The Dissertation Committee for Jie Yu
certifies that this is the approved version of the following dissertation:

Data-Driven Approach for Control Performance Monitoring and Fault Diagnosis

Committee:

S. Joe Qin, Supervisor

Thomas F. Edgar

Charles B. Mullins

Glenn Y. Masada

Robert H. Flake

Maruthi R. Akella

Data-Driven Approach for Control Performance Monitoring and Fault Diagnosis

by

Jie Yu, B.S., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2007

Dedicated to Yanling and my family.

Acknowledgments

I would like to acknowledge many people for helping me during my doctoral work. I would especially like to thank my supervisor, Dr. S. Joe Qin, for his generous time and commitment. Throughout my doctoral work, he encouraged me to develop innovative thinking and technical skills. Meanwhile he continually stimulated my research potential and greatly assisted me with academic writing. His enthusiastic guidance and support enable the completion of my dissertation at this point. I am also very grateful for having an exceptional doctoral committee and wish to thank Dr. Thomas F. Edgar, Dr. Charles B. Mullins, Dr. Glenn Y. Masada, Dr. Robert H. Flake and Dr. Maruthi R. Akella for their continual support and encouragement over the past four years.

I owe a special note of gratitude to Mr. Gary Gover and the management team at Weyerhaeuser, who offered me valuable opportunities to explore the challenging industrial world. My thanks also go out to Dr. Chris McNabb, Mr. Dell Cooper, Dr. John Watkins, Dr. Anthony Swanda and Dr. Apostolos Rigopolous for teaching me how to bridge the gap between academia and industry. I would say that the internship at Weyerhaeuser is the most precious treasure in my career path.

I extend my thanks to the members of Qin and Edgar Research Groups

who gave me helpful suggestions and warm encouragement along the journey. I am fortunate to have spent such an enjoyable time with them. Wish them all the best in completing their graduate work and achieving greater success in their professional career.

Finally, I want to express my deepest appreciation to my parents on the other side of the Pacific for their understanding and encouragement in my oversea study. My wife Yanling Wu deserves special thanks for letting me spend her time. She is always willing to share my pressure and joys. Without her love and support, I would have never finished my PhD degree. I look forward the opportunity to repay the debt.

JIE YU

The University of Texas at Austin

March 2007

Data-Driven Approach for Control Performance Monitoring and Fault Diagnosis

Publication No. _____

Jie Yu, Ph.D.
The University of Texas at Austin, 2007

Supervisor: S. Joe Qin

Due to the industrial value, control performance and process monitoring have attracted increasing attention in recent years. However, there still exist challenges that restrict the industrial applications of monitoring technology. This dissertation presents some innovative solutions to the monitoring issues.

To avoid the interactor requirement of minimum variance control (MVC) benchmark, a data-driven covariance monitoring framework is established. Relative to a user-defined benchmark, generalized eigenvalue analysis is employed to extract the directions with the worst performance. A statistical inference strategy is then developed to identify the worse or better performance directions and subspace. The covariance based indices are further derived to assess the performance degradation or improvement. To diagnose the controlled variables causing the performance change, two types of multivariate contribution methods are proposed. One is to evaluate the significance of

the eigenvector loadings while the other to examine the angle between each variable and the worse/better subspace.

Complementary to the data-driven performance monitoring scheme, a simplified solution to MVC benchmark is also developed. A right diagonal interactor is first factorized from process time delays and the corresponding MVC benchmark is derived with numerical simplicity. For more general MIMO processes, left and right diagonal interactors are integrated to characterize the more complex delay structure. The MVC estimation using the left/right diagonal interactors are presented. To further improve multivariable control performance, an iterative strategy of output weighting selection is proposed. Eigenvalue decomposition is implemented on the output covariance to find the directions with the largest variance inflation. A nondiagonal weighting matrix is then designed with respect to the eigendirections and more importance proportional to the corresponding eigenvalues is assigned.

In addition to control performance monitoring, process monitoring is also investigated with focus on fault detection and diagnosis of multistage overlay lithography processes. In our work, a multistage state-space model for the misalignment errors is developed from the physical principles and then formulated into the general mixed linear model. Subsequently, variance component analysis is employed to estimate the mean and variance components of the potential fault sources. A hypothesis testing procedure is adopted to detect the active faults in different layers while the mean/variance estimates are used to diagnose their magnitude and orientation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Dissertation Outline

There are hundreds to thousands of control loops operating under varying conditions in a typical industrial plant. The large-scale control systems in industrial facilities may include various controllers, from the univariate PID control loops to multivariable model predictive controllers (MPC) (Harris *et al.*, 1999). Eastman Chemical Company, for example, has reported a large-scale web based system to monitor the control performance of $14,000$ loops in multiple plants (Paulonis and Cox, 2003). In HVAC systems, Johnson Control has implemented over half a millions control monitors based on a pattern recognition technique (Seem, 1998; Seem, 2006). For such complicated control systems, the regular maintenance service is a challenging task for control engineers. Even if those controllers work properly in the initial period, a surprisingly high percentage of controllers will encounter different kinds of performance deterioration after a period of operation time (Huang and Shah, 1999). It has been reported that as many as 60% of industrial controllers have performance problems (Bialkowski, 1993; Ender, 1993; Rinehart and Jury, 1997; Ender, 1999; Harris and Seppala, 2002*a*). Poor control performance of industrial processes may be caused by various factors such as poor controller tuning, unappropriate control structure design, process/disturbance

dynamics changes, instrument (e.g. sensor, actuators) failure and equipment malfunction. Therefore, it is necessary yet challenging to develop automated and effective techniques for control performance assessment and monitoring (CPA/CPM) (Kozub, 1996; Hoo *et al.*, 2003). The major steps of CPA/CPM framework usually involve: (i) Design of a benchmark against which the current control performance is evaluated; (ii) Detection of performance degradation relative to the benchmark; (iii) Diagnosis of underlying loops or controlled variables leading to the performance deterioration; (iv) Control performance improvement through control retuning, redesign or process maintenance (Harris and Seppala, 2002*b*; Jelali, 2006). Generally, CPM methods should not disturb the normal operation of industrial processes and can be implemented under closed-loop conditions (Huang, 1997).

The research on control performance monitoring has been very active and prosperous over the past decade. It is estimated that several hundreds of papers have been published in this area. A few review articles or book chapters have summarized the latest developments of CPM theory and applications (Qin, 1998; Harris *et al.*, 1999; Huang and Shah, 1999; Shah *et al.*, 2002; Jelali, 2006; Qin and Yu, 2007). The earliest research work in the field of control performance monitoring can be attributed to Åström (Åström, 1970), who proposed to use the minimum variance control benchmark as a standard for evaluating and assessing the current control loop performance. The MVC theory was originally established by Åström (Åström, 1970) and Box and Jenkins (Box and Jenkins, 1970). Åström in his CPC-2 paper (Åström, 1976) noted

the following:

> In the special case of minimum variance control ... it is known that the covariance function will vanish for lags greater than the sum of the sampling interval and transport delay of the system. It is then sufficient to record output only and to compute its covariance function.

Process time delay serves as the most fundamental limitation on the achievable control performance (Huang and Shah, 1998). The MVC benchmark can provide a theoretical optimal performance bound under the time-delay restriction only. Greater research interest in MVC based performance monitoring was aroused by Harris's remarkable contribution to this area (Harris, 1989). He demonstrated the applicability of minimum variance index to assess the control loop performance of SISO process with routine operating data. Later Desborough and Harris extended the univariate MVC benchmark to feedback control and feedforward/feedback control schemes (Desborough and Harris, 1992; Desborough and Harris, 1993). A fast online algorithm for estimating the univariate MVC benchmark was also proposed in their work. Stanfelj et al. applied Harris's index to monitor and diagnose the performance of SISO control loops. Depending on simple but rigorous statistical analysis, this method is able to identify the causes of poor performance, which comes from feedback or feedforward loop and is due to plant/model mismatch or improper tuning (Stanfelj *et al.*, 1993). Following the similar principles as MVC benchmark, Kozub and

Garcia defined another so-called closed-loop potential (CLP) index for performance monitoring (Kozub and Garcia, 1993). Tyler and Morari attempted to detect control performance deterioration by using a likelihood based hypothesis testing method, which they also compared with the minimum variance estimation approach (Tyler and Morari, 1996). Lynch and Dumont presented three different estimators for the achievable minimum variance, the process time delay and the static input-output relationship of SISO process, respectively (Lynch and Dumont, 1996).

More recently, increasing research attention in CPM area was turned to the more complicated multivariable control systems (Qin and Yu, 2007). For MIMO processes, the time delay structure is far beyond a simple scalar in the univariate systems. It can be factorized as an interactor matrix, which is even more than the pairwised input-output time delay matrix. The difficulty in deriving the MVC benchmark of multivariable systems lies in the estimation of the interactor matrix (Huang *et al.*, 2005; Xia *et al.*, 2006). The concept of the interactor matrix was originally proposed by Wolovich and Falb (Wolovich and Falb, 1976), who have shown the analogy between the SISO time delay and the MIMO interactor matrix. The common feature therein is the feedback control invariance which can lead to the MVC benchmark. Generally, the interactor matrix of MIMO processes is not unique and could have different forms. Rogozinski et al. developed an algorithm for the factorization of the nilpotent interactor matrix (Rogozinski *et al.*, 1987). Peng and Kinnaert proved the existence of the unitary interactor matrix through the

solution to the singular linear quadratic regulation problem (Peng and Kinnaert, 1992). Unlike triangular interactor matrix, the unitary interactor matrix has the desired property of output-ordering invariance. That is, the MVC law does not change as the order of output variables is rearranged (Dugard *et al.*, 1984; Goodwin and Sin, 1984). An estimation algorithm of the unitary interactor matrix using closed-loop routine data was established by Huang et al. (Huang *et al.*, 1997*a*). Harris et al. investigated the procedures for assessing MIMO linear feedback control system based on MVC benchmark (Harris *et al.*, 1996). More specifically, the lower triangular interactor matrix and spectral factorization were adopted to acquire the MVC benchmark. Ko and Edgar presented the MVC based performance monitoring of feedback control systems (Ko and Edgar, 2001*b*), cascade control loops (Ko and Edgar, 2000) and MPC systems (Ko and Edgar, 2001*a*), respectively. For cascade control loops, they employed multivariate time series models to derive the primary and secondary MVC benchmarks sequentially. In MPC performance monitoring, a moving horizon approach was used to obtain a constrained MVC benchmark, which was proven to converge to the unconstrained MVC bound when constraints become inactive. McNabb and Qin (McNabb and Qin, 2003) proposed a subspace projection based MIMO control performance monitoring method. A multivariate time delay (MTD) matrix, which is equivalent to the interactor matrix in state-space framework, was derived and used to estimate the MVC benchmark. Then the output covariance was used to monitor the control performance and diagnose the performance suboptimality. Later

they extended the projection based performance monitoring scheme to include measured disturbances and setpoint changes. An approach was provided to separate the suboptimal performance caused by measured or unmeasured disturbances (McNabb and Qin, 2005*b*).

As the most popular benchmark, MVC benchmark has gained great attention in control performance monitoring community. MVC benchmark is a theoretical lower bound of the output variance and thus represents the optimal control performance under time delay constraint. In real processes, however, other types of restrictions on the achievable control performance may exist so that the MVC benchmark becomes irrelevant (Huang, 1997). Furthermore, its lack of robustness to model uncertainty and use of excessive input actions constitute other concerns for the applications of MVC benchmark (Åström and Wittenmark, 1997). Even for the controllers where MVC benchmark is adopted, it needs to be stressed that MV performance is probably unachievable. Therefore, some alternative benchmarks are also proposed to serve as appropriate measures to evaluate the control performance. Huang and Shah (Huang and Shah, 1998) attempted to modify the MVC benchmark by specifying the noise decay rate after the interactor order to improve the robustness. Horch and Isaksson proposed a modification of MVC performance index through a free closed-loop pole. Such an index is less sensitive to uncertain or time-varying time-delays, which is more advantageous for industrial applications (Horch and Isaksson, 1999). Huang and Shah designed a new linear quadratic Gaussian (LQG) benchmark to reduce the output variance without

affecting the input variance (Huang and Shah, 1999). Since both CV and MV are taken into account, more process information is needed, i.e., the full plant and disturbance models. In return, it provides a trade-off curve to balance the minimum input and output variances rather than the minimum output variance only for MVC benchmark. The LQG problem can be solved by using an infinite general predictive control (GPC) approach. Grimble introduced a simpler generalized minimum variance (GMV) optimal control law for performance monitoring. The objective is to minimize the variance of a fictitious output that includes the dynamically penalized error and control signals. Since the GMV control law has the same type of characteristics as the MVC law, the same procedures can thus be applied to estimate the GMV performance index from operating data (Grimble, 2002).

In addition to control performance monitoring, process monitoring is another critical aspect to ensure the normal operation and economic yields of industrial plants. The major difference between control performance monitoring and process monitoring is that the former is concerned with the control system operation while the latter is to deal with instruments (e.g. sensors), process components and operations. The primary task of process monitoring is to find out the abnormal process operations as well as the root causes leading to the process faults. A typical process monitoring system involves fault detection, fault diagnosis, fault estimation and fault reconstruction (Qin, 2003). Parallel to the research progress of control performance monitoring, process monitoring has also experienced great development over the last two decades.

Thus far, multivariate statistical methods such as principal component analysis (PCA) (Tong and Crowe, 1995; Kourti and MacGregor, 1996; Dunia *et al.*, 1996), partial least squares (PLS) (MacGregor *et al.*, 1994), Fisher discriminant analysis (FDA) (Chiang *et al.*, 2000; He *et al.*, 2005) and independent component analysis (Kano *et al.*, 2003; Lee *et al.*, 2006*a*) has been widely applied to process monitoring.

Despite the fruitful research activities in process and control performance monitoring, there still exist many limitations that need to be addressed to enrich the theoretical framework and facilitate industrial applications. The issues are summarized as follows:

1. The popular MVC benchmark is based solely on process time delay restrictions. However, there could exist other restrictions such as input constraints that may have significant impact on the achievable control performance.

2. The LQ objective function for MVC benchmark includes process outputs only and does not take into account the control effort explicitly. The minimum variance controller, although achievable under ideal situations, demands excessive control action inherently.

3. Poor robustness of MVC benchmark due to its sensitivity to model uncertainty may result in suboptimal performance bound under plant/model mismatch.

4. Interactor matrix is necessary to derive the MIMO MVC benchmark and its accurate estimation depends on a priori process knowledge such as the plant model or at least the first few Markov parameters, which are not desirable in industrial applications. Furthermore, the computation procedures of interactor matrix and MVC benchmark are really tedious.

5. Other alternative benchmarks like LQG and GMV benchmark can address some of the above issues for MVC benchmark. Nevertheless, the drawbacks of those benchmarks are equally challenging. LQG benchmark, for example, even requires both plant and disturbance models, while GMV benchmark needs appropriate dynamic input/output weightings.

6. A suitable weighting matrix is often desirable in the underlying LQ contol design of MVC benchmark. But unfortunately, there is no systematic method to optimize the selection of weighting matrix.

7. Current research effort of control performance monitoring is more focused on performance assessment aspect. In other words, it is to determine if the control system is working optimally, well or poorly. However, little has been reported with performance diagnosis emphasis, i.e., finding out the under-performing control loops or controlled variables from a multivariate perspective.

8. In process monitoring area, unlike the typical continuous or batch processes, the multistage processes such as overlay lithography process in

semiconductor manufacturing are difficult to deal with because of their complicated error propagation, which disables the traditional monitoring techniques.

This dissertation is concerned with providing solutions to some of these challenging issues from different perspectives. In one aspect, we establish a data-driven MIMO control performance monitoring and diagnosis framework. A user-defined historical benchmark with desirable performance is first proposed and integrated with the covariance monitoring scheme. Based on the data-driven covariance benchmark, multivariate statistical analysis used in process monitoring area can be further extended to control performance assessment and diagnosis. Not only the current performance may be assessed relative to the historical benchmark, the loops or variables responsible for performance degradation or improvement can also be diagnosed. Such data-driven performance monitoring approach does not require a priori process knowledge such as time delay information and is thus more convenient to implement in industry. Other underlying performance restrictions besides the process time delays are also taken into account implicitly. Hence it is more robust and realistic than MVC based performance monitoring method.

As a complementary part of the data-driven technique, on the other hand, the value of MVC benchmark should not be disregarded because it does point out the absolute lower bound of achievable control performance that cannot be obtained from historical benchmarks. In our study, a simplified estimation method of MVC benchmark is developed to alleviate the re-

quirements of process knowledge and avoid the computational intensity. The combined left/right interactors in the special forms of diagonal matrices are constructed to characterize the multivariate time delay structures and then the MVC benchmarks can be derived easily.

In addition to the above two major directions, the output weighting selection in LQ design for performance improvement and the multistage overlay lithography process monitoring are also investigated. As a summary, the specific research accomplishments are listed below:

1. Propose a covariance based data-driven benchmark for MIMO performance monitoring;

2. Apply generalized eigenvalue analysis and develop statistical inference strategy to identify the worse or better performance directions and subspaces;

3. Derive covariance based indices to evaluate the overall performance and quantify the performance degradation/improvement within the isolated worse/better subspaces;

4. Establish multivariate contribution methods to diagnose the degraded/improved loops or controlled variables leading to performance change;

5. Define and construct the combined left and right diagonal interactors to characterize the multivariate time delay structures of MIMO processes;

6. Develop the simplified solution to MVC benchmark under the combined left/right diagonal interactors;

7. Design a non-diagonal weighting matrix based on eigenvalue decomposition of output covariance to improve control performance of MIMO interacting systems;

8. Apply variance component analysis (VCA) method to fault detection and diagnosis of multistage overlay lithography processes;

The dissertation is organized as follows

In Chapter 2, a data-based covariance benchmark is proposed for control performance monitoring. Within the covariance monitoring scheme, generalized eigenvalue analysis is employed to extract the directions with the degraded or improved control performance against the benchmark. It is shown that the generalized eigenvalues and the covariance-based performance index are invariant to variance scaling of the data. A statistical inference method is further developed for the generalized eigenvalues and the corresponding confidence intervals are derived from asymptotic statistics. This procedure can be used to determine the directions or subspaces with significantly worse or better performance versus the benchmark. The covariance based performance indices within the isolated worse and better performance subspaces are then derived to assess the performance degradation and improvement.

The data-driven control performance monitoring framework is further extended to the performance diagnosis aspect with focus on variable identi-

fication in Chapter 3. To identify the control loops or controlled variables responsible for performance degradation or improvement, two types of multivariate contribution methods are proposed. One of the diagnostic methods is a loading based contribution chart to evaluate the significance of contribution of each loop/variable to the worse or better performance subspace. The bootstrap resampling procedure is conducted to estimate the probability distribution and statistics of the relevant eigenvector loadings. Then the confidence intervals are derived for the loadings. The other approach is to examine the angle between each individual loop/variable and the worse/better performance subspace. The cosine of the angle is defined as the contribution index and shown to be the canonical correlation coefficient between a unit vector and the worse/better subspace. The asymptotic statistics of canonical correlations is then utilized to derive the confidence limits for the angle based contributions. Two simulated examples (a multiloop control and a multivariable MPC system) and an industrial example from a power boiler unit are provided to demonstrate the effectiveness of the data-driven performance monitoring and diagnosis approaches in Chapters 2 and 3.

In Chapter 4, a simplified solution to MIMO MVC benchmark is developed. The MVC benchmark using a right diagonal interactor matrix is first derived for a class of well designed MIMO processes. Further, both the left and right diagonal interactors are integrated to characterize the more complex time-delay structure. The factorization of the combined left/right diagonal interactors and the corresponding MVC benchmark estimation are then pre-

sented. The properties of the combined diagonal interactors, i.e. the existence and uniqueness, are also investigated. The advantages of the new approach lie in the reduced a priori process knowledge and the simplified numerical procedures. In addition to MVC based control performance assessment, an adaptive strategy to select the weighting matrix in the LQ objective function is developed for performance improvement. The eigenvalue decomposition (EVD) on the output covariance is implemented to find out the directions with the largest variance inflation. Based on this information, the weighting matrix is constructed so that proportionally more importance is assigned to those directions with worse performance. This strategy enables the efficient performance boost in terms of covariance reduction for MIMO interacting systems. Simulated examples are used to illustrate the utility of the proposed performance monitoring and improvement approaches.

The problem of multi-stage overlay lithography process monitoring is explored in Chapter 5. First, a multi-stage state-space model for the misalignment errors of lithography processes is developed from the physical principles and the general mixed linear input-output model is then formulated to incorporate both deterministic and stochastic effects. Further, the minimum norm quadratic unbiased estimation (MINQUE) algorithm is adopted to estimate the mean and variance components of potential fault sources, and their asymptotic distributions are used to test the hypothesis concerning the statistical significance of each potential fault. Based on the above procedures, the root causes of misalignment errors in multi-layer overlay processes can be

detected and diagnosed with physical inference. A number of simulated examples are designed and tested to verify the validity of the presented approach in fault detection and diagnosis of multi-layer overlay processes.

Chapter 6 concludes the major contributions of this work and recommends some directions for future research.

Chapter 2

Statistical MIMO Control Performance Monitoring - Data-Driven Covariance Benchmark

2.1 Introduction

Despite the great success achieved in univariate control performance monitoring, the research of MIMO performance monitoring remains a challenge and has become a major focus in CPM/CPA area. The celebrated MVC bechmark suffers from the complicated multivariate time delay structure in terms of interactor matrix that constrains its applications to MIMO control performance monitoring. McNabb and Qin (McNabb and Qin, 2003) proposed a covariance monitoring scheme to account for MIMO variable correlation. The equivalent form of the interactor matrix, i.e. the MTD matrix, was defined to get minimum variance output, which also leads to minimum covariance for diagonal interactors. For general interactors, however, the covariance matrix based on minimum variance control is not guaranteed minimum in all directions. Given the higher demand of an interactor matrix for MVC benchmark, it appears to be more attractive by using data-driven benchmarks for control performance monitoring. Although Desborough and Harris (Desborough and Harris, 1992) studied data-based performance monitoring for SISO systems,

we are not aware of any work that extends it to the MIMO case.

In this study, a data-driven benchmark for control performance monitoring is proposed. This new benchmark is based on a historical period of operation data chosen by the user. The user-defined benchmark is usually a "golden" operation period achieving satisfactory control performance. Relative to the benchmark data, real-time control performance is monitored and the worse or better performance directions in the monitored period are identified using generalized eigenvalue analysis. To statistically determine the significance of performance degradation or improvement along every eigenvector direction, a statistical inference-based approach is developed. The confidence interval for each generalized eigenvalue is derived and calculated from the asymptotic statistics of the eigenvalue. The worse and the better eigendirections can thus be used to diagnose and further improve the control performance.

This chapter is organized as follows. The framework of data-driven covariance monitoring and generalized eigenvalue analysis is described in Section 2.2. A user-specified benchmark based on a period of historical data is proposed for assessing the control performance. A covariance-based index is also defined for assessing the overall control performance relative to the benchmark. Then a statistical inference strategy on the generalized eigenvalues is developed in Section 2.3. The numerical procedures for estimating and evaluating the confidence intervals of eigenvalues are derived from asymptotic distributions. The significant eigendirections with worse or better control per-

formance can thus be identified statistically. Further, the performance indices for the isolated worse/better subspaces are discussed. The proposed method is demonstrated in Section 2.4 using simulated examples and real data from an industrial power boiler system. Section 2.5 concludes the chapter.

2.2 Data-driven performance assessment benchmark

2.2.1 Covariance monitoring and generalized eigenvalue analysis

In the literature of control performance monitoring (Huang *et al.*, 1997*b*), the process output variance is an important parameter and the associated performance index may be defined as the ratio of minimum variance to actual variance. The sum of diagonal entries in the covariance matrix, i.e. the trace, serves as the performance index

$$\eta = \frac{E(y_t^T y_t)_{min}}{E(y_t^T y_t)} = \frac{tr\left\{cov(y_{mv})\right\}}{tr\left\{cov(y_t)\right\}} \tag{2.1}$$

where $cov(y_t)$ and $cov(y_{mv})$ represent the actual covariance and the covariance achieved under the MVC law, respectively. The value of variance index η is always between 0 and 1, where the upper bound 1 corresponds to the minimum variance output. In the above equation, however, only the diagonal elements of covariance matrix are taken into account and information from the off-diagonal entries is completely ignored. That is, the correlation between different variables is omitted. The calculation of minimum variance output, on the other hand, requires a priori knowledge of the plant and even the model of the process, which is not attractive in practice.

In this chapter, we propose to use a period of reference data as a user-specified benchmark. Generally, the reference data could be a period of "golden" operation data from the process with satisfactory control performance. For example, the benchmark data could be a period of operation data after a satisfactory controller tuning or updating. Let the benchmark period be I and the monitored period II, then the direction along which the largest variance ratio of the monitored period versus the benchmark period is attained should be

$$p = arg\ max \frac{p^T cov(y_{II})p}{p^T cov(y_I)p} \tag{2.2}$$

where $cov(y_I)$ and $cov(y_{II})$ denote the covariance matrices of the benchmark period I and the monitored period II, respectively. Define

$$\lambda(p) = \frac{p^T cov(y_{II})p}{p^T cov(y_I)p} \tag{2.3}$$

then we have

$$\frac{\partial \lambda(p)}{\partial p} = \frac{2cov(y_{II})p\big(p^T cov(y_I)p\big) - 2\big(p^T cov(y_{II})p\big)cov(y_I)p}{\big(p^T cov(y_I)p\big)^2} \tag{2.4}$$

Substituting Eqs. (2.3) into (2.4) yields

$$\frac{\partial \lambda(p)}{\partial p} = \frac{2cov(y_{II})p - 2\lambda(p)cov(y_I)p}{p^T cov(y_I)p} \tag{2.5}$$

Since the extremum point satisfies

$$\frac{\partial \lambda(p)}{\partial p} = 0 \tag{2.6}$$

the solution to Eq. (2.2) is equivalent to the following generalized eigenvalue analysis

$$cov(y_{II})p = \lambda cov(y_I)p \tag{2.7}$$

where λ is the generalized eigenvalue and p is the corresponding eigenvector. The eigenvector corresponding to the largest generalized eigenvalue λ_{max} represents the direction of the largest variance inflation in the monitored period against the benchmark period. This direction is referred to as worst performance direction (WPD).

In addition to the first generalized eigenvector, other subsequent eigenvectors with large enough eigenvalues (especially those much larger than one) are of remarkable suboptimality in control performance versus the benchmark. These large eigendirections actually constitute the subspace with worse performance of monitored period relative to the benchmark. The eigenvector corresponding to the smallest eigenvalue stands for the direction with the smallest variance ratio between the monitored period and the benchmark period. The eigendirections with eigenvalues significantly less than one span the subspace of improved performance over the benchmark. Since the data-driven benchmark is not a theoretical minimum, it is typical to have both degraded and improved performance directions.

