAX or ARX model using the actual controller inputs:

$$A^{\star}(q^{-1})y^{\star}_{t^{\star}} = \sum_{i=0}^{K-1} F_i(q^{-1})X_{i,t^{\star}} + C^{\star}(q^1)a^{\star}_{t^{\star}}$$

  where $X_{i,t^{\star}}$ is the exogenous sequence $X_{i,t^{\star}} = \{u_{Kt^{\star}-b-i}\}$. This regression model follows directly from Equation (3.9). In this approach all of the actual inputs are used. The parameters can be estimated in MATLAB^{TM} by treating this as an ARMAX or ARX model with K inputs. This method allows for the proper incorporation of the controller inputs, and does not presume that they are generated by a linear feedback controller. The disadvantage to this approach is that the model may be very much over-parameterized as K increases as seen from Equation (3.9).

- Approach 3: Fit an ARMAX or ARX model using an aggregated or averaged input:

$$A^{\star}(q^{-1})y^{\star}_{t^{\star}} = B^{\star}(q^{-1})u^{\star}_{t^{\star}-b^{\star}} + C^{\star}(q^{-1})a^{\star}_{t^{\star}}$$

  where $b^{\star} = 1 + \text{int}(\frac{b-1}{K})$. $u^{\star}_{t^{\star}}$ is the averaged input:

$$u^{\star}_{t^{\star}} = \frac{1}{K}\sum_{i=0}^{K-1} u_{Kt^{\star}-Kb^{\star}+m'_K+i}$$

  with $m'_K = Kb^{\star}_K - b$.

  This approximate regression model follows directly from Equation (3.9) by inserting $u^{\star}_{Kt}$ for the average of K controller inputs. When the re-sampled delay $b^{\star}$ is an event multiple of the original delay $b, Kb^{\star} = b, m'_K = 0$ and $u_{t^{\star}}$ is the average of the controller inputs in the interval $[Kt^{\star} - b, Kt^{\star} - b + K - 1]$. The concept of using the average value of the inputs is similar to the concept of aggregating (or summing) time-series values. As one might anticipate, this scheme will work best when the manipulated variable is not changing rapidly, otherwise model inaccuracy will be a potential problem.

In each of these cases the estimated CPA-MVB can be written as:

$$\sigma^2_{mv\star} = (1 + [\psi^\star_1]^2 + \cdots + [\psi^\star_{b^\star-1]}]^2)\sigma^2_{a^\star}$$

where $\psi^\star_j,\ j = 1 \cdots b^\star - 1$ are calculated from the series expansion of $\psi^\star(q^{-1}) = \frac{C^\star(q^{-1})}{A^\star(q^{-1})}$. The methods described in this section can only be used when the output sampling interval is an even multiple of the control interval. When the desired prediction horizon does not align with any available prediction horizons in the slowly-sampled system, the CPA-MVB can be estimated by interpolation of CPA-MVB values at nearby prediction horizons. When the process time delay is less than the down-sampled output period from the data historian, it is not possible to obtain a reliable estimation of the CPA-MVB.

3.6.1 Simulation Results

To illustrate the proposed approaches, the system described in Section 3.5 is used to study the down-sampling effect on the CPA-MVB. The output is now sampled at $K = 1, 2, 4, 8$. Note that the base sampling rate is 1 ms, and K is the multiple of the base sampling period, which is dimensionless. Before analyzing the results of the simulation study, we consider the theoretical results where the manipulated variable is generated by a linear feedback controller, and the case where all of the previous values of the manipulated variable are used in the regression analysis. The closed-loop transfer function is calculated for $K = 1$ and the sampled versions for $K = 2, 4$, and 8 using the method described. The spectral-factorization algorithm of Wilson [111] is used. Clearly, the results in Table (3.1) show a modest loss of efficiency.

In the second scenario, the sampled version of the disturbance model is required for $K = 2, 4$, and 8. For an AR(1) model, the sampled version is also an AR(1)

Table 3.1: **PI control with no constraints imposed (Theoretical CPA-MVBs)**

Multiple of base sampling time (1 ms)	1	2	4	8
CPA-MVB	1.42	1.42	1.44	1.50

model, with autoregressive parameter, ϕ^K [107]. The variance of the driving force $a^{\star}_{Kt}$ is readily calculated as $\frac{1-\phi^{2K}}{1-\phi^2}\sigma_a^2$. For an AR(1) model, in theory, there is no loss in efficiency by down-sampling.

Tables 3.2-3.5 present the simulation results for the PI and MVC controllers with maximum-change constraint $c = 0.1$ imposed as well as the hard constraint of 2 on the actuator as described in in Section 3.5.7. For this constraint value, approximately 40% of the control moves are constrained. In addition to a sample size $N = 10,000$, results are reported for $N = 100,000$. Due to the characteristics of fast sampling in a rolling process, this latter value of N is several times larger than the one encountered in a typical data analysis. The effect on sample size on the calculatged value for the CPA-MVB is also investigated.

Table 3.2: **PI control with maximum-change & hard constraint** ($N = 10,000$)

Multiple of base sampling time (1 ms)	Theoretical CPA-MVB	AR Model (Approach 1)	ARX Model (Approach 2)	ARX Model (Approach 3)
1	1.43	1.69	1.54	1.54
2	-	1.70	1.52	1.52
4	-	1.69	1.54	1.52
8	-	1.67	1.60	1.57

The results are shown in Tables 3.2-3.5, and the following observations can be made:

- As expected, the two approaches using ARX models, which include the manipulated variable, outperform the AR model. This indicates that inclusion of the input information in formulating the predictive models essential.

Table 3.3: **PI control with maximum-change & hard constraint** ($N = 100,000$)

Multiple of base sampling time (1 ms)	Theoretical CPA-MVB	AR Model (Approach 1)	ARX Model (Approach 2)	ARX Model (Approach 3)
1	1.43	1.55	1.43	1.43
2	-	1.56	1.43	1.43
4	-	1.56	1.44	1.43
8	-	1.55	1.46	1.46

Table 3.4: **MVC control with maximum-change & hard constraint** ($N = 10,000$)

Multiple of base sampling time (1 ms)	Theoretical CPA-MVB	AR Model (Approach 1)	ARX Model (Approach 2)	ARX Model (Approach 3)
1	1.43	1.59	1.53	1.53
2	-	1.59	1.52	1.52
4	-	1.59	1.54	1.55
8	-	1.63	1.63	1.59

Table 3.5: **MVC control with maximum-change & hard constraint** ($N = 100,000$)

Multiple of base sampling time (1 ms)	Theoretical CPA-MVB	AR Model (Approach 1)	ARX Model (Approach 2)	ARX Model (Approach 3)
1	1.43	1.48	1.43	1.43
2	-	1.48	1.43	1.44
4	-	1.50	1.44	1.45
8	-	1.55	1.49	1.51

- The ARX approach using averaged inputs gives essentially the same results as the ARX implementation using all of the inputs. In theory one might have anticipated the latter approach to do better. For this constraint value, the inputs are constrained, and the average values of the inputs over the down-sampled interval do not differ much from the actual values. One might also anticipate that the over parametrization of the model that uses all of the input data detracts from the performance of this model.

- The large sample results, $N = 100,000$, show that the CPA-MVB does not converge

to the theoretical value 1.43, when $K > 1$ and the manipulated variable information is excluded from the model. When the inputs are included, the results only show a small improvement compared to the case when $N = 10,000$. In addition, while the sampling interval is large, e.g., $K = 8$, the use of aggregated or averaged input (Approach 3) still gives more accurate results than when all controller inputs are used (Approach 2).

Sometimes mill control systems sample an output more frequently than a control interval. In this instance, it is important to remember that the CPA-MVB evaluates what a controller sees and does. In other words, a controller receives observations and computes control actions only at control intervals. Unless a controller has some special output averaging element that is used for collecting observations, a controller is essentially using output observations at control intervals to determine next moves. Consequently, more frequent output data should be down-sampled to obtain output observations at the same intervals as the control actions. The conventional CPA-MVB algorithm can then be readily applied to these down-sampled output observations.

3.7 Conclusion

Variance-based controller performance assessment (CPA-MVB) has been applied successfully to metal-rolling processes. These methods have great utility for the efficient allocation of resources for controller maintenance and development. Three modifications to implementing CPA in rolling mills have been resolved. While these modifications are not limited to rolling processes, they are major concerns in sheet-forming processes. In order to address changing speed during a rolling batch, the length-based CPA first introduced by Jelali [97] has been fully developed and evaluated. Next, the

incorporation of constraints - either on maximum changes to control inputs, or deadbands requiring a minimum change - has been described, and a solution for computing the CPA-MVB has been proposed. A third problem arise when the input and output are not sampled at the basic control intervals. The CPA-MVB can be estimated using simple modifications to the basic methodology. Analogies from the econometrics literature were used to develop the background theory for the latter two issues. The proposed approaches are illustrated using a range of examples using mill and simulation data.

One of the goals of this thesis is to draw more attention to the increasing need for controller and process monitoring in sheet-forming processes. While the applications have addressed feedback control systems with or without feedforward control, it is possible to extend these solutions to systems having more advanced control configurations including cascade and model-predictive control.

Chapter 4

Overview of Methods for Process Analysis

Process control, monitoring and diagnostics play an important role in ensuring that manufacturing industries operate safely, ensure environmental compliance, meet consumer demands, and are financially viable. Process understanding is central to meeting these objectives. Process analysis techniques are employed to answer questions concerning process understanding, and to monitor performance including the effectiveness of control systems. The phrase "process analysis" may refer to distinct terminologies in different areas, e.g., process modelling in the chemical industry and time series analysis in economics. Research is continuously underway to develop new analysis techniques, and to apply these techniques to industrial processes. These techniques can be data driven (empirical - e.g., multivariate statistical techniques), or mechanistic (making use of parameterized fundamental models). In Chapter 2, an overview was provided of techniques for assessing the performance of controllers, using data from closed-loop operation. In the current chapter, the focus is process

analysis, and the data are either open-loop or closed-loop. The underlying process is assumed to be continuous time, represented in either continuous time or discrete time (sampled data) forms. The purpose of the chapter is to provide a brief overview of approaches to process analysis and monitoring, in order to set the stage for the later chapters detailing the Singular Spectrum Analysis (SSA) approach.

Parametric approaches to process analysis and monitoring involve examining the values of parameters in models to infer changes in the condition of a process (e.g., fouling in a heat exchanger), and changes in dynamic behaviour (e.g., onset of oscillatory behaviour). The models are often mechanistic, derived from first-principles, however it is also possible to use parametric empirical models as well (e.g., transfer functions). The principle of parsimony is important - models should contain sufficient complexity to capture the important physical behaviour, without being unnecessarily complex. Overly complex models become difficult to solve numerically, particularly in real time, while excessive parameterization can negatively influence the precision of model predictions. The principle of parsimony is well established in time series model (see for example, [112]).

Non-parametric techniques do not require the specification of dynamic structures, although a few parameters, such as a maximum lag, must be sometimes specified. These techniques can be applied readily to plant data without the development of model equations. Amongst this class of techniques are multivariate statistical techniques (e.g., Principal Component Analysis (PCA), and Partial Least Square (PLS)) [113], frequency domain methods [114], as well as impulse response methods.

4.1 Parametric Modelling Methods

Parametric methods in process analysis characterize a system using model equations, with a limited set of parameters. The most well-known approaches in the process control community consist of mechanistic (first principles) modelling and system identification techniques applied to transfer function models.

4.1.1 Mechanistic Modelling

Mechanistic modelling, also described as first principles modelling or fundamental modelling, refers to physical modelling with differential or differential-algebraic equations. These equations are derived from conservation balances (e.g., mass, energy, and momentum balances) and constitutive equations (e.g., special physical relationships such as gas laws) [115]. Most applications lie in the field of engineering and science. While mechanistic modelling requires more effort up front, it does have many advantages including: 1) first principle models generally require less data, compensating instead by the specification of dynamic model structures; 2) fewer parameters need to be estimated; 3) models are more robust to process changes, and 4) models are easily converted to transfer function and state-space models for various purposes, e.g., control design, process monitoring, and optimization. Note that this modelling technique typically involves prior system knowledge from the engineering science fields such as transport phenomena, chemical kinetics, and thermodynamics. The underlying physical and chemical principles may not be thoroughly understood for a given process, making first principle models being less often available for industrial use. Model complexity can have an impact on numerical solution, as well as parameter estimation, both of which are concerns in practical use.

Parameters in mechanistic models typically appear nonlinearly in chemical process models, necessitating the use of nonlinear regression techniques (e.g., Least Squares, Maximum Likelihood, or Bayesian estimation [116]). Numerical optimization techniques are required, ranging from Gauss-Newton iteration to constrained nonlinear programming techniques (e.g., conjugate gradient or interior point methods [117, 118]). Basis function approaches, in which functions such as B-splines [119] are used to formulate the solution of differential equation models, can also be used to help regularizing the numerical solution and avoiding numerical difficulties.

4.1.2 System Identification Techniques

System identification refers to the process of estimating dynamic models, typically from input-output data [114, 120]. Strictly speaking, system identification comprises both parametric and non-parametric techniques, however for the purposes of the discussion in this chapter, we use the term to refer to parametric models. System identification models in the process control community are very often transfer function models in either discrete- or continuous-time forms, but state-space models are also included. In statistical society system identification techniques often refers to time series models (also named Box-Jenkin models [112]), which can be easily converted to transfer functions and state-space models. The advantage of using these parametric methods are: 1) less detailed information about process mechanisms is required, usually comprised of the dynamic order of a model; 2) only measured observations are required to identify mathematical models.

One common assumption for all of these methods is the existence of causal relationship between inputs and outputs. Prior knowledge of processes is needed to

specify model structures. To obtain good models, system identification methods require process data of high quality; accordingly, experimental design techniques are often introduced to prepare data for system identification, e.g., pseudo-random binary sequences (PRBSs), or other pulse response tests in open-loop. Experimental data for pulse experiments (pseudo-random binary or other types of pulse tests) can also be used to estimate non-parametric finite impulse response (FIR) and finite step response (FSR) models, and transfer function models estimated from data can also be used to generate FIR and FSR models. These models can be used for estimating various types of controllers, including Proportional-Integral-Derivative (PID) controllers and model predictive controllers (MPCs). Experimentation can be in either open- or closed-loop operation, depending on whether process regulation close to a setpoint is required (e.g., in the case of open-loop unstable processes, or processes requiring control to meet product specifications).

4.2 Non-parametric Modelling Methods

Non-parametric models differ from parametric models by not requiring a model structure (or prior process knowledge) to be specified (e.g., dynamic order). With no structure being specified, non-parametric models are characterized by large numbers of parameters (e.g., impulse response coefficients), in contrast to parametric models which have many fewer parameters to estimate. Within a statistical context, non-parametric models refer to distribution-free models in which it is not necessary to specify the nature of a probability distribution for a population being studied (e.g., Normal or Weibull). In this instance, non-parametric methods are sometimes referred to as distribution-free methods. Within system identification, non-parametric models

are sometimes referred to as black box models (e.g., see [114]), while hybrid models - consisting of both mechanistic elements and empirical elements - are sometimes referred to as grey-box models.

Prior process knowledge is often required by parametric methods, e.g., the dynamic order of a process, or the range of time delays. Non-parametric methods do not require model structure specification (beyond maximum lag information), and offer the following benefits: 1) they are model-free, i.e., no prior knowledge is needed for implementation. With limited knowledge of the underlying structure of a process, non-parametric methods are still applicable. Many non-parametric methods are easy to apply and to understand. Additionally, we are not constrained to making as many assumptions about the population that we are working with as what we have to make with a parametric method. 2) They are data-driven and can be more readily applied to process data without the development of model equations. Highly automated industrial processes provide massive amounts of routine process data, which can be either open- or closed-loop, and there is a strong need to acquire useful process information from recorded plant data. Non-parametric methods are growing in popularity and influence for the reasons discussed above. In this section non-parametric methods are discussed briefly in the scope of time and frequency domains.

4.2.1 Multivariate Analysis Techniques

In the time domain, multivariate analysis refers to any statistical technique examining multiple variables for process analysis purposes. One common use of multivariate analysis is to extract valuable process information through investigating the observations in a lower dimension space, i.e., the projection space formed by the original

process data. Often a projection space is of reduced dimension and can be constructed via a small number of basis functions. That is why many multivariate analysis techniques are referred to as reduced-rank techniques. Typical reduced-rank techniques can be classified into two groups in terms of different types of basis functions [113]: 1) predefined (user-defined) basis functions, e.g., square impulse functions [121], orthogonal polynomials [122, 123], splines [124], Fourier series [121], and wavelets [125]; 2) data-driven basis functions based on $Karhunen-Loève$ decomposition, e.g., Multiple Linear Regression (MLR) [126], Principal Component Analysis (PCA), and Partial Least Squares (PLS) [72]. These latter techniques decompose the covariance structure of the data to produce basis vectors that are linear combinations of the original process variables. Note that both open- and closed-loop data can be analyzed via multivariate analysis techniques.

4.2.2 Frequency Domain Methods

In the past few decades, two different but mathematically equivalent approaches, the time domain and frequency domain approaches, have been used for the current development of time series analysis. Ljung [114] demonstrates that time domain and frequency domain identification approaches are complementary. While time domain approaches are generally associated with parametric models, frequency domain approaches are associated with spectral analysis. Spectral analysis essentially decomposes the variation in a time series or process data into different frequency components. Readers are referred to some pioneering work in spectrum estimation done by Bartlett [127] and Blackman and Tukey [128]. The most well-known frequency domain methods are Fast Fourier Transform (FFT) and Discrete Fourier Transform

(DFT) applied to data. Generally, it can be assumed that frequency-dependent variations in many data demonstrate the primary characteristics of interest. Furthermore, a good understanding of periodic or systematic sinusoidal variations may yield more information about underlying physical mechanisms, which makes frequency domain methods more useful and practical.

4.3 Singular Spectrum Analysis

Singular spectrum analysis (SSA) is a non-parametric technique for process analysis that offers an alternative to established non-parametric techniques (e.g., multivariate statistical techniques such as PCA) for building models and performing process analysis. Conventional approaches such as PCA apply a covariance decomposition to data matrices consisting of lagged time series data for one or more process variables under consideration. Barkhuizen [7] have noted that a single time series typically contains information about key dynamics and properties of a process, because the time-evolution of the system leaves a footprint in the process output. In this thesis, instead of multivariate analysis, a time series analysis tool, Singular Spectrum Analysis (SSA), is proposed for process analysis. Recall that the property of dimensionality reduction in multivariate analysis often helps in dealing with masses of data. Similar to other data-driven multivariate analysis techniques, e.g., Principal Component Analysis (PCA), and Partial Least Square (PLS), SSA is also a reduced-rank technique based on $Karhunen - Loève$ decomposition of a specially-structured data matrix.

Aldrich and co-workers [5, 6, 7] have outlined clearly the feasibility of applying SSA analysis to plant data. Barkhuizen [7] has illustrated the application of SSA

to examples in the chemical and metallurgical engineering systems. The current thesis reports the promotion of the techniques to chemical process analysis. The promising properties and successful applications of the SSA technique detailed below motivate its use for performing process analysis and identifying the key features of the process through examining system outputs, and for performing minimum-variance-based controller performance assessment.

4.3.1 Brief Historical Remarks

In [129] it is claimed that the first appearance of Singular Spectrum Analysis (SSA) can be dated back to 1795 by Prony [130]. The first report of the basic SSA algorithm often refers to the publications by Pike [131], Broomhead and King [132, 133], and Fraedrich [134]. It is also worthwhile to mention the other version of SSA - the 'Caterpillar' method - developed by a research group at St. Petersburg University [135]. Golynandia et al. [129] states that the main difference between the Caterpillar and conventional SSA methods are in the research focus, instead of in the algorithm itself. While the Caterpillar SSA focuses attention on the separability concept, the conventional SSA emphasizes its dedication to the model building in signal-plus-noise structure. Regarding the widespread use of SSA, Vautard, Ghil, and their coworkers [136, 137, 138, 139] have published a series of SSA papers demonstrating the potential of SSA that have attracted significant attention.

The basic elements of SSA and some variations are described in detail in [140, 141]. Golyandina [129] provides an up-to-date summary of the SSA methodology, and additional details are provided in recent review papers in [142, 143, 144, 145]. The major research groups in the field of SSA are located in St. Petersburg University

in Russia, Cardiff University and Bournemouth University in the UK, and University of California, Los Angeles in the USA, where Professors Nina Golyandina, Anatoly Zhigljafsky, Hossein Hassani, and Michael Ghil are the key active contributors.

4.3.2 Advantages of SSA

Hassani [142, 143] summarizes a number of advantages of SSA that motivate its use:

1. SSA is a nonparametric technique. There are no statistical or distributional assumptions in the application of SSA, such as stationarity of an analyzed series and normality of a residual.

2. SSA can deal with seasonal, stationary, and non-stationary time series. Unlike conventional Autoregressive Integrated Moving Average (ARIMA) models, differencing is not required for SSA. SSA can accommodate non-stationary, integrating time series.

3. SSA decomposes a series into interpretable components, such as trends, harmonic series, and residuals. Additionally, these decomposed components can be used for time-series analysis, data forecasting, and process monitoring.

4. SSA can be applied to short time series, meaning that process information can be extracted without requiring many data. The requirement of a small data size in analysis is especially beneficial to nonlinear dynamics analysis.

   Process analysis is often considered as a prerequisite for modelling and forecasting [139]. With the fact that nonlinear processes are abundant in reality, the extension of successful linear dynamic analysis techniques to understand nonlinear processes is a natural move. Vautard comments in [138] that as least tens of thousands of

data points are required to get truly nonlinear information in nonlinear dynamics. Instead of a great number of quality lab experiment data, observations of physical systems are often applied to find numerical solutions of differential equations, ordinary and partial; however, shorter and noisier data in nature increase the difficulties encountered in applications. Traditional time series analysis techniques have been introduced to study nonlinear dynamic systems. In this thesis the data-adaptive SSA technique is promoted due to the success of applications in nonlinear geophysics systems [136].

Without prior knowledge of physical dynamics, SSA extracts reliable information from short and noisy data. Vautard et al. [138] demonstrates that SSA can provide nonlinear physical insights via using only few hundred data points typically available for geophysical and other natural systems. As well, given short data, SSA is proven to be an efficient noise reduction methodology, especially in the analysis of nonlinear processes with a large quasi-periodic component or an oscillation. Notice that singular spectrum analysis, similar to principal component analysis, is essentially a linear technique, which may result in overlooked nonlinear dynamics in analysis [7]. Nonlinear principal component analysis (NLPCA) and nonlinear singular spectrum analysis (NLSSA) are developed to broaden the application of data-adaptive methods in nonlinear dynamic systems [146, 147]. So far the applications of NLPCA and NLSSA have been applied mainly to geophysical time series.

4.4 Summary

This chapter presents a brief review of process analysis techniques, categorized into parametric and non-parametric groups. A state-of-the-art time series analysis technique, Singular Spectrum Analysis (SSA), is introduced, together with a brief overview of the history and advantages of this technique, and the motivation for using it to perform process analysis. In subsequent chapters, the SSA work will focus on: 1) understanding the algorithm from a filtering and frequency viewpoint; 2) applications in process analysis and controller performance assessment; 3) extension to two-dimensional analysis; and 4) industrial applications in the chemical process industries and sheet-forming processes.

Chapter 5

Filtering and Frequency Interpretations of SSA[3]

5.1 Introduction

Singular Spectrum Analysis (SSA) has proven to be a flexible method for the analysis of time-series data. Applications are reported in diverse areas such as climate change and geophysical phenomena [136, 138, 149], mineral processing [5] and telecommunication applications [150, 151]. The basic SSA method has been combined with the maximum entropy method (MEM) [152] and with multi-taper methods [153] to enhance the spectral analysis of data. Extensions to cope with missing data [154] and multi-scale applications have also been developed [155].

The basic elements of SSA were first reported in [131, 132]. Widespread use of SSA followed a series of papers by Vautard and Ghil [136] and Vautard et al. [138]. The

[3]This chapter has been published in *Physica D: Nonlinear Phenomena, 239:1958-1967, 2010* [148]. Please refer to the preface of the thesis for the nomenclature used in this Chapter.

monograph by Golynandia et al. [141] describes the basic algorithm plus a number of variations. A recent overview is given in [142].

The purpose of this paper is to give a number of interpretations of SSA from a signal processing perspective by addressing issues related to filtering interpretations, spectrum evaluation and recovery of harmonic signals. In particular, in the case where the trajectory matrix has many more rows than columns, the eigenvalues and eigenvectors are nearly identical to those of an associated symmetric Toeplitz matrix. The eigenvalues and eigenvectors of this latter matrix are highly structured [156, 157, 158]. These structured properties lead to a number of interesting filtering interpretations. Additionally, we note that the reconstruction phase in SSA can be interpreted as a forward and reverse filtering of the original data. This provides for a number of additional interpretations for the filtered series and their spectra.

The paper is organized as follows. In the next section we state the basic SSA algorithm and some variations. This is followed by filtering and spectral interpretations of the SSA algorithm. These interpretations make extensive use of symmetry properties of the eigenfilters that are used in the filtering. These in turn are derived from symmetry properties of the eigenvectors of the trajectory matrix. Two examples are then analyzed to illustrate the theoretical results.

5.2 Basic SSA and Some Variations

5.2.1 Basic SSA Algorithm

The basic SSA algorithm consists of the following steps [132, 136, 138].

1. Choose an embedded dimension K and define $L = N + 1 - K$, where N is the

number of observations in the time series.

K should be long enough to capture the transient response of a process (i.e., the memory of a process). For example, if there is prior operating experience indicating the settling time of the process, this can be used to come up with a suitable value for K. Alternatively, historical knowledge of the approximate time constant for the process can also be used, and K could be chosen to 4 time constants (for example). Another consideration that can be used is the knowledge of typical frequencies presented in the process. K should be sufficiently long to capture a full period of the slowest sinusoid.

These guidelines make reference to historical understanding of the process in a manner similar to the considerations often used to select a sampling period.

2. Form the $L \times K$ Hankel matrix $\boldsymbol{A}$ using mean-corrected data, $y_t,\ t = 1, 2, \cdots, N$.

$$\boldsymbol{A} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \cdots, \boldsymbol{y}_K] = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & \cdots & y_{K+1} \\ y_3 & y_4 & y_5 & \cdots & \cdots & y_{K+2} \\ \vdots & \vdots & \vdots & & & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & \cdots & y_{K+L-1} \end{bmatrix}$$

where $\boldsymbol{y}_i = (y_i, y_{i+1}, \cdots, y_{i+L-1})^T$. This matrix $\boldsymbol{A}$ is often referred to as the trajectory matrix. In most applications of SSA, $L > K$ [141, 136].

3. Determine the eigenvalues and eigenvectors of $\boldsymbol{A^T A}$. Denote the eigenvalues by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0$. For each eigenvalue λ_i there is a corresponding eigenvector $\boldsymbol{v}_i$.

$$(\boldsymbol{A^T A})\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \tag{5.1}$$

4. Define K new series, $\boldsymbol{w}_i = \boldsymbol{A}\boldsymbol{v}_i,\ i = 1, 2, \cdots, K$. Each series is of length L. Once the new series are constructed, the analysis then focuses on the new series, which are sometimes referred to as the latent variables. The individual series may be analyzed, or subsets may be grouped together.

   The utility of the method is derived from the following properties and interpretations of the eigenvalue analysis:

   (a) The eigenvectors $\boldsymbol{v}_i$ are orthonormal, i.e., $\boldsymbol{v}_i^T\boldsymbol{v}_j = 0$ $(i \neq j)$ and $\boldsymbol{v}_i^T\boldsymbol{v}_i = 1$.

   (b) The latent variables $\boldsymbol{w}_i$ are orthogonal, and

$$\|\boldsymbol{w}_i\|^2 = \boldsymbol{w}_i^T\boldsymbol{w}_i = (\boldsymbol{A}\boldsymbol{v}_i)^T\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{v}_i^T(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A})\boldsymbol{v}_i = \boldsymbol{v}_i^T\lambda_i\boldsymbol{v}_i = \lambda_i \tag{5.2}$$

   (c) Consequently,

$$\sum_{i=1}^{K}\boldsymbol{w}_i^T\boldsymbol{w}_i = \sum_{i=1}^{K}\boldsymbol{w}_i^T\sum_{i=1}^{K}\boldsymbol{w}_i = \sum_{i=1}^{K}\lambda_i \tag{5.3}$$

Often, the interesting features of a time series are found by analyzing the first few latent variables. A number of methods have been proposed to choose the number of latent variables for analysis. Most often, the construction of a scree plot [159], which is a plot of λ_i versus i, will indicate a knee or bend. This can be used to select the number of latent variables. Other methods have been proposed when the break points are not clear [160].

Scree plots are also useful for identifying harmonics in the data. As discussed in [132, 142, 136], if N and L are large enough, each harmonic results in two eigenvalues that are closely paired for a purely harmonic series. A harmonic component may produce a periodic component in the autocorrelation and partial autocorrelation function. However, the number of periodic components cannot be easily extracted from these functions. In addition, a slowly decreasing sequence of eigenvalues can

be produced by a pure noise series [141, 142]. These two observations suggest that a break or knee in the scree plot can be used to separate the signals that arise from harmonics and signals from noise or aperiodic components [141].

The eigenvalues of $\boldsymbol{A^T A}$ are most often calculated by undertaking a singular value decomposition (SVD) of $\boldsymbol{A}$. The right singular vectors of $\boldsymbol{A}$ are identical with the eigenvectors of $\boldsymbol{A^T A}$ and the eigenvalues of this latter matrix are the squares of the corresponding singular values of $\boldsymbol{A}$ [141, 161].