It needs to be pointed out that the generalized eigenvectors are not necessarily orthogonal to one another. Suppose the covariance matrices $cov(y_I)$ and $cov(y_{II})$ are both full rank, which is true in most cases, then generalized eigenvalue analysis can be simplified to the regular eigenvalue problem as following

$$\Big\{cov^{-1}(y_I) \cdot cov(y_{II})\Big\}p = \lambda p \tag{2.8}$$

where the matrix $cov^{-1}(y_I) \cdot cov(y_{II})$ is not necessarily a symmetric matrix

although $cov(y_I)$ and $cov(y_{II})$ are both symmetric. Subsequently, the orthogonality among the eigenvectors p cannot be guaranteed. They are, however, linearly independent because the combined matrix $cov^{-1}(y_I) \cdot cov(y_{II})$ is full rank. To further explore the relative orthogonality, let the eigenvalue decomposition of $cov(y_{II})$ be

$$cov(y_{II}) = D_{II}\Lambda_{II}D_{II}^T \tag{2.9}$$

where D_{II} is orthogonal satisfying $D_{II}D_{II}^T = I$ and Λ_{II} is diagonal. The generalized eigenvalue decomposition is then equivalent to

$$cov^{-1}(y_I)D_{II}\Lambda_{II}D_{II}^T p = \lambda p \tag{2.10}$$

or

$$\left\{\Lambda_{II}^{1/2}D_{II}^T cov^{-1}(y_I)D_{II}\Lambda_{II}^{1/2}\right\}\left\{\Lambda_{II}^{1/2}D_{II}^T p\right\} = \lambda\left\{\Lambda_{II}^{1/2}D_{II}^T p\right\} \tag{2.11}$$

Let $u = \Lambda_{II}^{1/2}D_{II}^T p$ and $M = \Lambda_{II}^{1/2}D_{II}^T cov^{-1}(y_I)D_{II}\Lambda_{II}^{1/2}$, Eq. (5.24) can be rewritten as

$$Mu = \lambda u \tag{2.12}$$

Since M is a symmetric matrix, the transformed eigenvectors u in Eq. (2.12) should be orthogonal to one another, i.e.

$$\begin{aligned} U^T U = & \left(P^T D_{II}\Lambda_{II}^{1/2}\right)\left(\Lambda_{II}^{1/2}D_{II}^T P\right) \\ = & \quad P^T\left(D_{II}\Lambda_{II}D_{II}^T\right)P \\ = & \quad P^T cov(y_{II})P = I \end{aligned} \tag{2.13}$$

where $U = \left[u_1, u_2, \ldots, u_q\right]$ and $P = \left[p_1, p_2, \ldots, p_q\right]$. Therefore, the generalized eigenvectors p_i are actually orthogonal relative to the covariance matrix $cov(y_{II})$.

2.2.2 Covariance-based performance index

To assess the overall control performance of a process in the monitored period versus the benchmark period, a covariance performance index can be defined similar to McNabb and Qin (McNabb and Qin, 2003) as follows

$$I_v = \frac{|cov(y_{II})|}{|cov(y_I)|} \tag{2.14}$$

where $|\cdot|$ is the determinant. In the geometric sense, the determinants give the volume of the hyper-ellipsoids formed by the covariance matrices of the benchmark data y_I and the monitored data y_{II}. This index accounts for the covariance, which is different from the typical variance ratio. Since

$$cov(y_{II})P = cov(y_I)P\Lambda \tag{2.15}$$

then we have

$$|cov(y_{II})| \cdot |P| = |cov(y_I)| \cdot |P| \cdot |\Lambda| \tag{2.16}$$

Noticing that P is nonsingular, we further derive the performance index as

$$I_v = \frac{|cov(y_{II})|}{|cov(y_I)|} = |\Lambda| = \prod_{i=1}^{q} \lambda_i \tag{2.17}$$

Therefore, the volume ratio is equivalent to the product of all generalized eigenvalues. The power of the covariance ratio method is in identifying directions with eigenvalues significantly different from one, which points to significant performance difference. In the next section, we develop confidence intervals for the eigenvalues and define covariance based performance metrics for the degraded and improved performance subspaces. If the volume-ratio index I_v is

significantly greater than one, the monitored performance is in general worse than the benchmark. In this case, the worse performance subspace of those large eigenvalues contains the overall performance degradation. By examining and retuning within the worse performance subspace, the control performance of the process would be improved with great potential. If, on the contrary, the index value is significantly less than one, the monitored period is considered to be of better overall performance than the benchmark. If the volume index value is close to one, the monitored performance might be treated as the same as or similar to the benchmark.

2.2.3 Invariance to scaling

It should be pointed out that scaling of the original data is usually performed. A reasonable scaling approach, known as auto-scaling, is to scale all data based on the sample means and standard deviations of the benchmark data. Let y_I' and y_{II}' be the scaled data vectors in the benchmark and monitored periods, respectively, and $\Phi = diag\Big\{\phi_1, \phi_2, \ldots, \phi_q\Big\}$ be the sample standard deviations of the benchmark data. Then we have

$$\begin{aligned} cov(y_I') &= cov\left(\Phi^{-1} y_I\right) \\ &= \Phi^{-1} cov(y_I) \Phi^{-T} \end{aligned} \tag{2.18}$$

and

$$cov(y_{II}') = \Phi^{-1} cov(y_{II}) \Phi^{-T} \tag{2.19}$$

The generalized eigenvalue analysis of the scaled data between the monitored and benchmark periods can be expressed as

$$cov(y'_{II})p' = \lambda' cov(y'_I)p' \tag{2.20}$$

where λ' and p' are the generalized eigenvalue and eigenvector. Since Φ is a nonsingular diagonal matrix, substituting Eqs. (2.18) and (2.19) into (2.20) yields

$$cov(y_{II})\Phi^{-1}p' = \lambda' cov(y_I)\Phi^{-1}p' \tag{2.21}$$

Comparing Eqs. (2.21) to (2.7), it can be found that the generalized eigenvalue is invariant to data scaling, i.e. $\lambda = \lambda'$, although the corresponding eigenvector is rotated as $p = \Phi^{-1}p'$ or $p' = \Phi p$. Therefore, the covariance based performance index in Eq. (2.17) is also invariant to data scaling.

It should be noted that in controller design one often chooses a linear quadratic objective which is a weighted sum of variances. With this design objective, one also needs to select appropriate weights, but there is no systematic method to choose the optimal weights. By using the covariance ratio as an assessment objective, we can achieve scale-invariance, which is analogous to dimensionless numbers. In fact, the generalized eigenvalues are invariant even for non-diagonal matrix Φ.

2.3 Statistical inference of worse/better performance directions

As stated before, the generalized eigenvalue λ_i is a quantitative index to measure the variance ratio of the monitored period against the benchmark period along each eigenvector direction. In practice, however, we only have sample covariances from limited data which yield sample eigenvalues l_i. The sample eigenvalues l_i from the sample covariance matrices are only statistical estimates of the real population eigenvalues λ_i. It is inevitable that error is present between λ_i and l_i to some extent. Therefore, the estimated eigenvalue l_i over one does not necessarily mean that the population eigenvalue λ_i is also larger than one statistically. To examine the significance of population eigenvalues λ_i with respect to the threshold value one, a statistical inference strategy is developed and the confidence intervals for the population eigenvalues are derived on the basis of their asymptotic distribution.

2.3.1 Statistical inference for generalized eigenvalues

First consider a special case in which there is negligible temporal autocorrelation among the observations. Let

$$S = \frac{1}{n-1}\sum_{k=1}^{n}(y_k - \overline{y})(y_k - \overline{y})^T \tag{2.22}$$

be the sample covariance matrix with y_k denoting the observation vector and n being the total number of samples. If $y_1, y_2, \cdots, y_n$ are assumed to be independent with a normal distribution $N(\mu, \Sigma)$, then $(n-1)S$ obeys the Wishart

distribution termed as $W(\Sigma, n-1)$, where Σ is the population covariance matrix (Anderson, 2003). For the generalized eigenvalue problem, the asymptotic distribution of the eigenvalues can be deduced from two Wishart distributions. Let

$$S_I = \frac{1}{m-1}\sum_{j=1}^{m}(y_{I,j} - \overline{y}_I)(y_{I,j} - \overline{y}_I)^T \tag{2.23}$$

and

$$S_{II} = \frac{1}{n-1}\sum_{k=1}^{n}(y_{II,k} - \overline{y}_{II})(y_{II,k} - \overline{y}_{II})^T \tag{2.24}$$

be the sample covariance matrices with the measurement vectors of the benchmark and monitored periods, and the number of samples are m and n, respectively. Similarly, $y_{I,j}$ and $y_{II,k}$ are assumed to be independent and distributed according to $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$. Then, the distributions of $(m-1)S_I$ and $(n-1)S_{II}$ are in the form of $W(\Sigma_I, m-1)$ and $W(\Sigma_{II}, n-1)$. Let λ_i and l_i represent the generalized eigenvalues from the population and the sample covariance matrices, respectively. If we define $d_i = \sqrt{n-1}(l_i - \lambda_i)$, the following theorem holds.

Theorem 2.3.1. *As $m \to \infty$,$n \to \infty$ and $\frac{m-1}{n-1} \to \eta > 0$,the limiting distribution of d_i is normal with mean 0. The asymptotic variance is*

$$var(d_i) = 2\frac{\lambda_i^2(1+\eta)}{\eta} \tag{2.25}$$

Proof. See reference (Anderson, 2003).

According to Theorem 2.3.1, the variance of sample eigenvalue l_i can

be deducted as

$$var(l_i) = \frac{1}{n-1} var(d_i) = 2\lambda_i^2 (\frac{1}{m-1} + \frac{1}{n-1}) \tag{2.26}$$

For many dynamic control systems, however, there exists significant autocorrelation among the output signals which cannot be neglected in the statistical inference of generalized eigenvalues. Therefore, the above statistical analysis is further extended to more common case of autocorrelated time series (Desborough and Harris, 1992; Harris, 2004). Let $z_I^{(i)} = p_i^T y_I$ and $z_{II}^{(i)} = p_i^T y_{II}$, which are the projected scores of benchmark data vector y_I and monitored vector y_{II} along the ith generalized eigendirection p_i, respectively. Then the ith population eigenvalue can be written as

$$\lambda_i = \frac{p_i^T E\left[(y_{II} - \overline{y}_{II})(y_{II} - \overline{y}_{II})^T\right] p_i}{p_i^T E\left[(y_I - \overline{y}_I)(y_I - \overline{y}_I)^T\right] p_i} = \frac{E\left[(z_{II}^{(i)} - \overline{z}_{II}^{(i)})(z_{II}^{(i)} - \overline{z}_{II}^{(i)})^T\right]}{E\left[(z_I^{(i)} - \overline{z}_I^{(i)})(z_I^{(i)} - \overline{z}_I^{(i)})^T\right]} \tag{2.27}$$

Let $\sigma_{I(i)}^2 = var(z_I^{(i)})$ and $\sigma_{II(i)}^2 = var(z_{II}^{(i)})$, then the above population eigenvalue becomes $\lambda_i = \frac{\sigma_{II(i)}^2}{\sigma_{I(i)}^2}$. Further define

$$s_{I(i)}^2 = \frac{1}{m-1} \sum_{i=1}^{m} (z_I^{(i)} - \overline{z}_I)^2 \tag{2.28}$$

and

$$s_{II(i)}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_{II}^{(i)} - \overline{z}_{II})^2 \tag{2.29}$$

then the ith sample eigenvalue l_i can be expressed as $l_i = \frac{s_{II(i)}^2}{s_{I(i)}^2}$, which indicates that l_i is the function of $s_{I(i)}^2$ and $s_{II(i)}^2$ as $l_i = f(s_{I(i)}^2, s_{II(i)}^2)$. Using Taylor series

expansion of l_i around $(\sigma^2_{I(i)}, \sigma^2_{II(i)})$, we can get

$$l_i = \lambda_i + \frac{1}{\sigma^2_{I(i)}}(s^2_{_{II(i)}} - \sigma^2_{II(i)}) - \frac{\sigma^2_{II(i)}}{\sigma^4_{I(i)}}(s^2_{_{I(i)}} - \sigma^2_{I(i)}) \tag{2.30}$$

It easily follows that

$$E(l_i - \lambda_i) = 0 \tag{2.31}$$

and

$$\mathrm{var}(l_i) = \mathrm{var}\left(\frac{1}{\sigma^2_{I(i)}}(s^2_{_{II(i)}} - \sigma^2_{II(i)}) - \frac{\sigma^2_{II(i)}}{\sigma^4_{I(i)}}(s^2_{_{I(i)}} - \sigma^2_{I(i)})\right) \tag{2.32}$$

Since there is no cross-correlation between the benchmark and monitored data sets, then it can be derived as

$$\begin{aligned} \mathrm{var}(l_i) &= \mathrm{var}\left(\tfrac{s^2_{II(i)} - \sigma^2_{II(i)}}{\sigma^2_{I(i)}}\right) + \mathrm{var}\left(\tfrac{\sigma^2_{II}(s^2_{I(i)} - \sigma^2_{I(i)})}{\sigma^4_{I(i)}}\right) \\ &= \tfrac{1}{\sigma^4_{I(i)}}\mathrm{var}(s^2_{II(i)}) + \tfrac{\lambda_i^2}{\sigma^4_{I(i)}}\mathrm{var}(s^2_{I(i)}) \end{aligned} \tag{2.33}$$

Using the asymptotic quadratic statistics given by Desborough and Harris (Desborough and Harris, 1992),

$$\mathrm{var}(s^2_{I(i)}) = 2\frac{\sigma^4_{I(i)}}{m-1} f_I^{(i)} \tag{2.34}$$

and

$$\mathrm{var}(s^2_{II(i)}) = 2\frac{\sigma^4_{II(i)}}{n-1} f_{II}^{(i)} \tag{2.35}$$

where

$$f_I^{(i)} = 1 + 2\sum_{j=1}^{m}(1 - \tfrac{j}{m})\rho^2_{I(i),j} \tag{2.36}$$

and

$$f_{II}^{(i)} = 1 + 2\sum_{j=1}^{n}(1 - \tfrac{j}{n})\rho^2_{II(i),j} \tag{2.37}$$

with $\rho_{I(i),j}$ and $\rho_{II(i),j}$ representing the autocorrelation coefficients of benchmark and monitored data along the ith eigendirection at lag j, respectively. Substituting Eqs. (2.34) and (2.35) into (2.33) can yield

$$\text{var}(l_i) = 2\lambda_i^2(\frac{f_I^{(i)}}{m-1} + \frac{f_{II}^{(i)}}{n-1}) \tag{2.38}$$

which includes the autocorrelation terms from both benchmark and monitored data. The autocorrelations can be estimated from the samples. For unautocorrelated signals, $f_I^{(i)} = f_{II}^{(i)} = 1$ and the variance of sample eigenvalue becomes

$$\text{var}(l_i) = 2\lambda_i^2(\frac{1}{m-1} + \frac{1}{n-1}) \tag{2.39}$$

which is the same as the result from the limiting distribution given in Theorem 2.3.1. It is also obvious that autocorrelation serves to inflate the variance of generalized eigenvalues.

2.3.2 Detection strategy for worse/better eigendirections

Based on the calculated eigenvalues from the sample covariance matrices of the monitored and the benchmark data, the remaining issue is to determine if the population eigenvalues from the population covariance matrices are larger or smaller than one. According to the statistical inference of generalized eigenvalues, the quantity $l_i - \lambda_i$ follows $N\big(0, 2\lambda_i^2(\frac{f_I}{m-1} + \frac{f_{II}}{n-1})\big)$ asymptotically. Therefore, we have

$$Z = \left[\sqrt{2\big(\frac{f_I}{m-1} + \frac{f_{II}}{n-1}\big)}\right]^{-1} \cdot \left(\frac{l_i - \lambda_i}{\lambda_i}\right) \tag{2.40}$$

follows the standard normal distribution. At a given confidence level $(1-\alpha)\times 100\%$, e.g. $\alpha = 0.05$, the confidence interval of each eigenvalue λ_i is in the form of $L(\lambda_i) \leq \lambda_i \leq U(\lambda_i)$ that satisfies the following probability

$$Pr\Big\{L(\lambda_i) \leq \lambda_i \leq U(\lambda_i)\Big\} = 1-\alpha \tag{2.41}$$

The above equation means there is $1-\alpha$ probability for the true value of λ_i to be inside the confidence interval. From Eq. (2.40), we have

$$Pr\left\{-z_{\alpha/2} \leq \left[\sqrt{2(\frac{f_I}{m-1}+\frac{f_{II}}{n-1})}\right]^{-1} \cdot \left(\frac{l_i-\lambda_i}{\lambda_i}\right) \leq z_{\alpha/2}\right\} = 1-\alpha \tag{2.42}$$

where $z_{\alpha/2}$ is the $\alpha/2$ critical value of a standard normal variate. It follows that

$$Pr\left\{\left[1+z_{\alpha/2}\sqrt{2\left(\frac{f_I}{m-1}+\frac{f_{II}}{n-1}\right)}\right]^{-1} \leq \frac{\lambda_i}{l_i} \leq \left[1-z_{\alpha/2}\sqrt{2\left(\frac{f_I}{m-1}+\frac{f_{II}}{n-1}\right)}\right]^{-1}\right\} = 1-\alpha \tag{2.43}$$

The confidence interval of λ_i can readily be derived with the upper and lower bounds as

$$U(\lambda_i) = \left[1-z_{\alpha/2}\sqrt{2\left(\frac{f_I}{m-1}+\frac{f_{II}}{n-1}\right)}\right]^{-1} \cdot l_i \tag{2.44}$$

and

$$L(\lambda_i) = \left[1+z_{\alpha/2}\sqrt{2\left(\frac{f_I}{m-1}+\frac{f_{II}}{n-1}\right)}\right]^{-1} \cdot l_i \tag{2.45}$$

The confidence limits for every population eigenvalue λ_i can be computed from the corresponding sample eigenvalue l_i. As a result, the decision rules to determine the worse/better performance eigendirections are summarized below:

(i). If the lower bound $L(\lambda_i) > 1$, then the population eigenvalue λ_i is statistically greater than one and the control performance of the monitored period is worse than that of the benchmark period along the corresponding eigendirection;

(ii). If the upper bound $U(\lambda_i) < 1$, then the population eigenvalue λ_i is statistically less than one and the control performance of the monitored period is better than that of the benchmark period along the corresponding eigendirection;

(iii). If the confidence bounds satisfy $L(\lambda_i) \leq 1 \leq U(\lambda_i)$, then the population eigenvalue λ_i is statistically the same as one and there is no significant difference between the monitored control performance and the benchmark performance along the corresponding eigendirection, which is termed as marginal direction.

2.3.3 Performance measure in the worse/better subspace

Suppose that the first w and the last b generalized eigenvectors are identified as the worse and the better performance eigendirections, respectively. The worse and the better performance subspaces are spanned by $P_w = [p_1, p_2, \ldots, p_w]$ and $P_b = [p_{q-b+1}, p_{q-b+2}, \ldots, p_q]$, respectively, where $p_1, p_2, \ldots, p_q$ are generalized eigenvectors of the corresponding eigenvalues in descending order. The projections of any sample y onto the worse and the better subspaces are given by $z_w = (P_w^T P_w)^{-1} P_w^T y$ and $z_b = (P_b^T P_b)^{-1} P_b^T y$.

Within the worse performance subspace, the projections of the benchmark data vector y_I and the monitored vector y_{II} are $z_I^{(w)} = (P_w^T P_w)^{-1} P_w^T y_I$ and $z_{II}^{(w)} = (P_w^T P_w)^{-1} P_w^T y_{II}$, respectively. Therefore,

$$cov(z_I^{(w)}) = (P_w^T P_w)^{-1} P_w^T cov(y_I) P_w (P_w^T P_w)^{-1} \tag{2.46}$$

and

$$cov(z_{II}^{(w)}) = (P_w^T P_w)^{-1} P_w^T cov(y_{II}) P_w (P_w^T P_w)^{-1} \tag{2.47}$$

It is also known from Eq. (2.7) that

$$cov(y_{II}) P_w = cov(y_I) P_w \Lambda_w \tag{2.48}$$

where $\Lambda_w = diag\{\lambda_1, \lambda_2, \ldots, \lambda_w\}$.

Equation (2.47) thus becomes

$$\begin{aligned} cov\big(z_{II}^{(w)}\big) &= (P_w^T P_w)^{-1} P_w^T cov(y_I) P_w \Lambda_w (P_w^T P_w)^{-1} \\ &= (P_w^T P_w)^{-1} P_w^T cov(y_I) P_w (P_w^T P_w)^{-1} (P_w^T P_w) \Lambda_w (P_w^T P_w)^{-1} \\ &= cov\big(z_I^{(w)}\big) (P_w^T P_w) \Lambda_w (P_w^T P_w)^{-1} \end{aligned} \tag{2.49}$$

Therefore, the covariance based performance index within the worse performance subspace can be derived as

$$\begin{aligned} I_{vw} &= \frac{|cov\big(z_{II}^{(w)}\big)|}{|cov\big(z_I^{(w)}\big)|} \\ &= |P_w^T P_w| \cdot |\Lambda_w| \cdot |P_w^T P_w|^{-1} \\ &= |\Lambda_w| = \prod_{i=1}^{w} \lambda_i \end{aligned} \tag{2.50}$$

Similarly, the performance index within the better subspace can be expressed as

$$I_{vb} = |\Lambda_b| = \prod_{i=q-b+1}^{q} \lambda_i \tag{2.51}$$

with $\Lambda_b = diag\{\lambda_{q-b+1}, \lambda_{q-b+2}, \ldots, \lambda_q\}$. The performance indices I_{vw} and I_{vb} represent the performance ratios of the monitored data versus the benchmark data within the worse and the better subspaces, respectively. In the geometric sense, they correspond to the volume ratios of the projected data in the worse/better performance subspace. The indices provide quantitative measures on how much the control performance is degraded or improved along the worse or better eigendirections.

The overall MIMO performance monitoring procedures based on covariance matrices can be summerized in Fig. 2.1. The performance index I_v can be monitored to indicate overall performance changes. Statistical inference on the generalized eigenvalues identifies worse or better performance subspace. Monitoring of indices I_{vw} or I_{vb} gives an overall measure of performance degradation or improvement within the respective subspace. If the data has non-zero mean, the monitoring scheme can be modified by replacing the covariance with the second-order moments.

2.4 Case studies

2.4.1 Simulated multiloop example

First consider the 4×4 process presented by McNabb and Qin (McNabb and Qin, 2005*a*). The open-loop process transfer function matrix $G(q^{-1})$ and

disturbance transfer function matrix $N(q^{-1})$ are given below

$$G(q^{-1}) = \begin{bmatrix} \frac{0.05q^{-3}}{1-0.95q^{-1}} & 0 & \frac{0.7q^{-3}}{1-0.3q^{-1}} & 0 \\ \frac{0.02966q^{-3}}{1-1.627q^{-1}+0.706q^{-2}} & \frac{0.0627q^{-6}}{1-0.937q^{-1}} & 0 & 0 \\ 0 & \frac{0.235q^{-5}}{1-0.765q^{-1}} & \frac{0.5q^{-2}}{1-q^{-1}+0.25q^{-2}} & 0 \\ \frac{0.5q^{-5}-0.4875q^{-6}}{1-1.395q^{-1}+0.455q^{-2}} & 0 & 0 & \frac{0.2q^{-6}}{1-0.8q^{-1}} \end{bmatrix} \tag{2.52}$$

and

$$N(q^{-1}) = \begin{bmatrix} \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 & 0 & 0 \\ 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 & 0 \\ 0 & 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 \\ 0 & 0 & 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} \end{bmatrix} \tag{2.53}$$

For this example, the disturbance, $a(t)$, is a four-dimensional normally distributed white noise sequence. The multi-loop controllers are four PI controllers of the discrete form $K_c\{1 + (1/T_r)[\Delta T/(1 - q^{-1})]\}$, where K_c is the proportional gain, ΔT is the controller sampling time and T_r is the integral time. The controller parameters for the benchmark period are $K_c = [0.816 \quad 0.625 \quad 0.184 \quad 0.370]$ and $T_r = [20 \quad 16 \quad 2.86 \quad 5]$. During the monitored period, the controller gains are changed to $K_c = [3.07 \quad 0.316 \quad 0.518 \quad 0.127]$ to alter the performance. In our simulation, $3,000$ samples are generated in the benchmark and the monitored periods. The closed-loop process output in the benchmark period I and the monitored period II is shown in Fig. 2.2.

Through generalized eigenvalue analysis between the covariance matrices of the benchmark data and the monitored data, the eigenvalues and the corresponding eigendirections are obtained and displayed in Fig. 2.3(a) and Fig. 2.4. The calculated overall performance index I_v is 1.925, and thus the

volume of the monitored data is almost twice as much of the benchmark data. It implies that the overall control performance of the monitored period is inferior to the performance of the benchmark period. When looking at each individual eigenvalue and eigendirection in Fig. 2.3(a) and Fig. 2.4, however, it can be found that the control performance is degraded along some directions and improved along the others. For example, the maximum eigenvalue is over 4, which means the variance along the first eigenvector direction is increased by a factor of 4. Therefore, the control performance of the monitored period is much worse than that of the benchmark period along this eigendirection. On the other hand, the minimum eigenvalue is far below one which indicates better monitored performance over the benchmark in the last eigendirection.

It is readily seen from Fig. 2.3(b) that the lower bounds of the first and the second eigenvalues are both larger than one. Consequently, the first two eigendirections are determined as worse directions. The upper bounds of the last two eigenvalues, on the contrary, are below one. Thus the corresponding two eigendirections are better directions with improved control performance in the monitored period against the benchmark period. Similarly, these two directions span the better performance subspace. The results are consistent with the loop tuning change in the simulation because some loops are detuned while the others have increased gain. The performance indices in the worse and better subspaces are $I_{vw} = 11.614$ and $I_{vb} = 0.1658$, respectively. These index values also indicate that the overall monitored performance is seriously deteriorated against the benchmark within the worse subspace while it is sig-

nificantly improved from the benchmark to the monitored period within the better subspace. In this way, the worse and the better performance subspaces can be identified. While the eigendirections in Fig. 2.4 show the contributions to the loops, it is inappropriate to use these directions to determine which loops are responsible for the worse or better control performance because the eigenvectors are not even orthogonal. Appropriate contributions and the confidence intervals must be developed to diagnose the loops with degraded or improved control performance.

2.4.2 Simulated MPC example

To illustrate the use of the proposed methodology on a multivariable control system, the following 2×2 process is taken into account. The process model from an industrial fractionation column (Harris *et al.*, 1996) is given by

$$G(q^{-1}) = \begin{bmatrix} \frac{-1.5q^{-3}}{1-0.659q^{-1}} & \frac{-0.167q^{-1}}{1-0.923q^{-1}} \\ \frac{-0.519q^{-5}}{1-0.784q^{-1}} & \frac{0.154q^{-3}+0.144q^{-4}}{1-0.874q^{-1}} \end{bmatrix} \tag{2.54}$$

The disturbance model is expressed as

$$N(q^{-1}) = \begin{bmatrix} \frac{1-0.5q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} & \frac{-0.2q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} \\ \frac{-0.5q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} & \frac{1-0.8q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} \end{bmatrix} \tag{2.55}$$

which is nonstationary. The constrained MPC design is discussed by Ko and Edgar (Ko and Edgar, 2001*a*). It is assumed that all the controlled and the manipulated variables have equal weightings with the manipulated variables constrained between ± 5. The prediction horizon and the control horizon are tuned to 20 and 5, respectively. A white noise sequence $a(t)$ with the covariance matrix of $0.1I$ is used. For this example, the effect of plant/controller model

mismatch is simulated. In the benchmark period I, no model mismatch is present and $1,000$ samples are simulated. In the monitored period II, a plant model mismatch is implemented with the process gain of G_{12} increased by 30%, and another $1,000$ samples are generated. The simulated closed-loop data of both CVs and MVs from the constrained MPC system are shown in Fig. 2.5.

By performing the generalized eigenvalue analysis on the covariance matrices, the two eigenvalues and the corresponding eigenvectors are exhibited in Fig. 2.6. The maximum eigenvalue is over 1.5 and thus the monitored performance seems to be worse than the benchmark performance. The corresponding largest eigenvector has a dominant loading coefficient in CV1, as shown in Fig. 2.6(b). Further carrying out the statistical inference method can result in the confidence intervals of the eigenvalues, as shown in Fig. 2.6(d). The first eigendirection is statistically determined as the worse direction since its lower bound is larger than one. The last eigendirection, on the other hand, is classified as marginal direction because the threshold value one is within its confidence bounds. It means that there is no significant difference between the benchmark and the monitored performance along this direction. The calculated overall performance index I_v is 1.334. Therefore, the overall performance in the monitored period is worse than that of the benchmark. By examining the first eigendirection, it appears that the major contributing variable to the worse performance is CV1 which is consistent with the model mismatch introduced in the simulation. To confirm this, statistical analysis of the contributors

will be discussed in Chapter 3.

2.4.3 Industrial example

An industrial wood waste burning power boiler process is investigated to further verify the validity of the performance assessment approach. In this example, the routine operating data are collected from the power boiler system with overall ten loops under closed-loop operation. The sampling time of the process is five seconds. The process and instrumentation diagram of the power boiler system is shown in Fig. 2.7. The data processing is applied to the controller error sequences between process variables (PV) and set points (SP), i.e. PV-SP. The detailed description for these control loops is given in Table 2.1.

A set of $4,000$ samples are extracted during the routine operation. The first $2,000$ samples are set as the benchmark data and the remaining $2,000$ samples are used as the monitored data. The controller error signal plots for the ten loops in the power boiler unit are shown in Fig. 2.8. The generalized eigenvalue analysis based covariance monitoring is firstly performed on the data set. The computation results from the presented monitoring procedures are depicted in Fig. 2.9. Fig. 2.9(a) exhibits the maximal and minimal eigenvalues, from which we can tell that the largest variance inflation of the monitored period over the benchmark period is close to 3 along the corresponding eigenvector direction. Hence, the control performance of the monitored period II along this direction is significantly worse compared to the benchmark period

I. The smallest eigenvalue, on the other hand, indicates that the variance of the monitored period II is much smaller than that of the benchmark along the corresponding eigendirection, which actually represents improved performance direction. The loading plots of the first and the last eigendirections are given in Fig. 2.9(b) and (c), respectively. It can be seen from Fig. 2.9(b) that Loop 6 contributes most significantly in the first eigendirection, but many other loops have large loadings as well. Chapter 3 gives confidence limits to detect significant contributing loops. Similarly, Fig. 2.9(c) shows that Loop 8 is most outstanding in the last eigenvector direction, which implies improved performance of Loop 8. The full spectrum of sample eigenvalues in descending order and the corresponding cumulative percentages are shown in Fig. 2.9(d).

The original samples from both the benchmark period I and the monitored period II are projected to the first and the last generalized eigendirections, which correspond to the largest and the the smallest eigenvalues, respectively. In Fig. 2.10(a), the monitored Period II exhibits larger variation than the benchmark Period I along the first eigendirection. The corresponding largest eigenvalue $l_{max} = 2.8517$ reflects the variance ratio of the projected data along this direction. The opposite situation, on the other hand, can be readily seen in Fig. 2.10(b) where the variance of the monitored Period II is much smaller than that of the benchmark Period I. The smallest eigenvalue $l_{min} = 0.1047$ indicates the variance ratio of projected points between the monitored and the benchmark periods. These variance changes cannot be easily seen in the original data in Fig. 2.8, which shows the effectiveness of the

proposed performance monitoring method.

The computational results for the confidence intervals of population eigenvalues are shown in Fig. 2.11. It can be observed that the lower bounds of eigenvalues for the first four eigendirections exceed the threshold value line. Hence, the first four eigendirections are worse directions with poorer control performance for the monitored operation data. These eigenvectors span the four-dimensional subspace of worse control performance and there is significant margin to further improve the performance within this subspace. The last three eigendirections, on the contrary, are better directions with superior control performance relative to the benchmark, because their upper bounds are all below one. Accordingly the better performance subspace is composed of the last three eigenvectors, and trying to maintain the loop operation within the three-dimensional subspace would benefit the overall system performance. As far as the middle part of the eigenvalue spectrum, i.e. the $5th$, $6th$ and $7th$ eigendirections, the corresponding upper and lower bounds are located on both sides of one. Therefore, these three directions are marginal with statistically the same control performance in the benchmark and monitored periods.

The calculated performance index values for the worse and better subspaces are $I_{vw} = 11.674$ and $I_{vb} = 0.036$, respectively. It means that the covariance volume of the monitored data within the worse subspace is 11 times larger than that of the benchmark, while the covariance volume in the better subspace is only around one thirtieth of the benchmark. The large value of I_{vw} also implies that there is great potential to further improve the monitored

performance within the worse subspace. On the other hand, the small I_{vb} values represents the remarkable performance improvement occurring within the better subspace.

In addition to the monitored period II, five additional periods of operating data are monitored for the power boiler system and each period includes $2,000$ samples. The new periods are marked as III, IV, V, VI and VII. The control performance of the monitored periods II-VII is assessed against the benchmark period I sequentially and the control error signals for all the periods are shown in Fig. 2.12. The worse and the better performance subspaces are identified using statistical inference strategy and the confidence interval charts are given in Fig. 2.13. The worse subspace of the monitored period III versus the benchmark I, for example, contains the first three eigendirections. The better subspace, on the other hand, is spanned by the last five eigendirections. The calculated overall performance index I_v values for the monitored periods II-VII are 0.421, 0.862, 13.680, 1.246, 4.497 and 1.089, respectively, as depicted in Fig. 2.14(a). Apparently, the overall control performance in the monitored periods IV and VI is significantly worse than that of the benchmark period I. On the other hand, the overall performance of the monitored period II is even better than the benchmark performance. Regarding the monitored periods III, V and VII, it can be seen that the performance index values are close to the reference line one, and therefore the overall performance in these periods is similar to the benchmark performance. The performance indices within the isolated worse/better subspace are shown in Fig. 2.14(b) and

Fig. 2.14(c). The large I_{vw} value in Period IV, for instance, means that the performance in Period IV degrades significantly versus the benchmark within the worse subspace. Thus there is remarkable margin to improve the control performance in the worse subspace of Period IV. The small I_{vb} values in Periods II and VII, on the contrary, suggest that the corresponding monitored periods have significantly improved performance within the respective better subspaces.

2.5 Summary

Multivariate statistical analysis has been successfully introduced into the MIMO control performance assessment and monitoring in this chapter. First, a data-driven benchmark is proposed as a user-defined benchmark. The user-defined benchmark is chosen from historical operation data that have desirable and representative performance and thus avoids a priori process knowledge such as the interactor matrix needed for the MVC benchmark. Generalized eigenvalue analysis is then performed to find the directions with the worst or best control performance in the monitored period versus the benchmark period. Statistical inference methods are developed to determine the directions with significantly worse or better control performance. For the MIMO performance assessment purpose, a covariance-based index is defined to evaluate the overall control performance in the monitored period against the benchmark, and performance indices for the worse/better subspace are also derived to assess the extent of performance degradation/improvement within the respective

subspace. The generalized eigenvalues as well as these performance indices are shown to be invariant to data scaling.

The present approach is applied to two simulated examples, a multi-loop control and a multivariable MPC system, and an industrial power boiler example. The results demonstrate the effectiveness of the data-driven benchmark based statistical performance monitoring approach. The directions and subspaces with worse and better control performance in the monitored period against the benchmark period are identified. Further results on the performance diagnosis aspect, i.e. identifying the responsible loops or CVs for performance degradation or improvement of MIMO processes, will be described in Chapter 3.

The proposed performance monitoring method provides an overall measure of performance changes that are observed in the control error signals. In principle, any causes that result in a change in the control error covariance can be detected and identified, including any process changes, active constraints, and disturbance changes. Therefore, it is regarded as a holistic or umbrella method. Further diagnosis should be integrated into this covariance based monitoring method, such as detection of significant process changes, significant disturbance changes, and change in active constraint sets. For significantly different operating modes which are for very different active constraint sets, it is arguable that multiple benchmarks, one for each mode, should be used. Dramatic changes in disturbance dynamics should call for an updated benchmark, whether or not it is data-based or MVC-based. If offset is accept-

able in the control error, it is straightforward to revise the method by replacing the sample mean with the corresponding setpoint.

Table 2.1: The description of ten control loops from the power boiler system

Loop ID	Category	Description
Loop 1	Flow Control	Power boiler feed water flow control
Loop 2	Flow Control	Oil burner air flow control
Loop 3	Flow Control	Bark-air flow control
Loop 4	Flow Control	Bark feed rate control
Loop 5	Flow Control	Bark air firing control
Loop 6	Level Control	Power boiler drum level control
Loop 7	Pressure Control	Combustion air pressure
Loop 8	Pressure Control	Furnace pressure control
Loop 9	Pressure Control	Over-fire air pressure control
Loop 10	Pressure Control	Steam head pressure control

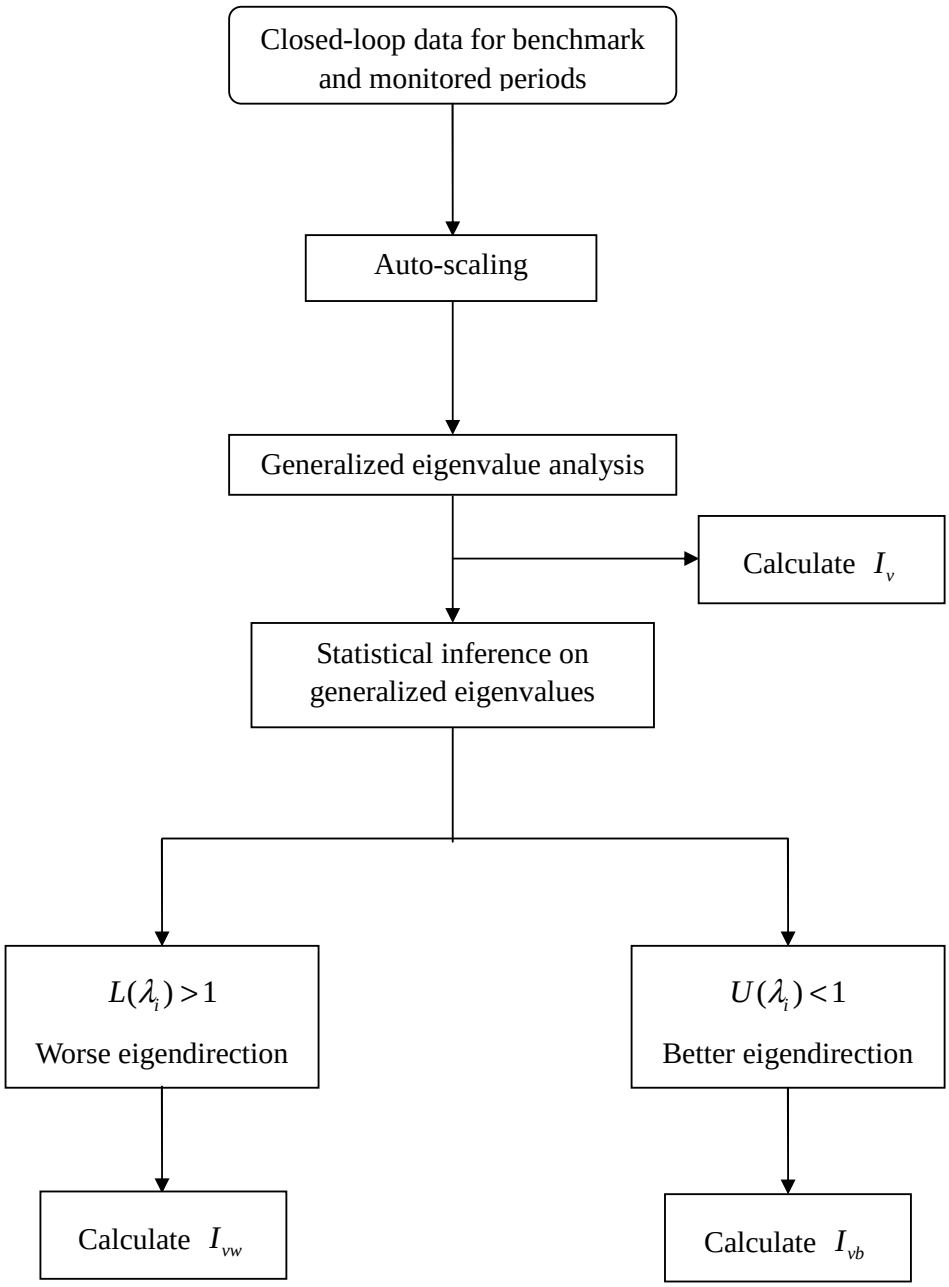

Figure 2.1: The scheme of the proposed control performance monitoring approach

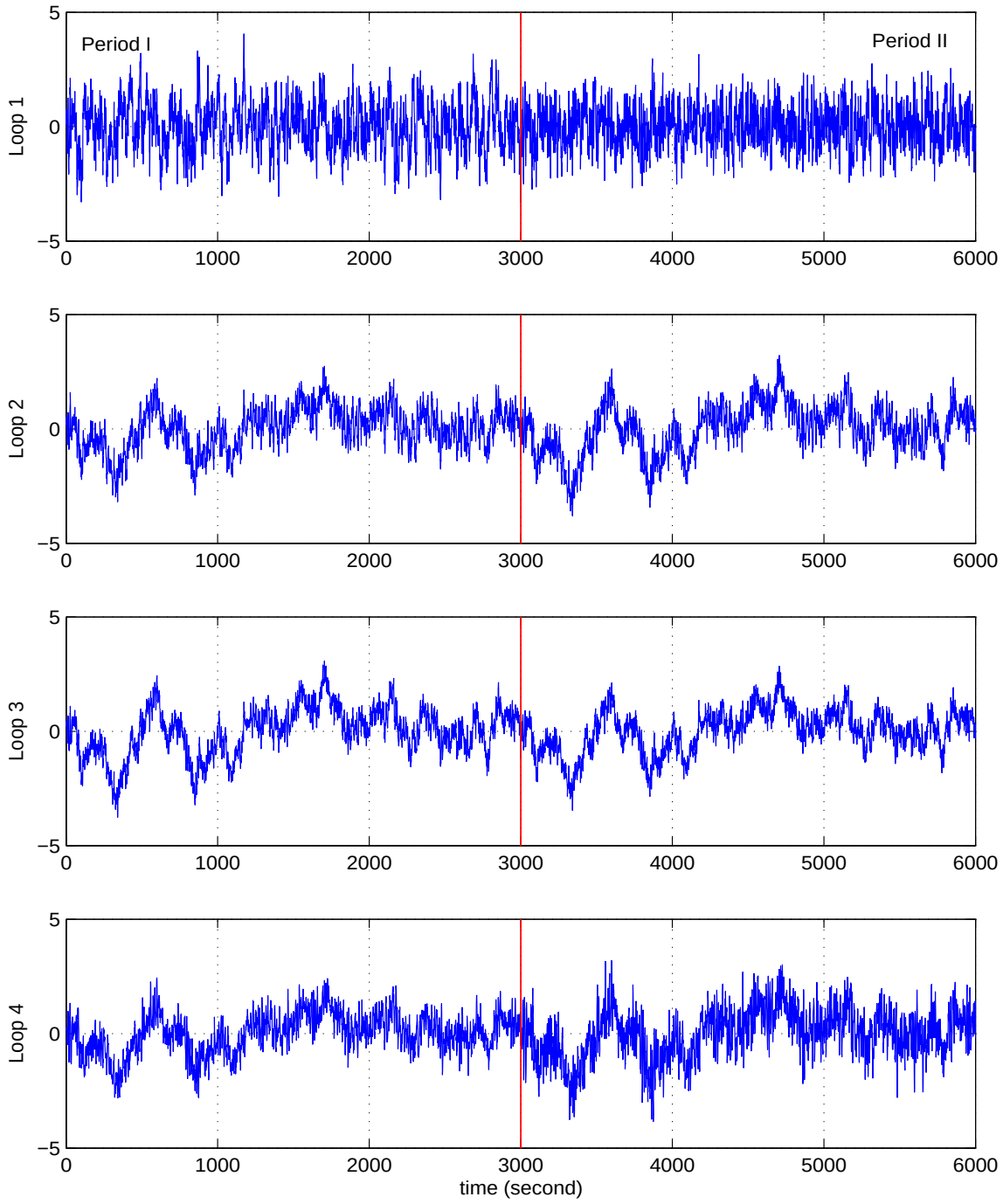

Figure 2.2: Simulated 4×4 example: The closed-loop process output in the benchmark and monitored periods

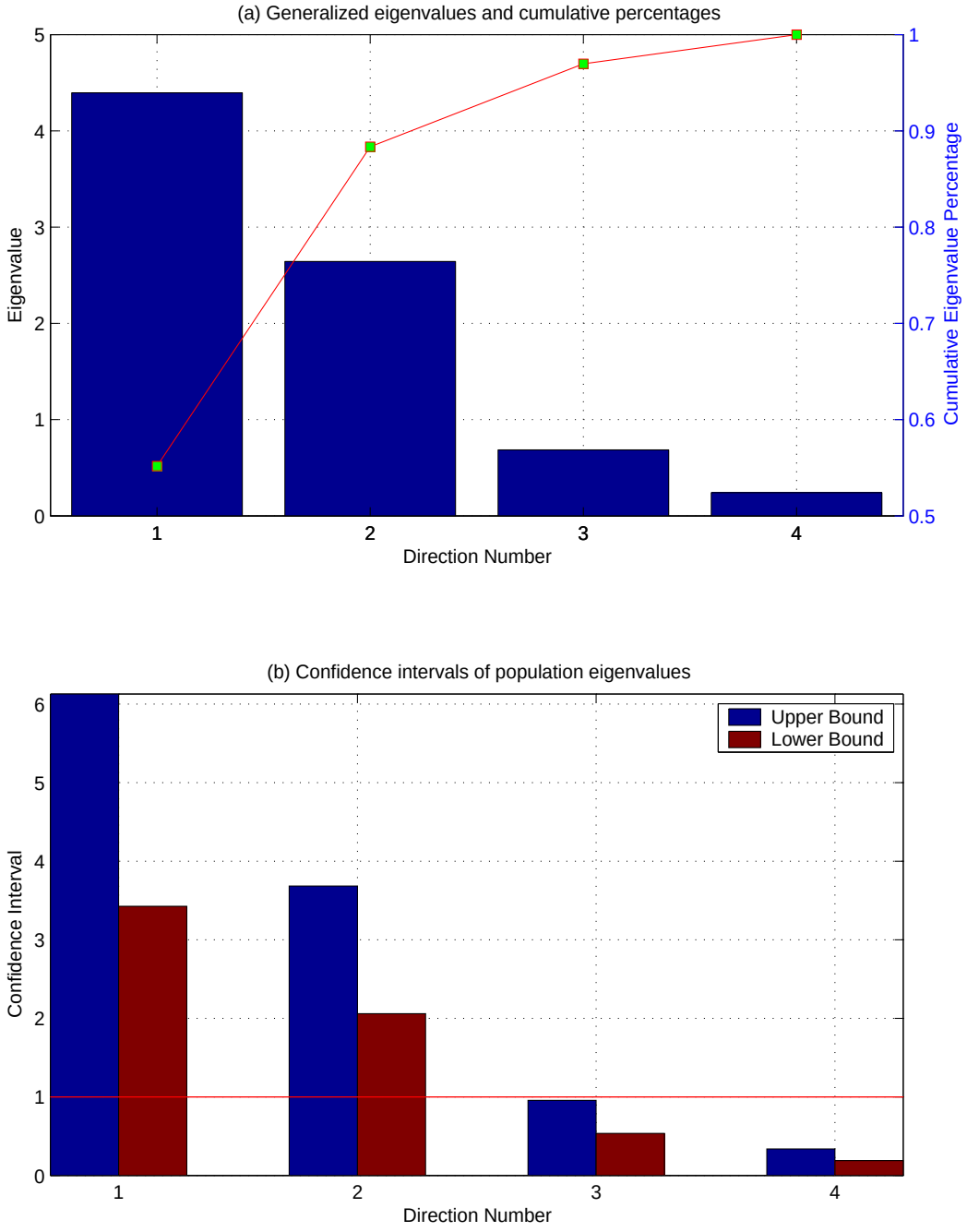

Figure 2.3: Simulated 4×4 example: (a) The full eigenvalue spectrum and the corresponding cumulative percentages; (b)the 95% confidence intervals for population eigenvalues

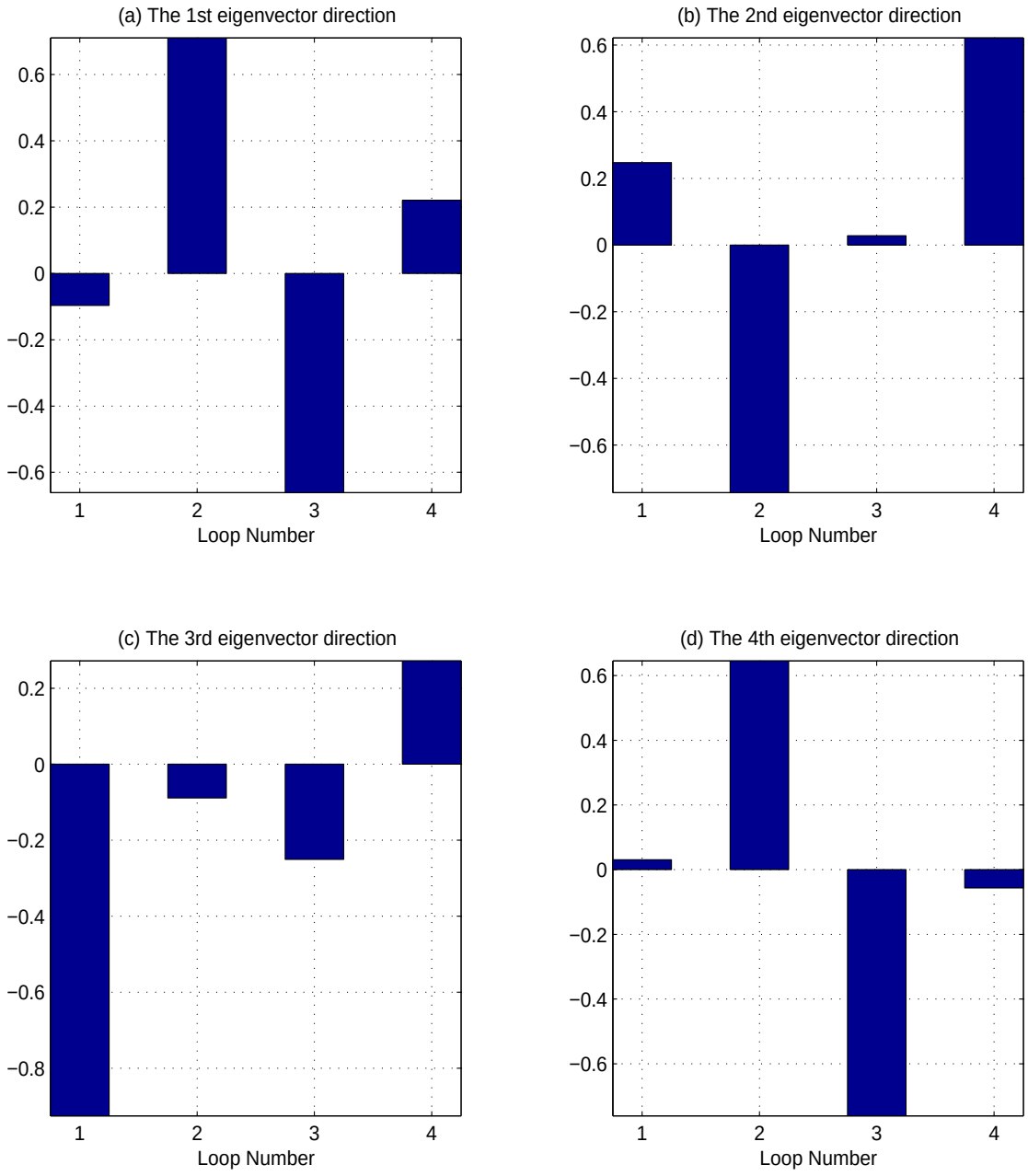

Figure 2.4: Simulated 4×4 example: Covariance based generalized eigenvalue analysis results for the monitored period II against the data-driven benchmark period I with (a) the 1st eigenvector direction; (b) the 2nd eigenvector direction; (c) the 3rd eigenvector direction; (d) the 4th eigenvector direction