5.2.2 Variation: Toeplitz Approximation to $\boldsymbol{A^T A}$

$\boldsymbol{A^T A}$ is symmetric and positive semi-definite. It can be written as

$$\boldsymbol{A^T A} = \begin{bmatrix} \boldsymbol{y}_1^T \boldsymbol{y}_1 & \boldsymbol{y}_1^T \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_1^T \boldsymbol{y}_K \\ \boldsymbol{y}_2^T \boldsymbol{y}_1 & \boldsymbol{y}_2^T \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_2^T \boldsymbol{y}_K \\ \vdots & \vdots & & \vdots \\ \boldsymbol{y}_K^T \boldsymbol{y}_1 & \boldsymbol{y}_K^T \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_K^T \boldsymbol{y}_K \end{bmatrix}$$

In situations when $L >> K$, we have

$$\frac{1}{L}\boldsymbol{y}_1^T \boldsymbol{y}_1 \simeq \frac{1}{L}\boldsymbol{y}_2^T \boldsymbol{y}_2 \simeq \frac{1}{L}\boldsymbol{y}_3^T \boldsymbol{y}_3 \simeq \cdots \simeq \frac{1}{L}\boldsymbol{y}_K^T \boldsymbol{y}_K \simeq \frac{1}{N}\sum_{t=1}^{N} y_t^2 = c_0$$

$$\frac{1}{L}\boldsymbol{y}_1^T \boldsymbol{y}_2 \simeq \frac{1}{L}\boldsymbol{y}_2^T \boldsymbol{y}_3 \simeq \cdots \simeq \frac{1}{L}\boldsymbol{y}_{K-1}^T \boldsymbol{y}_K \simeq \frac{1}{N-1}\sum_{t=1}^{N-1} y_t y_{t-1} = c_1$$

$$\vdots$$

$$\frac{1}{L}\boldsymbol{y}_1^T \boldsymbol{y}_K = \frac{1}{L}\sum_{t=1}^{L} y_t y_{t-(K-1)} = c_{K-1}$$

where c_i is the sample autocovariance at lag i (We have previously assumed that the data have been mean corrected). Consequently

$$\frac{\boldsymbol{A^T A}}{L} \simeq \boldsymbol{C} = \begin{bmatrix} c_0 & c_1 & \cdots & c_{K-1} \\ c_1 & c_0 & \cdots & c_{K-2} \\ \vdots & \vdots & & \vdots \\ c_{K-1} & c_{K-2} & \cdots & c_0 \end{bmatrix}$$

where $\boldsymbol{C}$ is the sample covariance matrix of the observations. The sample autocorrelation matrix $\boldsymbol{R} = \boldsymbol{C}/c_0$ is often used instead of $\boldsymbol{C}$ for analysis. This is appropriate when the data have been centered and normalized [132, 139].

5.2.3 Variation: Hankel Approximation and Diagonal Averaging

A singular value decomposition of the matrix is undertaken

$$\boldsymbol{A} = \sum_{i=1}^{\mu} \boldsymbol{X}_i = \sum_{i=1}^{\mu} \sqrt{\lambda_i} \boldsymbol{u}_i \boldsymbol{v}_i^T \tag{5.4}$$

where $\mu = \min(L, K)$, λ_i and $\boldsymbol{v}_i$ are the eigenvalues and eigenvectors of $\boldsymbol{A^T A}$ as described in Equation (5.1), and $\boldsymbol{u}_i$ are the eigenvectors of $\boldsymbol{A A^T}$, i.e., the solution to [162]

$$(\boldsymbol{A A^T}) \boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i \tag{5.5}$$

The orthogonal vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are related by

$$\boldsymbol{A v}_i = \sqrt{\lambda_i} \boldsymbol{u}_i, \quad i = 1, 2, \cdots, \mu \tag{5.6}$$

where the singular values of $\boldsymbol{A}$ are $\sqrt{\lambda_i}$, i.e., the square root of the eigenvalues of $\boldsymbol{A^T A}$. Clearly, $\boldsymbol{u}_i = \boldsymbol{w}_i / \sqrt{\lambda_i}$.

Each of the $\boldsymbol{X}_i$ in Equation (5.4) is of rank 1. A new series $\boldsymbol{x}_i$ of length N is

reconstructed by averaging each of the N anti-diagonals in $\boldsymbol{X}_i$. Attention is then focused on the new series $\boldsymbol{x}_i$, or groupings of these series. Guidelines for grouping of variables are typically based on the clustering and separation of the eigenvalues. A 'separability index' has been proposed in [142] to assist with grouping.

This diagonal averaging is computationally equivalent to calculating $\boldsymbol{x}_i$ using [163]

$$\boldsymbol{x}_i = \boldsymbol{D}^{-1}\boldsymbol{W}_i\boldsymbol{v}_i \tag{5.7}$$

where

$$\boldsymbol{W}_i = \begin{pmatrix} w_1 & 0 & \cdots & \cdots & \cdots & 0 \\ w_2 & w_1 & 0 & \cdots & \cdots & 0 \\ w_3 & w_2 & w_1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ w_{K-1} & \cdots & w_3 & w_2 & w_1 & 0 \\ w_K & \cdots & \cdots & w_3 & w_2 & w_1 \\ w_{K+1} & \cdots & \cdots & \cdots & w_3 & w_2 \\ \vdots & & & & & \vdots \\ w_{L-1} & w_{L-2} & \cdots & \cdots & \cdots & w_{L-K} \\ w_L & w_{L-1} & w_{L-2} & \cdots & \cdots & w_{L-K+1} \\ 0 & w_L & w_{L-1} & w_{L-2} & \cdots & w_{L-K+2} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & w_L & w_{L-1} & w_{L-2} \\ 0 & \cdots & \cdots & 0 & w_L & w_{L-1} \\ 0 & \cdots & \cdots & \cdots & 0 & w_L \end{pmatrix} \tag{5.8}$$

and $\boldsymbol{D}$ is an $N \times N$ diagonal matrix, whose diagonal elements are

$$\begin{pmatrix} 1 & 2 & \cdots & \mu-1 & \mu & \cdots & \mu & \mu-1 & \cdots & 2 & 1 \end{pmatrix} \tag{5.9}$$

and $\boldsymbol{w}_i = \begin{pmatrix} w_1 & w_2 & \cdots & w_L \end{pmatrix}^T = \boldsymbol{A}\boldsymbol{v}_i$, and $\boldsymbol{v}_i$ is one of the eigenvectors. To simplify the nomenclature, double subscripting on $\boldsymbol{w}_i$ has been avoided. It is understood that the elements of this vector depend upon the specific eigenvector used in the reconstruction.

Recall that $N = L + K - 1$ and $\mu = min(L, K) = K$ in most SSA applications. Then there are $N - 2(K-1) = 2L - N$ 'complete' rows with diagonal elements μ in $\boldsymbol{D}$. The matrix $\boldsymbol{W}_i$ is of dimension $N \times K$. The first and last $K-1$ rows are 'incomplete', which leaves $N - 2(K-1)$ 'complete' rows.

The latent variables $\boldsymbol{w}_i$ are orthogonal, and have squared norm $\| \boldsymbol{w}_i \|_2^2 = \lambda_i$. The squared norm of a reconstructed series using diagonal averaging is

$$\begin{aligned} \| \boldsymbol{x}_i \|_2^2 &= \| \boldsymbol{D}^{-1}\boldsymbol{W}_i\boldsymbol{v}_i \|_2^2 \\ &\leq \| \boldsymbol{D}^{-1} \|_2^2 \cdot \text{trace}(\boldsymbol{W}_i^T\boldsymbol{W}_i) \\ &= K \| \boldsymbol{D}^{-1} \|_2^2 \cdot \| \boldsymbol{w}_i \|_2^2 \\ &= K\lambda_i \| \boldsymbol{D}^{-1} \|_2^2 \\ &= \frac{K\lambda_i}{2(1 + 1/4 + 1/9 + \cdots 1/(K-1)^2) + (2L-N)/K^2} \\ &\simeq \frac{K}{\pi^2/3 + (2L-N)/K^2}\lambda_i, \qquad L \gg K \end{aligned} \tag{5.10}$$

and

$$\| \sum_{i=1}^{d} \boldsymbol{x}_i \|_2^2 \neq \sum_{i=1}^{d} \| \boldsymbol{x}_i \|_2^2 \tag{5.11}$$

Calculations indicate that the upper bound may be quite conservative. The reconstructed series $\boldsymbol{x}_i$ are not orthogonal making it impossible to calculate the variance

of grouped variables from the variance of the individual-reconstructed series.

The use of reduced-rank approximations to assist in the extraction of harmonic signals from additive noise has been considered extensively in the signal processing literature [164, 165, 166]. The trajectory matrix has a central role in these algorithms. Extensive research indicates that extraction of these signals is considerably enhanced when the trajectory matrix is replaced by a structured-low rank approximation [167, 168]. The SVD leads to an unstructured approximation, because the $\boldsymbol{X}_i$ are not Hankel. A structured-approximation is obtained when the trajectory matrix is calculated using the reconstructed series (or groupings of variables). While the SVD does not preserve the Hankel structure, the Hankel matrix constructed from the reconstructed series does not preserve the rank property. Cadzow [169] developed a simple iterative algorithm to preserve both the rank and Hankel structure. He has shown that this iteration will converge to reduced-rank approximation that has the Hankel structure for certain classes of signals including sinusoids and damped sinusoids corrupted by white noise. The reconstruction of the $\boldsymbol{x}_i$ corresponds to one iteration of Cadzow's algorithm. Other approaches for obtaining reduced-rank approximations with appropriate structure involve the use of structured total least squares [170, 171]. These methods are computationally intense, requiring the use of a nonlinear optimizer in high dimension.

5.3 Filtering Interpretation of SSA

Before discussing filtering interpretations, we state a number of definitions and properties.

Definition 1. *A column vector* $\boldsymbol{c}$ *of length* n *is symmetric if* $\boldsymbol{c}$ *equals the vector obtained by reversing the rows of* $\boldsymbol{c}$*, i.e.,* $c_i = c_{n+1-i},\ i = 1, 2, \cdots, n$*. Mathematically,* $\boldsymbol{c}$ *is symmetric, if* $\boldsymbol{J}\boldsymbol{c} = \boldsymbol{c}$*, where* $\boldsymbol{J}$ *is a* $n \times n$ *matrix with ones on the main anti-diagonal.* $\boldsymbol{J}$ *is known as the exchange matrix.*

A vector $\boldsymbol{c}$ *is skew-symmetric if the vector obtained by reversing the rows of* $\boldsymbol{c}$ *equals* $-\boldsymbol{c}$*, i.e.,* $c_i = -c_{n+1-i},\ i = 1, 2, \cdots, n$*. A skew-symmetric vector satisfies* $\boldsymbol{J}\boldsymbol{c} = -\boldsymbol{c}$*.*

Definition 2. *An* $n \times n$ *matrix* $\boldsymbol{X}$ *is persymmetric if it is symmetric about both its main diagonal and main anti-diagonal. A symmetric Toeplitz matrix is persymmetric.*

Property 1. *If a persymmetric matrix has* K *distinct eigenvalues, then there are* $[(K+1)/2]$ *symmetric eigenvectors, and* $[K/2]$ *skew-symmetric eigenvectors, where* $[x]$ *denotes the integer part of* x *[156, 157, 158, 172]. The eigenvectors appear in the pattern shown in Table 5.1.*

K	symmetric eigenvector	skew-symmetric eigenvector
odd	$\begin{pmatrix} \boldsymbol{\alpha}_i^T\boldsymbol{J} & \alpha_0 & \boldsymbol{\alpha}_i^T \end{pmatrix}^T$	$\begin{pmatrix} -\boldsymbol{\beta}_i^T\boldsymbol{J} & 0 & \boldsymbol{\beta}_i^T \end{pmatrix}^T$
even	$\begin{pmatrix} \boldsymbol{\alpha}_i^T\boldsymbol{J} & \boldsymbol{\alpha}_i^T \end{pmatrix}^T$	$\begin{pmatrix} -\boldsymbol{\beta}_i^T\boldsymbol{J} & \boldsymbol{\beta}_i^T \end{pmatrix}^T$

Table 5.1: **Eigenvector patterns for persymmetric matrices.** $\boldsymbol{J}$ **is the** $[K/2] \times [K/2]$ **exchange matrix.**

Property 2. *The eigenvectors of a persymmetric matrix* $\boldsymbol{X}$ *can be computed from matrices of lower dimension [156, 157]. For* K *even,* $\boldsymbol{X}$ *can be written as*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{J}\boldsymbol{X}_{21}\boldsymbol{J} \\ \boldsymbol{X}_{21} & \boldsymbol{J}\boldsymbol{X}_{11}\boldsymbol{J} \end{pmatrix} \tag{5.12}$$

where $\boldsymbol{X}_{11}$ *and* $\boldsymbol{X}_{21}$ *are* $[K/2] \times [K/2]$ *and* $\boldsymbol{X}_{11}^T = \boldsymbol{X}_{11},\ \boldsymbol{X}_{21}^T = \boldsymbol{J}\boldsymbol{X}_{21}\boldsymbol{J}$*.*

$[K/2]$ eigenvalues p_i, and associated symmetric eigenvectors $\tilde{\boldsymbol{\alpha}}_i$, are obtained from the eigenvalue problem

$$(\boldsymbol{X}_{11} + \boldsymbol{J}\boldsymbol{X}_{21})\tilde{\boldsymbol{\alpha}}_i = p_i\tilde{\boldsymbol{\alpha}}_i \tag{5.13}$$

$\boldsymbol{\alpha}_i$ in Table 5.1 is given by $\boldsymbol{\alpha}_i = \frac{1}{\sqrt{2}}\tilde{\boldsymbol{\alpha}}_i$ and $i = 1, 2, \cdots, [K/2]$.

The remaining $[K/2]$ skew-symmetric eigenvectors $\tilde{\boldsymbol{\beta}}_i$, and corresponding eigenvalues q_i, are determined from the eigenvalue problem

$$(\boldsymbol{X}_{11} - \boldsymbol{J}\boldsymbol{X}_{21})\tilde{\boldsymbol{\beta}}_i = q_i\tilde{\boldsymbol{\beta}}_i \tag{5.14}$$

$\boldsymbol{\beta}_i$ in Table 5.1 is obtained from $\boldsymbol{\beta}_i = \frac{1}{\sqrt{2}}\tilde{\boldsymbol{\beta}}_i$ and again $i = 1, 2, \cdots, [K/2]$.

The eigenvectors have considerable structures. This structure will be exploited in Section 5.3.2. Expressions for K odd can be found in [156].

Definition 3. *Let $b(x)$ be a polynomial of the form $b(x) = \sum_{k=1}^{n} b_k x^{k-1}$. $b(x)$ is a palindromic polynomial if $b_k = b_{n-k+1},\ k = 1, 2, \cdots, n$. $b(x)$ is an antipalindromic polynomial if $b_k = -b_{n-k+1},\ k = 1, 2, \cdots, n$ [173].*

Definition 4. *For a polynomial $b(x)$ with real coefficients, the reciprocal polynomial of $b(x) = \sum_{k=1}^{n} b_k x^{k-1}$ is obtained by reversing the coefficients, i.e., $\tilde{b}(x) = \sum_{k=1}^{n} b_{n-k+1} x^{k-1}$.*

Definition 5. *An eigenfilter is a polynomial formed filter whose polynomial coefficients are obtained from an eigenvector. Denoting an eigenvector by $\boldsymbol{v}$ the associated eigenfilter is $v(z^{-1}) = \sum_{k=1}^{K} v_k z^{-(k-1)}$, where z^{-1} is interpreted as the backshift operator, i.e., $z^{-1}y_t = y_{t-1}$.*

Property 3. *The eigenfilters constructed from a persymmetric matrix are either palindromic polynomials or anti-palindromic polynomials. This follows immediately*

*from the symmetric and skew-symmetric properties of the eigenvectors of symmetric Toeplitz matrices [**Property 2**].*

Property 4. *The roots of the eigenfilters constructed from a persymmetric matrix have unit magnitude, or they appear in reciprocal pairs [158, 173].*

Property 5. *The distribution of roots at $z = \pm 1$ of the eigenfilters constructed from a symmetric Toeplitz matrix is given in Table 5.2.*

Table 5.2: **Distribution of roots at $z = \pm 1$ for eigenfilters of a symmetric Toeplitz matrix**

Type	Order (K-1)	Root at $z^{-1} = 1$	Root at $z^{-1} = -1$
Antipalindrome	Odd	$\surd$	-
	Even	$\surd$	$\surd$
Palindrome	Odd	-	$\surd$
	Even	-	-

These results are established by substituting $z = \pm 1$ into those palindromic and antipalindromic polynomials. It is known that an antipalindromic polynomial always has an odd number of roots located at $z^{-1} = 1$ [173].

Property 6. *When the eigenvalues of a symmetric Toeplitz matrix are unique, the roots of the eigenfilter associated with the minimum/maxixum eigenvalue all lie on the unit circle. For the other eigenfilters, this property may or may not be satisfied [158, 174].*

5.3.1 Filtering Interpretation of Latent Variables

The latent variable $\boldsymbol{w}_i$ is readily interpreted as a filtered value of the original variables. To simplify the notation, let $\boldsymbol{v}$ be one of the eigenvectors and $\boldsymbol{w}$ be the corresponding

filtered value instead of $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$. Then, the t^{th} element of $\boldsymbol{w}$ can be written as

$$w_t = \sum_{m=1}^{K} v_m y_{t+m-1}, \qquad t = 1, 2, \cdots, L \tag{5.15}$$

Alternatively, it can be written in the form

$$w_t = \sum_{m=1}^{K} \tilde{v}_m y_{t+K-m}, \qquad t = 1, 2, \cdots, L$$

where the coefficients $\tilde{v}_m$ are obtained by simply reversing the order of the coefficients v_m. This can be expressed mathematically as $\tilde{\boldsymbol{v}} = \boldsymbol{J}\boldsymbol{v}$, where $\boldsymbol{J}$ is again a $K \times K$ exchange matrix [**Definition 1**].

5.3.2 Filtering Interpretations Using the Toeplitz Approximation

Let the eigenvectors be obtained from the Toeplitz approximation to $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$, i.e., $L\boldsymbol{C}$. Based on the definitions, the covariance matrix $\boldsymbol{C}$ is a symmetric Toeplitz matrix and is persymmetric. For the moment, let K be odd. From **Property 1**, the filtered values for the symmetric eigenvectors can be written as

$$w_t = \alpha_0 y_{t+[K/2]} + \sum_{m=1}^{[K/2]} \alpha_m (y_{t+[K/2]+m} + y_{t+[K/2]-m}), \quad t = 1, 2, \cdots, L \tag{5.16}$$

The filtered values using the skew-symmetric eigenvectors are calculated as

$$w_t = \sum_{m=1}^{[K/2]} \beta_m (y_{t+[K/2]+m} - y_{t+[K/2]-m}), \quad t = 1, 2, \cdots, L \tag{5.17}$$

$\boldsymbol{w}$ can be interpreted as an aligned or time-shifted value, which consists of either a weighted average of $[K/2]$ observed values adjacent to $y_{t+[K/2]}$ in the case of a symmetric eigenvector, or as a weighted difference for a skew-symmetric eigenvector. When plotting y_t and w_t, it is imperative to properly align the original data by aligning $y_{t+[K/2]}$ with w_t.

The filters in Equations (5.16) and (5.17) are recognized as noncausal Finite Impulse Response (FIR) filters, also called acausual or non-recursive filters [175, 176]. Equation (5.16) describes a zero-phase filter. While the filter may attenuate or amplify the data, it introduces no phase shift in the filtered values. If a series is described by purely harmonic components, these will appear at the same time point in the filtered data as in the original data. Equation (5.17) describes a differentiating filter. This filter introduces a phase lag of $\pm\pi$ radians. In a series with purely harmonic components, the filtered series will either lead or lag the original series. There are many classical design techniques for zero-phase filters that act as either averaging filters or differentiating filters [176].

The relationship between the even and odd filters can be expanded by using the alternate calculation method for the symmetric and skew-symmetric eigenvectors that follows from **Property 2**. When $L >> K$, the symmetric eigenvectors, and corresponding eigenvalues, can be calculated from

$$\boldsymbol{C}_{11} + \boldsymbol{J}\boldsymbol{C}_{21} \simeq \frac{2}{L}\boldsymbol{A}_{\mathbf{1}}^{\boldsymbol{T}}\boldsymbol{A}_{\mathbf{1}} \tag{5.18}$$

where $\boldsymbol{A}_{\mathbf{1}}$ is the $L \times ([K/2] + 1)$ matrix.

$$\boldsymbol{A}_{\mathbf{1}} = \begin{pmatrix} \frac{1}{2}(\boldsymbol{y}_1 + \boldsymbol{y}_K) & \frac{1}{2}(\boldsymbol{y}_2 + \boldsymbol{y}_{K-1}) & \cdots & \frac{1}{2}(\boldsymbol{y}_{[\frac{K}{2}]} + \boldsymbol{y}_{[\frac{K}{2}]+2}) & \boldsymbol{y}_{[\frac{K}{2}]+1} \end{pmatrix} \tag{5.19}$$

The skew-symmetric eigenvectors are obtained by replacing the sum of the variables by their differences in Equation (5.19).

$$\boldsymbol{A}_{\mathbf{2}} = \begin{pmatrix} \frac{1}{2}(\boldsymbol{y}_1 - \boldsymbol{y}_K) & \frac{1}{2}(\boldsymbol{y}_2 - \boldsymbol{y}_{K-1}) & \cdots & \frac{1}{2}(\boldsymbol{y}_{[\frac{K}{2}]} - \boldsymbol{y}_{[\frac{K}{2}]+2}) & \mathbf{0} \end{pmatrix} \tag{5.20}$$

For an even K, the similar patterns can be obtained.

5.3.3 Filtering Interpretation of Reconstructed Series

From Equations (5.7) and (5.8), we can show that

$$x_t = \frac{1}{K}\left(y_t + \sum_{m=1}^{K-1} \zeta_m (y_{t-m} + y_{t+m})\right), \quad t = K, K+1, \cdots, N-K+1 \tag{5.21}$$

The filter coefficients are recognized as $\frac{\boldsymbol{v} \circ \tilde{\boldsymbol{v}}}{K}$, where $\circ$ denotes convolution and $\tilde{\boldsymbol{v}}$ is again obtained by reversing the order of the coefficients of $\boldsymbol{v}$.

$$\begin{aligned} \zeta_1 &= v_1 v_2 + v_2 v_3 + \cdots + v_{K-2} v_{K-1} + v_{K-1} v_K \\ \zeta_2 &= v_1 v_3 + v_2 v_4 + \cdots + v_{K-2} v_K \\ &\vdots \\ \zeta_{K-1} &= v_1 v_K \end{aligned} \tag{5.22}$$

For the designated values of t, x_t is obtained by weighting y_t and $(K-1)$ values of $y_{t \pm m}$ on either side of y_t. The weights are symmetric. For values of t outside of the indicated range, we can construct a filtering interpretation as well. However, edge effects or end effects are observed as the filter coefficients are no longer symmetric.

Equation (5.21) is recognized as a noncausal FIR filter, sometimes referred to as an a-casual or non-recursive filter [176]. Since the filter coefficients are symmetric, this is a zero-phase filter as well. It is known that a zero-phase a-casual filter can be obtained by a two-step filtering algorithm [175]. First, filter the data $y_t,\ t = 1, 2, \cdots, N$ with the FIR filter $v(z^{-1})$ to produce the series $\tilde{w}_t$. Second, filter the series $\tilde{w}_t$ using the FIR filter $\tilde{v}(z^{-1})$, where $\tilde{\boldsymbol{v}} = \boldsymbol{J}\boldsymbol{v}$. This is equivalent to reversing the order of $\tilde{w}_t$ and filtering with $\tilde{\boldsymbol{v}}$, and then reversing the order of the resulting series. In both cases the filtering algorithm uses $y_t = 0$ when $t \leq 0$ and $\tilde{w}_t = 0$ when $t > N$. Once the double-filtered series is obtained, it is multiplied by the diagonal weighting matrix $\boldsymbol{D}$. A very close approximation to any of the individuals via rank-1 approximations,

$\boldsymbol{x}_i$, is obtained by using the MATLAB^{TM} command `filtfilt(`$\boldsymbol{v}$`,1,y)` and then multiplying the result by $\frac{1}{K}$. Except for a short transient at the beginning and at the end of the reconstructed series, the results coincide with those obtained by diagonal averaging. There are other choices for initial conditions [175] that may be advantageous to reduce edge effect transients. However, due to the length of series and small K value typically encountered in SSA analysis, edge effects are most often small.

The results in this section apply only to signals reconstructed from rank-1 approximations, say, $\boldsymbol{X}_i$. The grouping of several rank-1 reconstructions before averaging is equivalent to calculating the individual averaging operators and then grouping the resulting reconstructions. Consequently, if a rank-p approximation is desired, then p of these filters are arranged in a parallel configuration and the results are summed [163]. A similar idea was also developed in [177].

5.4 Spectral Interpretation of SSA

5.4.1 Spectral Interpretation of Latent Variables

The spectrum of the filtered series in Equation (5.15) is given by

$$S_{\boldsymbol{w}}(f) \;\simeq\; |v(e^{-j2\pi f})|^2 \cdot S(f) \tag{5.23}$$

where f is the normalized frequency, $0 \leq f \leq 0.5$, $S(f)$ is the spectrum of y_t, and $|\cdot|$ denotes the magnitude of the quantity '$\cdot$'. The approximation arises from the assumption that the spectrum of $y_t,\ t = 1, 2, \cdots, N$, is the same as the spectrum of $y_t,\ t = K, K+1, \cdots, N$, i.e., edge or end effects have been neglected.

5.4.2 Spectral Interpretations Using the Toeplitz Approximation

The interesting spectral features of the filtered signals arise from the structured nature of the eigenfilters.

Using the Toeplitz approximation, the eigenfilters are either palindromic or anti-palindromic. Any eigenfilter $v(z^{-1})$ can be factorized as [173]

$$v(z^{-1}) = c(z^{-1}-1)^{k_1}(z^{-1}+1)^{k_2}\prod_{i=1}^{k_3}(z^{-2}-2cos(\omega_i)z^{-1}+1)$$
$$\prod_{i=1}^{k_4} e_4(\zeta_i, z^{-1})\prod_{i=1}^{k_5} e_5(\tau_i, z^{-1}) \tag{5.24}$$

The term $(z^{-2}-2cos(\omega_i)z^{-1}+1)$ accounts for the complex roots (except ± 1) of unit magnitude, $e_4(\cdot)$ accounts for all real roots except those at ± 1, and $e_5(\cdot)$ accounts for the complex roots, which are neither purely real nor purely imaginary. $\{c, \zeta_i, \omega_i\} \in \mathbb{R}$, $\{\tau_i\} \in \mathbb{C}$, and $K = k_1 + k_2 + 2k_3 + 2k_4 + 4k_5$.

The limiting cases for the spectrum of the filtered signal $\boldsymbol{w}$, at $f = 0$ and $f = 0.5$, are

$$\begin{aligned}\lim_{f\to 0} S_{\boldsymbol{w}}(f) &\simeq \lim_{f\to 0} |v(e^{-j2\pi f})|^2 \cdot S(f) \\ &= |c0^{k_1}2^{k_2}\prod_{i=1}^{k_3} 2(1-\cos(\omega_i))\prod_{i=1}^{k_4} e_4(\Delta_i, 1)\prod_{i=1}^{k_5} e_5(\tau_i, 1)|^2 \cdot S(0) \\ &= 0, \qquad k_1 > 0\end{aligned} \tag{5.25}$$

$$\begin{aligned}\lim_{f\to 0.5} S_{\boldsymbol{w}}(f) &\simeq \lim_{f\to 0.5} |v(e^{-j2\pi f})|^2 \cdot S(f) \\ &= |c(-2)^{k_1}0^{k_2}\prod_{i=1}^{k_3} 2(1+\cos(\omega_i))\prod_{i=1}^{k_4} e_4(\Delta_i, -1)\prod_{i=1}^{k_5} e_5(\tau_i, -1)|^2 \cdot S(0.5) \\ &= 0, \qquad k_2 > 0\end{aligned} \tag{5.26}$$

Using the results in **Property 5** and Table 5.2, the low and high frequency characteristics of the eigenfilters are shown in Table 5.3. The frequency characteristics are much different from typical digital filters, and are more reminiscent of the filtering characteristics of Sleppian functions [178, 179].

Table 5.3: **The frequency characteristics for eigenfilters of Toeplitz matrix**

Type	Order (K-1)	$\lim_{f\to 0} \lvert v(e^{-j2\pi f})\rvert^2$	$\lim_{f\to 0.5} \lvert v(e^{-j2\pi f})\rvert^2$
Antipalindrome	Odd	0	-
	Even	0	0
Palindrome	Odd	-	0
	Even	-	-

When $v(z^{-1})$ corresponds to the minimum or maximum eigenvalue, all roots are on the unit circle [158, 174]. Consequently, $|v(e^{-2j\pi f})|$ will be zero at most $(K-1)$ values of f in the interval $0 \leq f \leq 0.5$, resulting in complete attenuation of $S_{\boldsymbol{w}}(f)$ at these frequencies. We also note that an antipalindromic eigenfilter can always be written as $v(z^{-1}) = (z^{-1}-1)^{k_1}\overline{v}(z^{-1})$, where k_1 is odd and $\overline{v}(z^{-1})$ is a palindromic eigenfilter [173]. Thus, an antipalindromic filter is always equivalent to palindromic filtering of the differenced variable $(z^{-1}-1)^{k_1}y_t$.

5.4.3 Spectral Interpretation of Reconstructed Series

As shown in Section 5.3, a series produced by the diagonal-averaging approach is fundamentally different from the latent variable. The spectral characteristic of a new reconstructed series $\boldsymbol{x}$ follows immediately from Equation (5.21).

$$S_{\boldsymbol{x}}(f) \;\simeq\; \frac{1}{K^2}|v(e^{-j2\pi f})|^4 \cdot S(f), \qquad 0 \leq f \leq 0.5 \tag{5.27}$$

The approximation again arises from the end or edge effects, which are expected to be small when $L \gg K$. The spectral properties of this associated filter follow immediately from the previous discussion.

5.5 Examples

5.5.1 Example 1: Harmonic Series with Two Real Sinusoids

In this section, we consider the following process

$$y_t = \sum_{k=1}^{2} \alpha_k \sin(2\pi f_k t) + \xi_t \tag{5.28}$$

where $\alpha_k = (4, 2)^T$, $f_k = (0.1, 0.4)^T$, and $\xi_t \sim \boldsymbol{\mathcal{N}}(0, 1)$, ξ_t is normally distributed with mean 0 and variance 1. The simulated data length, N, is 1024. Both the signal and the noise are mean-corrected for purposes of analysis. The signal-to-noise ratio (SNR) is 10.