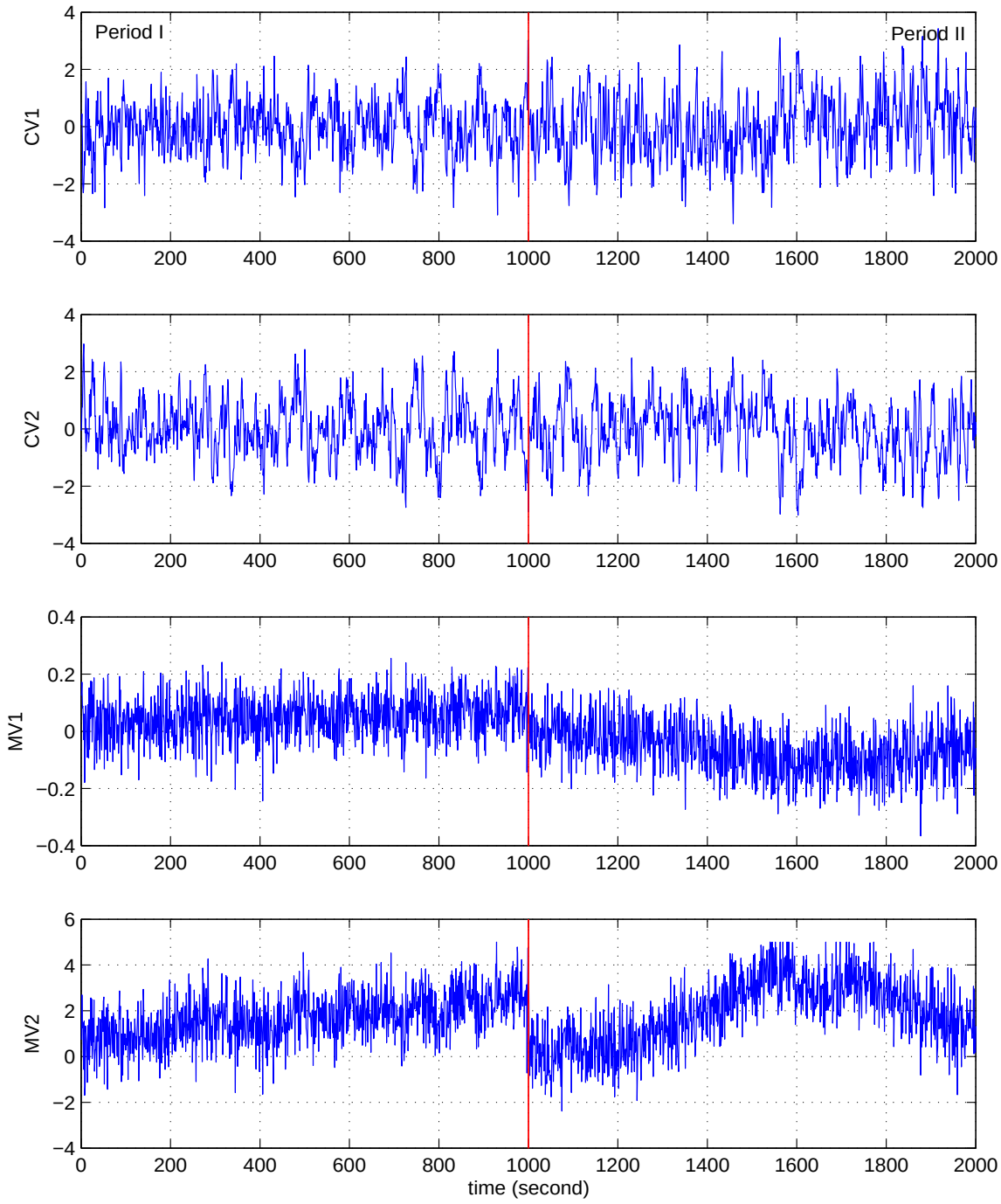

Figure 2.5: Simulated MPC example: The closed-loop data in the benchmark and monitored periods

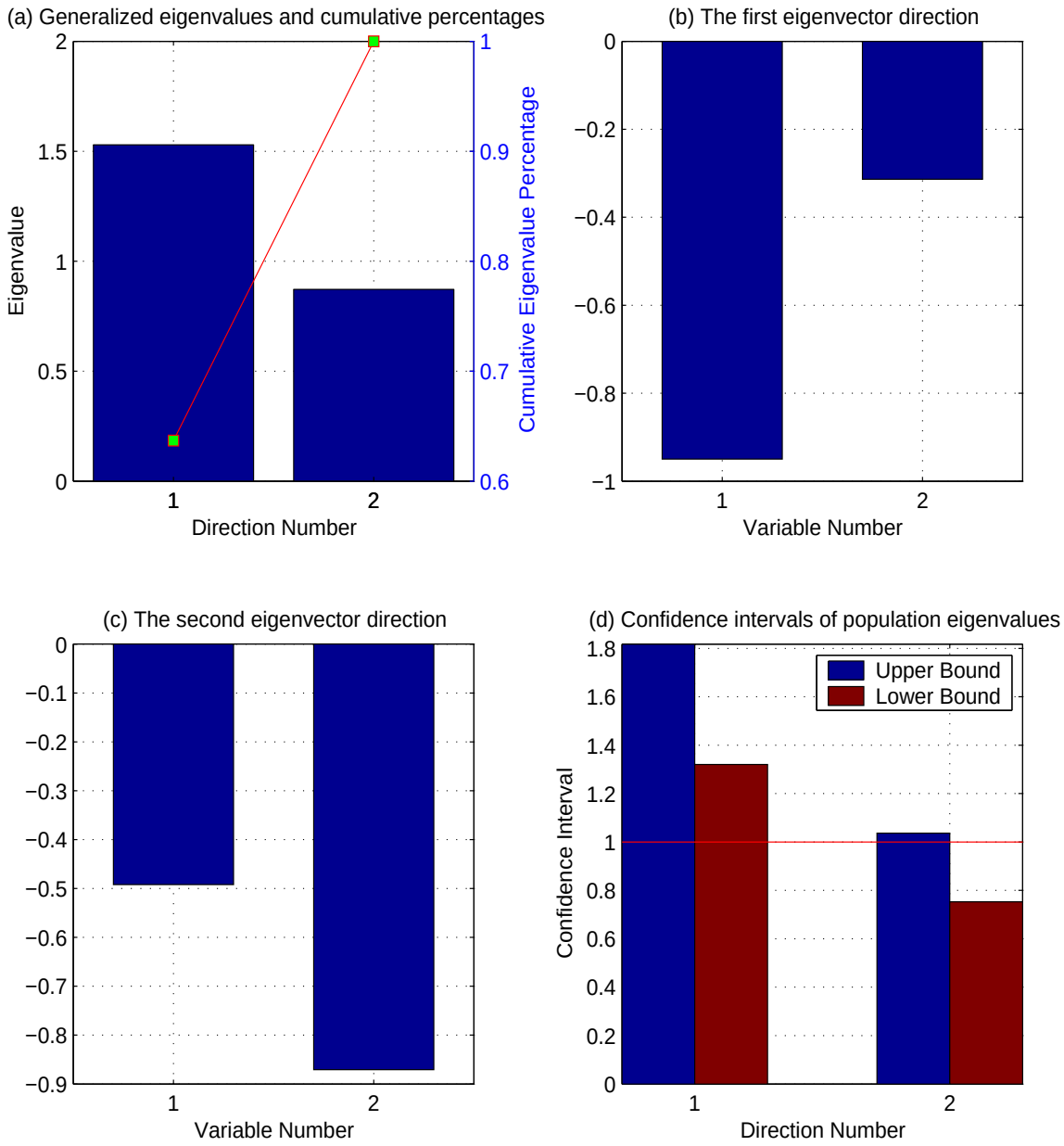

Figure 2.6: Simulated MPC example: Covariance based generalized eigenvalue analysis results for the monitored period II against the data-driven benchmark period I with (a) the eigenvalue spectrum and the cumulative percentages; (b) the first eigenvector direction; (c) the second eigenvector direction; (d) the 95% confidence intervals for population eigenvalues

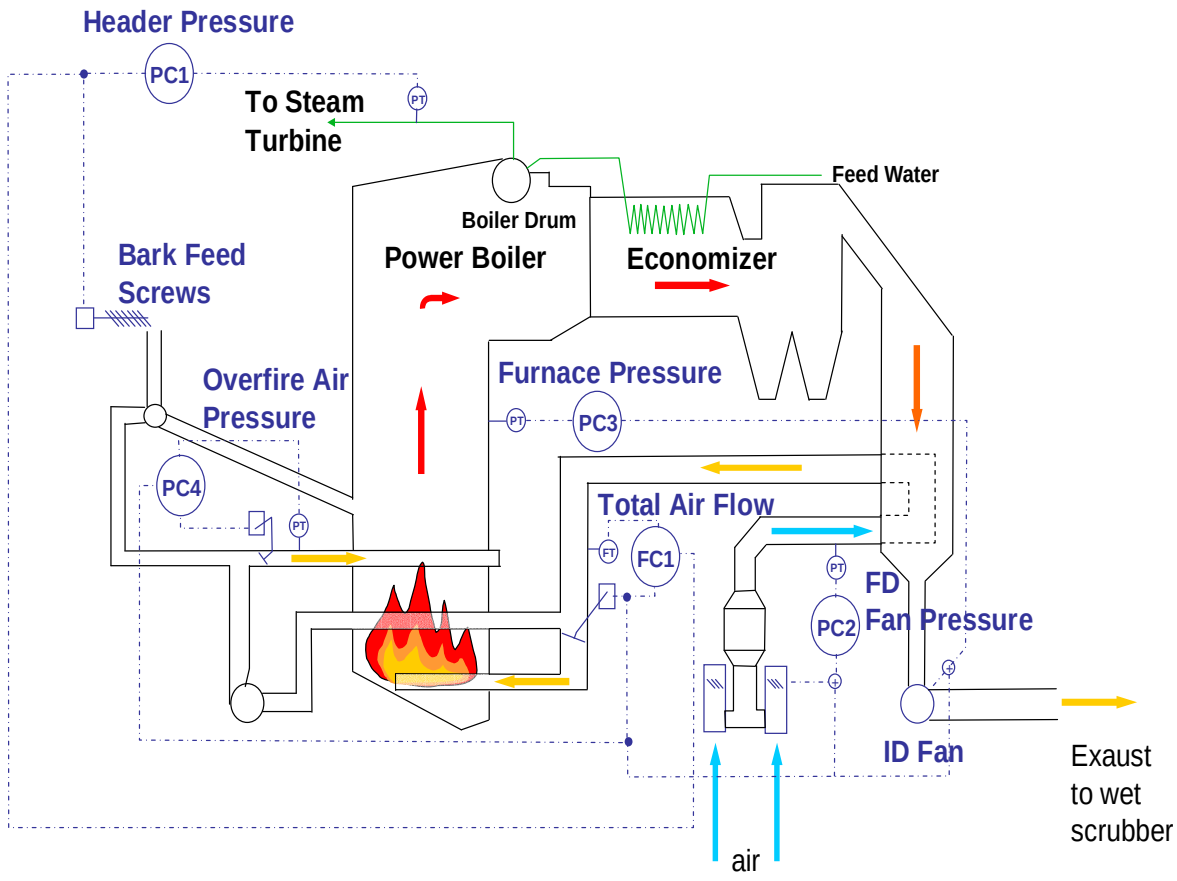

Figure 2.7: The process and instrumentation diagram of the power boiler system

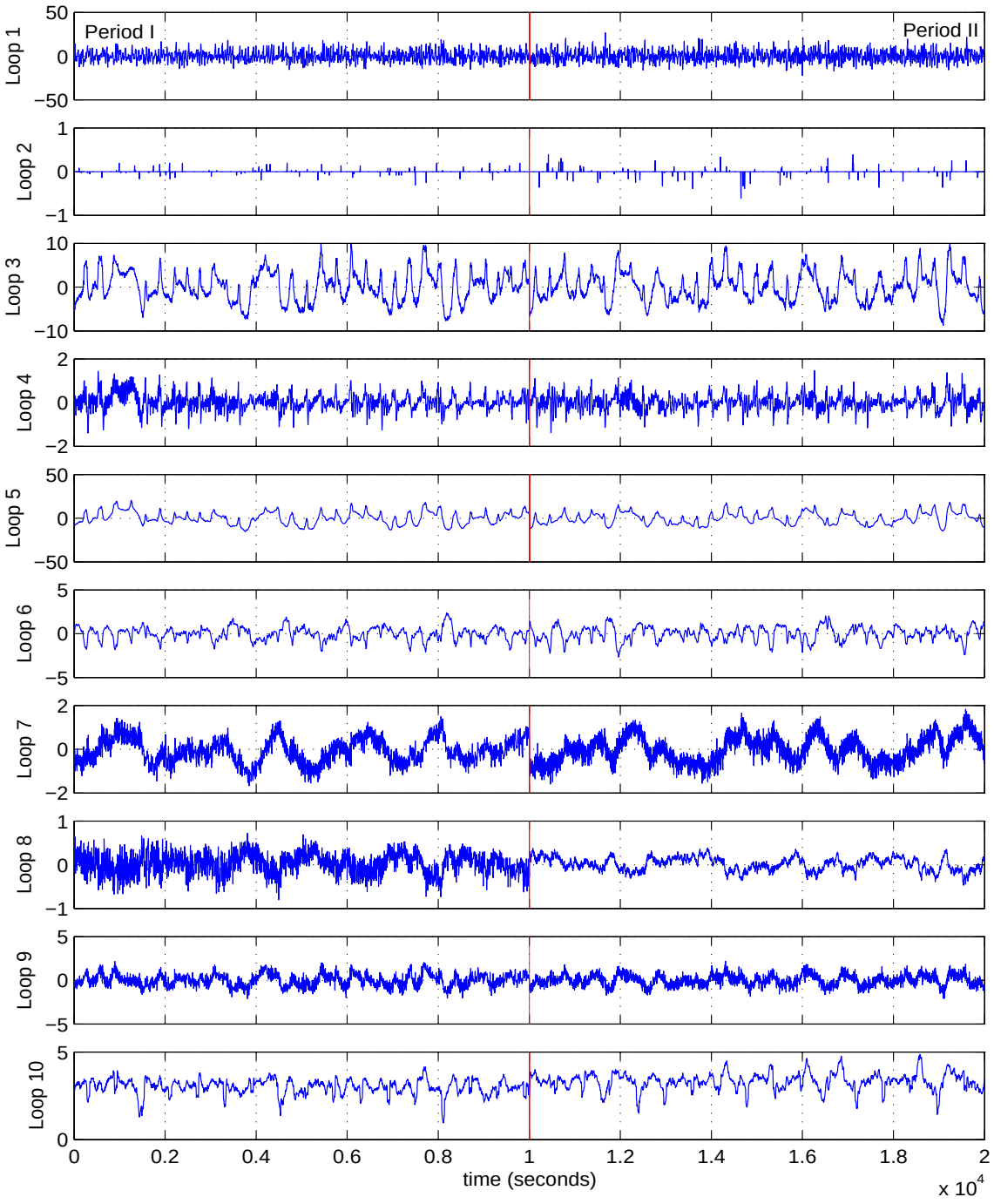

Figure 2.8: Signal plots of controller error for the ten control loops of the power boiler system

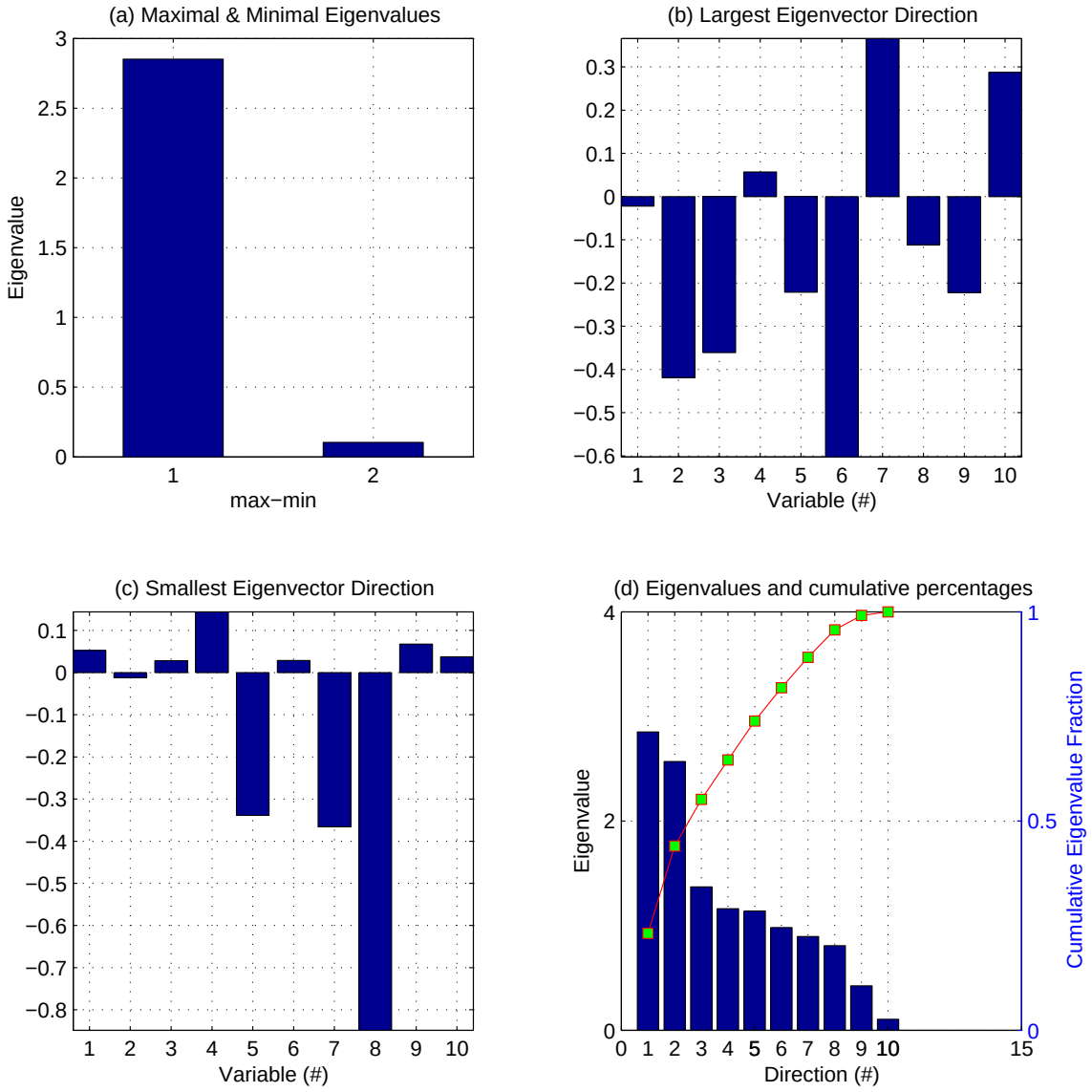

Figure 2.9: Industrial boiler example: Covariance based generalized eigenvalue analysis results for the monitored period II against the data-driven benchmark period I with (a) the maximal and minimal eigenvalues; (b) the eigenvector direction corresponding to the maximal eigenvalue; (c) the eigenvector direction corresponding to the minimal eigenvalue; (d) the full eigenvalue spectrum and the corresponding cumulative percentages

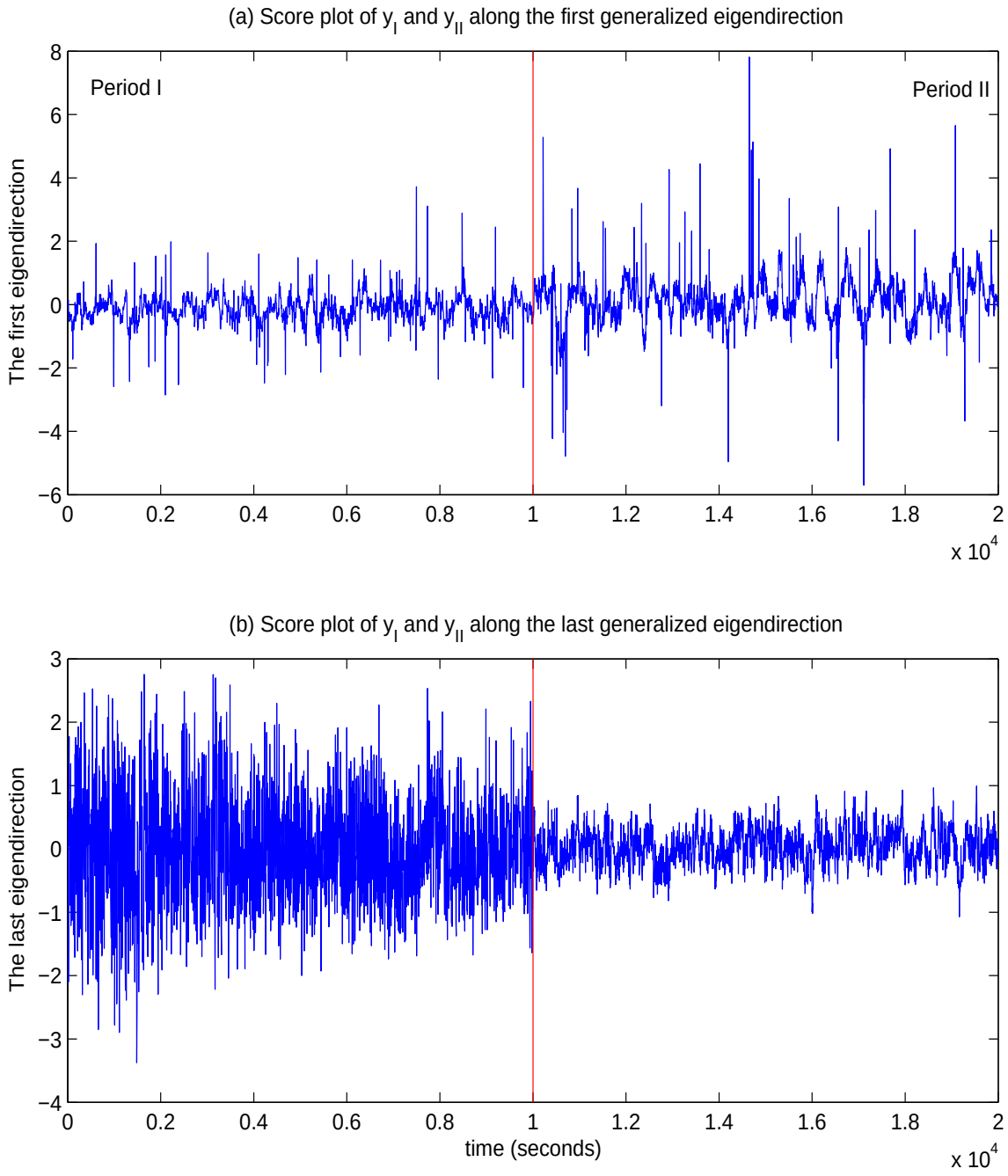

Figure 2.10: Industrial boiler example: Projected score plots of y_I and y_{II} along (a) the first generalized eigendirection and (b) the last generalized eigendirection

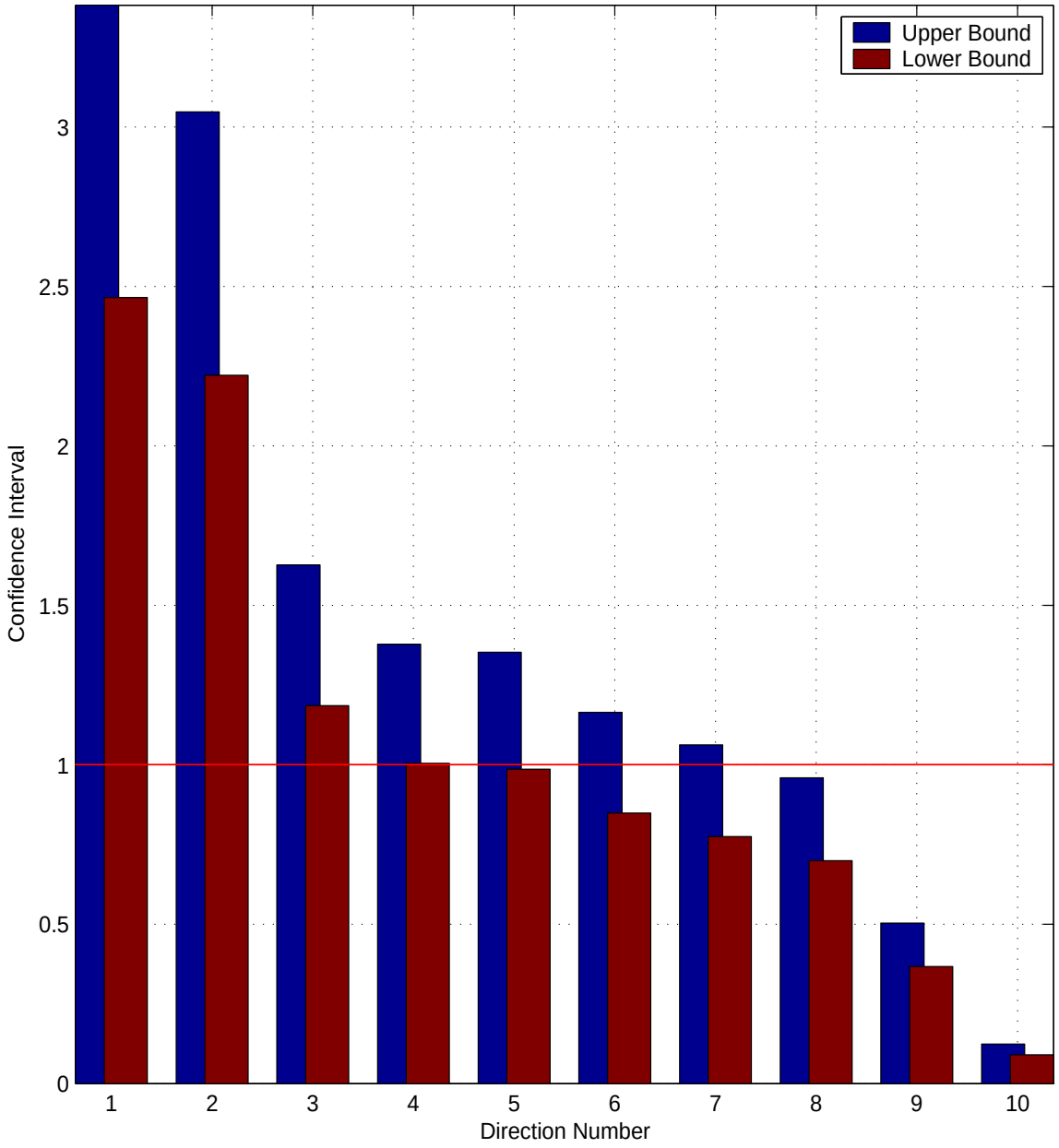

Figure 2.11: Industrial boiler example: The 95% confidence intervals for population eigenvalues