5.5.1.1 Scree Plot and Eigenvector Analysis

In this example, we choose $K = 25$. The scree plot (not shown) gives four significant eigenvalues, grouped into two pairs. One might anticipate that these two groups of eigenvalues arise from the presence of the two harmonics in the data that correspond to frequencies 0.1 and 0.4. The presence of additive noise produces a large number of much smaller eigenvalues.

We use the Toeplitz approximation $\boldsymbol{A^T A} \simeq \boldsymbol{LC}$ to calculate the SVD. This is a reasonable assumption, as indicated by the Frobenius norm ratio.

$$\frac{||\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{LC}||_F}{||\boldsymbol{A}^T\boldsymbol{A}||_F} = \frac{389.7 - 388.2}{388.2} = 0.031 \tag{5.29}$$

The first four eigenvectors are shown in the first column of Figure 5.1. As discussed in Section 5.3, all the eigenvectors are either symmetric or skew-symmetric. In this example, there are 13 symmetric eigenvectors, and 12 skew-symmetric eigenvectors.

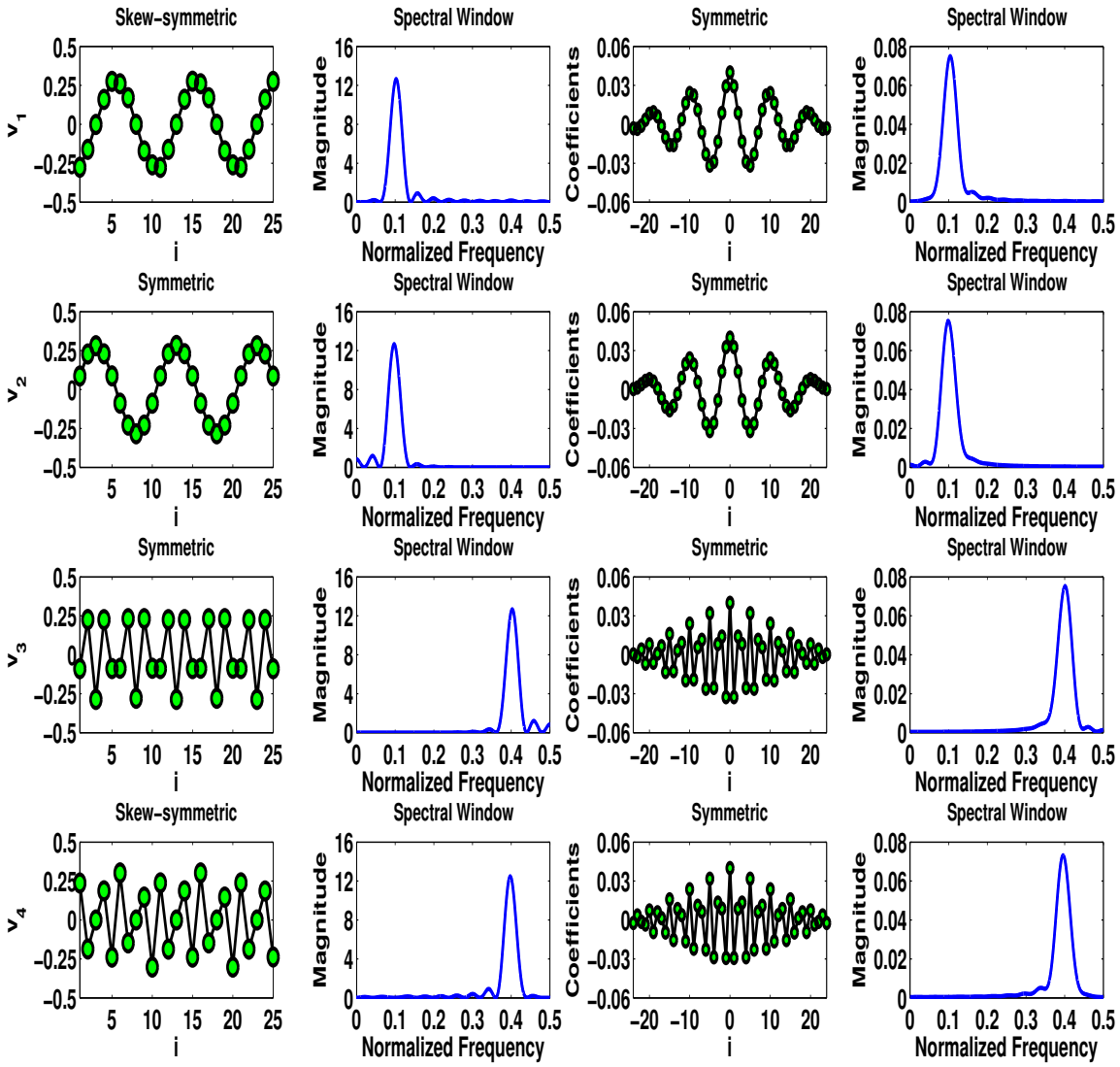


Figure 5.1: **Filtering and spectral interpretations of example 1** (a) First four eigenvectors (first column); (b) Magnitude of eigenfilter spectrum (second column); (c) Convolution filter coefficients (third column); (d) Magnitude of convolution filter spectrum (fourth column)

5.5.1.2 Eigenfilter Analysis

From **Property 5**, the eigenfilters obtained from the first four eigenvectors above are either palindromic or antipalindromic. The magnitude of the associated spectral windows of these eigenfilters is shown in the second column in Figure 5.1. The spectral windows show strong peaks at the harmonic frequencies, indicating that the filtering algorithm behaves similar to a notch-filter [176].

The roots for these eigenfilters are shown in Figure 5.2. (The first panel in Figure 5.2 corresponds to the largest eigenvalue.) The number of roots located at $z^{-1} = \pm 1$ for each eigenfilter is noted on the top of each subfigure. The roots at ± 1 follow the properties shown in Table 5.2. The eigenvalues are unique, so the zeros of eigenfilters associated with the smallest and largest eigenvalues all lie on the unit circle.

5.5.1.3 Convolution Filter Analysis

The third column in Figure 5.1 depicts the convolution filter coefficients. These are all zero-phase filters due to the symmetry of the filter coefficients. The magnitude of the spectral windows of corresponding convolution filters is shown in the fourth column of this figure. The peaks in these spectral plots correspond exactly to the significant frequencies in the signal, i.e., f = 0.1 and 0.4. The properties in Table 5.3 also hold.

5.5.1.4 Reconstructed Components (RCs)

The top four panels in Figure 5.3 show the first four RCs obtained by diagonal averaging. The original data is also included in each of these plots. Several observations can be made: i) the RCs are paired, and each pair has almost the same pattern, and

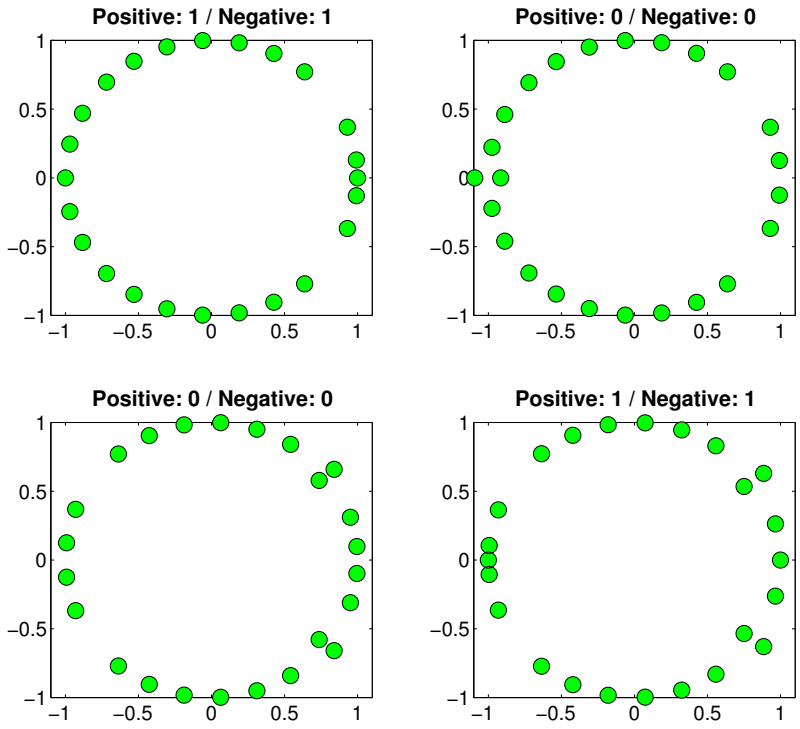


Figure 5.2: **Roots for the eigenfilters of example 1. The number of roots located at $z^{-1} = \pm 1$ is shown on the top of each panel**

ii) the first two RCs are much larger than the second two RCs. This is expected as the first two RCs correspond to the harmonic whose frequency is 0.1 and whose power magnitude is four times that of the second harmonic (See Equation (6.1)). The last panel in Figure 5.3 shows the grouped reconstructed series from the first four RCs. The group reconstructed series closely matches the original data. There is no discernable phase lag, which is to be expected.

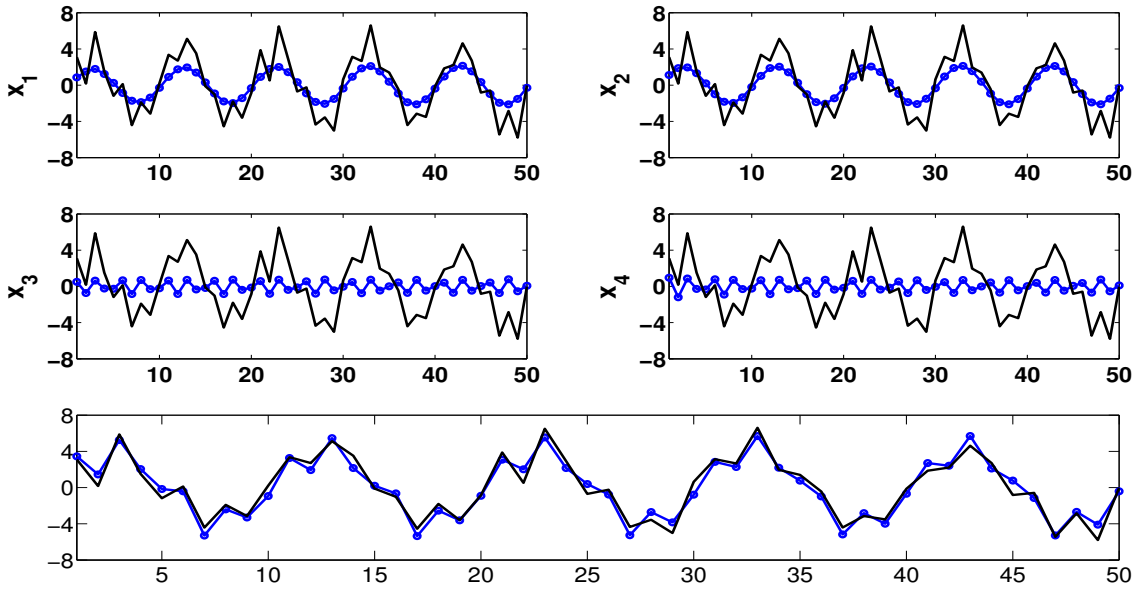


Figure 5.3: **First four RCs and the grouped series (dotted line) versus original series (solid line) of example 1**

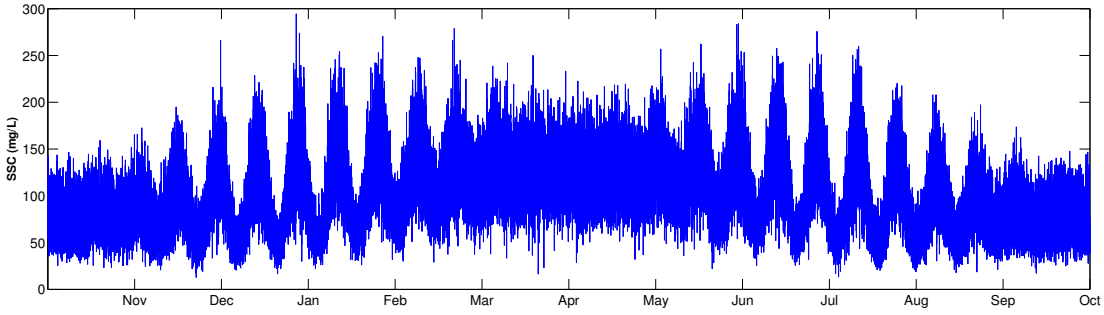


Figure 5.4: **Synthetic SSC time series**

5.5.2 Example 2: Synthetic SSC Data

As an example, we consider a synthetic suspended-sediment concentration (SSC) series shown in Figure 5.4 and analyzed by Schoellhamer [154]. A 15-min SSC time series with mean 100 mg/L was generated using Equations (6) and (7) in [154].

$$y(t) = 0.2\epsilon(t)c_s(t) + c_s(t) \tag{5.30}$$

and

$$\begin{aligned} c_s(t) =& 100 - 25\cos\omega_s t + 25(1 - \cos 2\omega_s t)\sin\omega_{sn} t \\ &+ 25(1 + 0.25(1 - \cos 2\omega_s t)\sin\omega_{sn} t)\sin\omega_a t \end{aligned} \tag{5.31}$$

where ϵ_t is normally distributed random variate with zero mean and unit variance. The seasonal angular frequency $\omega_s = 2\pi/365$ day^{-1}, the spring/neap angular frequency $\omega_{sn} = 2\pi/14$ day^{-1} and the advection angular frequency $\omega_a = 2\pi/(12.5/24)$ day^{-1}.

Equation (5.31) was simulated for one water year, giving $35,040$ 'observations' for analysis. Prior to analysis the data was mean corrected.

5.5.2.1 Scree Plot and Eigenvector Analysis

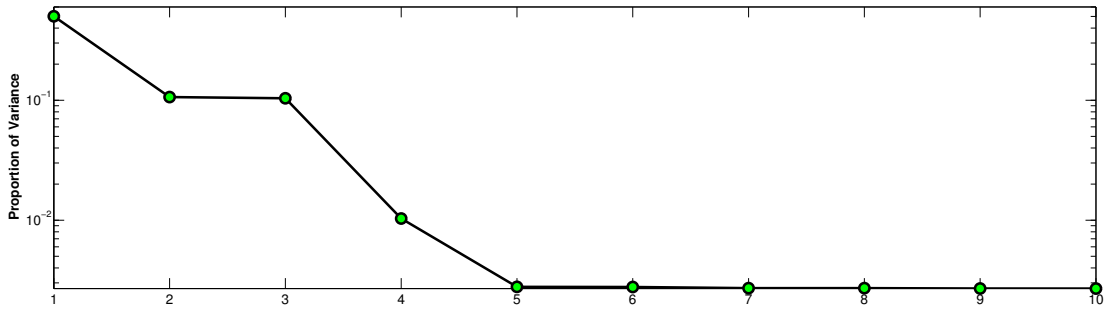


Figure 5.5: **Scree plot of synthetic SSC series**

In this example, we choose window length $K = 121$. We note that Schoellhamer

[154] used $K = 120$. Our analysis is based on the Toeplitz matrix. This provides a very good approximation to $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$ as indicated by the Frobenius norm ratio which has the value 0.002.

The scree plot for the eigenvalues, Figure 5.5, has been normalized by the sum of the eigenvalues since the eigenvalues are very large. Note the log scale on the vertical axis. The first three eigenvalues are an order of magnitude larger than the remaining ones. The presence of many small eigenvalues suggests that the data contains aperiodic or random components. Only one group of paired eigenvalues is evident in Figure 5.5. One might anticipate three groups of paired eigenvalues, given the structure of the model. An explanation for this behaviour is given in the next subsection.

By using the Toeplitz approximation, 61 symmetric and 60 skew-symmetric eigenvectors are obtained. The first column of Figure 5.6 shows the first three eigenvectors.

5.5.2.2 Eigenfilter Analysis

The magnitude of the associated spectral windows of these eigenfilters is shown in the second column in Figure 5.6. The spectral window corresponding to the largest eigenvalue has the characteristics of a low-pass filter. The second and third eigenfilters correspond to a notch-filter with normalized frequency of 0.02. This frequency corresponds to the advection angular frequency $\omega_a = 2\pi/(12.5/24)$ day^{-1}. Given the data and model, we may expect three groups of paired eigenvalues. However, by only simulating one year's data, the subtidal (annual) cycle shows up as a low frequency component. Thus, this effect is captured in the first eigenfilter. However, this eigenfilter also smooths out the effect of the fortnight component, which appears at the

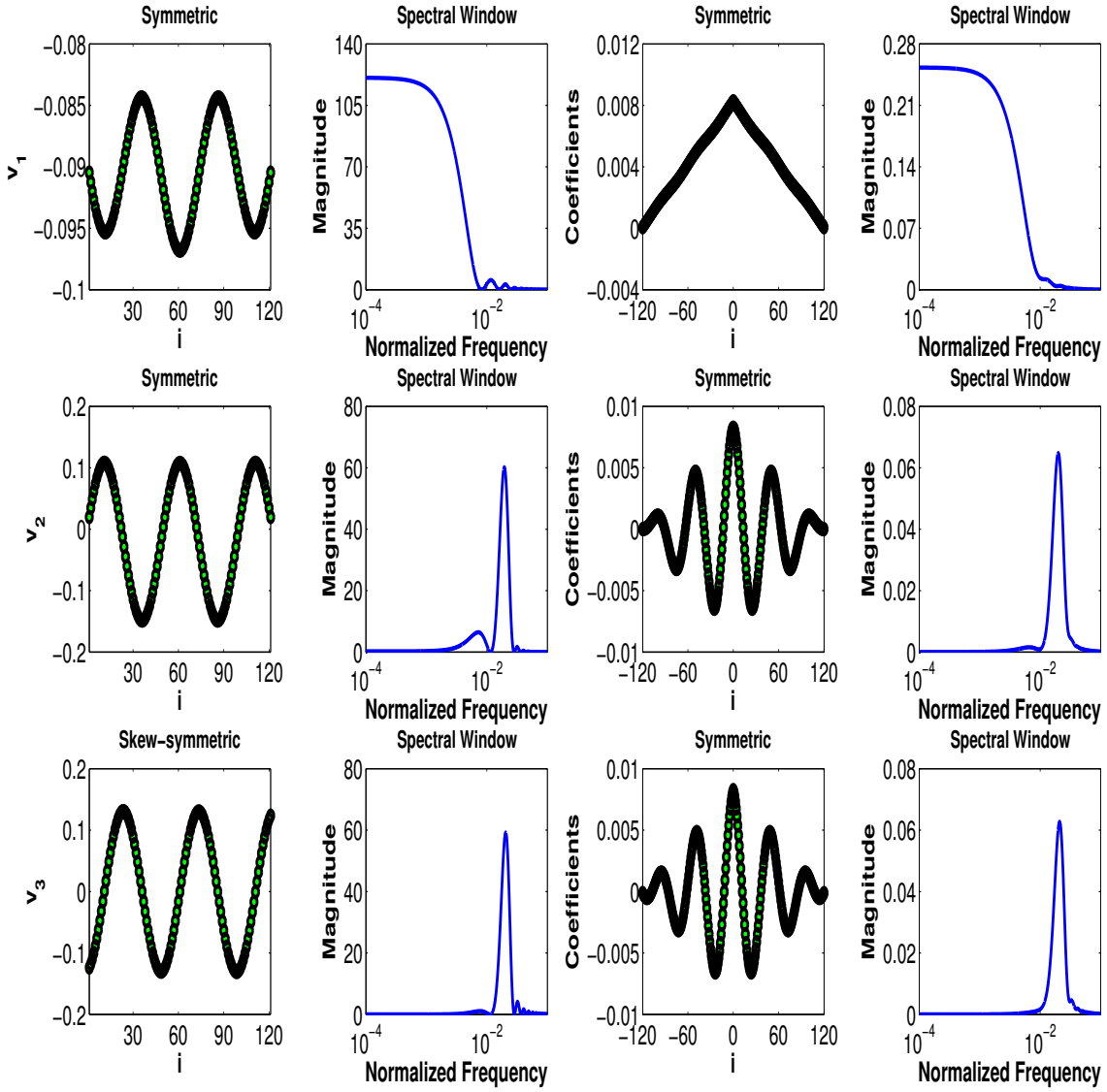


Figure 5.6: **Filtering and spectral interpretations of synthetic SSC series**, (a) First three eigenvectors (first column); (b) Magnitude of eigenfilter spectrum (second column); (c) Convolution filter coefficients (third column); (d) Magnitude of convolution filter spectrum (fourth column)

normalized frequency 7.44×10^{-4}. Advanced spectral methods such as Thompson's multi-taper method [178, 179] readily separate the fortnight/advection components. Additional insights into the spectrum are often obtained by using several spectral methods [180, 181]. In addition, harmonic frequencies are often specified explicitly as a seasonal lag in seasonal time series models, while in SSA they are presented in the coefficients of the eigenfilters.

The roots of the eigenfilters follow the same properties in Table 5.2 and example 1. Additionally, all the roots of eigenfilters related to smallest and largest eigenvalues are located on the unit circle. Due to similarity, the root plots are not shown.

5.5.2.3 Convolution Filter Analysis

The convolution filter coefficients are shown in the third column of Figure 5.6 and the corresponding spectral characteristics of these filters are shown in the fourth column of this figure. The convolution filter corresponding to the largest eigenvalue has a distinctively triangular shape and acts as a low-pass filter. The second and third convolution filters have their power concentrated at a normalized frequency of 0.02.

5.5.2.4 Reconstructed Components (RCs)

Table 5.4: **Variance summary of first three components**

a	$\frac{\lambda_i}{\sum_{i=1}^{K}\lambda_i}$	$var(\boldsymbol{x}_i)$	$\sum_{i=1}^{a} var(\boldsymbol{x}_i)$	$var(\sum_{i=1}^{a}\boldsymbol{x}_i)$
1	0.5011	745.1	745.1	745.1
2	0.1064	81.2	826.3	838.1
3	0.1039	79.2	905.5	1086.1

The first three RCs are plotted in the last three panels of Figure 5.7. Individually

they account for 48.7%, 5.1%, and 5.0% of the data variation. (The data variance is 1546.7.) Table 5.4 confirms that the RCs are not orthogonal, because the sum of the variances of the first three components is not equal to the variance of their sum.

The first RC is a very smooth signal, reflecting that it is obtained by a low-pass filtering of the data. By comparing the patterns of the other two RCs, it is clear that the identification of the harmonic component has been done correctly, as the convolution filter acts like a notch filter.

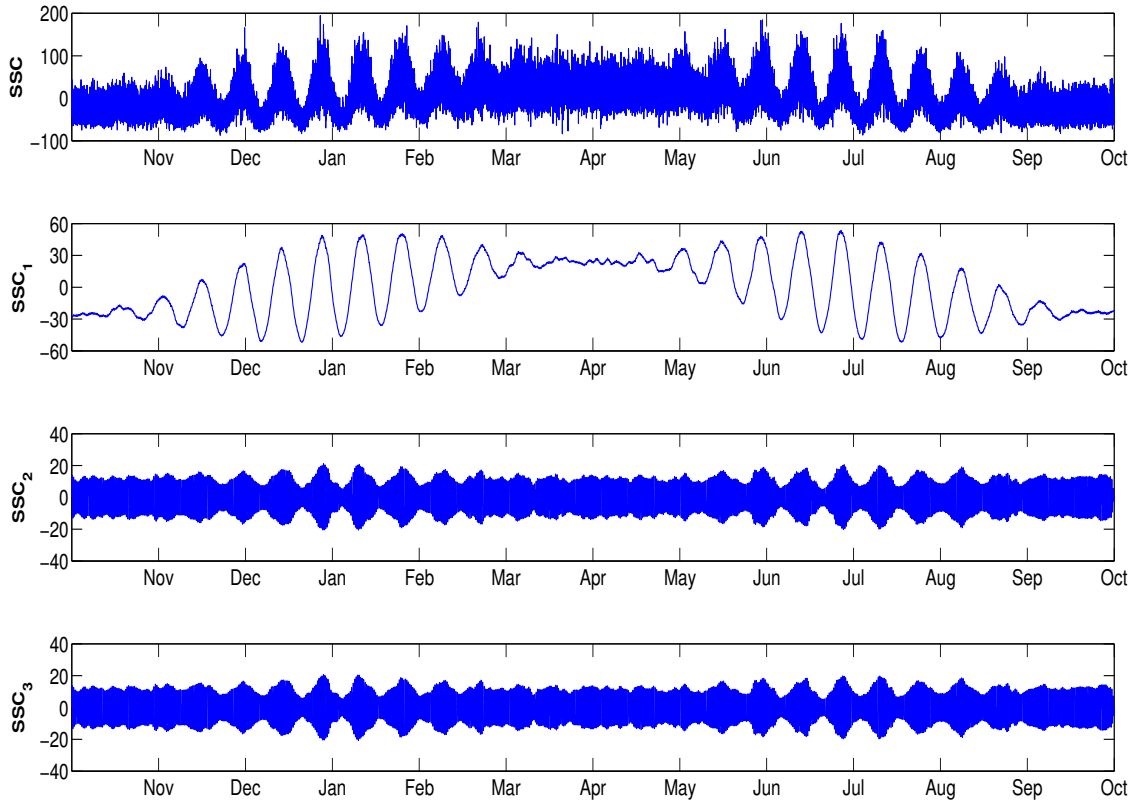


Figure 5.7: **Synthetic SSC series along with the first three RCs**

5.6 Conclusion and Discussion

Singular spectrum analysis is a flexible and useful method for time-series analysis. The primary contributions of this paper have been to provide additional insights into the filtering and spectral characteristics of SSA technology and the enhancements that arise by using diagonal averaging. These new filtering and spectral results are derived from the properties of symmetric Toeplitz matrices and the properties of the resulting eigenfilters and convolution filters. Filtering and spectral interpretations for the reconstructed series from diagonal averaging were derived. The symmetric and skew-symmetric behaviour of the eigenfilters was exploited to derive a number of these properties. It was shown that the reconstructed series could be interpreted as zero-phase filtered responses, obtained by a particular implementation of forward and reverse filtering of the original data. It was also shown that whereas the latent variables are orthogonal, the reconstructed series are not orthogonal. The results in this paper should enable a more thorough comparison of SSA with other filtering methods.

Multichannel extensions of SSA (MSSA) have been proposed [182, 183, 184]. MSSA produces data-adaptive filters that can be used separate patterns both in time and space. It is necessary to construct a 'grand' block matrix, a multichannel equivalent to $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}/L$. This matrix also has a block Toeplitz approximation. This approximation gives a symmetric, but not persymmetric grand matrix, although every sub-block matrix is a persymmetric matrix. Exploitation of the properties of these highly structured matrices to ascertain filtering and spectral properties, in a manner similar to that employed in this paper, should be possible.

Chapter 6

Application of Singular Spectral Analysis in Chemical Processes

6.1 Introduction

As established in Chapter 5, Singular Spectrum Analysis (SSA) is a powerful and flexible technique for time-series analysis. Pike [131] and Broomhead [132] first proposed the basic SSA algorithm, while Vautard and Ghil [136, 138] published a series of SSA papers demonstrating the applications of SSA that have attracted significant attention. The basic SSA method with details can be found in [140] and [141], and is summarized in the preceding chapter. A number of benefits of SSA that motivate its use can be summarized as [143] :

1. SSA can deal with seasonal, stationary, and non-stationary time series. Unlike conventional Autoregressive Integrated Moving Average (ARIMA) models in which an integer degree of non-stationarity must be specified, differencing is not an issue

for SSA.

2. SSA is a nonparametric technique with no statistical assumptions in the application, such as stationarity of an analyzed series and normality of residuals or random shocks.

3. SSA decomposes a series into interpretable components, such as trends, harmonic series, and residuals. Additionally, these decomposed components can be used for time-series analysis, data forecasting, and process monitoring.

4. SSA can be applied to short series, meaning that process information can be extracted without requiring many data. The requirement of a small data size in analysis is especially beneficial to nonlinear dynamics analysis.

6.1.1 Applications of SSA

Given the advantages above, practical applications of SSA have been extended to solve problems in the following.

6.1.1.1 Feature extraction

The major advantage of SSA is to decompose a time series into a number of independent and interpretable components, e.g., a slowly varying trend, oscillatory components and a structureless residual. With relatively little computational effort, the SSA decomposition provides feature extraction, e.g., trend extraction [185, 186] periodicity/seasonality detection [187, 188], and denoising (or prefiltering) [189, 190]. In addition, SSA applications in feature extraction have been provided with a smoothing and filtering interpretation [148].

6.1.1.2 Forecasting

The attraction of SSA applications in forecasting can be outlined as: 1) In SSA there is no statistical assumption concerning either signal or noise, while traditional time series methods require structural and distributional assumptions, e.g., stationarity for the data, linearity for the model, and normality for the residuals. The only assumption adopted in SSA forecasting is to satisfy the Linear Recurrent Formulae (LRF) summarized in [141]. 2) SSA decomposition makes it feasible to forecast either the single component itself or the grouped reconstructed series. 3) SSA has proven to work well in forecasting short, noisy time series [138]. Improving forecasts via SSA is validated against time series in different areas, especially economic and financial time series, which recently have been analyzed by Hassani and his coworkers [143, 144, 191, 192, 193].

6.1.1.3 Gap filling

Missing data are a common problem in data analysis, and gap-free data are often required for standard analysis tools. Modifications to SSA have been proposed to fill in the missing data and continue the analysis procedure. The first gap-filling approaches have been validated in the geosciences [154].

SSA-based gap-filling methods in existence can be classified into two groups:

1) Direct methods: Schoellhamer [154] and Golyandina and Osipov [194] have proposed modified versions of SSA to fill in the missing data by means of featured components, i.e., trends and periodic components, effectively providing an element of interpolation based on featured components. The filled data can be viewed as a filtered version of the original data. Golyandina and Osipov also remark that the

SSA gap-filling technique can bring new solutions to the SSA forecasting, because the forecasting problem can be formed in terms of missing data located in the end.

2) Iterative methods: The essential idea is that missing data follow the same patterns in the available measurements, e.g., the trend and oscillatory modes. Two standard steps are implemented [195, 196]: (a) missing values are filled in iteratively by using the leading reconstructed components and updating self-consistent lag-covariance matrix. Note that covariance matrix represents temporal correlations in the univariate cases and spatio-temporal correlations in the multivariate cases; and (b) a cross-validation technique is applied to update the key parameters in the SSA algorithm, i.e., the window length and the number of components to execute the gap-filling. The successful implementation of this algorithm is conditional on the pattern, the gap length, and variance fraction of missing data. Regarding the extension of this gap-filling algorithm, Wang and Liang [197] slightly modify the iteration procedure in order to take into account some uncertainties, while Kondrashov et al. [198] combine extra process information, i.e., a gappy driver, to help the gap-filling application.

6.1.1.4 Change-point detection

Moskvina [199] proposes a sequential application of SSA to the problem of detecting changes in time series. Note that one of the key features in SSA is the dimensional reduction of the space formed by vectors in the trajectory matrix, i.e., a subspace in a lower dimension that approximates well the above defined space and describes well the signal in a time series. The problem of change-point detection can be considered as the problem of outlier identification in a time series. It can be assumed that the generation mechanism of a time series, i.e., the signal structure, stays unchanged, and

the residual is an i.i.d. random variable. This results in reasonably small distances between the vectors in a trajectory matrix and a reduced subspace. Whenever there is a change in the generation mechanism, outliers appear in the time series, and the corresponding vectors are moved out of the reduced subspace. The underlying idea is that distances between the corresponding vectors and the reduced subspace are expected to increase, which can be used for change-point detection. Different detection statistics are proposed based on the aforementioned distances (see Moskvina's PhD thesis [199] for further details). Note that standard SSA analysis is performed in an off-line manner, while SSA-based change-point detection is implemented in an on-line procedure. Accordingly, change-point detection can be further studied as a promising process monitoring tool in industry.

6.1.1.5 Multi-dimension SSA

Multi-dimension SSA is a natural extension of SSA to multi-dimensional time series. Two forms are multi-channel SSA (M-SSA) and two-dimensional SSA (2D-SSA).

The theoretical foundation of M-SSA can be traced back to [132] in the context of nonlinear dynamics analysis. In comparison with Principal Component Analysis (PCA) with emphasis on the temporal structure, M-SSA studies more temporal and some spatial structures. Note that extended empirical orthogonal function (EEOF) in the meteorological literature is synonymous with M-SSA. Two different approaches are presented to implement M-SSA, which differs in the way of forming the multi-channel trajectory matrix. For a more detailed description see the Toeplitz method in [136] and the trajectory-matrix method in [132]. M-SSA has many applications, especially in the analysis of economic and financial time series [188, 190, 193].

The other multi-dimension extension is 2D-SSA, which is introduced by Golyandina [200, 201] with special focus on the 2D analysis and physical interpretations. 2D-SSA can decompose 2D data into low- and high-frequency components and give temporal-spatial patterns. Applications of 2D-SSA can be found in digital terrain analysis [201] and digital image analysis [200].

6.1.2 Motivation

Diverse SSA applications have already been demonstrated in research fields such as meteorology, oceanology and geophysics [190]; finance, economics and business [193]; and biosciences [202]; telecommunication [203]; and image processing [204]. Successful applications have proven a promising future of SSA-based data analysis in various areas; however, Aldrich and co-workers [5, 6] have noted the lack of SSA analysis of data from process plants. The SSA application to plant data in the chemical and metallurgical engineering systems has been illustrated in Barkhuizen's thesis [7]. In this work SSA is further promoted to the chemical industry.

Introducing SSA techniques into the collection of techniques used for analysis of chemical processes raises a number of questions. The first question is the difference between SSA and other existing multivariate data analysis methods, such as dynamical principal component analysis (DPCA) [205], canonical variate analysis (CVA) [206], and spectral PCA (using power spectra or autocovariance functions) [207]. SSA is mainly used for time series analysis (or process analysis) and forecasting, while DPCA and CVA focus on system identification and process monitoring, and spectral PCA contributes in oscillation detection and root cause diagnosis. DPCA implicitly estimates parametric models as linear combinations of the lagged outputs.

The field of chemical engineering affords a tremendous amount of research possibilities to applied mathematicians. One of the main reasons is the existence of oscillations and multiple steady states in complex chemical processes that are caused by interactions between the transport, thermodynamic and kinetic phenomena. Possible reasons are: 1) hard constraints, e.g., dead-band and valve stiction; 2) aggressive control effects; and 3) oscillatory disturbance. The process characteristics result in featured data similar to the ones in aforementioned areas, e.g., geoscience, economics, and business. Consequently, successful demonstrations motivate prospective applications of SSA in the process analysis and monitoring of chemical data.

6.2 SSA in Chemical Process Analysis

SSA analysis is mainly motivated by properties of decomposed components, which can be interpreted from a filtering and spectrum perspective. In comparison with other time series analysis tools, improvements of SSA in process analysis come from: 1) the filtering properties of decomposed series; 2) simpler power spectrum of decomposed series focusing on smaller frequency ranges; and 3) limited noise content in few decomposed series due to the separation of signal and noise in different frequency ranges. Further, process analysis plays a key role in the root-cause investigation and performance improvement of control systems. The more characteristics known about a process, the better control can be achieved. Barkhuizen [7] states that a single time series can contain information about key dynamics and properties of a process, because the evolution of a system leaves a footprint in the process output. The emphasis of this section is on promoting SSA applications to the chemical industry, and an example below is used to demonstrate the feasibility.

In the discussion that follows, the two tanks in series example first presented by Aldrich and Barkhuizen [6] is revisited using the enhanced SSA algorithm described in Chapter 5. The differences between the results in [6] and the results described below are: 1) in the treatment of the data from the tank simulation described below, mean-centering is applied (Aldrich and Barkhuizen [6] do not use mean-centering in their analysis), and 2) different window lengths. Mean-centering enhances the predictive ability of the SSA approach, because a non-zero mean not corrected will dominate the low-frequency behaviour, resulting in a prediction of the average of the data. A similar result is seen in time series forecasting.

6.2.1 Example 1: Flow Between Two Tanks in Series

Aldrich and Barkhuizen [6] illustrates the use of basic SSA via the flow of two non-interacting tanks in series. The output is the measured flow out of the tank process, while the input is the inlet flow. To show more SSA features, this flow example is reconsidered in this work. Note that it is a second order process, and the process transfer function can be written as:

$$G(s) = \frac{1}{(0.38s + 1)(2.62s + 1)} = \frac{1}{s^2 + 3s + 1} \tag{6.1}$$

Figure 6.1 displays the process response (thin solid line) to a square signal (thick solid line). Note that the same input signal and noise characteristics as described in [6] are used in the simulation computing in the work presented in this chapter, and consequently the input and output signals shown in Figure 6.1 are the very similar. While the stimulus square signal is of a period of 100 seconds, i.e., 100 data points, the measured flow response ('+' markers) is simulated with a Gaussian noise with

zero mean and 0.1 standard deviation. The signal variance is 0.0547, and the signal-to-noise ratio (SNR) is 5.47.

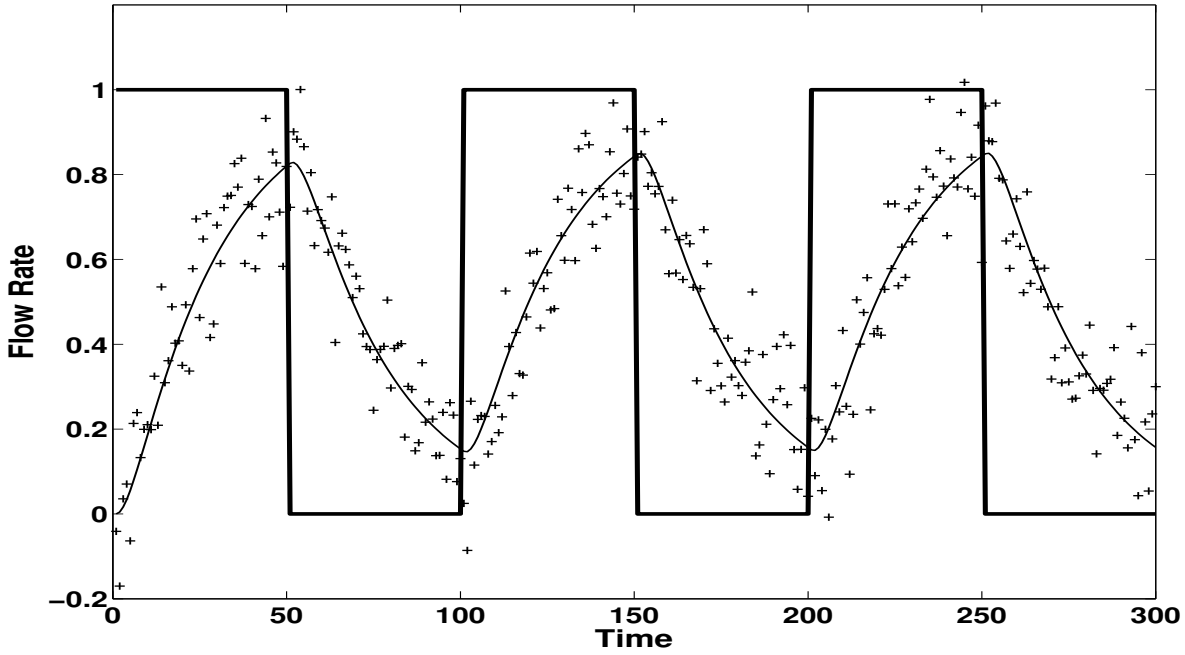


Figure 6.1: **Flow process response (after [6]): a square signal (think solid), an actual response (thin solid), and a simulated response ('+' markers)**

6.2.1.1 Scree Plot

The window length in this example is chosen as $K = 100$ (Notice that in [6] $K = 23$ is used). Figure 6.2 is a scree plot showing one pair of significant eigenvalues. Note that the SSA analysis is applied to the mean-corrected process measurements. In the scree plot the eigenvalues have been normalized by the summation, and only the first 20 eigenvalues are presented in a log scale. The first eight eigenvalues are considered to be more significant than the remaining ones. One might anticipate that: 1) the exceptional first two eigenvalues imply a harmonic as a pair of eigenvalues of similar

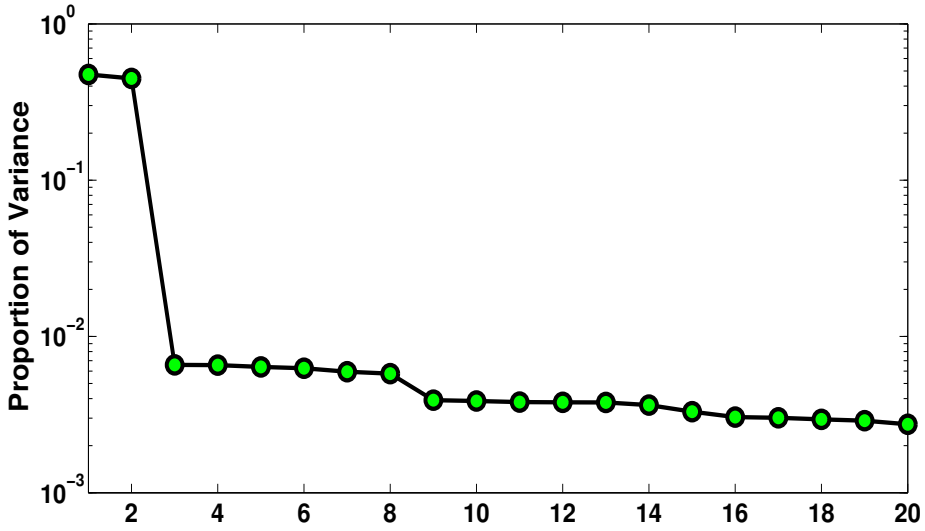


Figure 6.2: **Scree plot of Example 1**

magnitude are observed; 2) eigenvalues $3-8$ present smoothing drifts in the data; 3) a long tail of small eigenvalues indicate components with no specific patterns.

In SVD decomposition the Toeplitz approximation $\boldsymbol{X^T X} \simeq s\boldsymbol{C}$ in Chapter 5 is used. The Frobenius norm ratio can show that this is a reasonable assumption.

$$\frac{||\boldsymbol{A}^T\boldsymbol{A} - L\boldsymbol{C}||_F}{||\boldsymbol{A}^T\boldsymbol{A}||_F} = \frac{94.9}{750.5} = 0.126 \tag{6.2}$$

Note that the accuracy of Toeplitz approximation replies on the data and window length. In this analysis the number of measurements are relatively small, i.e., $N = 300$, while the window length is relatively large, i.e., $K = 100$. However, the Frobenius norm ratio still indicates a fairly good Toeplitz approximation.

6.2.1.2 Filtering Analysis

Chapter 5 demonstrates that by using the Toeplitz approximation the associated eigenvectors are either symmetric or skew-symmetric; thus, there are 50 symmetric and 50 skew-symmetric eigenvectors in this example. The first four eigenvectors are

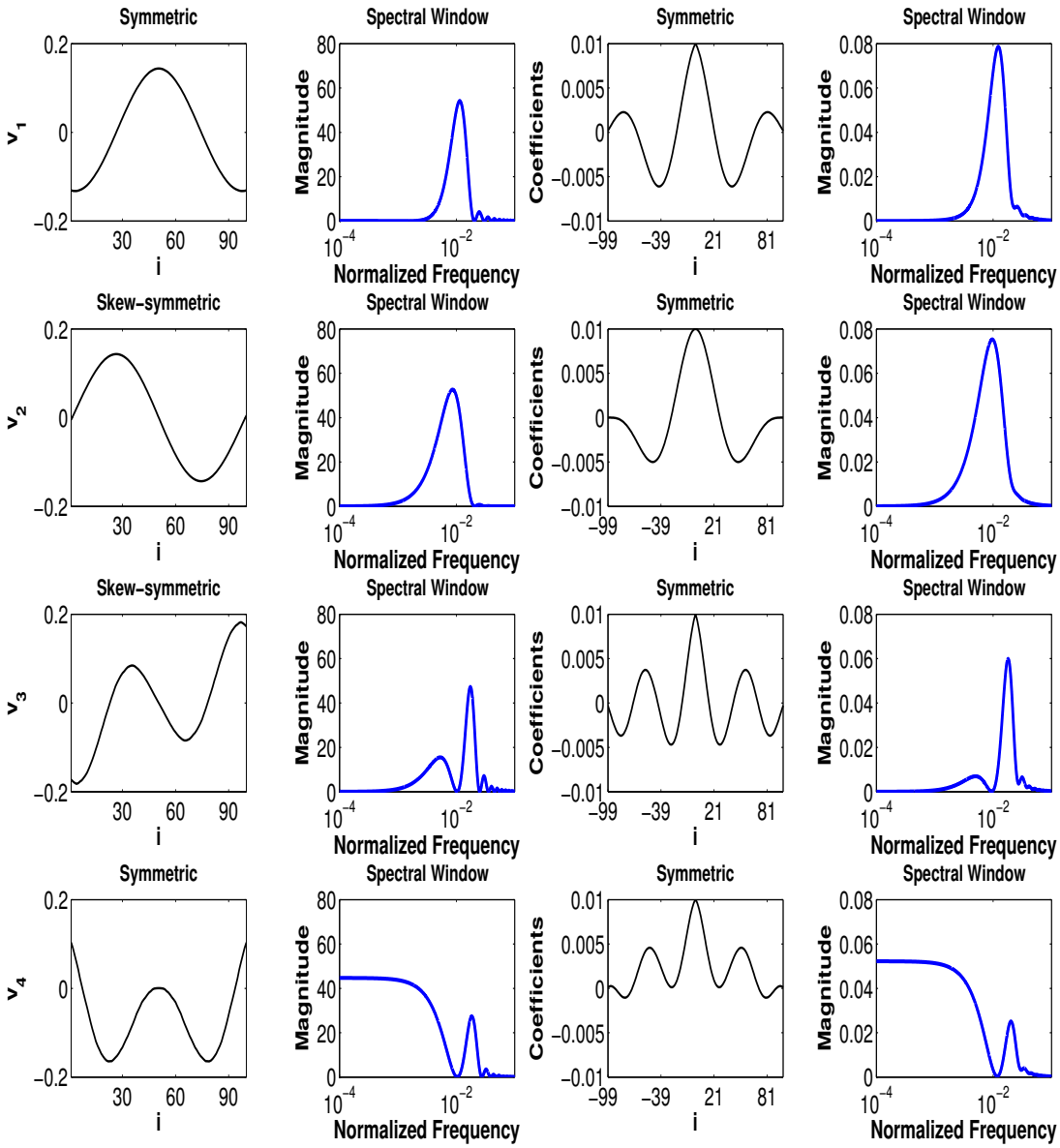


Figure 6.3: **Filtering and spectral interpretations of Example 1** (a) First four eigenvectors (first column); (b) Magnitude of eigenfilter spectrum (second column); (c) Convolution filter coefficients (third column); (d) Magnitude of convolution filter spectrum (fourth column)

plotted in the first column in Figure 6.3, and the magnitude of the associated spectrum plots of these eigenfilters are given in the second column. Several observations can be made: 1) the first two panels display strong peaks at frequency 0.01, which agrees well with the period of the stimulus square signal and the conclusion of the harmonics in the scree plot. This information is confirmed as well in the data plots in Figure 6.4. 2) the last panels present a mix of harmonics, implying that no specific patterns is discovered in the corresponding reconstructed components.

The convolution filter coefficients and the magnitude of the corresponding spectral windows are plotted in the last two columns in Figure 6.3. These filter coefficients are symmetric, which make zero-phase filters as described in Chapter 5. The frequency information gathered in column four is similar to the one in column two.

6.2.1.3 Reconstructed Components (RCs)

For illustration the first four RCs are plotted in the top four panels in Figure 6.4. By comparing the patterns of the first two RCs, we can conclude that the identification of a harmonic component in the scree plot has been done correctly. Spectrum plots in Figure 6.3 also confirm this harmonic at frequency 0.01. The third and fourth panels cover the end effects in SSA analysis. The bottom panel of Figure 6.4 draws the original series (solid line) versus the grouped series (dashdot line) composed of the first eight RCs. The grouped series clearly follows the main trend in the analyzed data. Furthermore, there is a good agreement between the grouped series and the real process response signal (not shown here).

The difference between the results in this work and the ones in [6] comes from: 1) the mean-correction of the data in this work - mean centering was not used in [6];

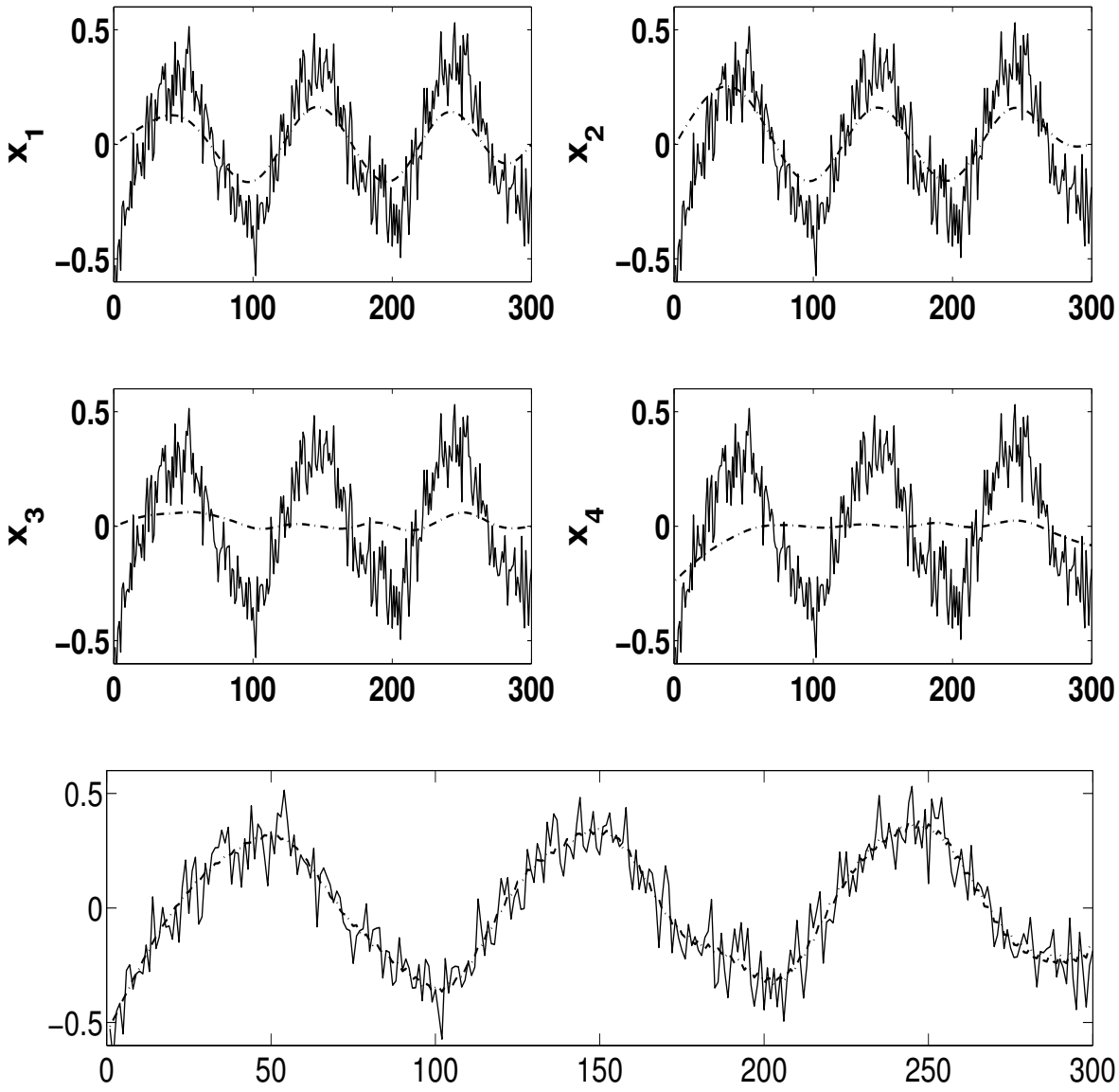


Figure 6.4: **The first four RCs and grouped series (dashdot line) versus original series (solid line) of Example 1**

2) different window lengths, and 3) the use of the new filtering theories described in Chapter 5. The importance of this flow example is to illustrate how periodic components appear in SSA analysis. In situations where there are periodic components in process data, one would expect to see pairs of eigenvalues associated with these components. This is a more direct and straightforward way to detect such behaviour in contrast to fitting time series models that contain ARMA and seasonal components. While plots of power spectrum generated from estimated time series models, and the auto-correlation function (ACF) computed directly from data, are often used for oscillation detection in data analysis these methods can be influenced by characteristics such as noise and components of time-varying frequencies. In particular, oscillatory components appearing in ACF plots can be obscured.

6.3 SSA in Chemical Process Monitoring

As demonstrated in Chapter 3, an explicit model identification procedure is a key step in computing MVC-based benchmarks in most performance assessment approaches. Parametric and nonparametric models are fitted in this procedure. Parametric models often cover transfer function models and state space models, which are expected to be of a low order [10]. Nonparametric models include impulse/step response models and frequency-domain based models [19]. Parametric identification approaches require the specification of model structure, which can be challenging for closed-loop data, whereas nonparametric approaches do not have this requirement. SSA is a nonparametric approach that offers the potential of some model reduction through the use of singular value decomposition (SVD) of the lagged process data. In addition, the SVD approach in SSA can deal with problems of poor numerical conditioning

that are often encountered in conventional impulse response estimation. The central contribution of this work is to introduce a new nonparametric SSA technique to the area of process monitoring in chemical industry. A new approach to compute the standard MVC performance benchmark is proposed. Besides, SSA provides a number of orthogonal singular vectors, which are potentially helpful for further investigating process behaviour and making predictions without much efforts. The developed SSA-based CPA is presented below in detail.

6.3.1 Basic CPA algorithm

In this work, univariate feedback systems without feedforward controllers are discussed for demonstration. For more details regarding the standard CPA algorithm, readers are referred to [14] and [15]. Harris [4] expresses a closed-loop system as:

$$y_t = \psi(B)a_t = e_{t|t-b} + \hat{y}_{t|t-b} = \psi_1(B)a_t + \psi_2(B)a_{t-b} \tag{6.3}$$

where the psi weights $\psi(B)$ can be split into two parts associated with the b-step ahead forecast error $e_{t|t-b}$ and b-step ahead forecast $\hat{y}_{t|t-b}$, and $\psi_1(B) = 1 + \psi_1 B + \cdots + \psi_{b-1}B^{b-1}$ (B is the backshift operator, i.e., $By_t = y_{t-1}$).

Since the random shocks $\{a_t\}$ are independent and identically distributed (I.I.D.), the minimum output variance in this case is:

$$\sigma_{mv}^2 = (1 + \psi_1^2 + \cdots + \psi_{b-1}^2)\sigma_a^2 \tag{6.4}$$

where σ_a^2 is the variance of a_t.

A normalized performance index between 0 and 1 is typically used to assess the performance of the overall control scheme, and is calculated as the ratio of the minimum variance to the mean square error (mse) in the data. This index can indicate

whether the current controller is working well or not.

$$\eta = \frac{\sigma_{mv}^2}{mse(y_t)} = \frac{\sigma_{mv}^2}{\sigma_y^2 + \mu_y^2} \tag{6.5}$$

where σ_y^2 denotes the output variance, and μ_y denotes the output mean (non-zero in the general case).

6.3.2 Proposed SSA-based CPA Algorithm

In practice, the linear regression approach provided in [14] is a straightforward way to compute the MVC-based performance index. The theory is to fit an AR model to an output data. The impulse response of the identified model can be used to calculate the psi weights, which are used in Equations 6.4 and 6.5 to calculate the performance index η. Since the analyzed closed-loop process is assumed to be stable (by controller design), the impulse weights ψ decay to 0, usually within a reasonable timeframe. A finite horizon is used for calculation, and the impulse response is truncated beyond a minimum threshold.

Starting with the ψ weight impulse response (moving average) representation of the output time series in Equation 6.3, $y_t = \psi(B)a_t$, we have

$$y_{t-b} = \psi(B)a_{t-b} \Rightarrow a_{t-b} = \frac{1}{\psi(B)}y_{t-b}$$

Accordingly,

$$y_t = \psi_1(B)a_t + \psi_2(B)a_{t-b} = \psi_1(B)a_t + \frac{\psi_2(B)}{\psi(B)}y_{t-b} = e_t + \frac{\psi_2(B)}{\psi(B)}y_{t-b}$$

resulting in an autoregressive representation for the output. As noted above, the assumption of a stable closed-loop system ensures the convergence of $\frac{\psi_2(B)}{\psi(B)}$. The tracking error over the time delay horizon, representing the MVC tracking error, is denoted as $e_t = \psi_1(B)a_t$.

The autoregressive model estimation can be posed as a lagged regression problem:

$$y_t = \sum_{i=1}^{\infty} \alpha_i y_{t-b-i+1} + e_t \approx \sum_{i=1}^{m} \alpha_i y_{t-b-i+1} + e_t$$

where m is the best model order when fitting an AR model to an output y_t.

A matrix form $\boldsymbol{y}_t = \boldsymbol{X} \cdot \boldsymbol{\alpha} + \boldsymbol{e}_t$ can be rewriten as:

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ \vdots \\ y_{b+m} \end{bmatrix}_{s\times 1} = \begin{bmatrix} y_{t-b} & y_{t-b-1} & \cdots & y_{t-b-m+1} \\ y_{t-b-1} & y_{t-b-2} & \cdots & y_{t-b-m} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ y_m & y_{m-1} & \cdots & y_1 \end{bmatrix}_{s\times m} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}_{m\times 1} + \begin{bmatrix} e_t \\ e_{t-1} \\ \vdots \\ \vdots \\ e_{b+m} \end{bmatrix}_{s\times 1}$$

where $s = t - b - m + 1$.

By standard linear regression, the estimated parameter vector can be obtained as $\hat{\boldsymbol{\alpha}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}_t$. Notice that it is quite possible that $\boldsymbol{X}^T\boldsymbol{X}$ will be poorly conditioned. Instead, a singular value decomposition (SVD) is applied to the matrix $\boldsymbol{X}$, and we have $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \sum_{i=1}^{m} \sqrt{\lambda_i}\boldsymbol{u}_i\boldsymbol{v}_i^T$.

Note that we can use an eigenvector decomposition for $\boldsymbol{X}^T\boldsymbol{X}$ as well:

$$(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i \Rightarrow \boldsymbol{v}_i = \lambda_i(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{v}_i$$

$$\Rightarrow (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{v}_i = \frac{1}{\lambda_i}\boldsymbol{v}_i \Rightarrow (\boldsymbol{X}^T\boldsymbol{X})^{-1} = \sum_{i=1}^{m} \frac{1}{\lambda_i}\boldsymbol{v}_i\boldsymbol{v}_i^T$$

With the previous results, we can obtain

$$\begin{aligned} \hat{\boldsymbol{y}}_{t|t-b} &= \boldsymbol{X} \cdot \hat{\boldsymbol{\alpha}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}_t \\ &= (\sum_{i=1}^{m} \sqrt{\lambda_i}\boldsymbol{u}_i\boldsymbol{v}_i^T)(\sum_{i=1}^{m} \frac{1}{\lambda_i}\boldsymbol{v}_i\boldsymbol{v}_i^T)(\sum_{i=1}^{m} \sqrt{\lambda_i}\boldsymbol{u}_i\boldsymbol{v}_i^T)^T\boldsymbol{y}_t \\ &= \sum_{i=1}^{m} (\boldsymbol{u}_i^T\boldsymbol{y}_t)\boldsymbol{u}_i = (\boldsymbol{u}_1^T\boldsymbol{y}_t)\boldsymbol{u}_1 + (\boldsymbol{u}_2^T\boldsymbol{y}_t)\boldsymbol{u}_2 + \cdots + (\boldsymbol{u}_m^T\boldsymbol{y}_t)\boldsymbol{u}_m \end{aligned}$$

where

$$(\boldsymbol{u}_i\boldsymbol{v}_i^T)(\boldsymbol{v}_j\boldsymbol{v}_j^T) = \boldsymbol{u}_i\boldsymbol{v}_i^T\boldsymbol{v}_j\boldsymbol{v}_j^T = \begin{cases} \boldsymbol{u}_i\boldsymbol{v}_i^T, & i = j \\ 0, & i \neq j \end{cases}.$$

In addition, $\hat{\boldsymbol{y}}_{t|t-b}$ can be rewritten in a matrix form.

$$\hat{\boldsymbol{y}}_{t|t-b} = \boldsymbol{U}(m)_{s\times m}\boldsymbol{U}^T(m)_{m\times s}\boldsymbol{y}_t = \boldsymbol{U}_{s\times s}\boldsymbol{I}(m)_{s\times s}\boldsymbol{U}_{s\times s}^T\boldsymbol{y}_t$$

where $\cdot(m)$ is a rank-m matrix, and $\boldsymbol{I}(m)_{s\times s}$ is a modified identity matrix having m non-zero entries (1) on the diagonal, with the remaining $s-m$ diagonal elements being set to 0.