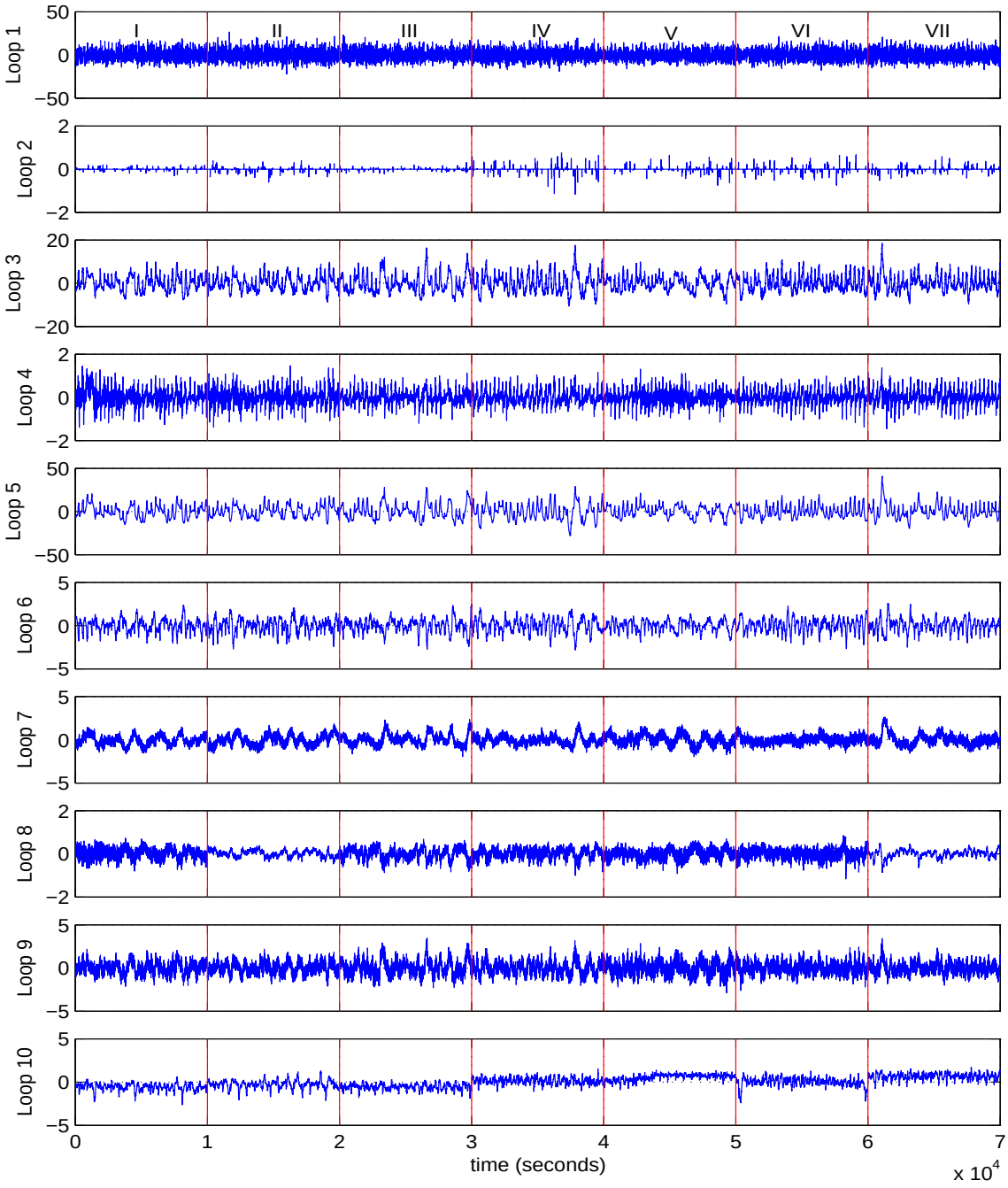

Figure 2.12: Industrial boiler example: Signal plots of controller error for the ten control loops in the periods I-VII

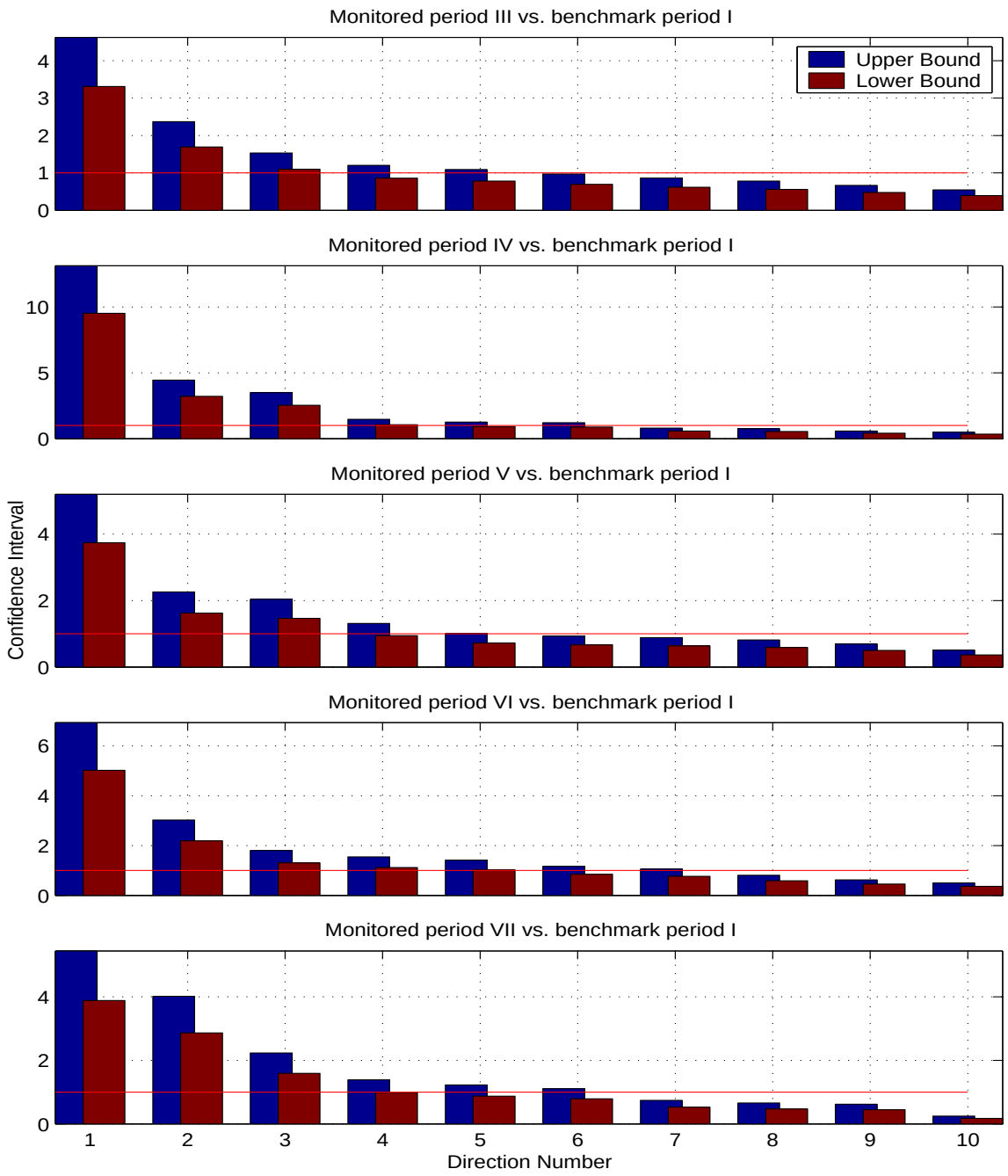

Figure 2.13: Industrial boiler example: The 95% confidence intervals of population eigenvalues for multi-period performance monitoring

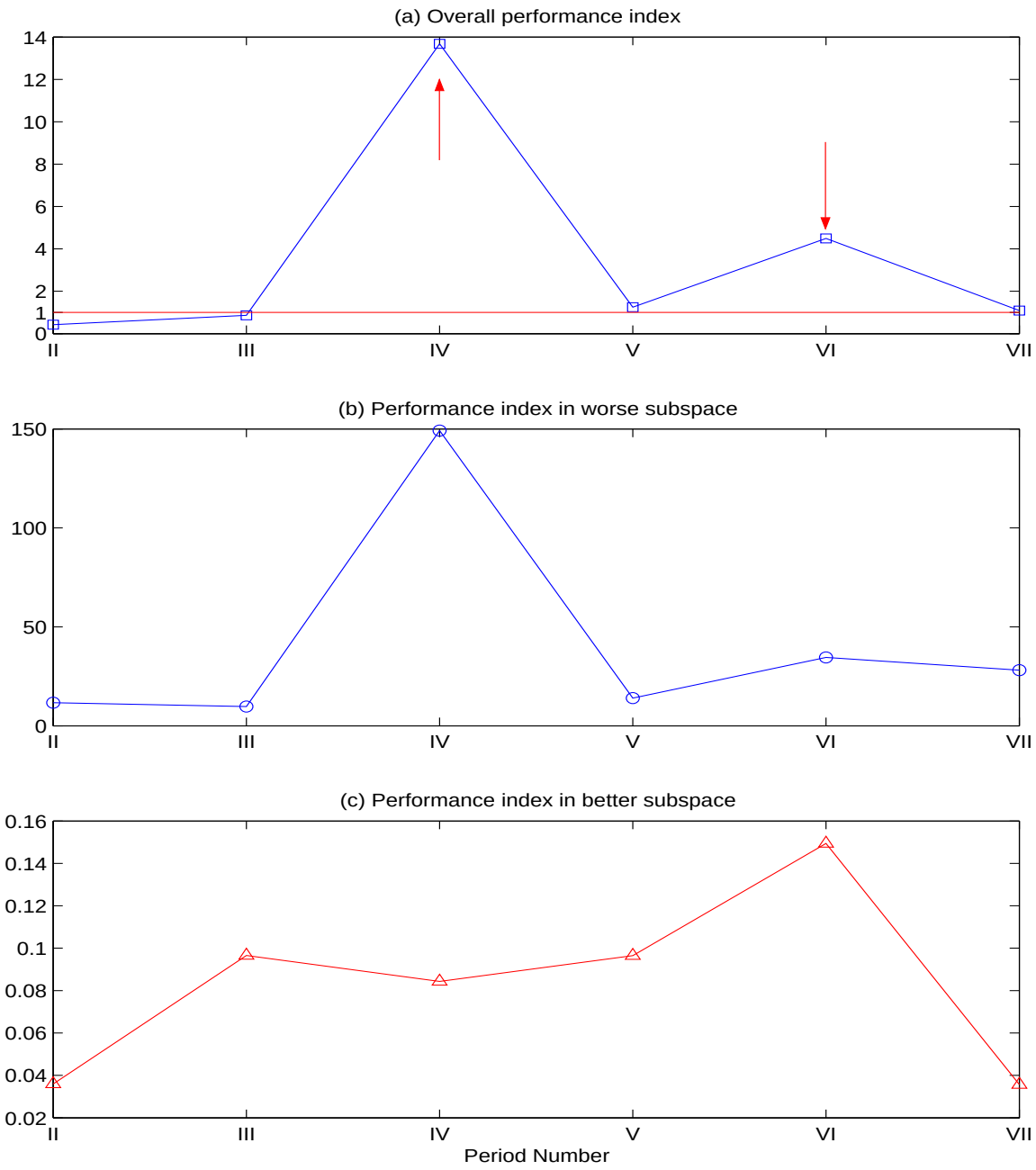

Figure 2.14: Industrial boiler example: Multi-period control performance assessment with (a) covariance-based overall performance index and isolated performance indices in (b) worse and (c) better subspaces

Chapter 3

Statistical MIMO Control Performance Monitoring - Performance Diagnosis

3.1 Introduction

Thus far, most research effort in CPM area has been focused on performance assessment. However, little work has been done in controller performance diagnosis. During performance assessment/monitoring, once a significant deterioration is detected in the system performance, a diagnosis step becomes necessary to isolate the driving sources and locate the controller problems. For MIMO interacting systems, the objective of performance diagnosis is to find out the control loops or controlled variables leading to the undesirable or degraded performance. With the identified loop/variable information, proper maintenance actions can be further taken to improve the performance by retuning the relevant loops/variables, modifying the control design or inspecting the related instruments. In MIMO performance diagnosis, a challenging issue lies in the fact that there exist complicated interactions among multiple loops or variables. Undesirable changes in a few control loops can cause performance degradations in many variables.

In contrast to the rare literature of control performance diagnosis, a

great deal of research has occurred and continues to grow for fault identification and diagnosis in process monitoring (Chiang *et al.*, 2001; Qin, 2003). Three types of methods, i.e., model based, data driven, and knowledge based, have been widely applied to fault detection and diagnosis in industrial processes. A popular category of tools for fault diagnosis is data driven multivariate statistical procedures, such as PCA and PLS. In addition, more recent work has attempted to address process monitoring issues through knowledge based approaches like Bayesian belief networks (BBN) (Mehranbod *et al.*, 2003; Mehranbod *et al.*, 2005), support vector machines (SVM) (Kulkarni *et al.*, 2005) and association rules (Lee *et al.*, 2006*b*).

In this work, we apply multivariate statistical techniques to the diagnosis of control performance change, similar to their use in process monitoring. Based on the data-driven benchmark, a statistical inference approach has been proposed and incorporated into covariance monitoring scheme for control performance assessment in Chapter 2. The eigendirections and subspaces with worse or better control performance in the monitored period against the benchmark period can be identified in this way. However, the directions are linear combinations of multiple control loops or controlled variables in the case of MIMO control systems. It is highly desirable to further develop a methodology to find out the root causes resulting in the system performance degradation or improvement. The diagnosis results could be in terms of individual control loops or controlled variables which are most relevant to the identified eigendirections and subspaces. The major task of this work is thus to iden-

tify and diagnose the degraded/improved control loops or controlled variables from the worse/better eigendirections or subspaces. Two types of multivariate contribution charts are proposed for control performance diagnosis. One is the loading based contribution chart which uses a bootstrap resampling technique (Martin and Morris, 1996; Davison and Hinkley, 1997; Politis, 1998). By the bootstrapping procedure, the probability distribution of the loading coefficients from each eigendirection can be estimated and the corresponding means and variances are thus obtained. The confidence intervals for the loading based contribution from each loop or variable on the worse/better eigendirections are then derived. As a result, the control loops or controlled variables responsible for the performance degradation or improvement can be determined. The other type of contribution is the cosine of the angle between each loop or variable and the worse/better performance subspace. Geometrically, it is tantamount to the projection of a unit vector to the worse/better subspace. The angle based contribution (ABC) is shown to be equivalent to the canonical correlation between each unit vector and the worse/better subspace. The confidence limit of the contributions is further deduced from the asymptotic distribution of the canonical correlations. Consequently, the degraded/improved control loops or controlled variables can be diagnosed from the contribution chart in the worse/better performance subspace. The results from these two types of contribution charts are also compared to demonstrate the validity of the methods. The flowchart of the proposed control performance diagnosis framework is exhibited in Fig. 3.1.

The rest of this chapter is organized as follows. The loading based contribution and its statistical inference is described in Section 3.2. The confidence intervals of the loadings are derived, and bootstrap resampling technique is employed to estimate the probability distributions and statistics of the loadings. Subsequently, the multiple contribution charts along the identified worse or better eigendirections are generated and the statistical decision rules are applied to identify the responsible control loops or controlled variables. In Section 3.3, the angle based contribution index is proposed and formulated. The geometric and statistical properties are presented, and the corresponding confidence limit is derived from the asymptotic statistics of canonical correlation. The angle based contribution charts are further developed and the statistical decision rules are established. Both methods are illustrated and compared in Section 3.4 using those two simulated examples given in Chapter 2. Different scenarios of control performance problems are simulated and tested. In Section 3.5, industrial data from the power boiler unit are utilized to further demonstrate the effectiveness of the methods. The high consistency in the performance diagnosis results between these two types of contribution charts is also highlighted, and the numerical features of two algorithms are discussed. The chapter ends with summary.

3.2 Loading based contribution for performance diagnosis

The population eigenvector, p, in Eq. (2.7) gives the loadings of each controlled variable in that eigendirection. Zero loadings indicate no contribution to the eigendirection. However, the population eigendirections are not known and only sample eigenvectors, $\hat{p}$, can be calculated from the sample covariances. If the distribution of $\hat{p}$ is known, confidence limits of the components of the loadings, $\hat{p}_j$, can be calculated and used to determine significant contributors to the eigendirection. To the author's knowledge, however, there is no theoretical result available for the distribution of the generalized eigenvectors.

In this work, bootstrap resampling method (Martin and Morris, 1996; Davison and Hinkley, 1997) is employed to estimate the probability distributions and the first and second-order statistics of the loadings. The basic idea of Bootstrapping consists of repeated random samples by replacement to generate a pool of data sets, which allow the estimation of the probability distribution of the random variables quantitatively. Let $X = \{x_{.,1}, x_{.,2}, \ldots, x_{.,m}\}$ be a set of measurements with vector $x_{.,i}$ denoting one observation on different variables. $x_{.,1}, x_{.,2}, \ldots, x_{.,m}$ are assumed to distribute independently and identically according to an unknown probability function $F(x)$. The nonparametric bootstrapping principle is to resample from the set of measurements X with equal probability (weight) on each measurement $x_{.,i}$. The distribution of the resampled data set would be close to $F(x)$ in statistical sense. Each

time an unordered collection $X^* = \{x^*_{.,1}, x^*_{.,2}, \ldots, x^*_{.,m}\}$ is produced with replacement using the rule $P(x^*_{.,i} = x_{.,j} \mid X) = 1/m$. Such procedure is repeated s times and a series of data sets $\{X^*_{(1)}, X^*_{(2)}, \ldots, X^*_{(s)}\}$ are generated and used to estimate the population distribution and statistics, e.g. mean and variance.

To apply the bootstrap resampling procedure to estimate the distribution of the loadings, the benchmark and monitored data sets Y_I and Y_{II} are resampled to generate a sequence of subsets $\{Y_I^{(1)}, Y_I^{(2)}, \ldots, Y_I^{(s)}\}$ and $\{Y_{II}^{(1)}, Y_{II}^{(2)}, \ldots, Y_{II}^{(s)}\}$, where s is the number of bootstrapping iterations and each pair of subsets $Y_I^{(i)}$ and $Y_{II}^{(i)}$ $(i = 1, 2, \ldots, s)$ are of the same structure as Y_I and Y_{II}, respectively. Then the generalized eigenvalue analysis is carried out sequentially between the covariance matrices of the pair $Y_I^{(i)}$ and $Y_{II}^{(i)}$, and a sequence of eigenvector loading matrices $p^{(1)}, p^{(2)}, \ldots, p^{(s)}$ can be calculated and used to estimate the statistics of loadings p. The bootstrap resampling procedure can be summarized as follows:

(i) From the benchmark data set Y_I, iterate s times and form a sequence of samples as $\{Y_I^{(1)}, Y_I^{(2)}, \ldots, Y_I^{(s)}\}$;

(ii) From the monitored data set Y_{II}, similarly iterate s times and generate a series of samples as $\{Y_{II}^{(1)}, Y_{II}^{(2)}, \ldots, Y_{II}^{(s)}\}$;

(iii) For each pair of resampled benchmark and monitored subsets $Y_I^{(i)}$ and $Y_{II}^{(i)}$ $(i = 1, 2, \ldots, s)$, conduct generalized eigenvalue analysis as

$$cov(Y_{II}^{(i)})p^{(i)} = l^{(i)} cov(Y_I^{(i)})p^{(i)}$$

to obtain a series of eigenvector loadings $p^{(1)}, p^{(2)}, \ldots, p^{(s)}$;

(iv) Approximate the probability distribution of loading p and estimate its statistics μ_p and σ_p. For the jth element of p, p_j, the mean and variance can be estimated from $p_j^{(1)}, p_j^{(2)}, \ldots, p_j^{(s)}$ $(j = 1, 2, \ldots, q)$ as follows:

$$\mu_{p_j} = \frac{1}{s} \sum_{i=1}^{s} p_j^{(i)}$$

and

$$\sigma_{p_j}^2 = \frac{1}{s-1} \sum_{i=1}^{s} \left(p_j^{(i)} - \mu_{p_j} \right)^2$$

Assuming the loadings are normally distributed, the confidence intervals of loadings can be written as

$$\begin{bmatrix} L(p_j) & U(p_j) \end{bmatrix} = \begin{bmatrix} p_j - z_{\alpha/2} \cdot \sigma_{p_j} & p_j + z_{\alpha/2} \cdot \sigma_{p_j} \end{bmatrix} \tag{3.1}$$

If both the upper and lower bounds have the same signs, it means the loading coefficient is statistically different from zero. Therefore, the corresponding loop/variable is of significant contribution to the worse or better performance eigendirection and is responsible for the performance change. On the contrary, if the upper and lower bounds have opposite signs, the loading contribution is statistically close to zero and the corresponding loop/variable can be regarded as a trivial contributor.

The statistical decision rules to diagnose the degraded or improved loops/variables are given as follows:

(i) For each of the identified worse performance eigendirections, if the confidence interval of loading of the jth loop/variable does not include zero, the corresponding jth loop/variable can be determined having significant contribution to the performance degradation;

(ii) For the better performance eigendirections, if the confidence interval of loading of the jth loop/variable does not include zero, the corresponding loop/variable is regarded as responsible for the performance improvement;

(iii) Loops or controlled variables belong to neither of the above categories are considered marginal.

The loading based contributions, similar to the T^2 based contributions in process monitoring (MacGregor *et al.*, 1994), produce multiple contribution charts and each corresponds to one eigendirection. Diagnosis based on these charts requires careful examination of the loadings in all significant eigendirections. This is sometimes inconvenient. In the next section, an angle based contribution approach is proposed which gives one contribution chart only and is simpler to use.

3.3 Angle based contribution for performance diagnosis

The angle based contribution is defined as the cosine of the angle between a variable and the worse/better performance subspace to quantify the contribution of each loop/variable within the subspace. If the ABC index is

close to one, it indicates that the angle approaches zero and the variable is virtually in the worse/better subspace. In this case, the corresponding control loop or controlled variable contributes significantly to the performance degradation or improvement. If ABC index is zero, on the other hand, the angle is 90 degrees and the variable is then perpendicular to the subspace. As a result, the control loop or controlled variable has no contribution to the worse/better subspace. As the ABC index is always between 0 and 1, the bigger it is, the more significant contribution the corresponding loop/variable has to the performance subspace. We can select, for example, a threshold value of the angle to be 45 degrees. An angle smaller than 45 degrees is considered more parallel than perpendicular, indicating that the variable contributes significantly to the performance degradation or improvement. On the contrary, an angle above 45 degrees implies weak correlation between the loop/variable and the performance subspace. In that situlation, the controlled variable contributes little to the performance change.

Without loss of generality, let $e_k = \left[\underbrace{0 \ldots 0}_{k-1} \ 1 \ 0 \ldots 0\right]^T$ be the kth unit vector and represent the kth control loop or controlled variable. Suppose that P spans the worse/better performance subspace, the cosine of the angle between the unit vector e_k and the subspace P (as depicted in Fig. 3.2) can be computed by

$$\cos(\theta_k) = \frac{\|\hat{e}_k\|}{\|e_k\|} = \|\hat{e}_k\| \tag{3.2}$$

where $\|\cdot\|$ denotes the norm of a vector and $\hat{e}_k$ is the projection of unit vector

e_k onto the worse/better subspace P, which is

$$\hat{e}_k = \Pi_P \cdot e_k = P(P^T P)^{-1} P^T e_k \tag{3.3}$$

where Π_P is the projection matrix to the column space of P. Since θ_k is always between 0 and 90 degrees, the angle based contribution index satisfies $0 \leq \cos(\theta_k) \leq 1$. Then Eq. (3.2) becomes

$$\begin{aligned} \cos(\theta_k) = & \quad \|P(P^T P)^{-1} P^T e_k\| \\ = & \quad \left[e_k^T P(P^T P)^{-1} P^T P(P^T P)^{-1} P^T e_k\right]^{\frac{1}{2}} \\ = & \quad \left[e_k^T P(P^T P)^{-1} P^T e_k\right]^{\frac{1}{2}} \\ = & \quad \left[e_k^T \Pi_P e_k\right]^{\frac{1}{2}} \end{aligned} \tag{3.4}$$

As discussed in Chapter 2, the generalized eigenvectors composing the subspace P are not orthogonal to one another in most cases. If P is transformed into a group of orthonormal basis $\tilde{P}$, i.e., $\tilde{P}^T \tilde{P} = I$, the computation of the contribution index can be simplified to

$$\begin{aligned} \cos(\theta_k) = & \quad \left[e_k^T \tilde{P}(\tilde{P}^T \tilde{P})^{-1} \tilde{P}^T e_k\right]^{\frac{1}{2}} \\ = & \quad \left(e_k^T \tilde{P} \tilde{P}^T e_k\right)^{\frac{1}{2}} \\ = & \quad \|\tilde{P}^T e_k\| = \left(\sum_{j=1}^{l} \tilde{p}_{kj}^2\right)^{\frac{1}{2}} = \|\tilde{p}_{k\cdot}\| \end{aligned} \tag{3.5}$$

where l is the dimension of the worse/better performance subspace P and $\tilde{p}_{k\cdot}$ is the kth row vector of $\tilde{P}$. The above equation reveals that the ABC index is entirely determined by the loadings.

Since $\tilde{P}$ is the orthogonal matrix, Eq. (3.5) can be rewritten as

$$\cos(\theta_k) = \left\| (\tilde{P}^T \tilde{P})^{-\frac{1}{2}} (\tilde{P}^T e_k)(e_k^T e_k)^{-\frac{1}{2}} \right\| \tag{3.6}$$

The above equation reveals that $cos(\theta_k)$ is the canonical correlation coefficient between $\tilde{P}$ and e_k (Anderson, 2003). Consequently, θ_k is the canonical angle between the unit vector e_k and the worse/better subspace P.

Since P and $cos(\theta_k)$ are calculated from sample covariances, the confidence limit for $cos(\theta_k)$ should be derived to help determine whether the contribution is statistically significant. Let r_k and ρ_k $(k = 1, 2, \ldots, q)$ be the sample and population canonical correlations between the kth loop/variable and the worse/better performance subspace, respectively. Further, suppose that the K non-zero sample canonical correlations are distinct and that all the remained correlations are zero. The following theorem then holds.

Theorem 3.3.1. *Let the random variable z_k be*

$$z_k = \sqrt{n} \frac{r_k^2 - \rho_k^2}{2\rho_k(1 - \rho_k^2)}, \qquad k = 1, 2, \ldots, K \tag{3.7}$$

where n is the number of samples. Then $z_1, z_2, \ldots, z_K$ are mutually independent and z_k has the limiting normal distribution of $N(0, 1)$ for $k = 1, 2, \ldots, K$.

Proof. See reference (Anderson, 2003).