The MVC-based CPA measure η can be computed as follows:

$$\begin{aligned}\eta = 1 - \frac{\hat{\boldsymbol{y}}_{t|t-b}^T\hat{\boldsymbol{y}}_{t|t-b}}{\boldsymbol{y}_t^T\boldsymbol{y}_t} &= 1 - \frac{\boldsymbol{y}_t^T\boldsymbol{U}\boldsymbol{I}(m)\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{I}(m)\boldsymbol{U}^T\boldsymbol{y}_t}{\boldsymbol{y}_t^T\boldsymbol{y}_t} \\ &= 1 - \frac{\boldsymbol{y}_t^T\boldsymbol{U}\boldsymbol{I}^2(m)\boldsymbol{U}^T\boldsymbol{y}_t}{\boldsymbol{y}_t^T\boldsymbol{U}\boldsymbol{I}^2(s)\boldsymbol{U}^T\boldsymbol{y}_t} \\ &= 1 - \frac{\sum_{i=1}^m(\boldsymbol{u}_i^T\boldsymbol{y}_t)^2}{\sum_{i=1}^s(\boldsymbol{u}_i^T\boldsymbol{y}_t)^2} = \frac{\sum_{i=m+1}^s(\boldsymbol{u}_i^T\boldsymbol{y}_t)^2}{\sum_{i=1}^s(\boldsymbol{u}_i^T\boldsymbol{y}_t)^2}\end{aligned}$$

It is clear that the MVC index can be calculated directly in terms of the singular vectors of the lagged output matrix $\boldsymbol{X}$, without the need to specify any model information beyond the time delay horizon b.

6.3.3 Discussion and Extensions

6.3.3.1 Process Analysis & Controller Performance Assessment

The basis of the SSA approach to CPA proposed in Section 6.3.2 is to expand the prediction $\hat{\boldsymbol{y}}_{t|t-b} = \sum_{i=1}^m(\boldsymbol{u}_i^T\boldsymbol{y}_t)\boldsymbol{u}_i$, which is a linear combination of singular vectors $\boldsymbol{u}_i$ defined by projection coefficients $\boldsymbol{u}_i^T\boldsymbol{y}_t$, $i = 1, 2, \cdots, m$. The SSA is applied to the data lagged by the delay horizon b, and the output observations $\boldsymbol{y}_t$ are then projected

onto the resulting singular value decomposition.

Information about the dynamic structure of the closed-loop process can be gleaned from both the SSA analysis on the lagged output measurements, and the projection of the output $\boldsymbol{y}_t$ on the singular vectors resulting from the SSA. The projection coefficients, $\boldsymbol{u}_i^T \boldsymbol{y}_t$, provide insight into possible sub-optimal behaviour of the existing controller, because they describe couplings between output values y_t and output values beyond the time delay horizon. If the controller was a minimum variance controller, there would be no such couplings because the MVC acts to eliminate these dependencies.

As well, the SSA of the lagged data provides insight into the dynamic structure from a process analysis perspective. If the closed-loop process is stationary, the correlation structure across time of the output measurements is invariant under shift of the absolute time index, i.e., the relationships between output measurements $y_t, \ldots, y_{t-L}$ is the same as the relationships between output measurements $y_{t-b}, \ldots, y_{t-b-L}$. Consequently, the SSA can be examined for evidence of dynamic elements such as harmonic behaviour that might indicate a poorly tuned controller such as aggressive control effects, or a process maintenance issue such as oscillatory disturbance.

In the example that follows, we demonstrate a series of graphical tests that can be used to diagnose dynamic behaviour. In particular, we propose the following sequence of diagnostics:

1. Computation of the SSA on the delayed outputs and projection of $\boldsymbol{y}_t$ onto the resulting left singular vectors to compute the MVC-based performance index.

2. Normalizing the absolute value of the projection coefficients $\boldsymbol{u}_i^T \boldsymbol{y}_t$ to place them in the range $[0, 1]$, and plotting versus the singular vector index.

3. Normalizing the singular values associated with the SSA and plotting in a scree plot to depict the progression of singular values in the SSA.

4. Computation and plotting of the autocorrelation function (ACF) for the closed-loop data;

5. Computation and plotting of the spectra of the process data $\boldsymbol{y}_t$, SSA-based prediction $\hat{\boldsymbol{y}}_{t|t-b}$, and associated prediction error.

This comprehensive set of plots and quantitative diagnostics provides a more complete picture of the closed-loop controller performance.

6.3.3.2 Proposed Spectrum Analysis

The orthogonal decomposition provided by SSA makes it possible to express the spectrum of the prediction, $\hat{\boldsymbol{y}}_{t|t-b}$, as the sum of spectra associated with each of the left singular vectors:

$$S_{\hat{\boldsymbol{y}}_{t|t-b}}(f) \;=\; \sum_{i=1}^{m} (\boldsymbol{u}_i^T \boldsymbol{y}_t)^2 \cdot S_{\boldsymbol{u}_i}(f)$$

where f is the frequency, and $S_{\boldsymbol{u}_i}(f)$ is the spectrum of $\boldsymbol{u}_i$. Note that $\boldsymbol{u}_i^T \boldsymbol{y}_t$ is a scalar.

Since the singular vectors $\boldsymbol{u}_i$ represent dynamic couplings (correlation) in the output time series, the spectra associated with these singular vectors can indicate the presence of certain periodic components and dynamic elements (e.g., low-frequency dynamics). Insights from such spectral decompositions can be used to identify maintenance issues and controller tuning issues in the closed-loop process. Moreover, if the matrix $\boldsymbol{X}^T \boldsymbol{X}$ is approximated by a covariance matrix, the eigenvectors have additional filtering and spectrum properties presented in Chapter 5.

6.3.3.3 Zero-padding Technique

In practical spectrum analysis, zero-padding is a fundamental tool. Zero-padding in the time domain results in spectral interpolation and higher resolution of the data spectrum. In this work, zero-padding can be used to eliminate end effects in an SSA decomposition. If the process output data vector is padded with $(m+b-1)$ zeros, the data vector becomes $\left[\mathbf{0}_{m+b-1}^T \; y_1 \; y_2 \; \cdots \; y_t\right]^T$. The matrix form $\boldsymbol{y}_t = \boldsymbol{X} \cdot \boldsymbol{\alpha} + \boldsymbol{e}_t$ in Section 6.3.2 can be expressed as:

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{b+m} \\ y_{b+m-1} \\ \vdots \\ y_{b+2} \\ y_{b+1} \\ y_b \\ \vdots \\ y_1 \end{bmatrix}_{N\times 1} = \begin{bmatrix} y_{t-b} & y_{t-b-1} & \cdots & y_{t-b-m+1} \\ y_{t-b-1} & y_{t-b-2} & \cdots & y_{t-b-m} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & y_1 \\ \vdots & \vdots & & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & y_1 & & 0 \\ y_1 & 0 & & 0 \\ 0 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{N\times m} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \vdots \\ \alpha_m \end{bmatrix}_{m\times 1} + \begin{bmatrix} e_t \\ e_{t-1} \\ \vdots \\ e_{b+m} \\ e_{b+m-1} \\ \vdots \\ e_{b+2} \\ e_{b+1} \\ e_b \\ \vdots \\ e_1 \end{bmatrix}_{N\times 1}$$

In this case, $\boldsymbol{y}_t$ and $\boldsymbol{u}_i$ are of size $N \times 1$ instead of size $s \times 1$. The zero-padding technique is shown above for the SSA-based CPA problem. It can be used in a similar manner in the SSA application of process analysis, in which the time delay is not imposed.

6.3.4 Example 2: A Feedback Control System

An example in [14] is simulated for demonstration. This example provides a progression of the types of analyses and information that can be gained from the SSA-based performance assessment. The process is

$$y_t = u_{t-b} + d_t = u_{t-b} + \frac{1 - 0.2B}{1 - B} a_t$$

where the disturbance d_t represents the effect of all unmeasured disturbances.

Using a simple controller $\nabla u_t = -K y_t$, the closed-loop system can be written as:

$$y_t = \frac{1 - 0.2B}{1 - B + KB^b} a_t$$

where the process delay $b = 3$, and the tuning parameter $K = 0.1$.

By long division, a psi weight polynomial is obtained:

$$\psi(B) = 1 + 0.80B + 0.80B^2 + 0.70B^3 + 0.62B^4 + 0.54B^5 + \cdots$$

	data variance	minimum variance	performance index
True Process	4.6455	2.2800	0.4908
Conventional CPA	4.6293	2.1908	0.4732
SSA-based CPA	4.6722	2.1892	0.4686

Table 6.1: **CPA results for the feedback control system**

In this example the process data mean is 0. As displayed in Table 6.1, results of the proposed SSA-based CPA give a good match with the theoretical ones of the true process and the ones of the conventional CPA. Accordingly, this example confirms the effectiveness of the proposed technique.

As discussed in Section 6.3.2, orthogonal components $\boldsymbol{u}_i$ are computed using the SSA technique. The prediction $\hat{\boldsymbol{y}}_{t|t-b}$ can be formed through a linear combination of the first m left singular vectors $\boldsymbol{u}_i$.

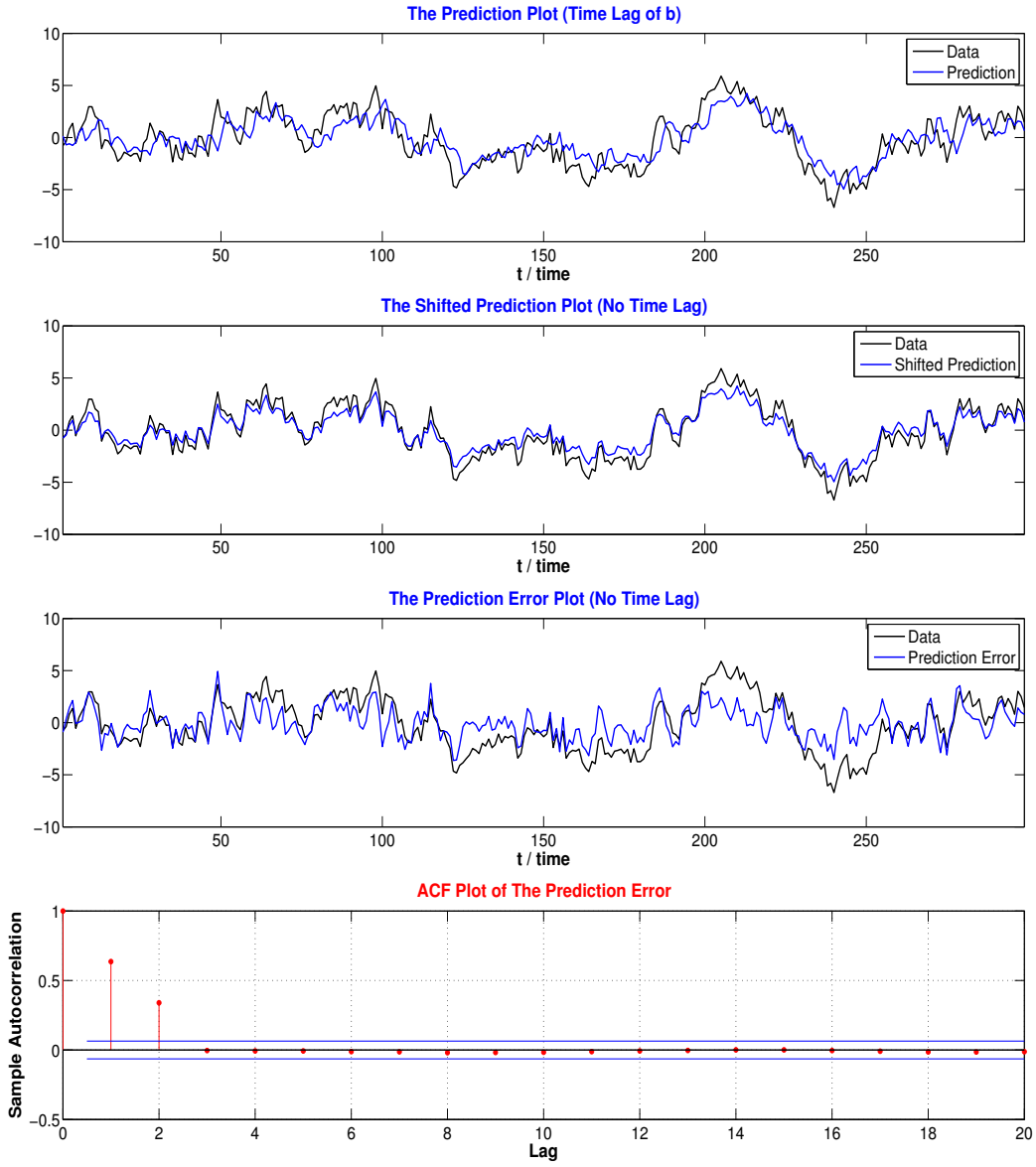


Figure 6.5: **Prediction and prediction error analysis for Example 2 (a) Prediction versus process data; (b) Shifted prediction versus process data; (c) Prediction error versus process data; (d) The ACF plot of the prediction error.**

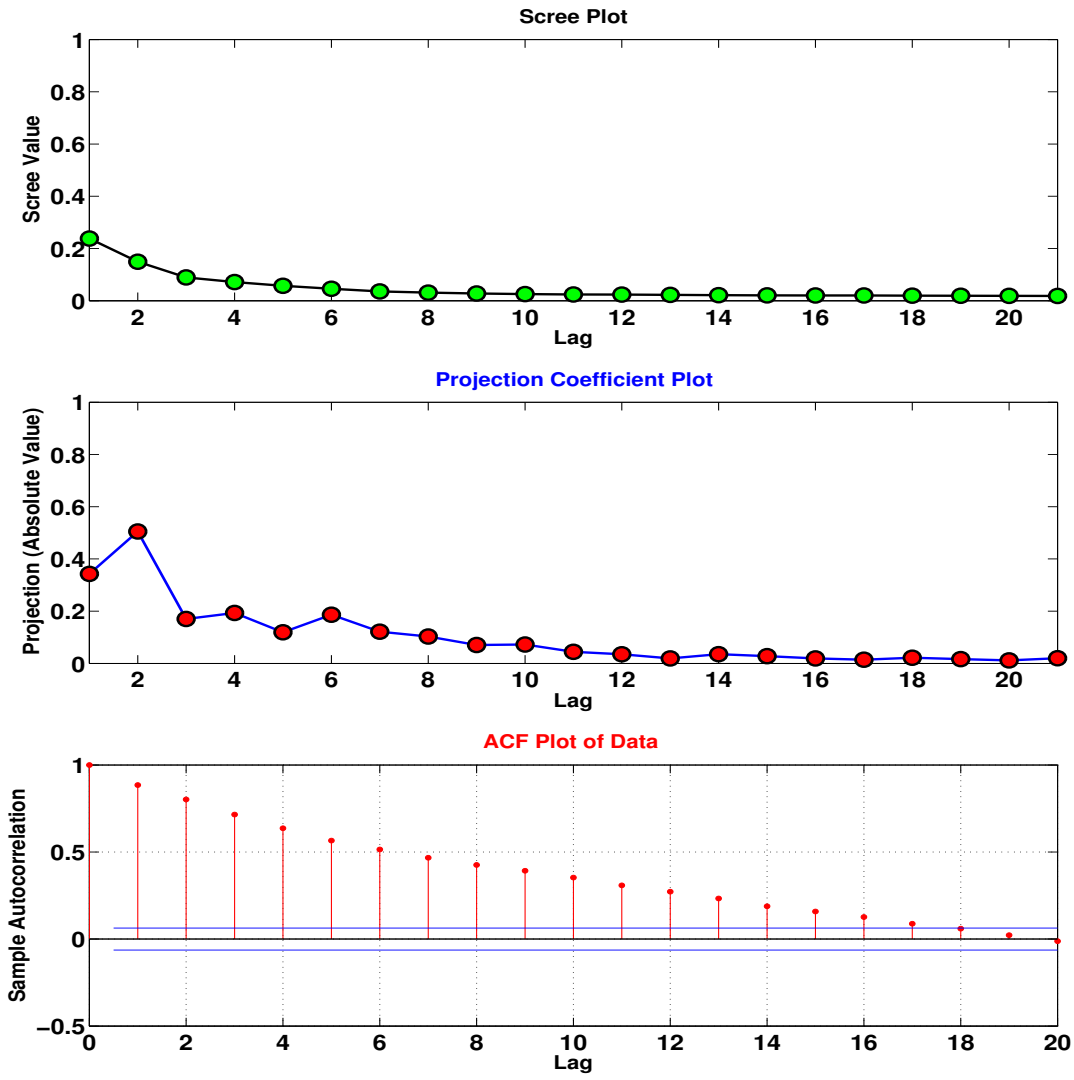


Figure 6.6: **Scree, projection coefficient, and ACF plots for Example 2**

The first panel in Figure 6.5 provides a plot of the process data $\boldsymbol{y}_t$ and the prediction $\hat{\boldsymbol{y}}_{t|t-b}$. One interesting observation is that $\hat{\boldsymbol{y}}_{t|t-b}$ lags $\boldsymbol{y}_t$ by b, i.e., lag 3 in this example. The ACF plot of the prediction error $\boldsymbol{e}_{t|t-b}$ in the fourth panel shows that the prediction error is an MA process of order $b-1$, as expected. Shifting the prediction by the delay lag b produces a prediction that follows exactly the original

data, as shown in the second panel. This indicates the reason why a low performance index, 49.08%, is obtained. In the meanwhile, the prediction error time series is shown in the third panel, together with the output data. Additionally, the prediction has been expressed as a summation of orthogonal components. Theoretically, these components can be eliminated through the manipulated variables, if under MVC.

Figure 6.6 depicts the scree, projection coefficient, and ACF plots for the SSA-based CPA analysis. The scree plot shows a tapered profile, suggesting that the system is not under MVC, which is further confirmed by the ACF plot which shows a slowly decaying sequence of autocorrelation representative of an autoregressive time series. Moreover, the projection coefficient plot indicates several fairly large values beyond lag 3, indicating that there are still correlations beyond the delay horizon associated with sub-MVC performance. These values are $30-50\%$ of the largest value, and are roughly twice the magnitude of the values that are approached asympotitically in the projection coefficient plot.

Given the results in Section 6.3.2, the closed-loop series can be broken into a summation of weighted singular vectors, i.e., $(\boldsymbol{u}_i^T \boldsymbol{y}_t)\boldsymbol{u}_i$. The four panels in Figure 6.7 show the first four weighted singular vectors together with the original series. Several observations can be made: 1) the first two weighted singular vectors are larger than the second two weighted singular vectors. This implies that the magnitudes in four plots have a good agreement with the values in the projection coefficient plot in Figure 6.6. 2) the first two weighted singular vectors follow considerable trends in the original series, and the vectors lag $\boldsymbol{y}_t$. This gives a good match with the prediction plot in Figure 6.5, because the prediction is a summation of the first m weighted singular vectors.

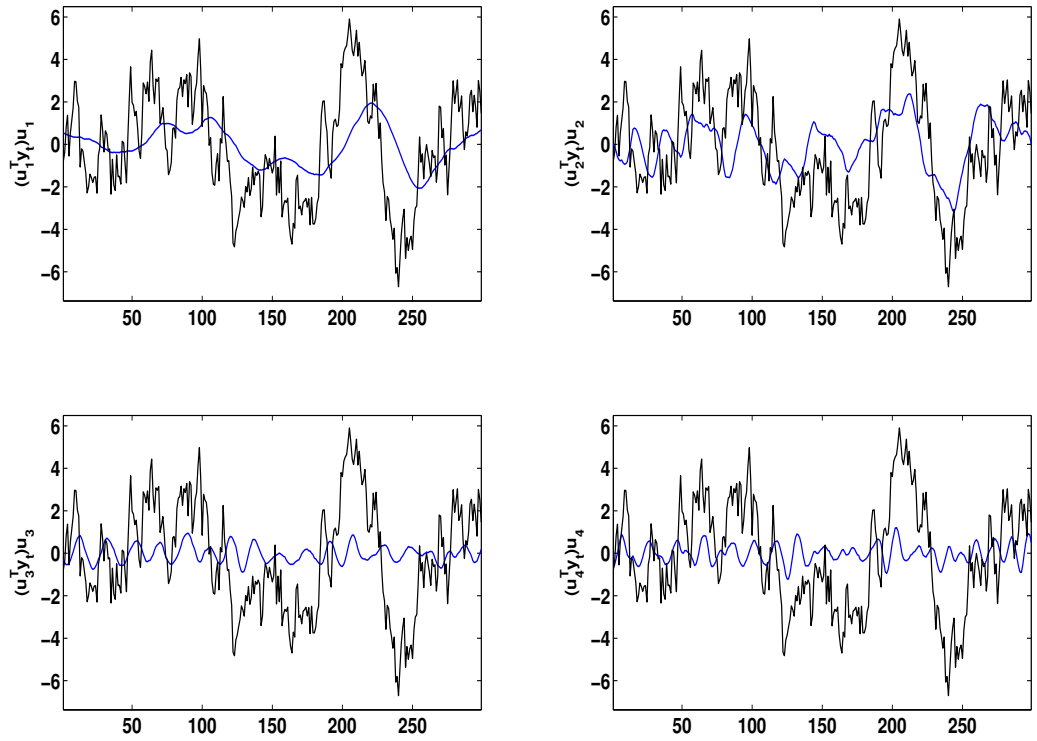


Figure 6.7: **First four weighted singular vectors (blue line) versus original series (black line) of Example 2**

In Figure 6.8, the spectra for the process data, prediction, and prediction error are provided. Note that the spectra in this Figure are not smoothed. The prediction spectrum depicts energy concentrated at the low-frequency range and only gets partial energy in process data in this range. Meanwhile, the prediction error does catch energy at both low-frequency and high-frequency parts. Note that the vertical axis scales are different. Recall that the prediction error is an MA process of order $b-1$ instead of a white noise. Because the decomposition is orthogonal, theoretically the variances are additive, i.e., the areas under the spectral curves are additive.

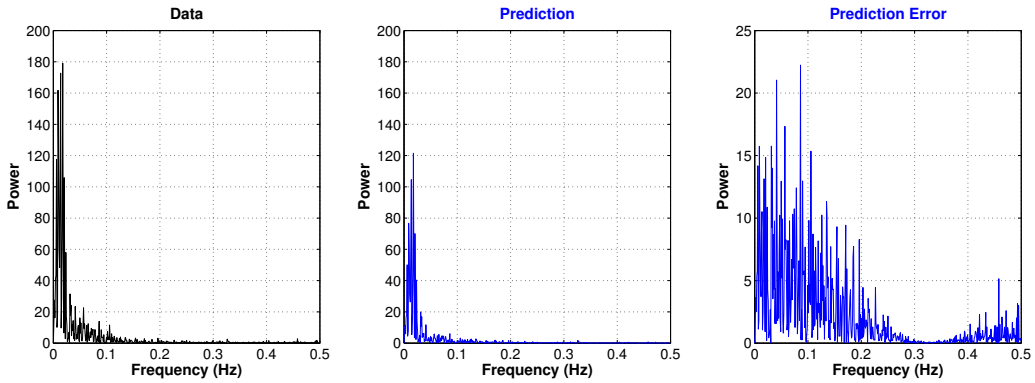


Figure 6.8: **Amplitude spectrum plots of Example 2**

6.4 Conclusion

Singular Spectrum Analysis (SSA) has been introduced as an effective technique for process analysis, and for CPA. Specifically:

1) chemical process analysis

   Structural insights of SSA in Chapter 5 are extended into elements in chemical process data from a filtering and spectrum perspective. An example validates the potential use of SSA in the chemical industry, and additional interpretations are provided based on the filtering interpretations in Chapter 5. This demonstration agrees well with the SSA work proposed by Aldrich's group [5, 6], and complements the work of Barkhuizen [7] .

2) controller performance assessment (CPA)

   The SSA-based technique for computing the MVC-based performance index provides a direct means for calculating the CPA index from data, requiring only knowledge of the delay horizon and the maximum delay to include in a trajectory matrix. As well, structural insights into elements in the closed-loop data are presented;

further, these elements can include dynamic coupling beyond the time-delay horizon, and periodic components. The results in this work produce a link between principal-component like analyses of process data and the MVC-based controller performance index. The application and interpretation of the SSA-based approach have been illustrated using an example.

Recently, a two-dimensional extension of SSA (2D-SSA) has been proposed in [200]. Similar to the standard SSA, obtaining filtering and frequency insights of 2D-SSA will be a natural extension. Furthermore, 2D-SSA provides process analysis in both temporal and spatial directions. It may be possible to extend 2D-SSA to plant data from cross-direction control systems and distributed control systems [208], e.g., sheet forming processes. Preliminary research results of 2D-SSA are presented in the next chapter.

Chapter 7

Two-dimensional SSA

Multi-dimensional SSA is a natural extension of SSA, and offers promise for multivariate analysis of manufacturing processes. Multi-dimensional SSA typically appears in two forms: multi-channel SSA (M-SSA) and two-dimensional SSA (2D-SSA). The discussion in this chapter focuses on 2D-SSA, which was recently introduced by Golyandina et al. in [194, 200]. This chapter is structured as follows: 1) the basic algorithm of 2D-SSA is covered; 2) 2D-SSA is compared to other data analysis techniques; 3) 2D spectral analysis is presented to assist in extracting further insights of 2D reconstructed series by 2D-SSA; 4) Preliminary results using simulated sheet forming data are shown to validate the feasibility of 2D-SSA analysis.

7.1 The Basic Algorithm

Golyandina and Usevich [194] first introduced the 2D extension of SSA. Recall that the basis of 1D-SSA is the singular value decomposition (SVD) of a trajectory matrix

that is placed in a Hankel matrix form. In 2D-SSA, a trajectory matrix is of Hankel-block-Hankel (HbH) form.

In the discussion that follows, we will assume that there are two spatial dimensions. Note however that 2D SSA can be applied to series having two indices, which could be two spatial indices, or a temporal and spatial index (e.g., measurements in time over a 1D spatial domain). In addition, terminology related to rolling operations will be used, namely machine direction (the direction of the sheet moving through the rolling mill), and cross-direction (the direction across the width of the sheet).

Let $\{y(i,j),\, i = 1, 2, \cdots, N_r,\ j = 1, 2, \cdots, N_c\}$ be a 2D array. Note that each row and column represents a series in the Machine Direction (MD) and Cross Direction (CD), respectively.

$$\boldsymbol{Y}_{t,x} = \begin{bmatrix} y(1,1) & y(1,2) & \cdots & y(1,N_c) \\ y(2,1) & y(2,2) & \cdots & y(2,N_c) \\ \vdots & \vdots & & \vdots \\ y(N_r,1) & y(N_r,2) & \cdots & y(N_r,N_c) \end{bmatrix}$$

Window sizes (L_r, L_c) are defined for row and column series, respectively. In addition, we have $K_r = N_r - L_r + 1$ and $K_c = N_c - L_c + 1$. An $(L_r L_c \times K_r K_c)$ HbH matrix can be formed:

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots & \boldsymbol{A}_{K_c} \\ \boldsymbol{A}_2 & \boldsymbol{A}_3 & \boldsymbol{A}_4 & \cdots & \boldsymbol{A}_{K_c+1} \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 & \boldsymbol{A}_5 & \cdots & \boldsymbol{A}_{K_c+2} \\ \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{A}_{L_c} & \boldsymbol{A}_{L_c+1} & \boldsymbol{A}_{L_c+2} & \cdots & \boldsymbol{A}_{N_c} \end{bmatrix},$$

where

$$\boldsymbol{A}_j = \begin{bmatrix} y(1,j) & y(2,j) & \cdots & y(K_r,j) \\ y(2,j) & y(3,j) & \cdots & y(K_r+1,j) \\ \vdots & \vdots & & \vdots \\ y(L_r,j) & y(L_r+1,j) & \cdots & y(N_r,j) \end{bmatrix}, \quad j = 1, 2, \cdots, N_c.$$

With an HbH matrix, SVD and grouping steps are the same as in 1D-SSA. The only difference is in a 2D-Hankelization step.

Assume

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_{1,1} & \boldsymbol{Z}_{1,2} & \boldsymbol{Z}_{1,3} & \cdots & \boldsymbol{Z}_{1,K_c} \\ \boldsymbol{Z}_{2,1} & \boldsymbol{Z}_{2,2} & \boldsymbol{Z}_{2,3} & \cdots & \boldsymbol{Z}_{2,K_c} \\ \boldsymbol{Z}_{3,1} & \boldsymbol{Z}_{3,2} & \boldsymbol{Z}_{3,3} & \cdots & \boldsymbol{Z}_{3,K_c} \\ \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{Z}_{L_c,1} & \boldsymbol{Z}_{L_c,2} & \boldsymbol{Z}_{L_c,3} & \cdots & \boldsymbol{Z}_{L_c,K_c} \end{bmatrix}$$

There are two steps in a 2D-Hankelization process. First, within-block Hankelization, which Hankelizes each block, is implemented. Second, Hankelization is applied to the whole matrix, a process named "between-block Hankelization". The order of these two steps can be reversed.

$$\mathcal{H}\boldsymbol{Z} = \mathcal{H} \begin{bmatrix} \mathcal{H}\boldsymbol{Z}_{1,1} & \mathcal{H}\boldsymbol{Z}_{1,2} & \mathcal{H}\boldsymbol{Z}_{1,3} & \cdots & \mathcal{H}\boldsymbol{Z}_{1,K_c} \\ \mathcal{H}\boldsymbol{Z}_{2,1} & \mathcal{H}\boldsymbol{Z}_{2,2} & \mathcal{H}\boldsymbol{Z}_{2,3} & \cdots & \mathcal{H}\boldsymbol{Z}_{2,K_c} \\ \mathcal{H}\boldsymbol{Z}_{3,1} & \mathcal{H}\boldsymbol{Z}_{3,2} & \mathcal{H}\boldsymbol{Z}_{3,3} & \cdots & \mathcal{H}\boldsymbol{Z}_{3,K_c} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathcal{H}\boldsymbol{Z}_{L_c,1} & \mathcal{H}\boldsymbol{Z}_{L_c,2} & \mathcal{H}\boldsymbol{Z}_{L_c,3} & \cdots & \mathcal{H}\boldsymbol{Z}_{L_c,K_c} \end{bmatrix}$$

After undertaking the SVD decomposition to the HbH matrix $\boldsymbol{W}$, we have:

$$\boldsymbol{W} = \sum_{i=1}^{\mu} \boldsymbol{W}_i = \sum_{i=1}^{\mu} \sqrt{\lambda_i} \boldsymbol{u}_i \boldsymbol{v}_i^T \tag{7.1}$$

where $\mu = min(L_r L_c, K_r K_c)$, λ_i and $\boldsymbol{v}_i$ are the eigenvalues and eigenvectors of $\boldsymbol{W}^T\boldsymbol{W}$, and $\boldsymbol{u}_i$ are the eigenvectors of $\boldsymbol{W}\boldsymbol{W}^T$.

Converting the vector $\boldsymbol{v}_i$ into a matrix form, a 2D eigenfilter can be obtained. To simplify the notation, let $\boldsymbol{v}$ be one of the eigenvectors instead of $\boldsymbol{v}_i$. Then, 2D eigenfilter coefficients can be written as:

$$\boldsymbol{v} = \begin{bmatrix} v_{1,1} & v_{1,2} & v_{1,3} & \cdots & v_{1,K_c} \\ v_{2,1} & v_{2,2} & v_{2,3} & \cdots & v_{2,K_c} \\ v_{3,1} & v_{3,2} & v_{3,3} & \cdots & v_{3,K_c} \\ \vdots & \vdots & \vdots & & \vdots \\ v_{K_r,1} & v_{K_r,2} & v_{K_r,3} & \cdots & v_{K_r,K_c} \end{bmatrix}$$

As seen, a number of 2D-arrays $\boldsymbol{W}_i$ are reconstructed. The challenge is to interpret patterns of 2D reconstructed components and associated 2D eigenfilters.

7.2 2D-SSA Compared to Other Multivariate Analysis Techniques

7.2.1 Conventional Multivariate Analysis Methods

Chemical processes are inherently multivariate by nature, with behaviours arising from the interplay between material and energy balances, chemical reactions, and other types of phenomena. For processes having spatially distributed behaviour, a fundamental model will typically consist of a system of partial differential equations (PDEs) arising from the material and energy balances. For applications in which high fidelity is required, numerical solution of the PDE models will be necessary in most applications. However, the computational burden of such models is in general higher,

so for real-time applications in particular, empirical models offer an attractive alternative. Within the array of empirical modelling approaches available, multivariate techniques have found widespread acceptance.

Recall that analysis of industrial processes can be often treated as a multivariate time series problem. Assuming linear process dynamics, a typical discrete transfer function model form is [209, 210]:

$$\boldsymbol{Y}_t = \boldsymbol{G}(q^{-1})\boldsymbol{U}_t + \boldsymbol{D}_t \tag{7.2}$$

where $\boldsymbol{Y}_t$ is a vector of n outputs, $\boldsymbol{U}_t$ is a vector of m actuator effects, $\boldsymbol{D}_t$ is a multivariate disturbance, and $\boldsymbol{G}(q^{-1})$ is a process transfer function matrix, where q^{-1} is a backshift operator, i.e., $q^{-1}\boldsymbol{Y}_t = \boldsymbol{Y}_{t-1}$.

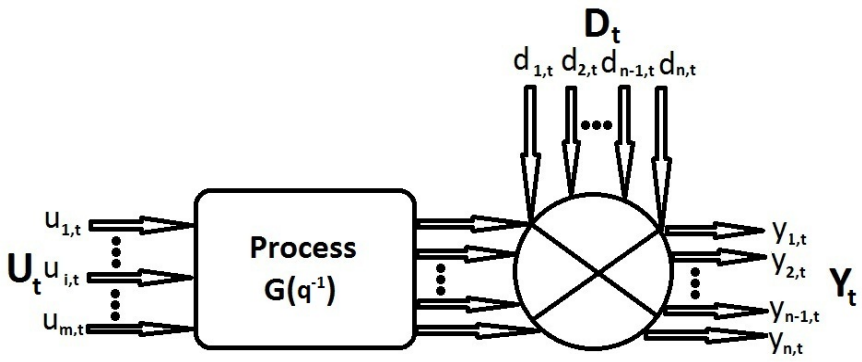


Figure 7.1: **Multivariate linear system with disturbance**

A schematic for a multivariate linear system with disturbance is illustrated in Figure 7.1. Models of the form in Equation (7.2) can be estimated using multiple regression and multivariate time series methods, and a range of techniques (e.g., subspace identification) from the system identification field.

The model Equation (7.2) contains explicit time dependence, but can be used to describe spatial variation as well. In this instance, elements of the output vector would

correspond to the value of a process variable at a specific location (e.g., two elements of $\boldsymbol{Y}_t$ can represent the temperature evolution in time at two different locations on an aluminum sheet). There is only one backshift operator involved, and it is associated with the time domain. This approach is similar to that used in nonlinear regression of parameters in mechanistic models, where it is common to have different responses representing concentrations at different locations, or weight fractions of polymers at different chain lengths.

Alternatively, it is possible to incorporate both spatial and temporal variation in a transfer function model, through the use of two (or more, depending on the dimension of the spatial domain) backshift operators. For example, Wellstead et al. [211] approximate PDEs as two-dimensional transfer function models, i.e., 2D-ARMAX models, shown in Equation 7.3. Instead of only one time index in VAR models [209, 210], both time and space indices are considered in 2D-ARMAX models. Here the subscript t denotes the measurement location in time, and the subscript x denotes the measurement position in the spatial direction. In addition, w^{-1} and q^{-1} are interpreted as backshift operators, i.e., $w^{-1}\boldsymbol{Y}(x,t) = \boldsymbol{Y}(x-1,t)$ and $q^{-1}\boldsymbol{Y}(x,t) = \boldsymbol{Y}(x,t-1)$. A 2D-ARMAX model can be expressed as:

$$\boldsymbol{A}(w^{-1},q^{-1})\boldsymbol{Y}(x,t) = q^{-v}\boldsymbol{B}(w^{-1},q^{-1})\boldsymbol{U}(x,t) + \boldsymbol{C}(w^{-1},q^{-1})\boldsymbol{E}(x,t) \tag{7.3}$$

where $\boldsymbol{Y}(x,t)$ is the output, $\boldsymbol{U}(x,t)$ is the actuator input, and $\boldsymbol{E}(x,t)$ is the disturbance.

2D-ARMAX models have some drawbacks. First, an underlying full-rank assumption needs to be made to a measurement matrix. This may not hold when measurements are strongly correlated. Even if a full-rank matrix is given, there might be some other process constraints, arising from for example mass and energy balance

constraints. Rapid sampling - e.g., of gauge across a sheet - may also lead to poor conditioning and rank deficiency. Second, a large number of actuators and output measurements in sheet forming processes make it a large scale problem. Similar to conventional VAR models, Ramarathnam [212] comments in his thesis that 2D-ARMAX models are computationally too expensive since too many parameters were involved. This motivates, as well, reduced-rank approaches, which simplify problems by dimensional reduction.

7.2.1.1 Open-Loop Data versus Closed-Loop Data

Data can be collected in industrial operations in either open-loop or closed-loop forms. For open-loop processes, a typical model form is $\boldsymbol{Y}_t = \boldsymbol{G}(q^{-1})\boldsymbol{U}_{t-b}$. For sheet forming processes the elements of the output vector can represent quantities such as thickness or basis weight at spatial locations across the sheet. In such a modelling approach, spatial interactions are represented by off-diagonal transfer function elements in $\boldsymbol{G}(q^{-1})$. A common assumption in such models, especially for paper machines, is to assume separability of temporal and spatial effects in $\boldsymbol{G}(q^{-1})$, so that we can write $\boldsymbol{G}(q^{-1}) = \boldsymbol{M}\boldsymbol{G}_0(q^{-1})$. $\boldsymbol{G}_0(q^{-1})$ is a common scalar temporal transfer function of a low order, and $\boldsymbol{M}$ is an interaction matrix which describes couplings in the spatial direction between the measurements in the output vector $\boldsymbol{Y}$. Implicit in the form of $\boldsymbol{M}$ is that the spatial (cross-direction) dynamics are fast relative to the temporal (machine direction) dynamics, so that $\boldsymbol{M}$ is a constant matrix [213]. This assumption is typically appropriate for paper machines, in which the scanning gauge is located further down the mill, by which time spatial dynamics have settled out. An interaction matrix of a special structure, e.g., Toeplitz symmetric and circulant symmetric,

can be given by assuming that each actuator has the same and symmetric effect. This leads to interaction-matrix based approaches, which are dominant in spatial control problems [121, 212, 213]. In essence, orthogonal basis functions are used to describe an open-loop interaction matrix. The separability assumption is typically not valid for metal-rolling processes, and consequently, a full transfer function matrix is required.

Good quality open-loop data requires testing, e.g., 'bump' tests and PRBS tests. In industrial operations, these tests are likely to happen infrequently, and open-loop data are often not available because of the need to balance production of off-specification products against the quality of information obtained.

Under closed-loop control, the system can be expressed as

$$\boldsymbol{Y}_t = [\boldsymbol{I} + \boldsymbol{G}(q^{-1})\boldsymbol{G}_c(q^{-1})]^{-1}\boldsymbol{D}_t + [\boldsymbol{I} + \boldsymbol{G}(q^{-1})\boldsymbol{G}_c(q^{-1})]^{-1}\boldsymbol{G}(q^{-1})\boldsymbol{G}_c(q^{-1})\boldsymbol{SP}_t \quad (7.4)$$

where $\boldsymbol{D}_t$ is a disturbance and $\boldsymbol{SP}_t$ represents the setpoint. The closed-loop structure is very likely different from the open-loop one, and how the structure changes depends on the controller structure. To get good quality closed-loop data to estimate process models, perturbations need to be introduced, typically in the form of dither signals to the controller output in the closed-loop control system. Analysis tools can be applied to routine operating data, which can be collected without disturbing processes, however in this instance, the model reflects the disturbance-output relationship.

One difference between process analysis based on open- versus closed-loop data lies in how we interpret the results. For example, if PCA is applied to an open-loop process where a separability assumption holds, PCA will provide information about spatial interactions, such as on a paper machine. However, separability may not be able to be assumed when applying PCA to closed-loop data, even if data are from a paper machine because a control system can introduce different kinds of coupling

between variables. Accordingly, interpretations of principal components are different and may not be as easy as in open-loop data.

7.2.2 Reduced-Rank Techniques

Reinsel and Velu [214] introduce the idea of reduced-rank coefficient models to multiple linear regression and multivariate time series modeling. Regardless of whether data are open- or closed-loop, reduced-rank techniques are applicable for analysis via basis functions. In reality, basis functions can be classified into two types. One type of is predefined, e.g., square impulse functions [121], orthogonal polynomials [122, 123], splines [124], Fourier series[121], and wavelets [125]. The other type is data-driven, based on $Karhunen - Loève$ decomposition, e.g., PCA and SSA. Chapter 5 also utilizes reduced-rank approximations to a trajectory matrix to provide filtering and frequency interpretations of 1D-SSA.

7.2.2.1 Predefined Basis Functions

As discussed in the previous section, an interaction matrix can have some special structures. Three matrix forms, centrosymmetric, Toeplitz symmetric and circulant symmetric, are most commonly used [98, 208, 212, 213, 215]. A band-diagonal Toeplitz symmetric form, where the same element is repeated in each matrix diagonal, is shown below for demonstration. One more assumption is that each actuator has a localized spatial effect in a nearby region.

$$\boldsymbol{M} = \begin{bmatrix} m_1 & m_2 & \cdots & m_h & 0 & \cdots & \cdots & 0 \\ m_2 & m_1 & m_2 & \cdots & m_h & & & \vdots \\ \vdots & m_2 & m_1 & m_2 & & \ddots & & \vdots \\ m_h & \vdots & m_2 & \ddots & \ddots & & m_h & 0 \\ 0 & m_h & & \ddots & \ddots & m_2 & \vdots & m_h \\ \vdots & & \ddots & & m_2 & m_1 & m_2 & \vdots \\ \vdots & & & m_h & \cdots & m_2 & m_1 & m_2 \\ 0 & \cdots & \cdots & 0 & m_h & \cdots & m_2 & m_1 \end{bmatrix}_{N_c \times N_c}$$

Basis function based approaches are suggested to address a rank-reduction problem. A basis function expansion is proposed to represent an interaction matrix [121, 122, 123]. By appropriately truncating higher order terms, spatial dynamics can be re-described in a finite-dimensional form. Typical basis functions, which are predefined with a prior process knowledge, cover square impulse functions, orthogonal polynomials, splines, Fourier series, and wavelets. For illustration a brief discussion about two basis functions is provided in this section, and the reader is referred to [121] for more details. Note that in practise discrete basis functions are used. Below basis functions are denoted by $\boldsymbol{\phi}_i(x)$.

Square Impulse Functions [121]

Assume that there are N_c positions in the spatial direction, and orthogonal impulse functions can be defined as:

$$\boldsymbol{\phi}_i(x) = \begin{cases} 1 & for\ x = i \\ 0 & for\ x \neq i \end{cases} \qquad i = 1, 2, \cdots, N_c$$

Each actuator response can be expressed via defined basis functions above. A physical plot for these basis functions is:

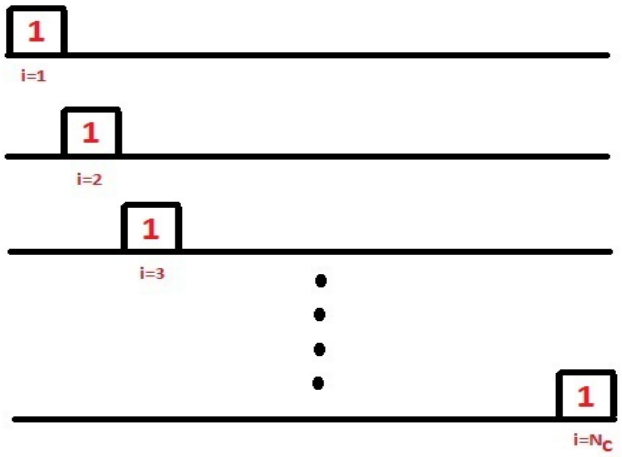


Figure 7.2: **Square impulse functions**

If both actuator responses and output measurements are expressed by the same basis functions, actuator setpoints can be given by solving a linear regression problem.

Orthogonal Polynomials [122, 123]

One class of discrete orthogonal polynomials that are often used in basis function representation are the Gram polynomials [123]. A general recursive form can be written as

$$\boldsymbol{\phi}_i(x) = \frac{(N_c-1)(2i-1)}{i(N_c-i)}(1-\frac{2(x-1)}{N_c-1})\boldsymbol{\phi}_{i-1}(x) - \frac{(i-1)(N_c-1+i)}{i(N_c-i)}\boldsymbol{\phi}_{i-2}(x)$$

where the subscript i denotes a polynomial in the i^{th} order, N_c is the number of spatial positions, and $x = 1, 2, \cdots, N_c$. In addition, the following basis function assignments are made: $\boldsymbol{\phi}_0(x) = 1$ and $\boldsymbol{\phi}_{-1}(x) = 0$. An example of polynomials up to the fifth order are plotted in Figure 7.3 for illustration, and there are 100 spatial positions in this example.

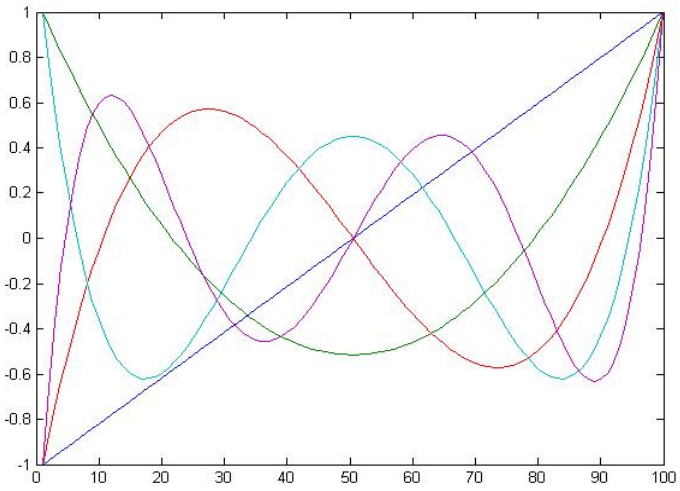


Figure 7.3: **Gram polynomials with the order up to five**

7.2.2.2 Data-driven basis functions

Data-driven basis functions have been widely used in many scientific and engineering applications. Early users of these techniques included atmospheric, meteorologic, and geophysical scientists. One class - the multivariate statistical approaches such as PCA and PLS - identify linear bases by decomposing the covariance structure of the data. The family of SSA methods - M-SSA and 2D-SSA - are little known within the manufacturing industries, and thus it is promising to apply these methods in this area. Four methodologies are chosen in this work for comparison, and the key features of these methods are summarized in Table 7.1. The category 'basis functions' refers to predefined functions (e.g., Gram polynomials), while the other three methods are data-adaptive and give orthogonal basis functions, typically eigenvectors. Note that all these reduced-rank methods presented here can be applied to both open-loop and closed-loop data.

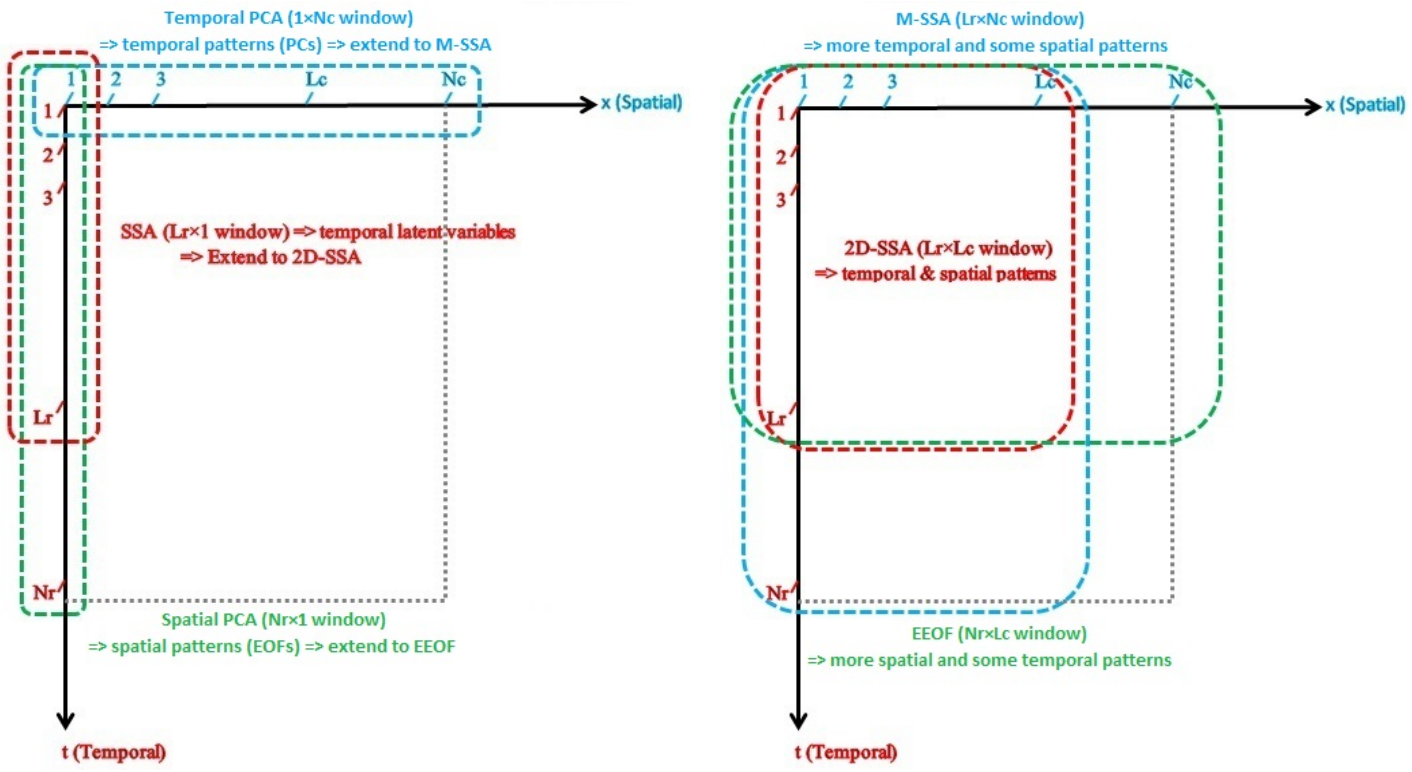
All methods except 2D-SSA essentially give 1D analysis and physical interpretations. 2D-SSA, which was introduced by Golyandina [194], is especially interesting

	Basis functions	PCA	M-SSA	2D-SSA
dimensional reduction	Yes	Yes	Yes	Yes
emphasis	spatial	temporal	more temporal and some spatial	temporal and spatial
problem size	short series	short and long series	short and long series	short and long series
correlation structure	-	temporal	temporal and spatial	temporal and spatial
prior knowledge (Yes) /data-adaptive (No)	Yes	No	No	No
physical interpretations	1D	1D	1D	2D
application areas	engineering	different areas	oceanography and meteorology	meteorology

Table 7.1: **Comparison table of reduced-rank methods**

due to the development of 2D physical interpretations. 2D-SSA can break down 2D data into low- and high-frequency components. Golyandina [194] applies 2D-SSA to digital terrain analysis in order to determine the continental, regional, and local components of a topographic surface.

Data-driven basis functions can be understood in practise in terms of a 'window', conceptually similar to a window function in signal processing. The left plot in Figure 7.4 shows a comparison of PCA and SSA. Standard PCA, also named temporal PCA, can be described as sliding a $1 \times N_c$ window along the temporal direction. Temporal PCA thus identifies high-variance temporal patterns, i.e., Principal Components (PCs). Spatial PCA (also named Empirical Orthogonal Functions (EOFs)) slides an $N_r \times 1$ window across N_c channels in the spatial direction and gives spatial patterns. Similarly, SSA is a time series technique which considers only one channel, and an $L_r \times 1$ window is used to identify temporal patterns.



(a) **PCA and SSA** (b) **EEOF, M-SSA and 2D-SSA**

Figure 7.4: **Windowing interpretations of PCA, SSA, EEOF, M-SSA, and 2D-SSA**. The axes are the temporal (t) and spatial (x) coordinates.

Windowing interpretations of PCA and SSA can be carried over into multi-channel and two-dimensional cases. In the first case, we can extend temporal and spatial PCA to multi-channel SSA (M-SSA) and extended EOF (EEOF), respectively. For M-SSA, as shown in the right plot of Figure 7.4, an $L_r \times N_c$ window can be moved along the t axis in order to get more temporal and some spatial patterns. By moving an $N_r \times L_c$ window along the x axis, EEOF focuses more on spatial patterns instead. The second conceptual route leads from SSA to 2D-SSA. To follow this route, 2D-SSA moves an $L_r \times L_c$ window in the (t, x)-plane, which gives temporal-spatial patterns. Windowing interpretations imply that the essential difference among these algorithms

is the trajectory matrix formed in each technique. In comparison, 2D-SSA is the only technique offering 2D insights into data, which has great potential to improve two-dimensional process analysis.

7.3 2D Spectral Analysis

The techniques discussed above are fundamentally time- and spatial-domain tools. In a manner similar to the 1D case, frequency domain approaches, i.e., 2D spectral tools, have been defined for the 2D case as well, and can be of great benefit by providing inside into the periodicity structure in the data and process. Multidimensional periodicity is present quite frequently in the metal rolling industry, due to a number of factors including eccentricities in rollers, and oscillations in hydraulic controllers. Such examples serve to motivate multivariate spectral analysis. Additional details on multivariate spectral analysis can be found in [216] and [217].

Intuitively, 2D spectral tools might be useful for processes encountering spatial characterizations. Garello [217] discusses 2D spectral tools, which are classified as nonparametric and parametric. For nonparametric spectral estimators, a generalized 2D periodogram is defined as:

$$S_{\boldsymbol{Y}}(f_t, f_s) \;=\; \frac{1}{N_r N_c} \Big| \sum_{k_1=1}^{N_r} \sum_{k_2=1}^{N_c} \boldsymbol{W}(k_1, k_2) \boldsymbol{Y}(k_1, k_2) e^{-j2\pi(k_1 f_t + k_2 f_s)} \Big|^2 \qquad (7.5)$$

where $\boldsymbol{Y}(k_1, k_2)$ is a 2D signal, $1 \leq k_1 \leq N_r$ and $1 \leq k_2 \leq N_c$; f_t and f_s are normalized frequencies at temporal and spatial directions, i.e., $0 \leq f_t \leq 0.5$ and $0 \leq f_s \leq 0.5$, and $\boldsymbol{W}(k_1, k_2)$ is defined as a 2D-window function.

Similar to well-known windows in 1D signal processing, such as Hanning and Hamming windows, a 2D-window function is used to reduce the power of sidelobes and

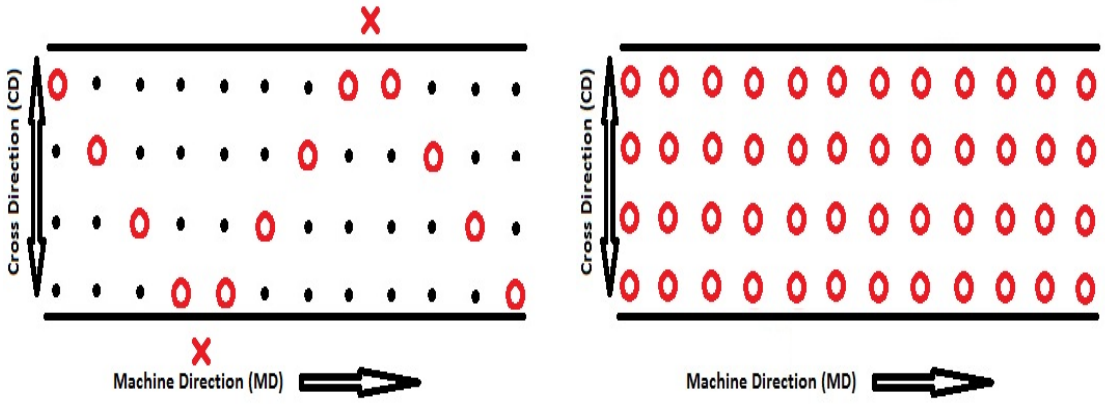
give a spectral estimator of higher resolution, however the cost of using a window is to widen the main lobe, which may instead cause a spectral estimate of lower resolution. Note that little research has been done to the area of 2D-window functions, which makes it an interesting area to explore.

Garello also investigates some parametric 2D spectral analysis methods, which are based on specific models, such as AR or MA models. For harmonic analysis, parametric spectral estimators are only desirable when a precise model can be fitted to a 2D signal; otherwise, these estimators are not reliable.

In this thesis, a 2D periodogram calculated by Equation 7.5 is used. In addition, the simplest window with all window coefficients equal to 1 is used, i.e., $\boldsymbol{W}(k_1, k_2) = 1$. In conjunction with prior process information, the 2D spectral technique can be used to provide physical interpretations in industrial processes.

7.3.1 Application of 2D Techniques to Sheet Forming Processes

Spatial characterizations are often encountered in sheet forming processes. Due to demands on high quality products, better process understanding and process control are required. Ammar [208] states that 30-300 actuators and 200-2000 measurements could exist in sheet forming processes. It is clear that different sensor types make modelling and control of systems fundamentally different. Cross-sheet sensors are typically classified as either scanning or fixed, depending on whether the scanning sensor remains fixed in a spatial location, or traverses the sheet. The data structure resulting from both types of sensors are shown in Figure 7.5 [218].



(a) **The Scanning Sensor** (b) **The Fixed Sensor**

Figure 7.5: **Typical sensors** (∘ = measurement, × = sensor off sheet)

As shown in Figure 7.5, scanning sensors have a number of shortcomings. Without data reconstruction (typically in scanning sensor applications), scanning sensors introduce missing data and can lead to the inseparability of temporal and spatial variations. Fixed sensors have the advantage of providing a full sheet measurement profile without missing data, which greatly reduces errors associated with process state estimators, at the cost of extra sensors. Accordingly, different sensors change process analysis considerably. Scanning sensors (e.g., for thickness or basis weight) are used frequently in paper machines. In contrast, fixed sensors are typically used in metal rolling operations. In this work, the focus is on data obtained using fixed sensors, consistent with the emphasis on analysis techniques for metal rolling operations.

To illustrate the capability of 2D spectral analysis for diagnosing metal rolling manufacturing processes, two scenarios of a rolling mill are presented below. Note that the programming code is written in MATLAB^{TM} script. The thickness values of the rolling sheets with or without defects in examples are assigned to be 1.5 and

2, respectively. Programming code of examples are provided in the appendix.