The above theorem associates the sample and population canonical correlations. Under the $(1-\alpha) \times 100\%$ confidence level, the confidence interval for the random variable $\frac{r_k^2 - \rho_k^2}{2\rho_k(1-\rho_k^2)}$ is given as

$$\left[-\frac{z_{\alpha/2}}{\sqrt{n}} \quad \frac{z_{\alpha/2}}{\sqrt{n}} \right] \tag{3.8}$$

Using the threshold value of the population correlation as $\rho_k = \cos 45^\circ = \frac{1}{\sqrt{2}}$, the confidence interval of the sample correlation r_k is

$$\sqrt{\frac{1}{2} \pm \frac{z_{\alpha/2}}{\sqrt{2n}}} \tag{3.9}$$

Therefore, if $r_k > \epsilon_r = \sqrt{\frac{1}{2} + \frac{z_{\alpha/2}}{\sqrt{2n}}}$, the kth loop/variable is considered to contribute significantly to the worse or better performance. Otherwise, the contribution from that loop/variable is insignificant. If the benchmark data and the monitored data have different numbers of samples, the geometric average is used for n. The statistical decision rule to determine the degraded or improved loops/variables based on ABC index can be summarized below:

(i) Relative to the worse performance subspace P_w, if the ABC index $\cos\theta_k^w > \epsilon_r$, the corresponding loop/variable can be determined as a contributor to the worse subspace.

(ii) Regarding the better performance subspace P_b, if the ABC index $\cos\theta_k^b > \epsilon_r$, the corresponding loop/variable is regarded as a significant contributor to the performance improvement;

(iii) Variables belong to neither of the above two categories are marginal loop/variable with little contribution to the performance change.

3.4 Simulated examples

3.4.1 Multiloop example

We reuse the 4×4 multiloop process investigated in Chapter 2. The open-loop process transfer function matrix $G(q^{-1})$ and disturbance transfer function matrix $N(q^{-1})$ are given below

$$G(q^{-1}) = \begin{bmatrix} \frac{0.05q^{-3}}{1-0.95q^{-1}} & 0 & \frac{0.7q^{-3}}{1-0.3q^{-1}} & 0 \\ \frac{0.02966q^{-3}}{1-1.627q^{-1}+0.706q^{-2}} & \frac{0.0627q^{-6}}{1-0.937q^{-1}} & 0 & 0 \\ 0 & \frac{0.235q^{-5}}{1-0.765q^{-1}} & \frac{0.5q^{-2}}{1-q^{-1}+0.25q^{-2}} & 0 \\ \frac{0.5q^{-5}-0.4875q^{-6}}{1-1.395q^{-1}+0.455q^{-2}} & 0 & 0 & \frac{0.2q^{-6}}{1-0.8q^{-1}} \end{bmatrix} \tag{3.10}$$

and

$$N(q^{-1}) = \begin{bmatrix} \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 & 0 & 0 \\ 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 & 0 \\ 0 & 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} & 0 \\ 0 & 0 & 0 & \frac{1-0.1875q^{-1}}{1-0.9875q^{-1}} \end{bmatrix} \tag{3.11}$$

The disturbance $a(t)$ in this example is normally distributed white noise sequence. The multiloop controllers for the process are four PI controllers with the proportional gain $K_c = [3.07 \quad 0.625 \quad 0.518 \quad 0.370]$ and integral time constant $T_r = [20 \quad 16 \quad 2.86 \quad 5]$ (McNabb and Qin, 2005*a*). In the benchmark period, the process gains of Loop 1 and Loop 3 are changed to 0.816 and 0.184, respectively. In the monitored period, the controller gains of Loop 2 and Loop 4 are adjusted to 0.316 and 0.127, respectively. These changes are done to simulate both the improved and degraded performance directions. In the simulation, $3,000$ samples are generated in the benchmark and monitored periods, respectively. The closed-loop process output in the benchmark period I and

the monitored period II is shown in Chapter 2, where we have identified the first two eigendirections as the worse performance directions and the last two eigendirections as the better directions. Therefore, the control loops responsible for degraded performance can be diagnosed using contribution charts along the first two eigendirections. The last two eigenvectors, on the other hand, are used to identify control loops with improved performance.

The loading based contribution charts are first implemented to diagnose the degraded and improved control loops. After 500 iterations of bootstrap resampling, the probability distribution of the loadings can be estimated. For instance, the bootstrap resampling distribution of the loading coefficient of Loop 1 along the first eigendirection is plotted in Fig. 3.3. It is observed that the loading coefficients follow an approximately normal distribution. The mean and variance of the corresponding loading coefficient can be estimated from the bootstrapping samples. The other loading coefficients are also found to follow the approximately normal distribution with different statistics. The loading based contribution charts under 95% confidence level along the worse and better performance eigendirections are drawn in Fig.3.4. Along the first and second eigendirections, the lower confidence bounds of Loop 2 and Loop 4 are larger than zero, respectively. These two loops are thus inferred to have significant contribution on the worse performance in the monitored period over the benchmark period. The upper confidence limits of Loop 1 and Loop 3, on the contrary, are below zero along the last two eigendirections. Therefore, these two loops have improved performance relative to the benchmark. The

diagnosis results agree with the simulated controller changes in this example.

In comparison, the ABC charts within the worse and better performance subspaces are depicted in Fig. 3.5. From the ABC chart within the worse subspace, it can be seen that the contribution index values of Loop 2 and Loop 4 both exceed the 95% control limit. Therefore, these two loops contribute significantly to the worse performance. In the contribution chart for the better subspace, on the other hand, the contributions of Loop 1 and Loop 3 go beyond the control limit. These two loops are thus contributors to the better performance subspace and determined as improved loops.

These two types of contribution charts lead to the same performance diagnosis results on the simulated multiloop example. The diagnosis findings are in agreement with the control tuning design in the simulation. Loop 1 and Loop 3 are detuned in the benchmark period, while Loop 2 and Loop 4 have decreased controller gains during the monitored period. This accounts for the performance degradation in Loop 2 and Loop 4 and performance improvement Loop 1 and Loop 3 in the monitored period against the benchmark. The multiloop control example demonstrates that the proposed multivariate contribution methods can diagnose the contributing loops effectively.

3.4.2 MPC example

To examine the utility of the proposed methods on multivariable control systems, the 2×2 MPC example used in Chapter 2 is reconsidered. The process and disturbance models from an industrial fractionation column (Harris *et*

al., 1996) are given by

$$G(q^{-1}) = \begin{bmatrix} \frac{-1.5q^{-3}}{1-0.659q^{-1}} & \frac{-0.167q^{-1}}{1-0.923q^{-1}} \\ \frac{-0.519q^{-5}}{1-0.784q^{-1}} & \frac{0.154q^{-3}+0.144q^{-4}}{1-0.874q^{-1}} \end{bmatrix} \tag{3.12}$$

and

$$N(q^{-1}) = \begin{bmatrix} \frac{1-0.5q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} & \frac{-0.2q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} \\ \frac{-0.5q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} & \frac{1-0.8q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} \end{bmatrix} \tag{3.13}$$

The constrained MPC design (Ko and Edgar, 2001*a*) is employed for the 2×2 process. In this example, we simulate the effect of plant/model mismatch that is one of the most possible reasons for MPC system performance deterioration. In the benchmark period I, no model mismatch is present and $1,000$ samples are produced. In the monitored period II, a plant/model mismatch is implemented by changing the process gain of G_{12} 20% from 0.167 to 0.200. The other $1,000$ samples are then collected in the monitored period II. The simulated closed-loop data from the constrained MPC system have been plotted in Chapter 2.

For the control performance monitoring purpose, only the operation data of controlled variables are taken into analysis. In Chapter 2, the statistical inference strategy is applied to the two-dimensional CV data and the first eigendirection is identified as the worse performance direction. The second eigendirection, however, is determined as the marginal direction with similar control performance achieved in the benchmark and monitored periods. Thereby only the first eigendirection, i.e. the worse direction, is needed to diagnose the degraded CV performance. The loading based contribution chart

along the worse eigendirection is shown in Fig. 3.6, where the number of iterations in bootstrap resampling procedure is set as 200. It is obvious that the upper confidence bound of CV1 is less than zero. Therefore, CV1 is determined as degraded variable with worse control performance. CV2, on the other hand, is a negligible contributor because its confidence bounds include zero. As a comparison, the angle based contribution chart for the worse performance subspace is plotted in Fig. 3.7. The contribution of CV1 is over the control limit while the contribution of CV2 is below the control limit. Therefore, CV1 is the variable of worse performance while CV2 is not. This conclusion is exactly the same as that from the loading based contribution charts. Furthermore, the fact that the plant/model mismatch takes place in the transfer function matrix entry G_{12} in the simulation implies that CV1 should be of deteriorated performance while the performance of CV2 is virtually unchanged. Therefore, these two types of contribution charts provide the correct performance diagnosis results for the plant/model mismatch effect of the MPC system.

The other plant/model mismatch case of this MPC system is simulated with the process gain of G_{22} increased by 30% during the monitored period II. $1,000$ samples are collected from the benchmark and monitored periods, respectively. The closed-loop data of CVs and MVs are plotted in Fig. 3.8. Likewise, the first eigendirection is statistically identified as the worse direction while the second eigendirection remains as the marginal direction. In this case, the plant/model mismatch happens in G_{22} and thus CV2 is inferred to be of degraded performance instead. As shown in Figs. 3.9 and 3.10, the lower

confidence bound of loading contribution for CV2 is larger than zero and its angle based contribution is also above the control limit line. Therefore, both loading and angle based contribution charts point to CV2 as the significant contributor to performance degradation in the monitored period against the benchmark in this case.

3.5 Industrial boiler application

In the industrial example, the power boiler system has overall ten control loops under closed-loop operation. The boiler process and instrumentation diagram and the loop description are given in Chapter 2.

The closed-loop process data in the user-defined benchmark period I and the monitored period II are taken into analysis. The sampling time of the control system is five seconds. Each period includes $2,000$ samples with process variables (PV) and their corresponding setpoints (SP). The controller errors, i.e. PV-SP, of the ten loops during the benchmark and monitored periods are shown in Chapter 2, where it is demonstrated that the first four eigendirections are statistically determined as the worse performance directions and the last three eigendirections are the better directions. Accordingly, the first four eigendirections span the worse performance subspace, while the last three eigendirections constitute the better subspace with improved control performance.

To implement loading based contribution charts, the first four as well as the last three eigenvectors are examined sequentially to find out the de-

graded or improved control loops. The bootstrap resampling procedure is iterated with 500 iterations to yield the approximate probability distribution of eigenvector loadings. As an illustration, the bootstrapping distributions of the loadings for Loop 1 along the first and the eighth eigendirections are plotted in Fig. 3.11. The histograms appear to be normally distributed for both loading coefficients along the worse and better directions. With the estimated loading statistics, the confidence intervals of the loading contributions can be further evaluated. The multiple contribution charts along the first four and the last three eigendirections under 95% confidence level are drawn in Figs. 3.12 and 3.13, respectively. Loop 6 is first diagnosed as the degraded loop because its upper confidence bound is less than zero along the $1st$ and $2nd$ eigendirections. Similarly, Loop 1, Loop 2 and Loop 7 are further determined as degraded loops responsible for the worse performance along the subsequent eigendirections. On the other hand, Loop 5, Loop 8 and Loop 9 contribute significantly to the last three eigendirections and are thus inferred as improved loops leading to the better control performance. The remained two loops, i.e. Loop 3, Loop 4 and Loop 10, are concluded as the marginal loops that contribute insignificantly to either worse or better performance subspace. It needs to be noted that a large magnitude of loadings does not necessarily mean the statistical significance of the corresponding loops and the confidence limits are required to make reliable conclusions.

As shown in Fig. 3.14, the ABC charts within the worse and better subspaces offer an alternative way to diagnose the control performance. It is

obvious that the contributions from Loop 1, Loop 2, Loop 6 and Loop 7 exceed the control limit line in Fig. 3.14. Hence these four loops are determined as degraded loops. On the other hand, the contributions of Loop 5, Loop 8 and Loop 9 go beyond the control limit line within the better subspace in Fig. 3.14. These three loops are thus inferred as the improved loops. The rest three loops, i.e. Loop 3, Loop 4 and Loop 10, are classified as marginal loops without significant performance change. Another interesting feature to be noticed is that the descending order of loop contributions relative to the worse subspace is Loop 2, Loop 1, Loop 10, Loop 6, Loop 7, Loop 3, Loop 5, Loop 4, Loop 9 and Loop 8, while the ascending order of loop contributions to the better subspace is Loop 2, Loop 1, Loop 10, Loop 6, Loop 4, Loop 7, Loop 3, Loop 5, Loop 9 and Loop 8. The ascending order of the loops within the better subspace is virtually the same as their descending order in the worse subspace and the slight difference lies in the position of Loop 4. The high symmetry in loop orders results from the fact that the worse and the better performance subspaces are nearly complementary with each other. The exception in Loop 4 is due to its marginal performance change.

The loop performance diagnosis results from the two types of contribution charts are summarized and compared in Table 3.1. In both types of contribution charts, the diagnosis results are exactly the same. The consistency in the diagnosis results demonstrates the validity of these two approaches in the control performance diagnosis. However, ABC charts involve a single chart for each subspace, which is more convenient to use. The other advantage

of ABC chart over loading based contribution chart lies in their numerical algorithm properties, where the latter needs bootstrap resampling. Therefore, the ABC method is recommended due to its simplicity and computational efficiency, although both methods give the same diagnosis results.

The power of the proposed data driven method is that it is applicable to both multiloop and multivariable controllers. Correlations induced by interactions and parallel CVs can be considered in the covariance based method. For this application with only several multiloop controllers, it is possible to assess the performance by univariate variance ratio. In this study, the output variance ratio of each loop between the benchmark and the monitored periods is calculated. For the degraded loops, the variance ratios vary from 2.6487 to 1.1351, which are all larger than one. The variance ratios for the improved loops are 0.8801, 0.3818 and 0.8569, respectively. These results agree with the diagnosis results using the proposed methods. However, the variance ratios of the marginal loops, i.e. Loop 3, Loop 4 and Loop 10, are 0.9386, 0.8587 and 1.3139 which may also be larger or smaller than one. The variance ratio index relies on the single loop output only and ignores the interactions among different loops completely. In contrast, the proposed MIMO performance diagnosis methods offer a more complete picture including both variance and covariance information.

The boiler process data in Chapter 2 include additional monitored periods III through VII, with Period VI having the worst overall performance relative to the benchmark Period I. Therefore, it is of interest to perform diag-

nosis of contributing loops in this period. The two types of contribution charts are generated following the worse/better eigendirections identified in Chapter 2. From the loading based contribution charts given in Figs. 3.15 and 3.16, it can be observed that Loop 2, Loop 3, Loop 5 and Loop 9 are degraded loops along the four worse eigendirections, while Loop 4, Loop 8 and Loop 10 are improved loops with significant contribution to the four better eigendirections. As a comparison, the ABC charts in the worse and better performance subspaces are shown in Fig. 3.17. It is obvious that Loop 2, Loop 3, Loop 5 and Loop 9 are over the 95% control limit line in the worse subspace and are thus detected as degraded loops. On the other hand, Loop 4, Loop 8 and Loop 10 contribute significantly in the better subspace and are determined as improved loops. Again the diagnosis results from two types of contribution charts are the same.

3.6 Summary

In this chapter, following the data-driven benchmark and covariance monitoring scheme presented in Chapter 2, two types of multivariate contribution methods are developed for control performance diagnosis. One is loading based contribution chart that utilizes the loading information, and the other is angle based contribution chart. Confidence limits are successfully developed for both types of contributions. The loading based contribution method requires bootstrapping to calculate the confidence limits and produce multiple contribution charts. The angle based contribution approach, however, is sim-

pler to use with a single chart and the confidence limit is derived from the asymptotic statistics of canonical correlation. The angle based contribution chart is preferred because a single chart is more convenient for visualization. However, for the simulated multiloop and MIMO MPC controllers, and an industrial boiler case study, the diagnosis results from both methods are highly consistent.

The diagnosis results from both types of contribution methods match well the simulation designs in simulated examples, indicating correct performance diagnosis findings. The diagnosis results from the industrial example agree with a simple variance ratio calculation, although the latter is unable to capture covariance and interaction information. These diagnosis approaches can be incorporated into the data-driven performance monitoring scheme presented in Chapter 2 to identify contributors to covariance changes. The proposed diagnosis methods are by no means final solutions for exact root cause diagnosis. Variance/covariance changes can be attributed to process dynamics changes, disturbance dynamics changes, changes in active constraint sets, and possibly abnormal changes in sensors, actuators or process equipments. Further study should focus on the integration of multiple methods to deal with these possible causes.

Table 3.1: Industrial example: Comparison of the loop diagnosis results from both contribution charts with single loop variance ratios for monitored Period II versus benchmark Period I

Loop ID	Variance Ratio	Decision from Loading	Decision from Angle
Loop 1	1.1351	Degraded	Degraded
Loop 2	2.6487	Degraded	Degraded
Loop 3	0.9386	Marginal	Marginal
Loop 4	0.8587	Marginal	Marginal
Loop 5	0.8801	Improved	Improved
Loop 6	1.2315	Degraded	Degraded
Loop 7	1.1951	Degraded	Degraded
Loop 8	0.3818	Improved	Improved
Loop 9	0.8569	Improved	Improved
Loop 10	1.3139	Marginal	Marginal

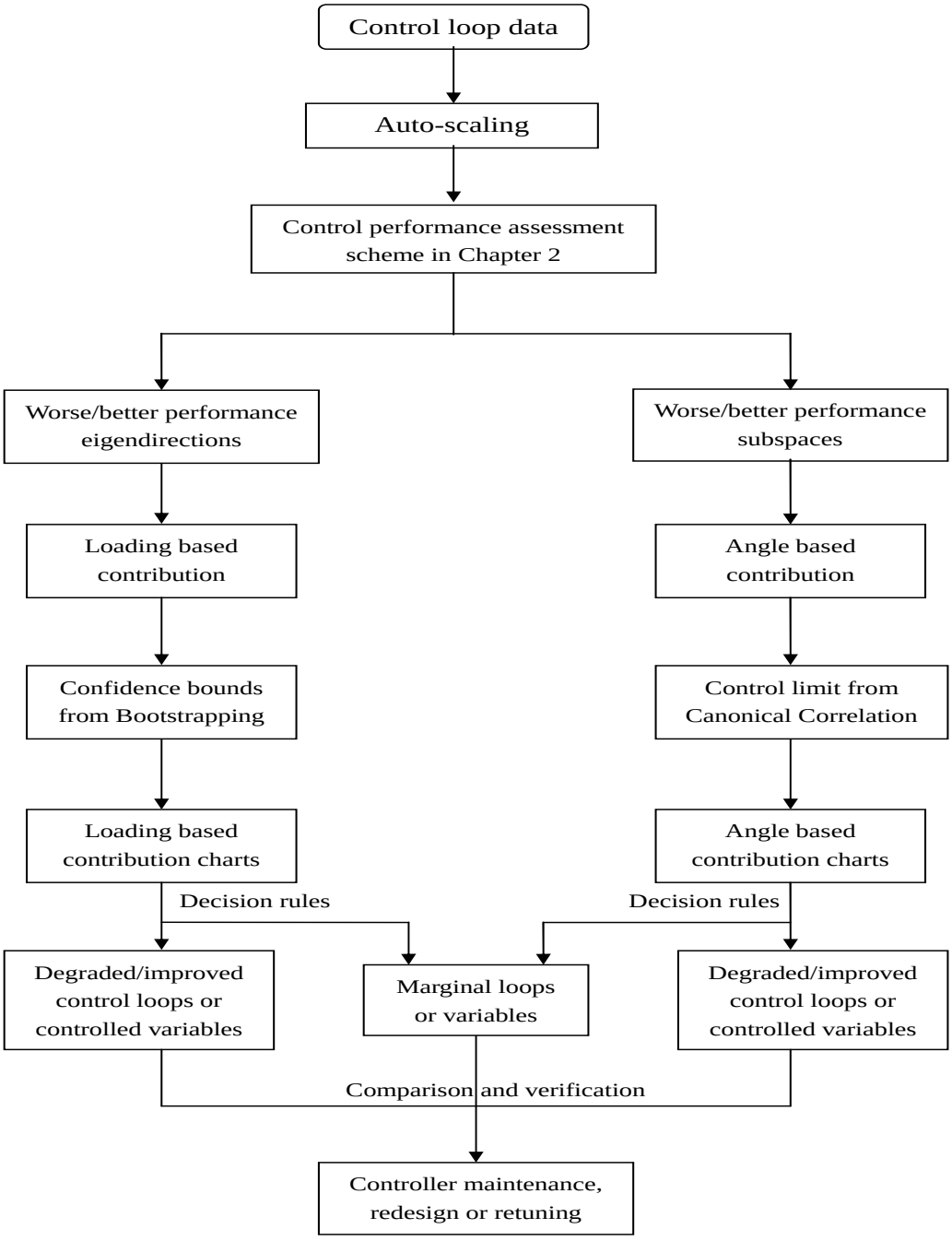

Figure 3.1: Schematic diagram of the control performance diagnosis procedures

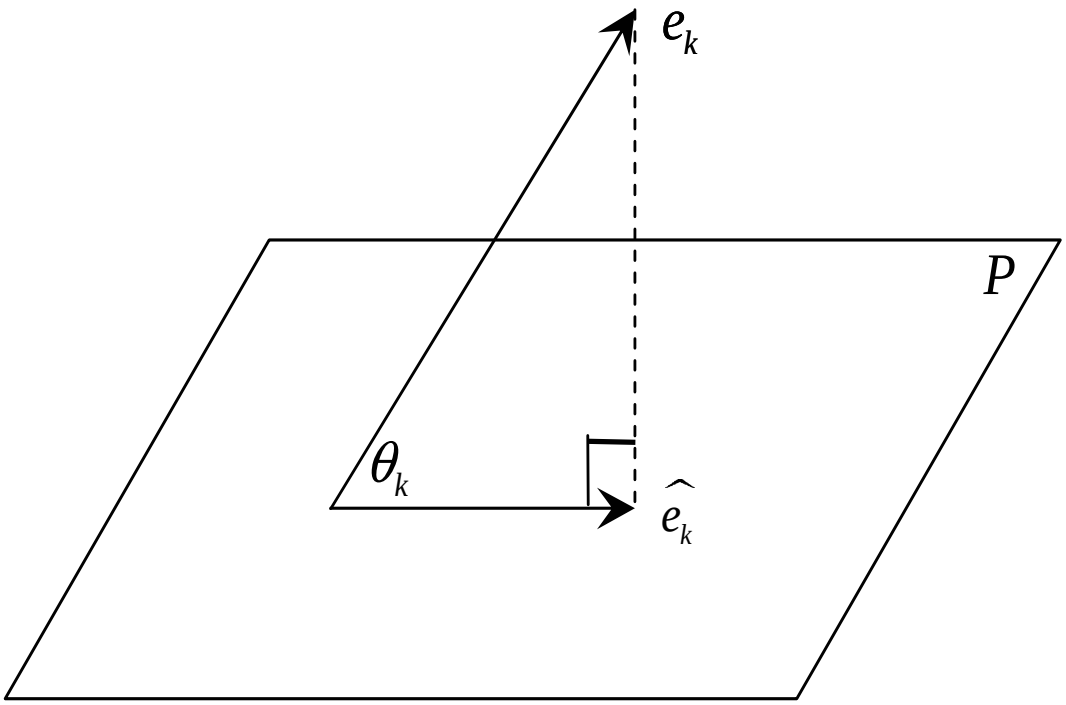

Figure 3.2: Graphical illustration of the angle based contribution index

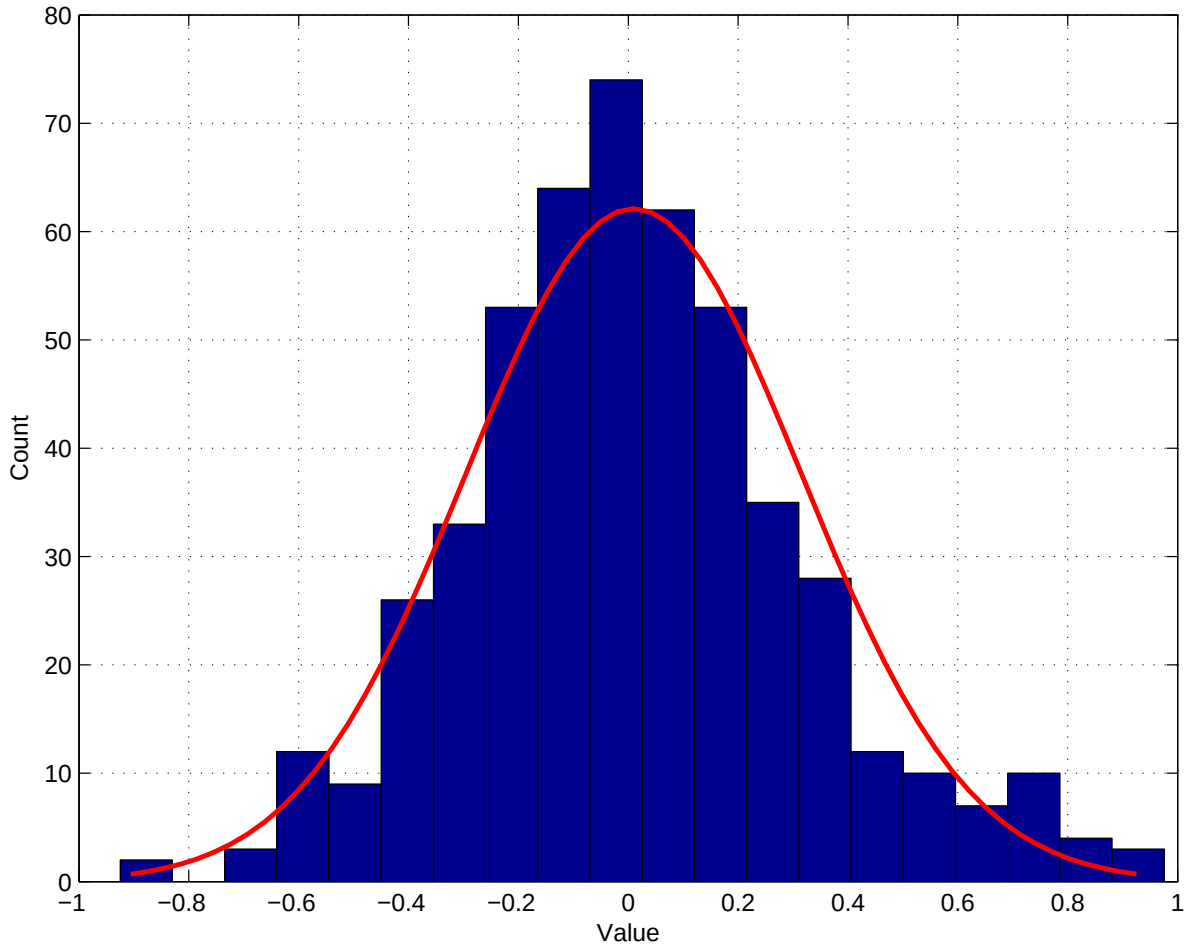

Figure 3.3: Simulated multiloop example: Bootstrapping probability distribution of the loading coefficient for Loop 1 along the first eigendirection