7.3.1.1 Example 1 - Uniform Ridge on the Work Roll

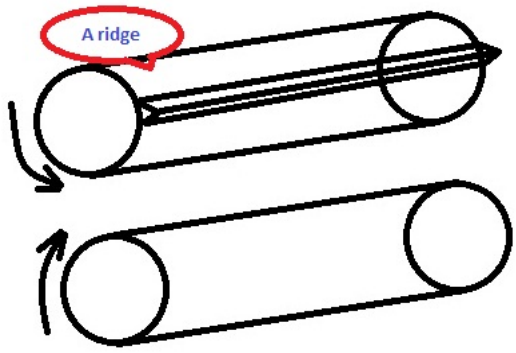


Figure 7.6: **Example 1: A mill with one ridge**

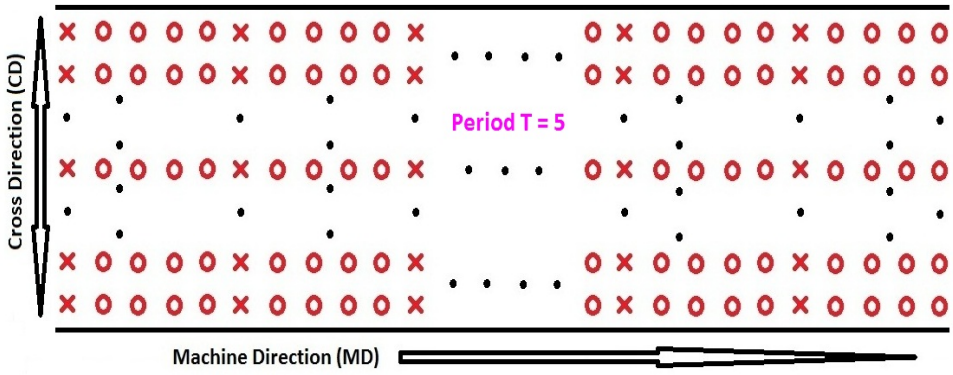


Figure 7.7: **The sheet profile for Example 1**

In this example, a mill with a uniform ridge defect across a work roll is shown in Figure 7.6, while Figure 7.7 displays a resulted sheet profile. The ridge introduces a temporal periodicity in the machine direction because it is uniform across the work roll. A temporal period of 5 seconds is assumed to be present in the simulated data, and no spatial periodicity in the machine direction should be present. The number of

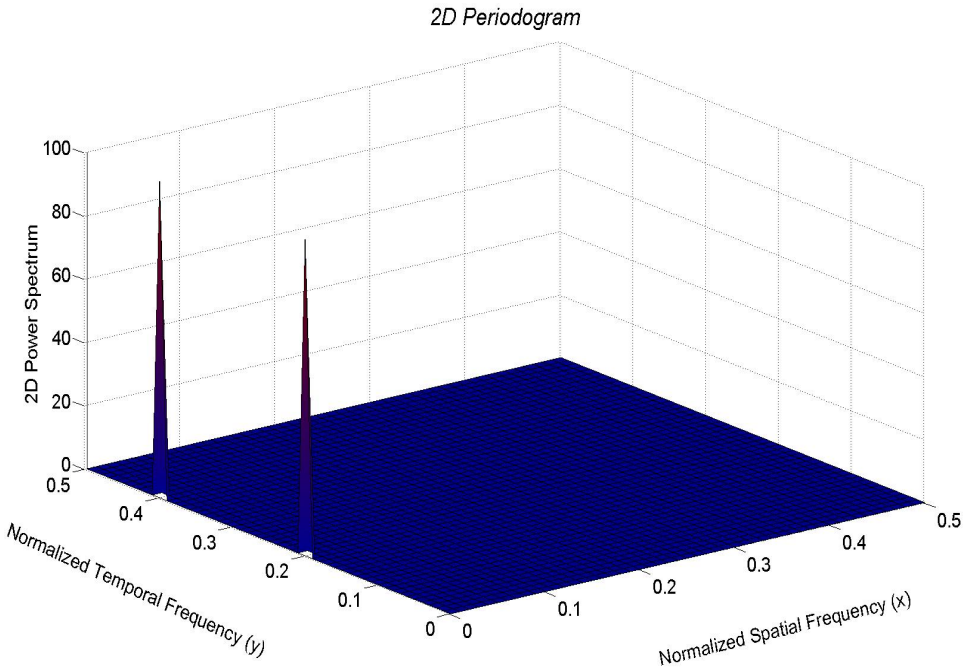


Figure 7.8: **2D periodogram for Example 1**

observations in the temporal and spatial direction are 100 and 50 respectively. With these hypotheses in mind, a 2D periodogram is calculated using Equation 7.5 from the simulated data, as depicted in Figure 7.8. Two peaks at normalized temporal frequencies of 0.2 Hz and 0.4 Hz are observed, as expected. Temporal effects are caused by periodic defects across the work roll. The temporal frequency at 0.2 Hz matches the period of 5 seconds; the one at 0.4 Hz is caused by the harmonic. Note that a spatial frequency only shows up at 0 Hz, i.e., a non-periodic component. The uniformity of the ridge defect on the work roll in the mill results in a profile in the spatial direction that is always uniform, and the results are consistent physically. These results can also be verified by theoretical derivation as well via using the mathematical representation of the defect, and calculating the 2D spectrum from its definition in terms of a fast Fourier transform (FFT).

7.3.1.2 Example 2 - Non-Uniform Ridge on the Work Roll

A non-uniform ridge on the work roll is considered in this example, in contrast to the uniform ridge of Example 1. The non-uniform ridge is illustrated in Figure 7.9. The resulting defects on the sheet profile conceptually are presented in Figure 7.10.

A non-uniform ridge on the work roll - one that moves along the circumference of the role as the cross-direction position changes - will still produce the same temporal periodicity at fixed location. This is because the a ridge at an arbitrary fixed location on the circumference of the role will cause a bump related to the rolling frequency - regardless of its spatial location. Consequently, we should find a temporal periodic component at the same frequency as in the previous example, and related to the roll diameter, and the rolling speed.

At a fixed moment in time, there will not be a spatial periodicity, because if the roll is frozen in time, there is only one bump across the roll in the cross direction. The spatial periodicity in this instance is not separable from the temporal periodicity: spatial periodicity occurs across time as well.

The period of a ridge defect shown in the temporal direction is assumed to be 100 seconds. For more precise results, 1000 observations in the temporal direction and 50 in the spatial direction are simulated. The calculated 2D periodogram in Figure 7.11 exhibits significant high-frequency peaks. The estimated 2D spectral information is misleading. In addition, a smaller low-frequency peak at a temporal frequency of 0.01 Hz and a spatial frequency of 0 Hz is observed. This peak is associated with the ridge defect, as expected. Example 2 will be rediscussed later in the application of the 2D-SSA technique.

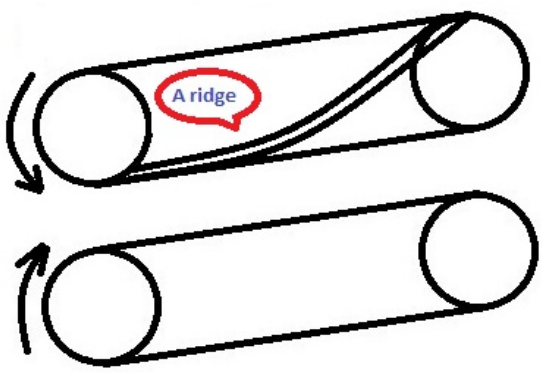


Figure 7.9: **Example 2: A rolling mill with a non-uniform ridge**

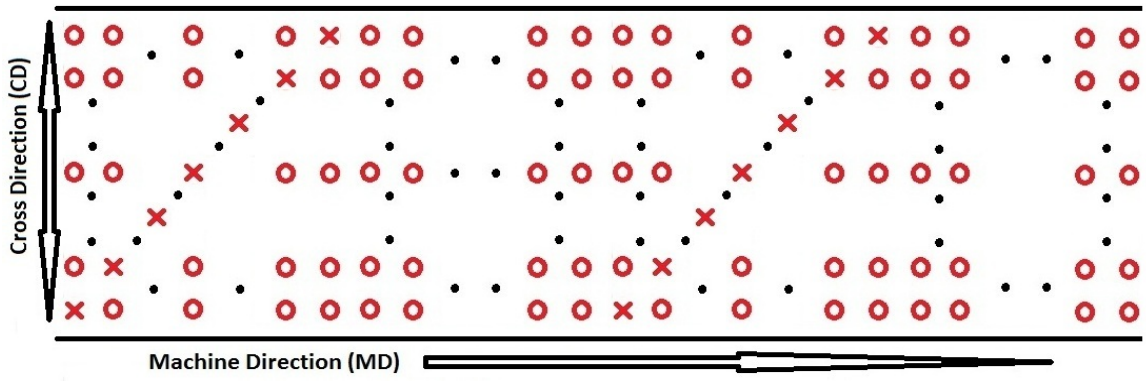


Figure 7.10: **The sheet profile for Example 2**

7.3.1.3 Discussion on 2D spectral tools

In the two examples, the 2D spectral tools do extract some useful information regarding synthetic rolling process data. The data considered in these examples have been noise-free. In industrial data, this will not be the case. In general, in order to obtain reliable spectral estimates, it will be necessary to clean the data, using filtering techniques. Two challenges are: 1) to form efficient filters; and 2) to better understand physical properties in processes. This motivates the combination of the 2D-SSA technique and 2D spectral tools, since 2D-SSA potentially offers 2D data cleaning filters.

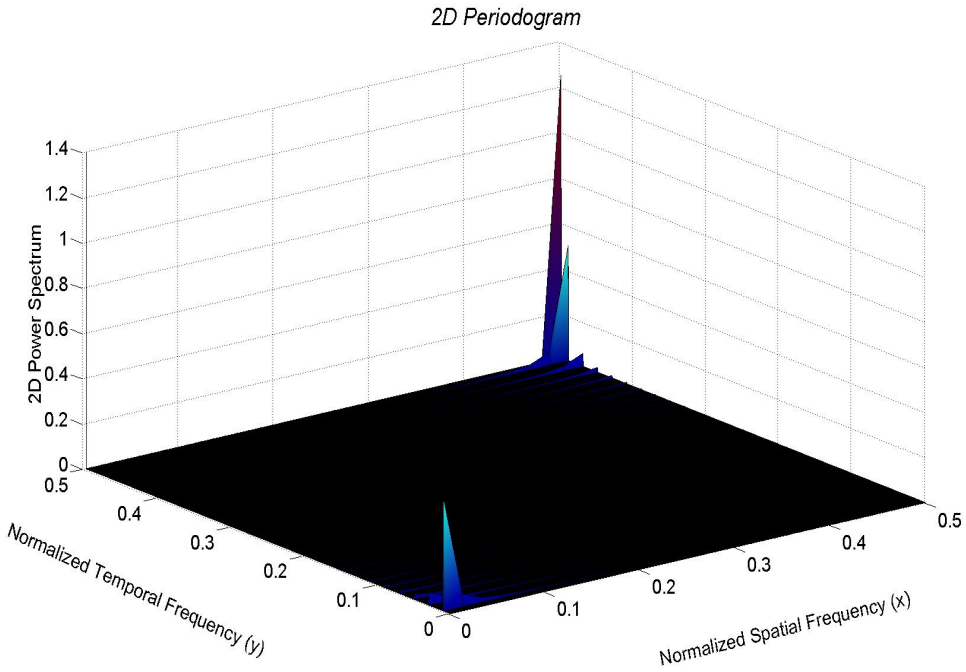


Figure 7.11: **2D periodogram for Example 2**

7.4 Preliminary Results of 2D-SSA Applications

The examples in Section 7.3.1 demonstrate some of the beneficial properties of 2D spectral tools, but in Example 2 it is clear that these tools can sometimes provide misleading information. 2D-SSA offers the promise of providing reconstructed components with spectral characteristics. To illustrate the application of the 2D-SSA technique, Example 2 is re-worked to provide continuity in discussion. The data are placed in a HbH matrix form. An SVD decomposition in Equation 7.1 is then applied, and the resulting singular vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are used to establish the reconstructed components $\boldsymbol{W}_i$. 2D spectral tools are expected to extract more useful and accurate information in the reconstructed components. In order to avoid computer memory constraints, a data set at 50 temporal locations for 200 spatial scans is used.

The results of the 2D-SSA analysis are presented in Figure 7.12. The top left plot

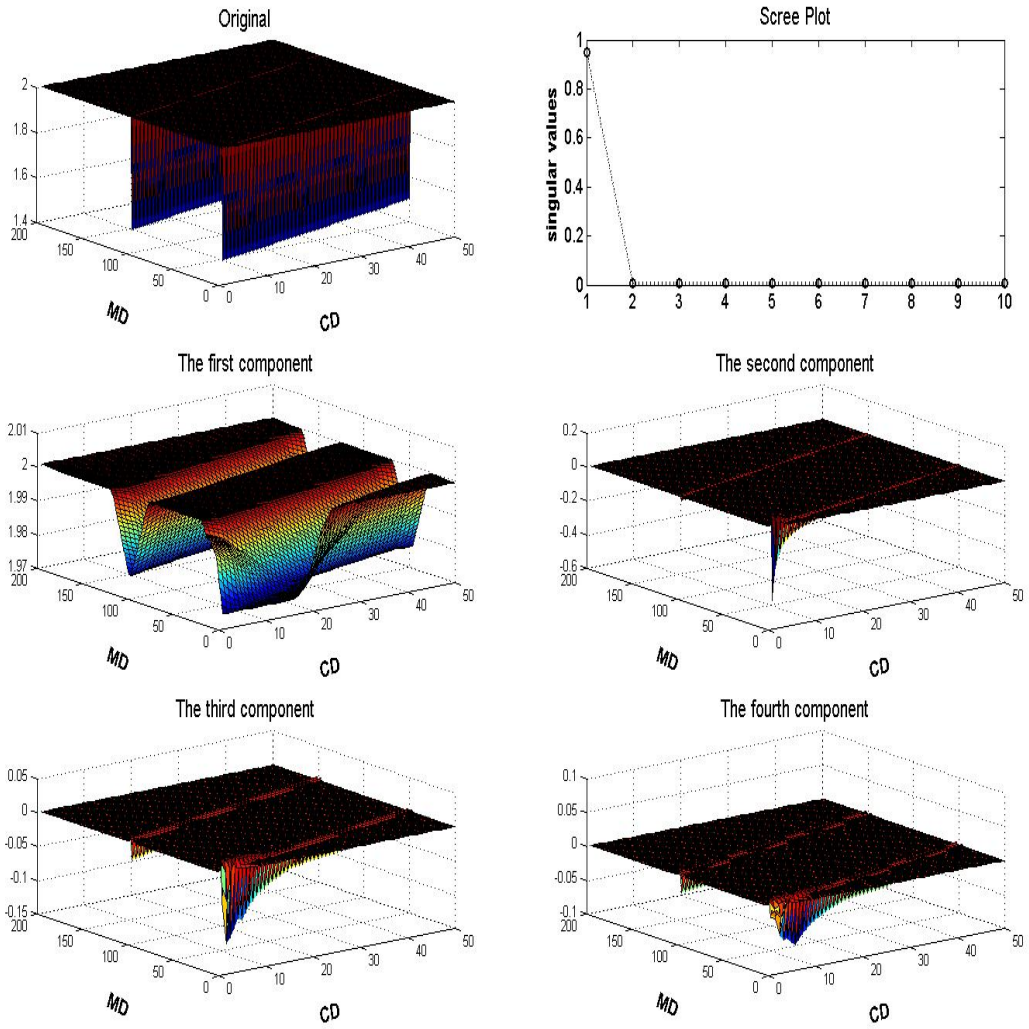


Figure 7.12: **2D-SSA analysis of Example 2** (a) Simulated data in Example 2 (top left); (b) The scree plot (top right); (c-f) First four reconstructed components (second and third rows)

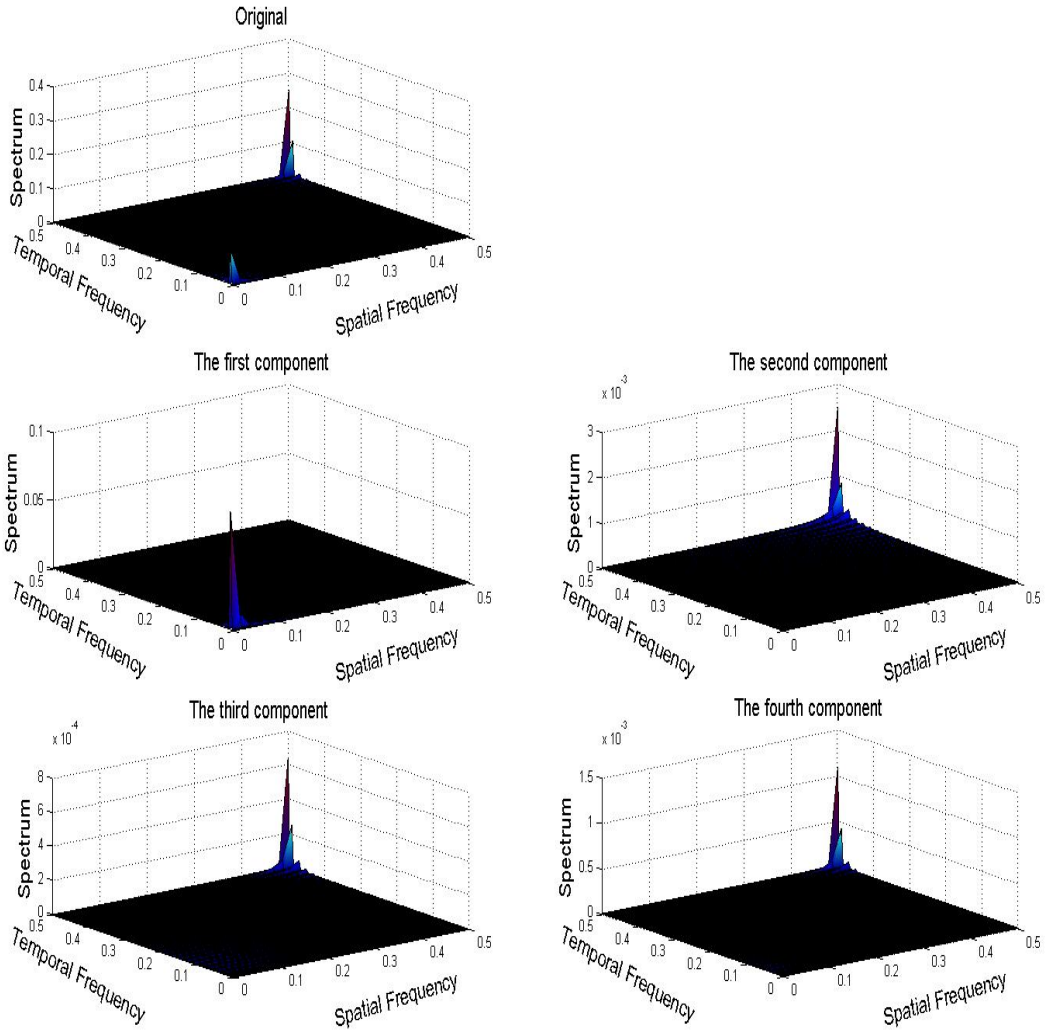


Figure 7.13: **2D periodogram for original data and the first four reconstructed components in Example 2**

charts the simulated data, the top right plot charts a scree plot, and the second and third rows of plots graph the first four reconstructed components. From the scree plot, only the first singular value is significant which is confirmed by the vertical axis scales in the reconstructed component plots. 2D periodograms for the original data and the first four reconstructed components are produced in Figure 7.13. The order of magnitude shows that the first component catches significant process characteristics with a low-frequency peak at a temporal frequency of 0.01 Hz and a spatial frequency of 0 Hz. This behaviour is expected due to the ridge defect. Insignificant high-frequency information can be observed in other components, illustrating the potential advantages of the 2D-SSA technique for providing more reliable physical interpretations of industrial processes. As well, reconstructed components can be explained as filtered versions of the original data.

7.5 Conclusion

In this chapter, multivariate time series tools are reviewed, setting the stage for comparison with the 2D-SSA technique for analyzing spectra in both temporal and spatial dimensions. In comparison with other basis function based techniques, 2D-SSA provides 2D insights from a windowing perspective. Furthermore, 2D spectral tools are briefly reviewed, and a 2D periodogram is suggested to be implemented together with 2D-SSA. The decomposition feature of 2D-SSA makes it possible to filter the original data, providing clearer results for the spectral analysis. Preliminary results in the analysis of the simulated data demonstrate the potential contribution of these tools in 2D process analysis.

Chapter 8

Conclusions and Recommendations

8.1 Conclusions

The goals of this thesis are: 1) to propose extensions of minimum-variance-based controller performance diagnostics for sheet forming processes that address a number of limitations of existing CPA techniques, 2) to propose interpretations and modifications to singular spectrum analysis and to demonstrate their applications in analyzing one-dimensional and two-dimensional problems, and 3) to propose an SSA-based extension of minimum-variance-based CPA that offers the promise of more detailed diagnostics.

The following specific conclusions and contributions arise from this work:

1) Length-based CPA for batch rolling

   Batch rolling process operation contains segments where the speed is constant, and startup/shutdown segments where the speed changes. The constant time

delay assumption on which CPA-MVB is based no longer holds during the transient parts in which speed is changing. Jelali [97] proposed a new length-based CPA for metal processing, but with no detailed implementation information. A length-based CPA is proposed in this thesis, with detailed implementation steps.The key step is to convert time-based data to length-based data via linear interpolation.

2) Constraint effects in CPA applications

Constraints on manipulated variables are often encountered, resulting in input nonlinearities. Conventional MVC-based CPA is not applicable to systems with nonlinear elements. Nonlinearities are introduced when the constraints are active. Harris and Yu [29] exploited the concept of a feedback invariant, which is not affected by certain classes of nonlinearity. A modified CPA-MVB algorithm is proposed to construct a feedback-invariant-performance bound that accommodates the presence of input constraints, i.e., constraints on controller output. A case study is presented for demonstration.

3) Resampling problem in CPA applications

Different input and output sampling intervals can make it difficult to compute the MVC-based CPA. This work discusses in detail the concerns of implementing conventional CPA with these data, and proposes solutions to compute CPA-MVB values in the scenarios of different input/output sampling rate. Industrial examples illustrate the impact of resampling on CPA.

4) Algorithm variations and interpretations of 1D-SSA

Singular spectrum analysis (SSA) in signal processing is introduced to the problem of process analysis. The key step in 1D-SSA is a singular value decomposition (SVD) on a lagged data matrix, i.e., a trajectory matrix. Eigenfilters and convolution filters are formed based on singular vectors of a trajectory matrix. This thesis proposes a modification of the trajectory matrix construction (symmetric Toeplitz matrices) that leads to filters of special structure and having a zero-phase lag property. New interpretations of 1D-SSA are developed from filtering and frequency perspectives. Examples are used to illustrate the application of these tools to process analysis problems.

5) Applications of SSA in chemical process analysis and performance assessment

The application of SSA for chemical process analysis is investigated by revisiting a published two-tank with periodic forcing example in more detail. Using the structure of a linear regression problem with SVD, a new process monitoring extension to CPA based on the SSA technique is proposed. The performance of the new approach is compared to the standard CPA-MVB approach results. In addition to providing a new non-parametric approach for computing the CPA index, the decomposed series provide supplementary process information for analysis use (e.g., periodic components within the delay horizon, or beyond the delay horizon). Extensions are identified for future work.

6) Interpretations of 2D-SSA

A recently proposed multi-dimensional SSA algorithms from the literature is applied to multi-dimensional data analysis. In particular, 2D-SSA is proposed

to detect temporal-spatial features in sheet forming processes. Windowing interpretations and preliminary results in case studies indicate the capability of 2D-SSA to perform 2D process analysis. Together with 2D spectral analysis and process knowledge, a case study illustrates observed process behaviours with physical features in reconstructed series. The final goal of 2D-SSA analysis is to get some meaningful physical interpretations of the 2D processes under investigation.

8.2 Recommendations for Future Work

Based on the work presented in this thesis, future research should include:

1) Extensions to CPA diagnostics to address industrial challenges

   In Chapter 3, three practical limitations in industrial CPA applications are addressed. Additional concerns in CPA applications include: 1) missing data can be often encountered in industry. How robust are MVC-based measures when dealing with missing data? 2) application results have demonstrated that different MVC-based performance assessments work well in different industrial processes, e.g., paper making versus chemical processes. However, few guidelines are available to select a proper MVC-based measure for applications in different industries.

2) One-dimensional process analysis tool

   New filtering and frequency interpretations offer more understanding of 1D-SSA and advantages for its applications. The difficulty in SSA analysis is to link process information to interpretations from the decomposed series. Additional

guidelines should be developed to facilitate this interpretation, leading to more widespread use of SSA as a process analysis tool.

3) Filtering and frequency interpretations of 2D-SSA

In Chapter 5, filtering and frequency interpretations of 1D-SSA are derived from properties of symmetric Toeplitz matrices and resulting filters. Eigenfilters and convolution filters can be developed from eigenvectors of these trajectory matrices. Golyandina [200] proposes the concept of 2D-SSA, which makes possible process analysis in both temporal and spatial directions. Chapter 7 presents a trajectory matrix in 2D-SSA of a Hankel-block-Hankel (HbH) matrix form. One future direction is to investigate properties of this HbH trajectory matrix, and develop modifications to the algorithm to make it more structured. The goal is to restructure this matrix so that an SVD decomposition of this matrix gives singular vectors with structural properties, e.g., symmetry and skew-symmetry, that can be used to provide 2D-filtering in signal processing and 2D data analysis. Essentially, the aim is to develop structural modifications of a trajectory matrix, 2D filters based on structural singular vectors, and filtering interpretations for 2D-SSA in a manner analogous to the 1D-SSA case proposed in Chapter 5 of this thesis.

4) 2D-SSA-based tools for assessing profile/flatness performance in sheet forming processes

2D-SSA has potential for analyzing process characteristics for sheet forming processes. It may be possible to interpret reconstructed series from the decomposition of a 2D trajectory matrix in an HbH structure (or variations of this

HbH structure) in terms of characteristics in the sheet profile. For example, in flatness control systems, eigenvector decompositions have been used as a basis for developing more effective control systems. Such tools would complement 2D spectral analysis for investigating the characteristics of sheet forming operation.

5) Two-dimensional CPA analysis

Hölttä [219] has introduced a new concept of 2D-CPA, however the proposed measure is essentially an MVC-benchmarked CPA measure across the sheet width, i.e., in the Cross Direction (CD). The work in Chapter 6 links the standard 1D-CPA benchmark with principal-component like decompositions, which makes the 2D-CPA concept more promising. It should be possible, using the 2D-SSA approach and keeping in mind the SSA-based computation of the 1D-CPA, to develop a fully 2D-CPA measure for controllers regulating both temporal and spatial properties, e.g., flatness control in sheet forming processes.

6) Application of SSA-based CPA techniques to plant and experimental data

In order to demonstrate the efficacy of SSA-based approaches for process analysis and CPA, and to provide further guidelines for interpretation, these techniques should be applied to larger-scale problems incorporating temporal and both temporal/spatial elements.

Bibliography

[1] A. Cinar, C. Undey, Statistical process and controller performance monitoring. a tutorial on current methods and future directions, Proceedings of the 2000 American Control Conference 4 (1999) 2625–2639.

[2] W. Bialkowski, Dreams versus reality: A view from both sides of the gap: Manufacturing excellence with come only through engineering excellence, Pulp & paper Canada 94 (1993) 19–27.

[3] M. Jelali, An overview of control performance assessment technology and industrial applications, Control Engineering Practice 14 (2006) 441–466.

[4] T. Harris, Assessment of control loop performance, The Canadian Journal of Chemical Engineering 67 (1989) 856–861.

[5] G. Jemwa, C. Aldrich, Classification of process dynamics with monte carlo singular spectrum analysis, Computers and Chemical Engineering 30 (2006) 816–831.

[6] C. Aldrich, M. Barkhuizen, Process system identification strategies based on the use of singular spectrum analysis, Minerals Engineering 16 (2003) 815–826.

[7] M. Barkhuizen, Analysis of process data with singular spectrum methods, Ph.D. thesis, University of Stellenbosch (2003).

[8] T. Harris, C. Seppala, L. Desborough, A review of performance monitoring and assessment techniques for univariate and multivariate control systems, Journal of Process Control 9 (1999) 1–17.

[9] N. Thornhill, M. Oettinger, P. Fedenczuk, Refinery-wide control loop performance assessment, Journal of Process Control 9 (1999) 109–124.

[10] Y. Shardt, Y. Zhao, F. Qi, K. Lee, X. Yu, B. Huang, S. Shah, Determining the state of a process control system: Current trends and future challenges, The Canadian Journal of Chemical Engineering 90 (2011) 217–245.

[11] A. Swanda, D. Seborg, Controller performance assessment based on setpoint response data, Proceedings of the 1999 American Control Conference 4 (1999) 2625–2639.

[12] K. Åström, Computer control of a paper machine - an application of linear stochastic control theory, IBM Journal 11 (1967) 389–405.

[13] K. Åström, Introduction to Stochastic Control Theory, Courier Dover Publications, 2012.

[14] L. Desborough, T. Harris, Performance assessment measures for univariate feedback control, The Canadian Journal of Chemical Engineering 70 (1992) 1186–1197.

[15] L. Desborough, T. Harris, Performance assessment measures for univariate feedforward/feedback control, The Canadian Journal of Chemical Engineering 71 (1993) 605–616.

[16] T. Harris, F. Boudreau, J. MacGregor, Performance assessment of multivariable feedback controllers, Automatica 32 (1996) 1505–1518.

[17] B. Huang, S. Ding, N. Thornhill, Practical solutions to multivariate feedback control performance assessment problem: Reduced a priori knowledge of interactor matrices, Journal of Process Control 15 (2005) 573–583.

[18] A. Ordys, D. Uduehi, M. Johnson, Process Control Performance Assessment: From Theory to Implementation, Springer, 2007.

[19] B. Huang, R. Kadali, Dynamic Modeling, Predictive Control and Performance Monitoring, Springer-Verlag, 2008.

[20] T. Harris, Interpretations of multivariate performance assessment indices, Journal of Process Control 19 (2009) 701–710.

[21] J. Yu, S. Qin, Mimo control performance monitoring using left/right diagonal interactors, Journal of Process Control 19 (2009) 1267–1276.

[22] B. Huang, Performance assessment of processes with abrupt changes of disturbances, The Canadian Journal of Chemical Engineering 77 (1999) 1044–1054.

[23] B. Huang, Minimum variance control and performance assessment of time-variant processes, Journal of Process Control 12 (2002) 707–719.

[24] F. Olaleye, B. Huang, E.Tamayo, Feedforward and feedback controller performance assessment of linear time-variant processes, Industrial & Engineering Chemistry Research 43 (2004) 589–596.

[25] F. Olaleye, B. Huang, E.Tamayo, Industrial applications of a feedback controller performance assessment of time-variant processes, Industrial & Engineering Chemistry Research 43 (2004) 597–607.

[26] F. Olaleye, B. Huang, E.Tamayo, Performance assessment of control loops with time-variant disturbance dynamics, Journal of Process Control 14 (2004) 867–877.

[27] F. Xu, B. Huang, Performance monitoring of SISO control loops subject to LTV disturbance dynamics: An improved LTI benchmark, Journal of Process Control 16 (2006) 567–579.

[28] F. Xu, B. Huang, E. Tamayo, Performance assessment of MIMO control systems with time-variant disturbance dynamics, Computers & Chemical Engineering 32 (2008) 2144–2154.

[29] T. Harris, W. Yu, Controller assessment for a class of non-linear systems, Journal of Process Control 17 (2007) 607–619.

[30] W. Yu, D. Wilson, B. Young, Control performance assessment for nonlinear systems, Journal of Process Control 20 (2010) 1235–1242.

[31] W. Yu, D. Wilson, B. Young, Nonlinear control performance assessment in the presence of valve stiction, Journal of Process Control 20 (2010) 754–761.

[32] W. Yu, D. Wilson, B. Young, A comparison of nonlinear control performance assessment techniques for nonlinear processes, The Canadian Journal of Chemical Engineering 90 (2012) 1442–1449.

[33] K. Hoo, M. Piovoso, P. Schnelle, D. Rowan, Process and controller performance monitoring: Overview with industrial applications, International Journal of Adaptive Control and Signal Processing 17 (2003) 635–662.

[34] J. Qin, Control performance monitoring - a review and assessment, Computers and Chemical Engineering 23 (1998) 173–186.

[35] J. Qin, Recent developments in multivariable controller performance monitoring, Journal of Process Control 17 (2007) 221–227.

[36] M. Jelali, Control Performance Management in Industrial Automation, Springer, 2013.

[37] M. Grimble, Controller performance benchmarking and tuning using generalised minimum variance control, Automatica 38 (2002) 2111–2119.

[38] J. Gao, R. Patwardhan, K. Akamatsu, , Y. Hashimoto, G. Emoto, S. Shah, B. Huang, Performance evaluation of two industrial MPC controllers, Control Eng. Pract. 11 (2003) 1371–1387.

[39] R. Julien, M. Foley, W. Cluett, Performance assessment using a model predictive control benchmark, Journal of Process Control 14 (2004) 441–456.

[40] S. L. Shah, R. Patwardhan, B. Huang, Multivariate controller performance analysis: Methods, applications and challenges, In Proceedings of the Chemical Process Control Conference (2001) 187–219.

[41] P. Eriksson, A. Isaksson, Some aspects of control loop performance monitoring, Proceedings of the Third IEEE Conference on Control Applications 2 (1994) 1029–1034.

[42] B. Ko, T. F. Edgar, Pid control performance assessment: The single-loop case, AIChE Journal 50 (2004) 1211–1218.

[43] A. Stack, F. D. III, The optimal control structure: An approach to measuring control-law nonlinearity, Computers & Chemical Engineering 21 (1997) 1009–1019.

[44] R. Haber, L. Keviczky, Nonlinear System Identification - Input-Output Modeling Approach, Springer, 1999.

[45] M. Guay, P. McLellan, D. Bacon, Measure of closedloop nonlinearity and interaction for nonlinear chemical processes, AIChE Journal 43 (1997) 2261–2278.

[46] M. Guay, R. Dier, J. Hahn, P. McLellan, Effect of process nonlinearity on linear quadratic regulator performance, Journal of Process Control 15 (2005) 113–124.

[47] J. Liu, Q. Meng, F. Fang, Minimum variance lower bound ratio based nonlinearity measure for closed loop systems, Journal of Process Control 23 (2013) 1097–1107.

[48] T. Hägglund, A control-loop performance monitor, Control Engineering Practice 3 (1995) 1543–1551.

[49] A. Horch, A simple method for detection of stiction in control valves, Control Engineering Practice 7 (1999) 1221–1231.

[50] N. Thornhill, T. Hägglund, Detection and diagnosis of oscillation in control loops, Control Engineering Practice 5 (1997) 1343–1354.

[51] N. Thornhill, S. Shah, B. Huang, Detection of distributed oscillations and root-cause diagnosis, In Proceedings of the CHEM-FAS (2001) 167–172.

[52] N. Thornhill, B. Huang, H. Zhang, Detection of multiple oscillations in control loops, Journal of Process Control 13 (2003) 91–100.

[53] T. Hägglund, Automatic detection of sluggish control loops, Control Engineering Practice 7 (1999) 1505–1511.

[54] T. Hägglund, Industrial implementation of on-line performance monitoring tools, Control Engineering Practice 13 (2005) 1383–1390.

[55] J. Gerry, Prioritizing and optimizing problem loops using a loop monitoring, In Proceedings of the ISA 422 (2002) 145–150.

[56] B. Huang, S. Shah, L. Badmus, A. Vishnubhotla, Control performance assessment: An enterprise asset management solution, Tech. rep., Matrikon (1999).

[57] I. Åkesson, Plant loop auditing in practice, Proceedings of the GMA-Kongress (2003) 927–934.

[58] A. Rakar, S. Zorzut, V. Jovan, Assessment of production performance by means of KPI, in: Proceedings of the Control, 2004, pp. 6–9.

[59] W. Shewhart, Economic Control of Quality of Manufactured Product, New York, 1931.

[60] J. MacGregor, Using on-line process data to improve quality: Challenges for statisticians, International Statistical Review 65 (1997) 309–323.

[61] J. Qin, Statistical process monitoring: Basics and beyond, Journal of Chemometrics 17 (2003) 480–502.

[62] S. Damarla, Multivariate statistical process monitoring and control, Ph.D. thesis, National Institute of Technology Rourkela (2011).

[63] J. MacGregor, T. Kourti, Statistical process control of multivariate processes, Control Engineering Practice 3 (1995) 403–414.

[64] J. Downs, J. Doss, Present status and future needs a view from north american industry, in: Proceedings of the Chemical Process Control IV Conference, 53, 1991.

[65] R. Woodward, P. Goldsmith, Cumulative Sum Techniques: Mathematical and Statistical Techniques for Industry, Monograph No. 3, Olive & Boyd, 1964.

[66] S. Roberts, Control chart tests based on geometric moving averages, Technometrics 1 (1959) 239–250.

[67] J. Hunter, The exponentially weighted moving averages, Journal of Quality Technology 18 (1986) 203–210.

[68] D. Montgomery, G. Runger, N. Hubele, Engineering Statistics, John Wiley & Sons, 2009.

[69] E. Martin, A. Morris, J. Zhang, Process performance monitoring using multivariate statistical process control, IEE Proceedings - Control Theory and Applications 143 (1996) 132–144.

[70] F. Alt, Economic design of control charts for correlated, multivariate observations, Ph.D. thesis, Georgia Institute of Technology (1977).

[71] C. Lowry, C. Douglas, A review of multivariate control charts, IIE transactions 27 (1995) 800–810.

[72] J. Jackson, A User's Guide to Principal Components, John Wiley & Sons, 1991.

[73] H. Hotelling, Relations between two sets of variates, Biometrika 28 (1936) 321–377.

[74] A. Izenman, Reduced-rank regression for the multivariate linear model, its relationship to certain classical multivariate techniques, and its application to the analysis of multivariate data, Ph.D. thesis, University of California, Berkeley (1972).

[75] S. Wold, Exponentially weighted moving principal components analysis and projections to latent structures, Chemometrics and Intelligent Laboratory Systems 23 (1994) 149–161.

[76] A. Raich, A. Cinar, Statistical process monitoring and disturbance diagnosis in multivariable continuous processes, AIChE Journal 42 (1996) 995–1009.

[77] L. Chiang, E. Russell, R. Braatz, Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis, Chemometrics and Intelligent Laboratory Systems 50 (2000) 243–252.

[78] R. Dunia, S. Qin, A unified geometric approach to process and sensor fault identification and reconstruction: The unidimensional fault case, Computers & Chemical Engineering 22 (1998) 927–943.

[79] R. Dunia, S. Qin, Subspace approach to multidimensional fault identification and reconstruction, AIChE Journal 44 (1998) 1813–1831.

[80] E. Martin, A. Morris, An overview of multivariate statistical process control in continuous and batch process performance monitoring, Transactions of the Institute of Measurement and Control 18 (1996) 51–60.

[81] P. Nomikos, J. MacGregor, Monitoring batch processes using multiway principal component analysis, AIChE Journal 40 (1994) 1361–1375.

[82] P. Nomikos, J. MacGregor, Multi-way partial least squares in monitoring batch processes, Chemometrics and Intelligent Laboratory Systems 30 (1995) 97–108.

[83] P. Nomikos, J. MacGregor, Multivariate spc charts for monitoring batch processes, Technometrics 37 (1995) 41–59.

[84] M. Kramer, Nonlinear principal component analysis using autoassociative neural networks, AIChE Journal 37 (1991) 233–243.

[85] W. DeVries, S. Wu, Evaluation of process control effectiveness and diagnosis of variation in paper basis weight via multivariate time-series analysis, IEEE Transactions on Automatic Control 23 (1978) 702–708.

[86] A. Horch, A. Isaksson, A modified index for control performance assessment, Journal of Process Control 9 (1999) 475–483.

[87] N. Stanfelj, T. Marlin, J. MacGregor, Monitoring and diagnosing process control performance: The single-loop case, Industrial & Engineering Chemistry Research 32 (1993) 301–314.

[88] B. Huang, S. Shah, E. Kwok, Good, bad or optimal? performance assessment of multivariable processes, Automatica 33 (1997) 1175–1183.

[89] B. Huang, S. Shah, Practical issues in multivariable feedback control performance assessment, Journal of Process Control 8 (1998) 421–430.

[90] T. Harris, C. Seppala, Recent developments in controller performance monitoring and assessment techniques, AICHE Symposium Series 326 (2002) 208–222.

[91] D. Kozub, Controller performance monitoring and diagnosis - industrial perspective, in: 15th Triennial World Congress, 2002.

[92] A. Horch, Condition monitoring of control loops, Ph.D. thesis, Royal Institute of Technology, Stockholm (2000).

[93] A. Ingimundarson, Dead-time compensation and performance monitoring in process control, Ph.D. thesis, Lund Institute of Technology, Lund (2003).

[94] A. Ingimundarson, T. Hägglund, Closed-loop performance monitoring using loop tuning, Journal of Process Control 15 (2005) 127–133.

[95] C. Lynch, G. Dumont, Control loop performance monitoring, IEEE Transactions on Control Systems Technology 4 (1996) 185–192.

[96] A. Stenman, F. Gustafsson, K. Forsman, A segmentation-based method for detection of stiction in control valves, International Journal of Adaptive Control and Signal Processing 17 (2003) 625–634.

[97] M. Jelali, Performance assessment of control systems in rolling mills - application to strip thickness and flatness control, Journal of Process Control 17 (2007) 805–816.

[98] J. VanAntwerp, Globally optimal robust control for large scale sheet and film processes, Ph.D. thesis, University of Illinois at Urbana-Champaign (1999).

[99] Novelis, Novelis rolling course notes (2008).

[100] T. Harris, Optimal controllers for nonsymmetric and nonquadratic loss functions, Technometrics 34 (1992) 298–306.

[101] D. Clarke, R. Hastings-James, Design of digital controllers for randomly disturbed systems, Proceedings of the Institution of Electrical Engineers 118 (1971) 1503–1506.

[102] J. MacGregor, P. Tidwell, Discrete stochastic control with input constraints, Electrical Engineers, Proceedings of the Institution of 124 (1977) 732–734.

[103] S. Skogestad, Simple analytic rules for model reduction and PID controller tuning, Journal of Process Control 13 (2003) 291–309.

[104] P. Moden, T. Soderstrom, On the achievable accuracy in stochastic control, Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on 17 (1978) 490–495.

[105] L. Telser, Discrete samples and moving sums in stationary stochastic process, Journal of American Statistical Association 62 (1967) 484–499.

[106] W. Wei, Effect of systematic sampling on arima models, Communications in Statistics-Theory and Methods 10 (1981) 2389–2398.

[107] B. Abraham, J. Ledolter, Forecast efficiency of sysematically sampled time series, Communications in Statistics - Theory and Methods 11 (1982) 2857–2868.

[108] C. Granger, P. Siklos, Systematic sampling, temporal aggregation, seasonal adjustment, and cointegration theory and evidence, Journal of Econometrics 66 (1995) 357–369.

[109] J. MacGregor, Optimal choice of the sample interval for discrete process control, Technometrics 18 (1976) 151–160.

[110] K. Brewer, Some consequences of temporal aggregation and systematic sampling for arima and armax models, Journal of Econometrics 1 (1971) 133–154.

[111] G. Wilson, The factorization of matricial spectral densities, SIAM Journal on Applied Mathematics 23 (1972) 420–426.

[112] G. Box, G. Jenkins, G. Reinsel, Time Series Analysis: Forecasting and Control, John Wiley & Sons, 2013.

[113] T. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, 2003.

[114] L. Ljung, System Identification: Theory for the User, Prentice Hall PTR, 1999.

[115] B. Huang, Y. Qi, A. Murshed, Dynamic Modelling and Predictive Control in Solid Oxide Fuel Cells: First Principle and Data-Based Approaches, John Wiley & Sons, 2013, Ch. First Principle Modelling for Chemical Processes.

[116] G. Seber, C. Wild, Nonlinear Regression, Wiley, New York, 2003.

[117] J. Nocedal, S. Wright, Numerical Optimization, Springer, 2006, Ch. Conjugate Gradient Methods.

[118] Y. Nesterov, A. Nemirovskii, Y. Ye, Interior-point Polynomial Algorithms in Convex Programming, Philadelphia: Society for Industrial and Applied Mathematics, 1994.

[119] L. Schumaker, Spline Functions: Basic Theory, Cambridge University Press, 1981.

[120] K. Keesman, System Identification: An Introduction, Springer, 2011.

[121] S. Duncan, W. Heath, A. Halouskova, M. K. R., Application of basis functions to the cross-directional control of web processes, UKACC International Conference 2 (1996) 1278–1283.

[122] W. Heath, Orthogonal functions for cross-directional control of web forming processes, Automatica 32 (1996) 183–198.

[123] K. Kristinsson, G. Dumont, Cross-directional control on paper machines using gram polynomials, Automatica 32 (1996) 533–548.

[124] A. Halouskova, M. Karny, I. Nagy, Adaptive cross-direction control of paper basis weight, Automatica 29 (1993) 425–429.

[125] A. Cinar, A. Palazoglu, F. Kayihan, Chemical Process Performance Evaluation, CRC Press, 2004.

[126] N. Draper, H. Smith, Applied Regression Analysis, John Wiley and Sons, 1981.

[127] M. Bartlett, An Introduction to Stochastic Processes, with Special Reference to Methods and Applications, CUP Archive, 1978.

[128] R. Blackman, J. Tukey, The measurement of power spectra from the point of view of communications engineering part i, Bell System Technical Journal 37 (1958) 185–282.

[129] N. Golyandina, A. Zhigljavsky, Singular Spectrum Analysis for Time Series, Springer, 2013.

[130] B. de Prony, Essai xperimental et analytique: Sur les lois de la dilatabilit de fluides lastique et sur celles de la force expansive de la vapeur de lalkool,a diffrentes tempratures., Journal De Lcole Polytechnique 1 (1795) 24–76.

[131] E. Pike, J. McWhirter, M. Bertero, C. Mol, Generalized information theory for inverse problems in signal-processing, Communications, Radar and Signal Processing, IEE Proceedings F 131 (1984) 660–667.

[132] D. Broomhead, G. King, Extracting qualitative dynamics from experimental data, Physica D 20 (1986) 217–236.

[133] D. Broomhead, G. King, On the qualitative analysis of experimental dynamical systems, Nonlinear Phenomena and Chaos 113 (1986) 114.

[134] K. Fraedrich, Estimating the dimensions of weather and climate attractors, Journal of the Atmospheric Sciences 43 (1986) 419–432.

[135] D. Danilov, A. Zhigljavsky, Principal Components of Time Series: The 'caterpillar Methods, University of St. Petersburg, 1997.

[136] R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Physica D: Nonlinear Phenomena 35 (1989) 395–424.

[137] M. Ghil, R. Vautard, lnterdecadal oscillations and the warming trend in global temperature time series, Nature 350 (1991) 324–327.

[138] R. Vautard, P. Yiou, M. Ghil, Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, Physica D: Nonlinear Phenomena 58 (1992) 95–126.

[139] M. Ghil, M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann, A. Robertson, A. Saunders, Y. Tian, F. Varadi, P. Yiou, Advanced spectral methods for climatic time series, Reviews of Geophysics 40 (2002) 1–41.

[140] J. Elsner, A. Tsonis, Singular Spectrum Analysis: A New Tool in Time Series Analysis, New York: Plenum Press, 2010.

[141] N. Golyandina, V. Nekrutkin, A. Zhigljavsky, Analysis of Time Series Structure: SSA and Related Techniques, Chapman & Hall / CRC, 2001.

[142] H. Hassani, Singular spectrum analysis: Methodology and comparison, Journal of Data Science 5 (2007) 239–257.

[143] H. Hassani, A. Zhigljavsky, Singular spectrum analysis: Methodology and application to economics data, Journal of Systems Science and Complexity 22 (2009) 372–394.

[144] H. Hassani, D. Thomakos, A review on singular spectrum analysis for economic and financial time series, Statistics and Its Interface 3 (2010) 377–397.

[145] A. Zhigljavsky, Singular spectrum analysis for time series: Introduction to this special issue, Statistics and Its Interface 3 (2010) 255–258.

[146] W. Hsieh, Nonlinear principal component analysis by neural networks, Tellus A 53 (2001) 599–615.

[147] W. Hsieh, A. Wu, Nonlinear multichannel singular spectrum analysis of the tropical pacific climate variability using a neural network approach, Journal of Geophysical Research: Oceans 107 (2002) 1–15.

[148] T. Harris, H. Yuan, Filtering and frequency interpretations of singular spectrum analysis, Physica D: Nonlinear Phenomena 239 (2010) 1958–1967.

[149] G. Castagnoli, C. Taricco, S. Alessio, Isotopic record in a marine shallow-water core: Imprint of solar centennial cycles in the past 2 millennia, Advances in Space Research 35 (2005) 504–508.

[150] G. Tzagkarakis, M. Papadopouli, T. Panagiotis, Singular spectrum analysis of traffic workload in a large-scale wireless LAN, Proceedings of the 10th ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (2007) 99–108.

[151] M. Papadopouli, G. Tzagkarakis, T. Panagiotis, Trend forecasting based on singular spectrum analysis of traffic workload in a large-scale wireless LAN, Performance Evaluation 66 (2009) 173–190.

[152] C. Penland, M. Ghil, K. Weickmann, Adaptive filtering and maximum entropy spectra with application to changes in atmospheric angular momentum, Journal of Geophysical Research Atmospheres 96 (1991) 22659–22671.

[153] SSA-MTM group, (mostly) UCLA, SSA-MTM toolkit for spectral analysis. URL `http://www.atmos.ucla.edu/tcd/ssa/`

[154] D. Schoellhamer, Singular spectrum analysis for time series with missing data, Geophysical Research Letters 28 (2001) 3187–3190.

[155] P. Yiou, D. Sornette, M. Ghil, Data-adaptive wavelets and multi-scale singular-spectrum analysis, Physica D 142 (2000) 254–290.

[156] A. Cantoni, P. Butler, Eigenvalues and eigenvectors of symmetric centrosymmetric matrices, Linear Algebra and Its Applications 13 (1976) 275–288.

[157] A. Cantoni, P. Butler, Properties of the eigenvectors of persymmetric matrices with applications to communication theory, IEEE Transactions on Communications COM-24 (1976) 804–809.

[158] J. Makhoul, On the eigenvectors of symmetric toeplitz matrices, Proceedings of the Acoustics, Speech, and Signal Processing 29 (1981) 868–872.

[159] R. Catell, Scree test for the number of factors, Multivariate Behavioral Research 1 (1966) 245–276.

[160] J. Hayton, D. Allen, V. Scarpello, Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis, Organizational Research 7 (2004) 191–205.

[161] L. Scharf, The SVD and reduced rank signal processing, Signal Processing 25 (1991) 113–133.

[162] G. Golub, C. VanLoan, Matrix Computations, John Hopkins University Press, 1996.

[163] P. Hansen, S. Jensen, FIR filter representations of reduced-rank noise reduction, IEEE Transaction on Signal Processing 46 (1998) 1737–1741.

[164] R. Kumaresan, D. Tufts, Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise, IEEE Transactions on Acoustics, Speech, and Signal Processing 30 (1982) 833–840.

[165] Y. Li, K. Liu, J. Razavilar, A parameter estimation scheme for damped sinusoidal signals based on low-rank hankel approximation, IEEE Transaction on Signal Processing 45 (1997) 481–486.

[166] J. Razavilar, Y. Li, K. Liu, A structured low-rank matrix pencil for spectral estimation and system identification, Signal Processing 65 (1998) 363–372.

[167] J. Razavilar, Y. Li, K. Liu, Spectral estimation based on structured low rank matrix pencil, Proceedings of the Acoustics, Speech, and Signal Processing 5 (1996) 2503–2506.

[168] M. Chu, R. Funderlic, R. Plemmons, Structured low rank approximation, Linear Algebra and Its Applications 366 (2003) 157–172.

[169] J. Cadzow, Signal enhancement - a composite property mapping algorithm, IEEE Transactions on Acoustics, Speech, and Signal Processing 36 (1988) 49–62.

[170] B. Moor, Total least squares for affinely structured matrices and the noisy realization problems, IEEE Transactions on Signal Processing 42 (1994) 3104–3113.

[171] I. Markovsky, J. Willems, S. Huffel, B. Moor, R. Pintelon, Application of structured total least squares for system identification and model reduction, IEEE Transactions on Automatic Control 50 (2005) 1490–1500.

[172] S. Reddi, Eigenvector properties of toeplitz matrices and their application to spectral anaysis of time series, Signal Processing 7 (1984) 45–56.

[173] I. Markovsky, S. Rao, Palindromic polynomials, time-reversible systems, and conserved quantities, 16th Mediterranean Conference on Control and Automation (2008) 125–130.

[174] E. Robinson, Statistical Detection and Estimation, Halfner, 1967.

[175] J. Kormylo, V. Jain, Two-pass recursive digital filter with zero phase shift, Proceedings of the Acoustics, Speech, and Signal Processing AS22 (1974) 384–387.

[176] C. Lindquist, Adaptive & Digital Signal Processing with Digital Filtering Applications, Stewart & Sons, 1989.

[177] I. Dologlou, G. Carayannis, Physical interpretation of signal reconstruction from reduced rank matrices, IEEE Transaction on Signal Processing 39 (1991) 1681–1682.

[178] D. Thomson, Spectrum estimation and harmonic analysis, Proceedings of the IEEE 70 (1982) 1055–1096.

[179] D. Percival, A. Walden, Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques, Cambridge University Press, 1993.

[180] P. Yiou, B. Baert, M. Loutre, Spectral analysis of climate data, Surveys in Geophysics 17 (1996) 619–663.

[181] M. Ghil, C. Taricco, Past and Present Variability of the Solar-terrestrial System: Measurement, Data Analysis and Theoretical Models, Societá Italiana Di Fisica, Bologna, & IOS Press, Amsterdam, 1997, Ch. Advanced Spectral Analysis Methods.

[182] G. Plaut, R. Vautard, Spells of low-frequency oscillations and weather regimes in the northern hemisphere, Journal of the Atmospheric Sciences 51 (1994) 210–236.

[183] L. Zotova, C. Shumb, Multichannel singular spectrum analysis of the gravity field data from GRACE satellites, AIP Conference Proceedings 1206 (2010) 473–479.

[184] S. Raynaud, P. Yiou, R. Kleeman, S. Speich, Using mssa to determine explicitly the oscillatory dynamics of weakly nonlinear climate systems, Nonlinear Processes in Geophysics 12 (2005) 807–815.

[185] D. Salgado, F. Alonso, Tool wear detection in turning operations using singular spectrum analysis, Journal Of Materials Processing Technology 171 (2006) 451–458.

[186] T. Alexandrov, A method of trend extraction using singular spectrum analysis, Statistical Journal 7 (2009) 1–22.

[187] I. Florinsky, R. Eilers, B. Wiebe, M. Fitzgerald, Dynamics of soil salinity in the canadian prairies: Application of singular spectrum analysis, Environmental Modelling & Software 24 (2009) 1182–1195.

[188] A. Groth, M. Ghil, S. Hallegatte, P. Dumas, Identification and reconstruction of oscillatory modes in us business cycles using multivariate singular spectrum analysis, Tech. rep., Environmental Research & Teaching Institute (2011).

[189] V. Oropeza, The singular spectrum analysis method and its application to seismic data denoising and reconstruction, Ph.D. thesis, University of Alberta (2010).

[190] V. Oropeza, M. Sacchi, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis, Geophysics 76 (2011) V25–V32.

[191] H. Hassani, S. Heravi, A. Zhigljavsky, Forecasting european industrial production with singular spectrum analysis, International Journal of Forecasting 25 (2009) 103–118.

[192] H. Hassani, A. Soofi, A. Zhigljavsky, Predicting daily exchange rate with singular spectrum analysis, Nonlinear Analysis: Real World Applications 11 (2010) 2023–2034.

[193] H. Hassani, S. Heravi, A. Zhigljavsky, Forecasting uk industrial production with multivariate singular spectrum analysis, Journal of Forecasting 32 (2013) 395–408.

[194] N. Golyandina, K. Usevich, I. Florinsky, Filtering of digital terrain models by two-dimensional singular spectrum analysis, International Journal of Ecology & Development 8 (2007) 81–94.

[195] D. Kondrashov, Y. Feliks, M. Ghil, Oscillatory modes of extended nile river records (AD 622-1922), Geophysical Research Letters 32 (2005) L10702 (1–4).

[196] D. Kondrashov, M. Ghil, Spatio-temporal filling of missing points in geophysical data sets, Nonlinear Processes in Geophysics 13 (2006) 151–159.

[197] D. Wang, S. Liang, Singular spectrum analysis for filling gaps and reducing uncertainties of MODIS land products, Geoscience and Remote Sensing Symposium 5 (2008) 558–561.

[198] D. Kondrashov, Y. Shprits, M. Ghil, Gap filling of solar wind data by singular spectrum analysis, Geophysical Research Letters 37 (2010) L15101 (1–6).

[199] V. Moskvina, Application of the singular spectrum analysis for change-point detection in time series, Ph.D. thesis, Cardiff University (2001).

[200] N. Golyandina, K. Usevich, Matrix Methods: Theory, Algorithms and Applications, World Scientific Publishing, 2010, Ch. 2D-extension Of Singular Spectrum Analysis: Algorithm And Elements Of Theory.

[201] N. Golyandina, E. Osipov, The "caterpillar-ssa method for analysis of time series with missing values, Journal of Statistical Planning and Inference 137 (2007) 2642–2653.

[202] F. Ghaderi, H. Mohseni, S. Sanei, Localizing heart sounds in respiratory signals using singular spectrum analysis, IEEE Transactions on Biomedical Engineering 58 (2011) 3360–3367.

[203] G. Tzagkarakis, M. Papadopouli, P. Tsakalides, Trend forecasting based on singular spectrum analysis of traffic workload in a large-scale wireless LAN, Performance Evaluation 66 (2009) 173–190.

[204] L. Rodrïguez-Aragön, A. Zhigljavsky, Singular spectrum analysis for image processing, Statistics and Its Interface 3 (2010) 419–426.

[205] W. Ku, R. Storer, C. Georgakis, Disturbance detection and isolation by dynamic principal component analysis, Chemometrics and Intelligent Laboratory Systems 30 (1995) 179–196.

[206] W. Larimore, System identification, reduced-order filtering and modeling via canonical variate analysis, IEEE in American Control Conference (1983) 445–451.

[207] N. Thornhill, S. Shah, B. Huang, A. Vishnubhotla, Spectral principal component analysis of dynamic process data, Control Engineering Practice 10 (2002) 833–846.

[208] M. Ammar, Cross-directional processes: Modeling, identification and robust control, Ph.D. thesis, The Univeristy of British Columbia (2009).

[209] H. Ltkepohl, New Introduction to Multiple Time Series Analysis, Springer, 2007.

[210] C. Seppala, T. Harris, D. Bacon, Time series methods for dynamic analysis of multiple controlled variables, Journal of Process Control 12 (2002) 257–276.

[211] P. Wellstead, M. Zarrop, S. Duncan, Signal processing and control paradigms for industrial web and sheet manufacturing, International Journal of Adaptive Control and Signal Processing 14 (2000) 51–76.

[212] J. Ramarathnam, Control and monitoring of sheet and film forming processes, Ph.D. thesis, University of Alberta (2009).

[213] G. Stewart, Two dimensional loop shaping controller design for paper machine cross-directional processes, Ph.D. thesis, The Univeristy of British Columbia (2000).

[214] G. Reinsel, R. Velu, Multivariate Reduced-rank Regression: Theory and Applications, Springer, 1998.

[215] J. VanAntwerp, A. Featherstone, R. Braatz, B. Ogunnaike, Cross-directional control of sheet and film processes, Automatica 43 (2007) 191–211.

[216] M. Priestley, Spectral Analysis and Time Series, London: Academic Press, 1981.

[217] R. Garello, Two-dimensional Signal Analysis, John Wiley & Sons, 2010.

[218] L. Bergh, J. MacGregor, Spatial control of sheet and film forming processes, The Canadian Journal of Chemical Engineering 65 (1987) 148–155.

[219] V. Hölttä, H. Koivo, Performance metrics for web-forming processes, Journal of Process Control 21 (2011) 885–892.

Appendix A

MATLAB^{TM} programming code of examples in Section 7.3.1

A.1 Example 1 - Uniform Ridge on the Work Roll

```
clear all
close all
clc

N1 = 1000;
N2 = 50;
T = 5;

Y = 2*ones(N1,N2);

for i = 1:N1/T+1
```

```
if (i-1)*T+1 > N1
break;
else
Y((i-1)*T+1,:) = 1.5;
end
end
```

A.2 Example 2 - Non-Uniform Ridge on the Work Roll

```
clear all
close all
clc

N1 = 1000;
N2 = 50;
T = 100;

A = 2*ones(N2,N2);

for i = 1:N2
for j = 1:N2
if i == j
A(i,j) = 1.5;
```

```
else
end
end
end

Y = 2*ones(N1,N2);

for i = 1:round(N1/T)
Y((i-1)*T+1:(i-1)*T+N2,1:N2) = A;
end
```