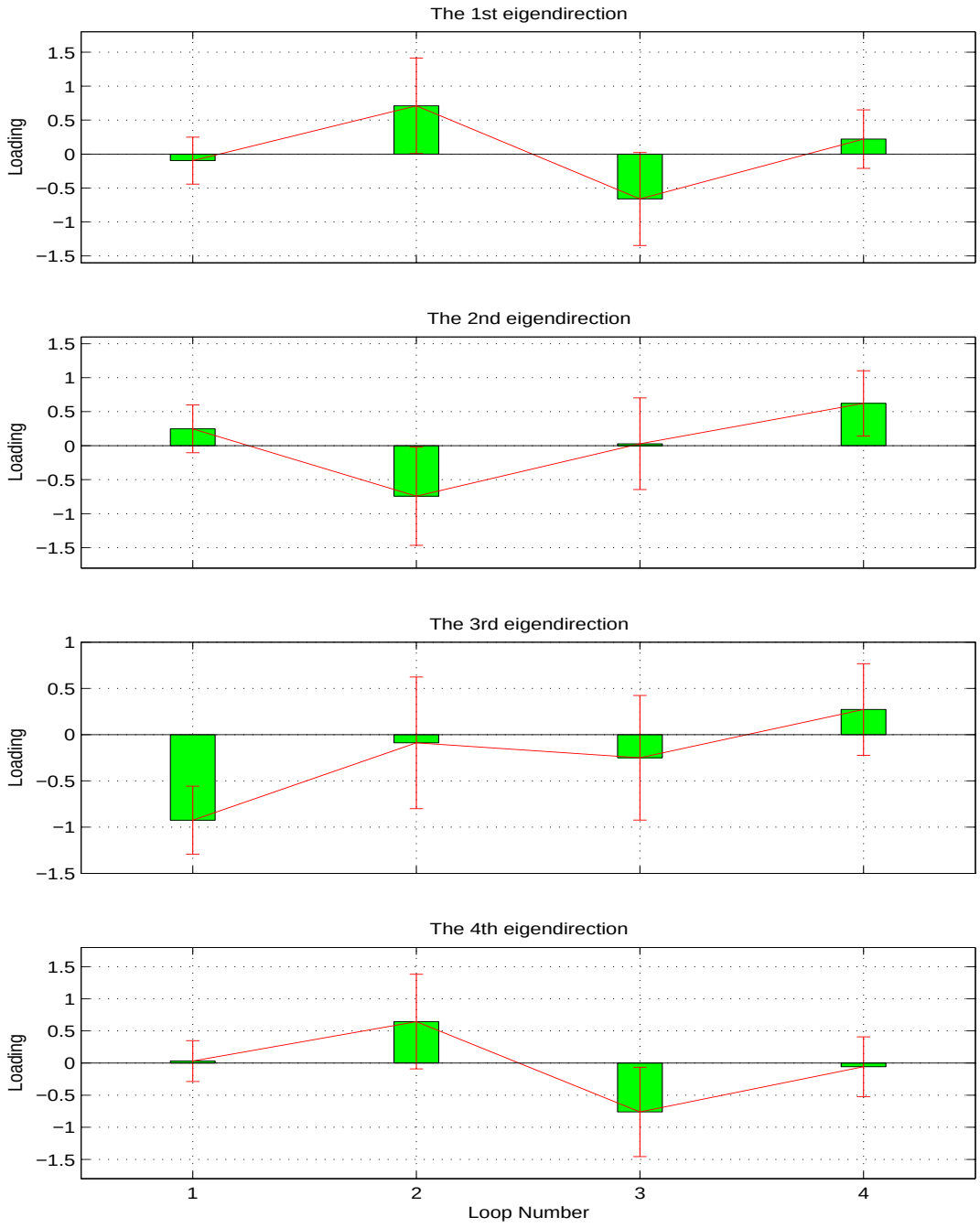

Figure 3.4: Simulated multiloop example: Loading based contribution charts with 95% confidence limits in the worse and better performance eigendirections

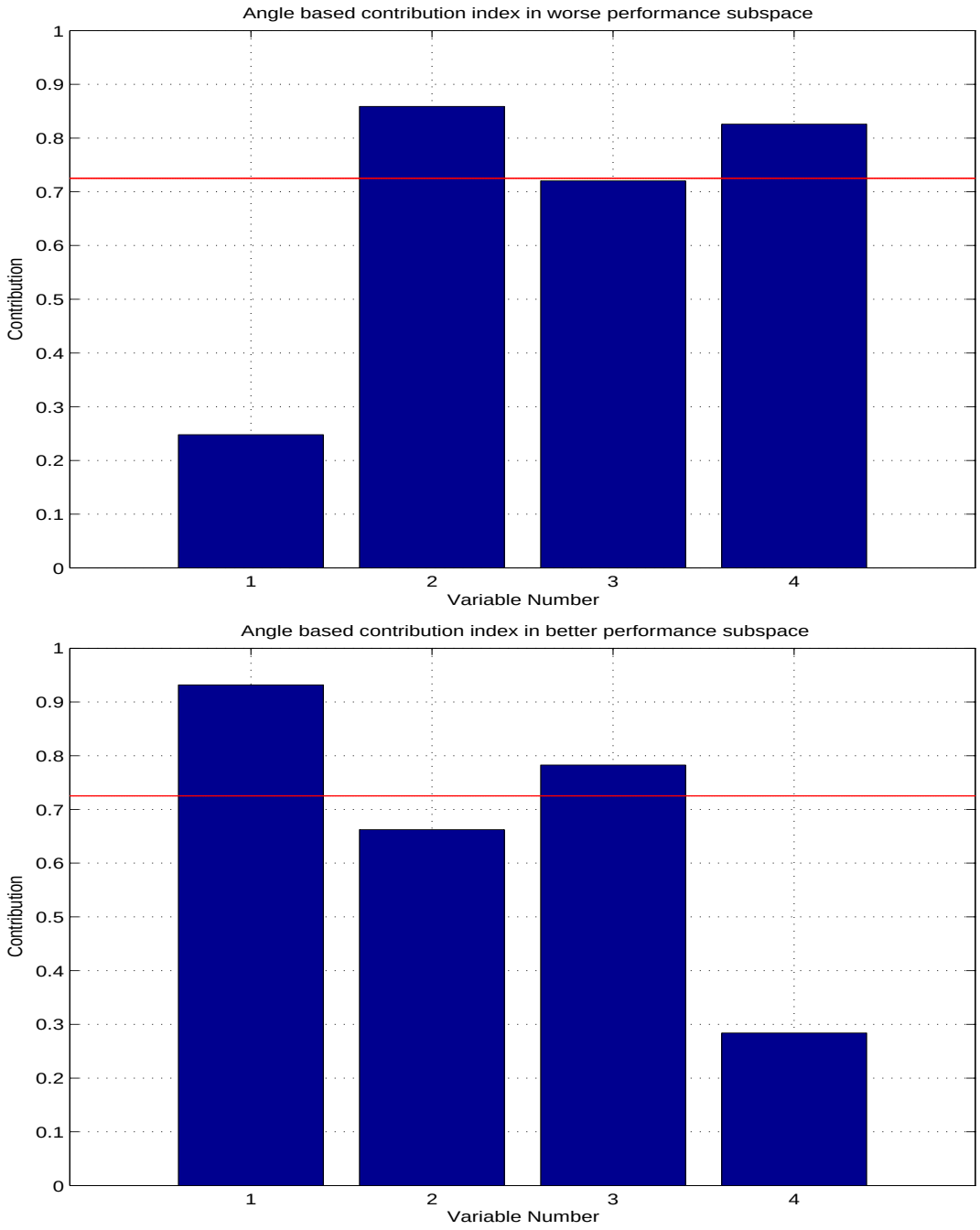

Figure 3.5: Simulated multiloop example: ABC charts with 95% confidence limits in the worse and better performance subspaces

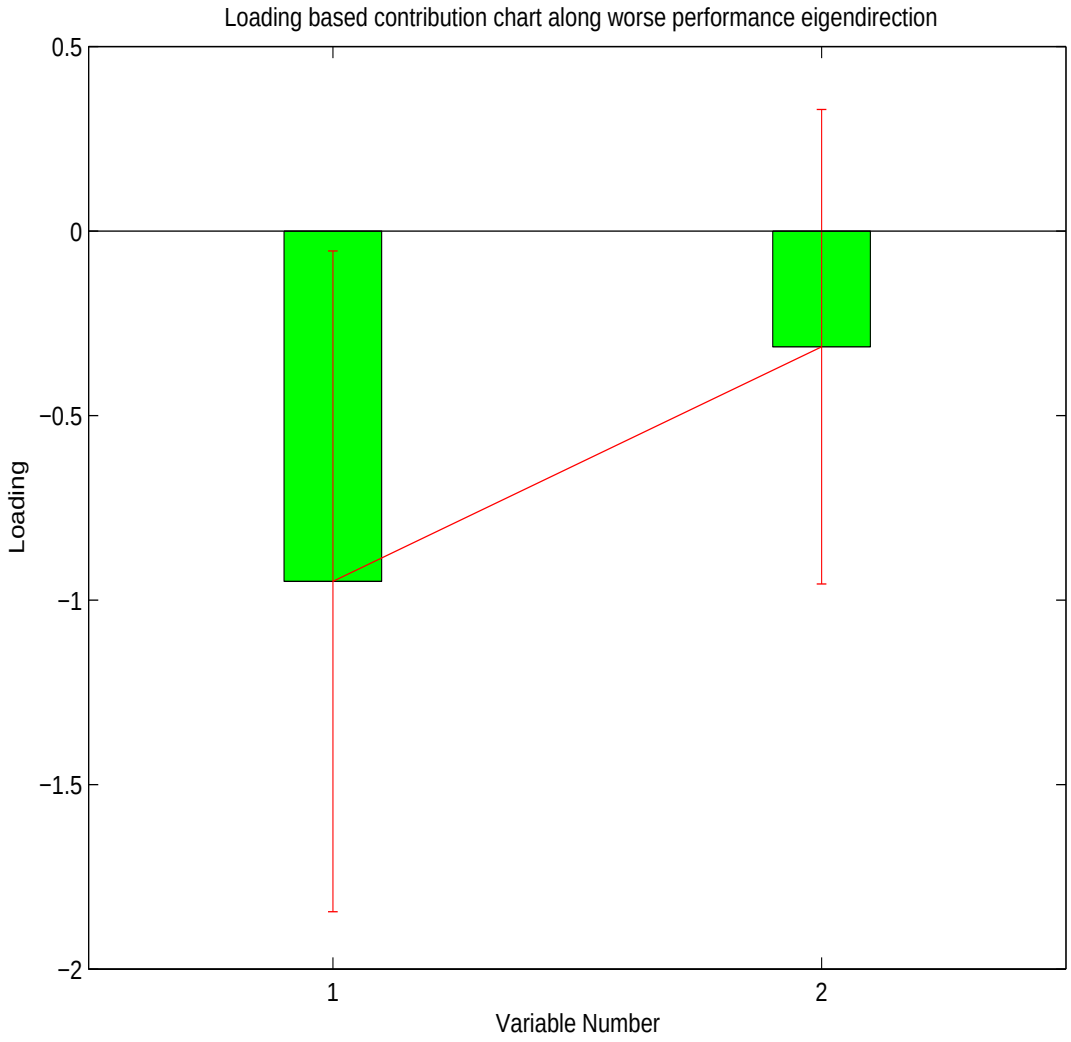

Figure 3.6: Case 1 of simulated MPC example: Loading based contribution chart with 95% confidence limits in the worse performance eigendirection

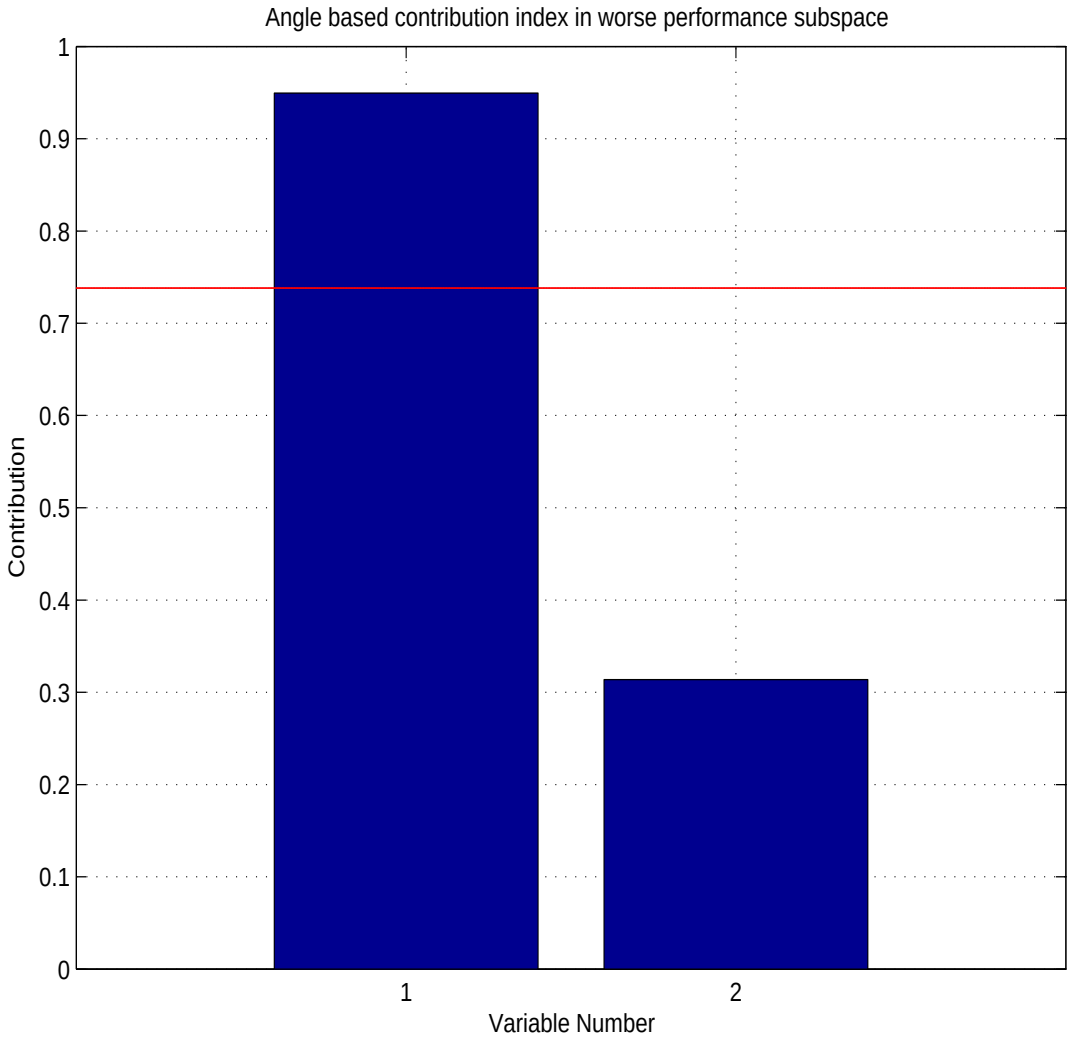

Figure 3.7: Case 1 of simulated MPC example: ABC chart with 95% confidence limit in the worse performance subspace

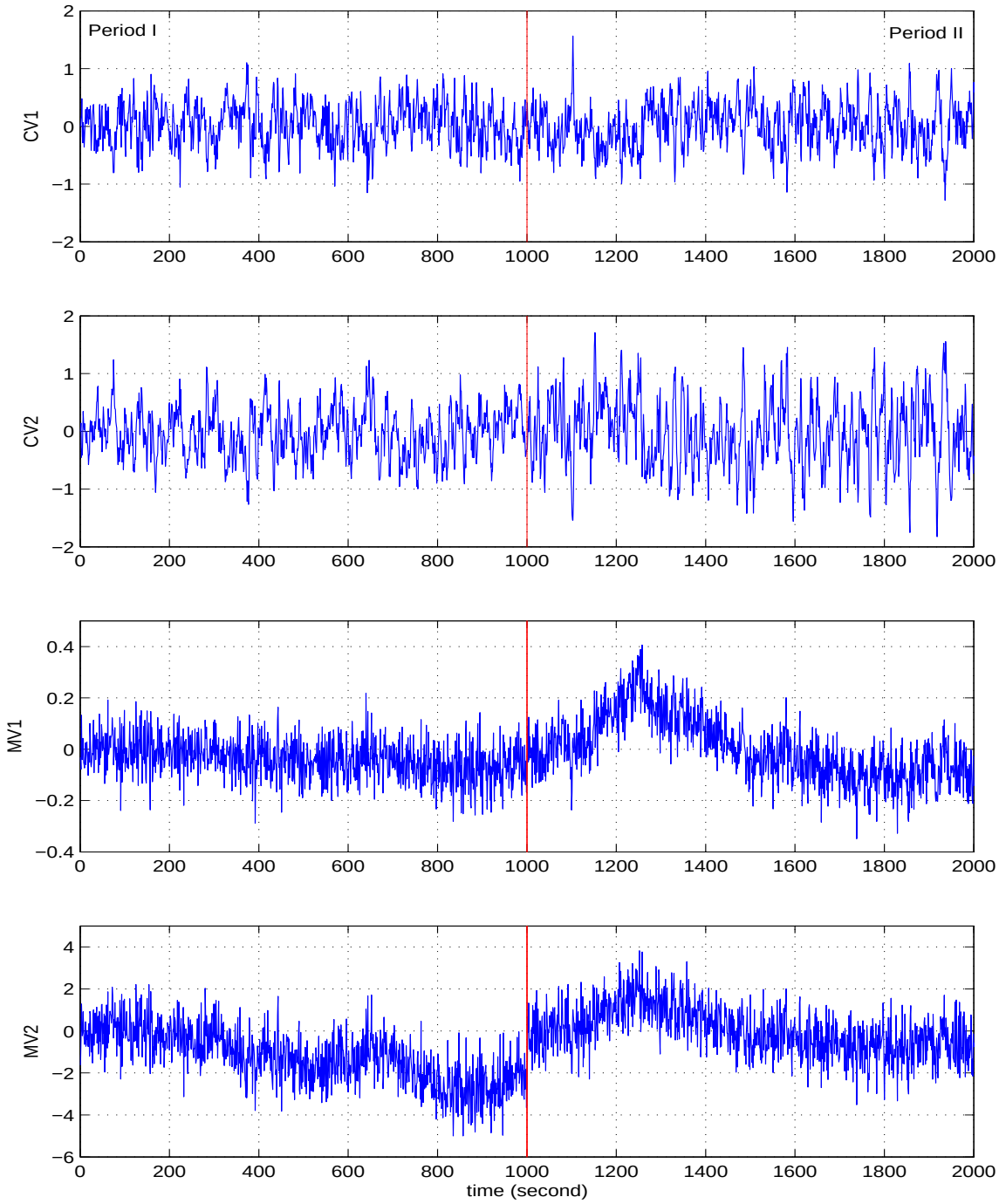

Figure 3.8: Case 2 of simulated MPC example: Closed-loop data of the benchmark and monitored periods

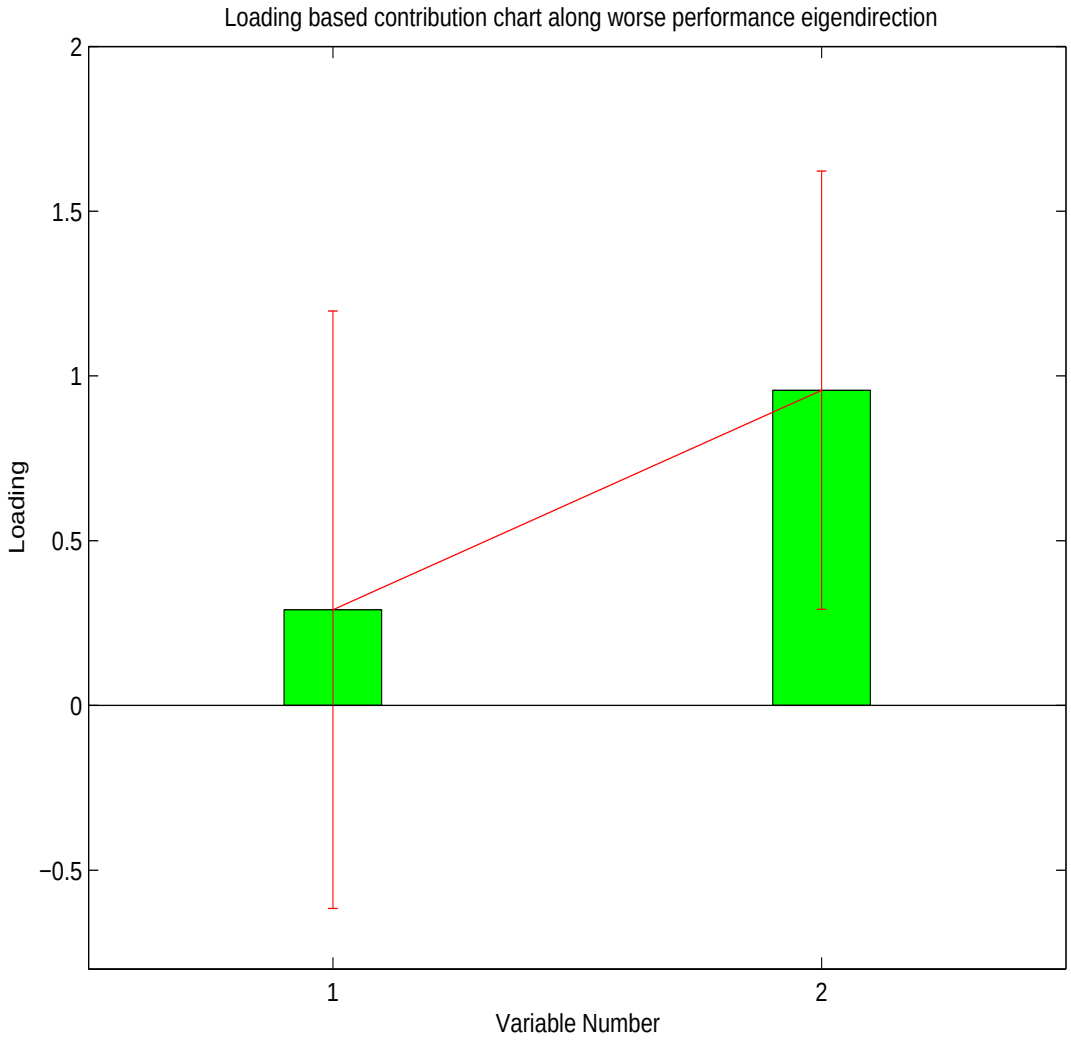

Figure 3.9: Case 2 of simulated MPC example: Loading based contribution chart with 95% confidence limits in the worse performance eigendirection

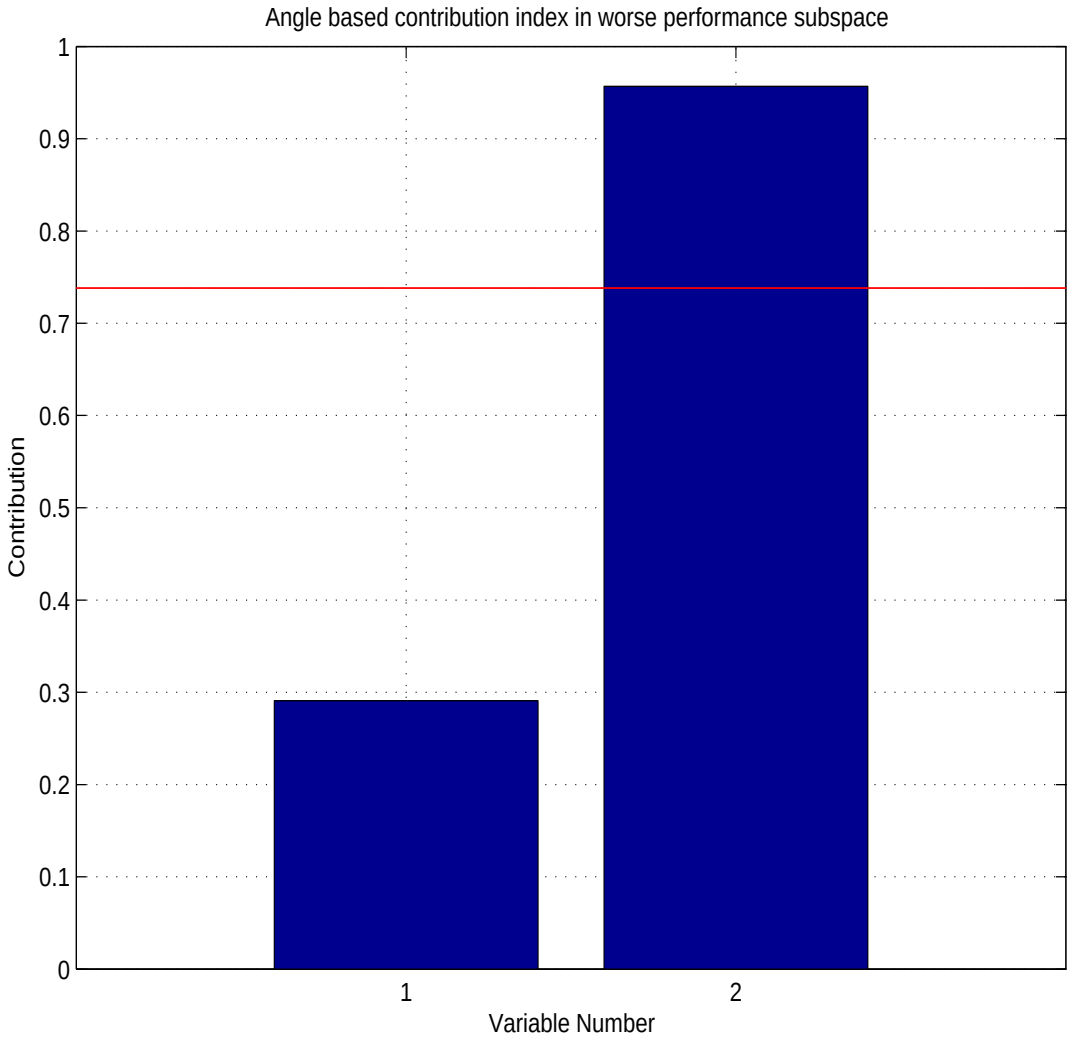

Figure 3.10: Case 2 of simulated MPC example: ABC chart with 95% confidence limit in the worse performance subspace

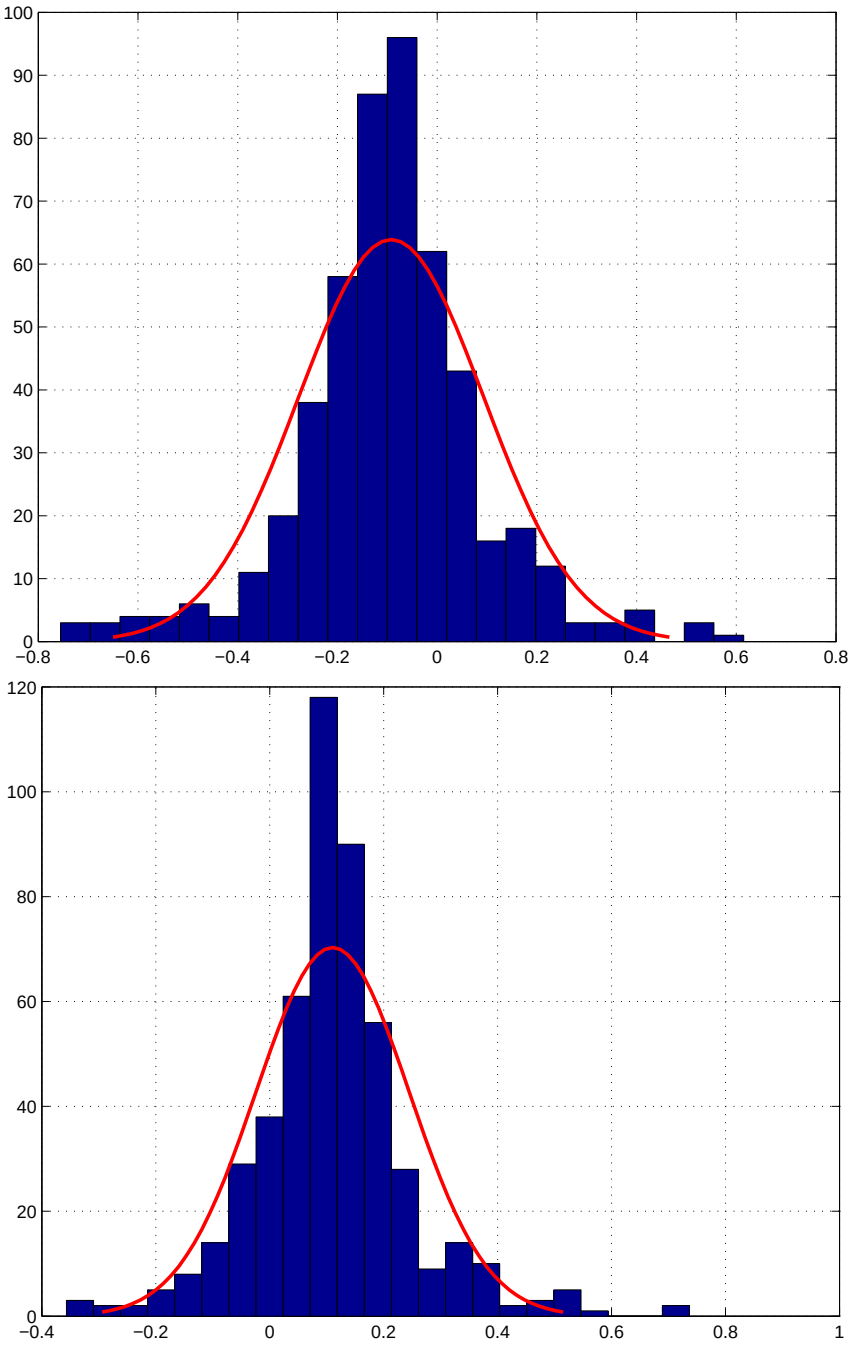

Figure 3.11: Industrial example: Bootstrapping probability distributions of the loading coefficients for Loop 1 along the first (worse) and the eighth (better) eigendirections

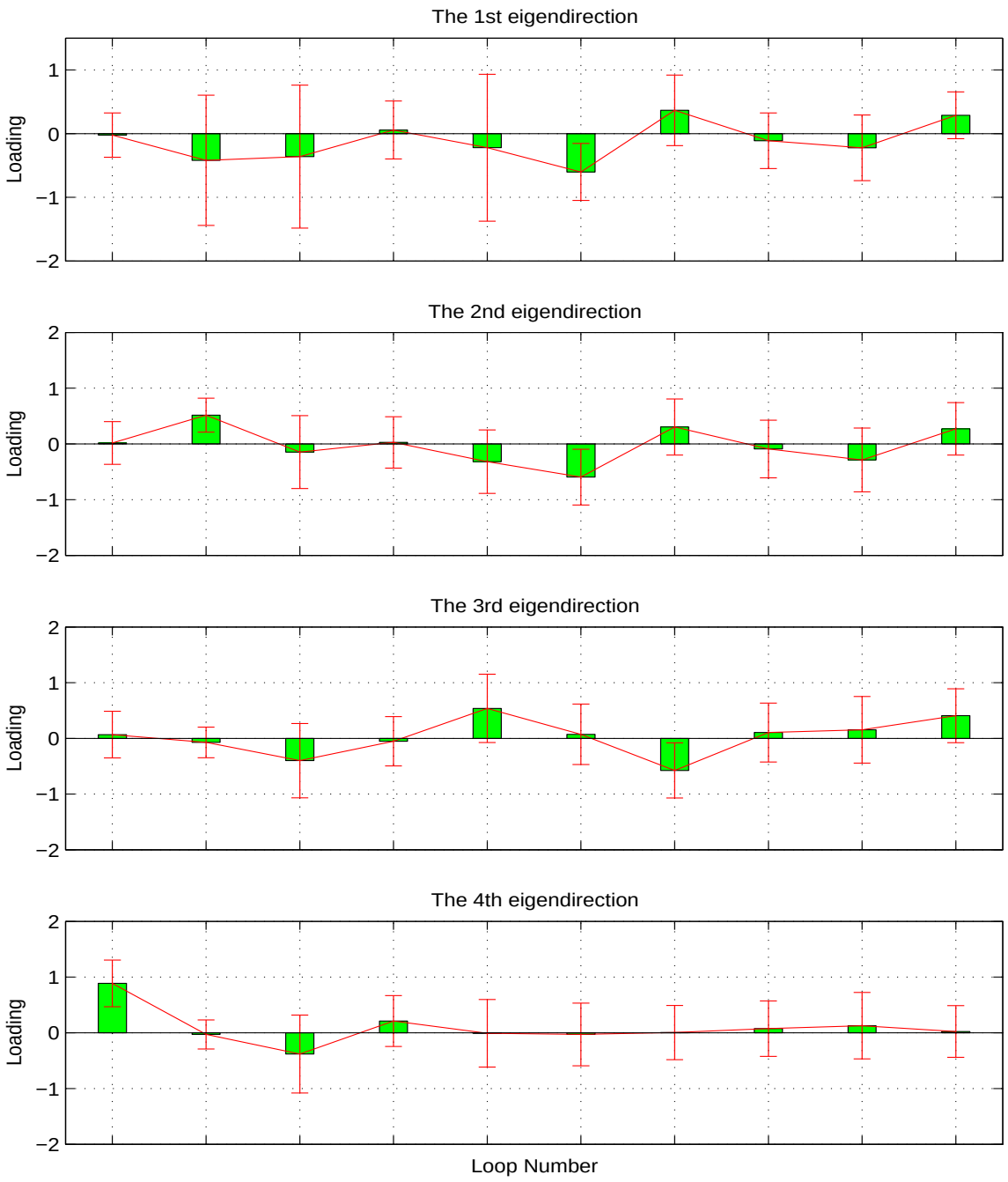

Figure 3.12: Industrial example: Loading based contribution charts with 95% confidence limits in the worse performance eigendirections of Period II over benchmark Period I

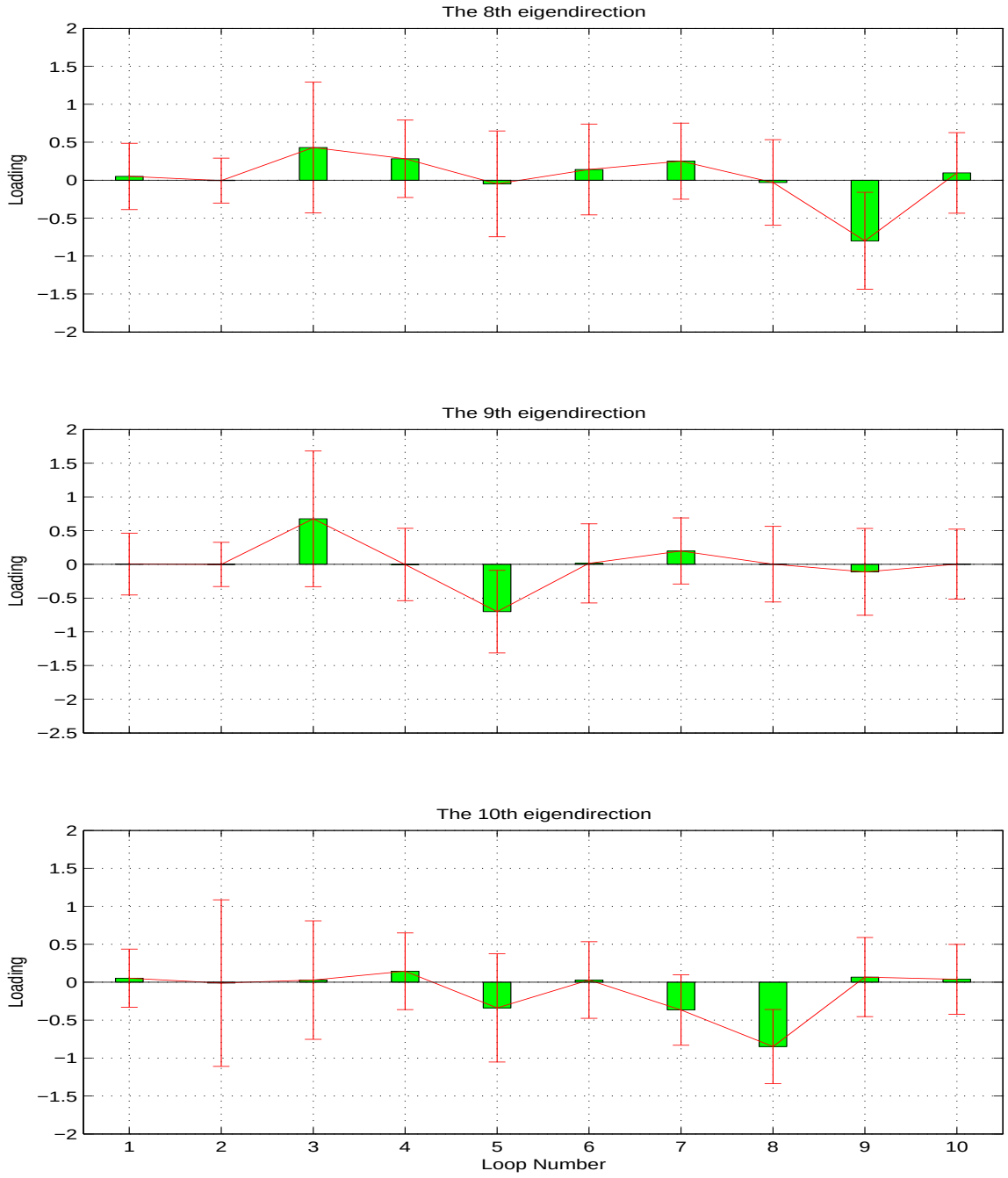

Figure 3.13: Industrial example: Loading based contribution charts with 95% confidence limits in the better performance eigendirections of Period II over benchmark Period I

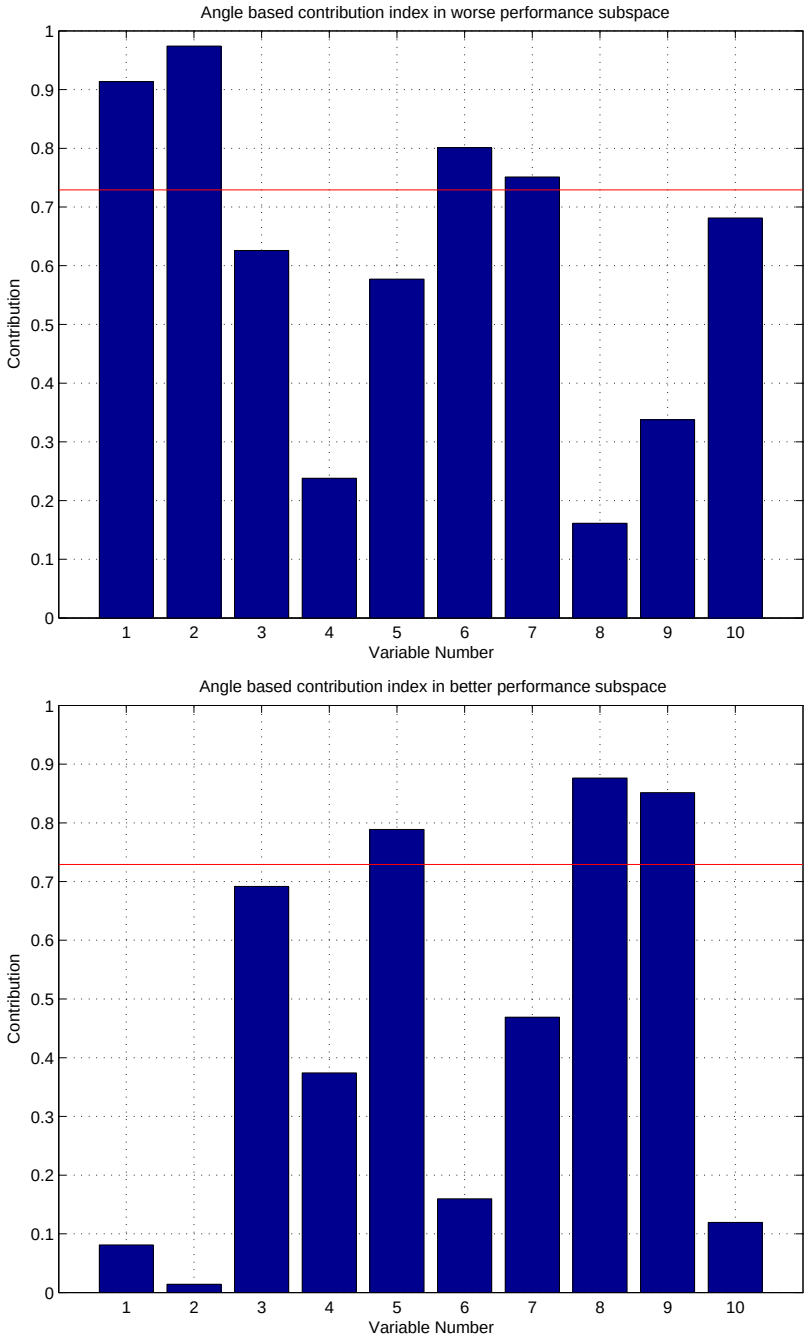

Figure 3.14: Industrial example: ABC charts with 95% confidence limits in the worse and better performance subspaces of Period II over benchmark Period I

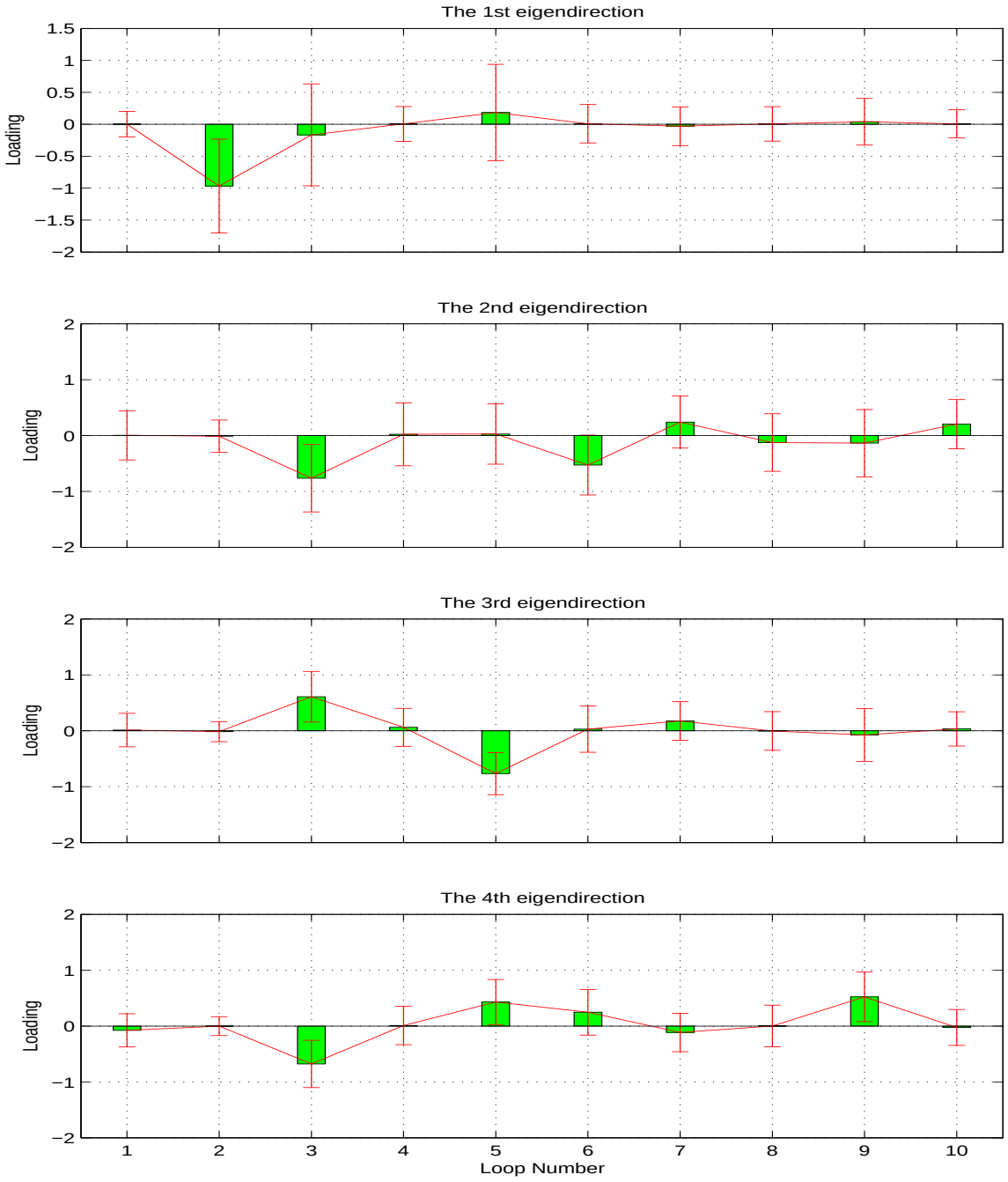

Figure 3.15: Industrial example: Loading based contribution charts with 95% confidence limits in the worse performance eigendirections of Period IV over benchmark Period I

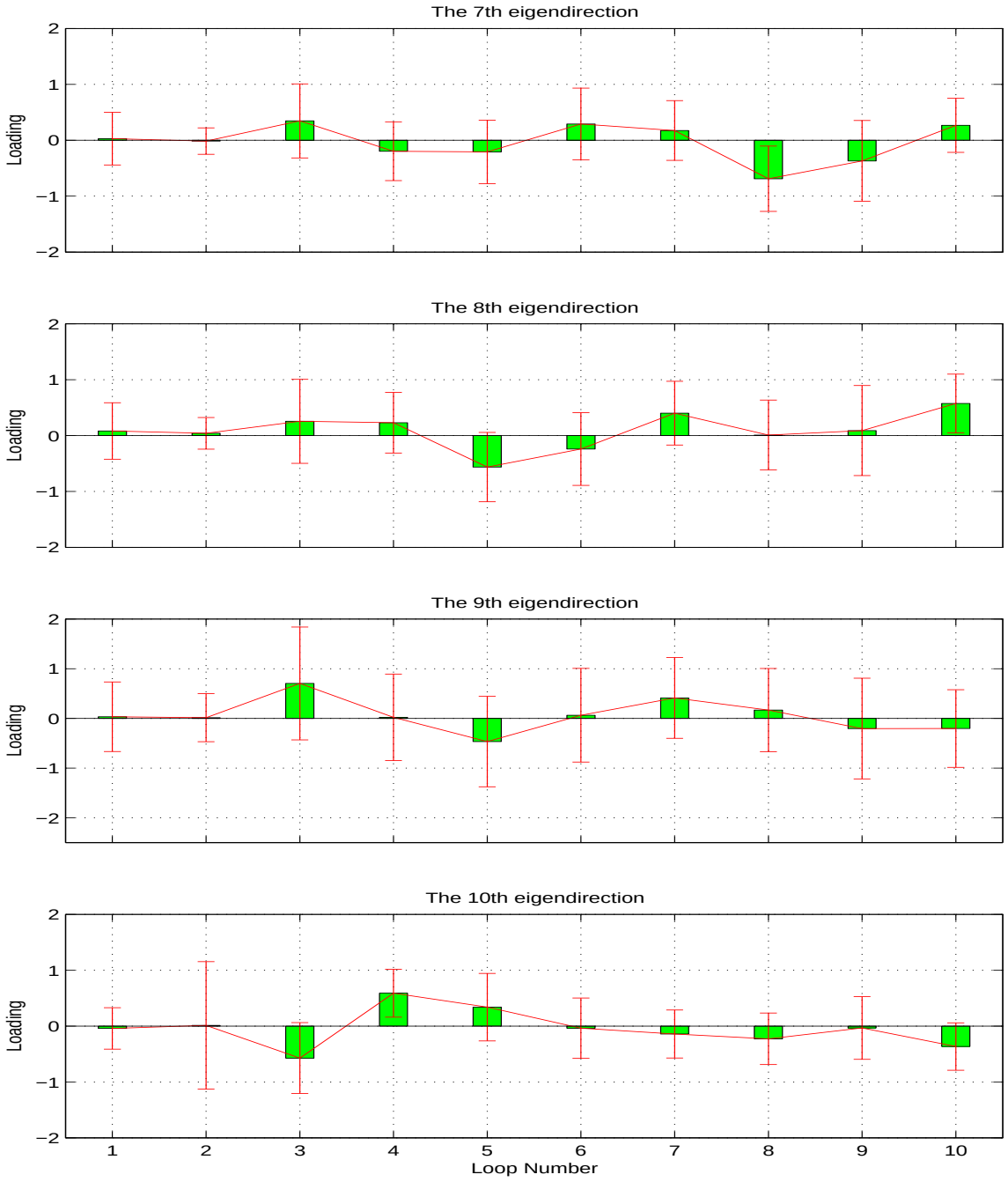

Figure 3.16: Industrial example: Loading based contribution charts with 95% confidence limits in the better performance eigendirections of Period IV over benchmark Period I

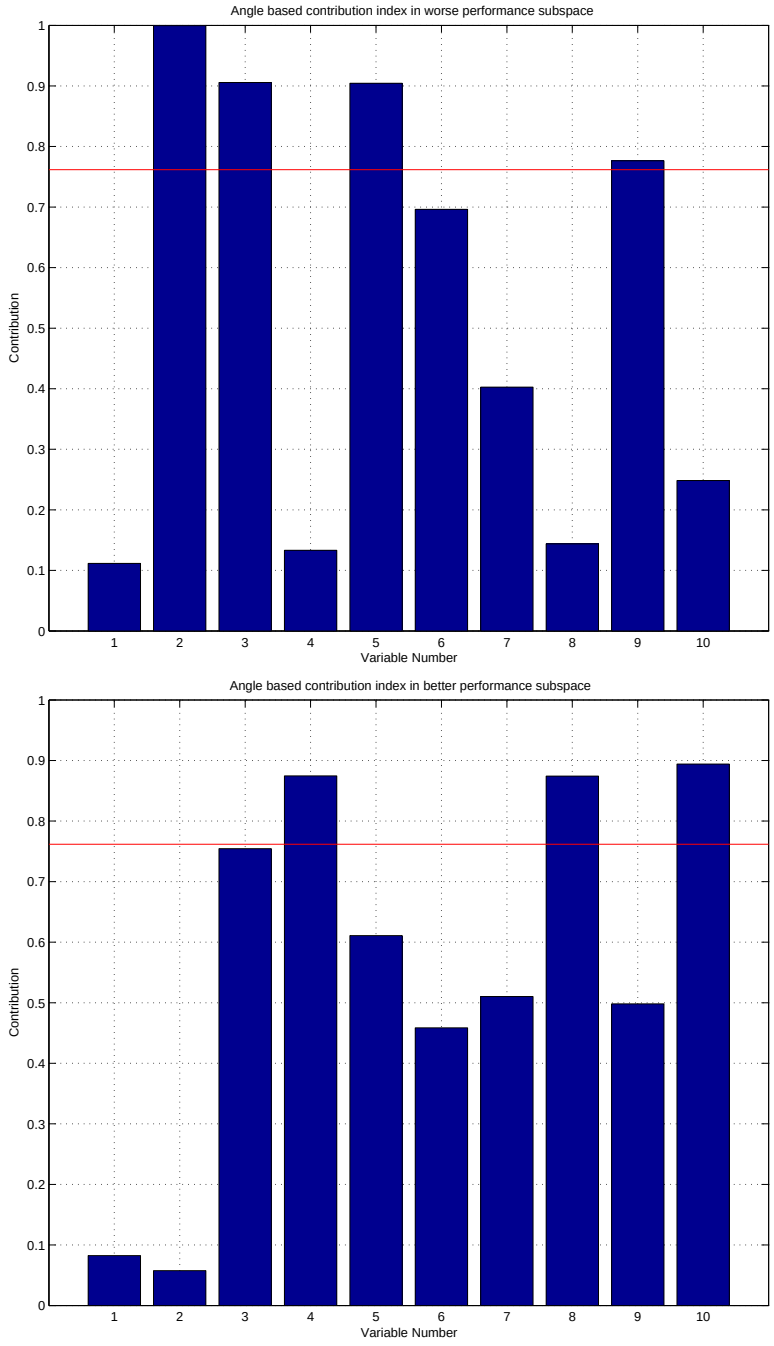

Figure 3.17: Industrial example: ABC charts with 95% confidence limits in the worse and better performance subspaces of Period IV over benchmark Period I

Chapter 4

Simplified solutions to MIMO MVC based performance monitoring and improvement - diagonal interactors and adaptive weighting

4.1 Introduction

As is known, the accurate estimation of interactor matrix and MVC benchmark for MIMO systems requires plenty of process knowledge such as full plant model or at least the first few Markov parameters. Furthermore, the estimation procedure is computationally intensive and not desirable for industrial implementations. Some research effort has been attempted to alleviate the computational complexity of MVC benchmark. Huang et al. proposed a filtering and correlation (FCOR) algorithm to compute the MIMO MVC benchmark by using closed-loop operating data (Huang *et al.*, 1997*b*). The first few Markov parameters (up to the order of system delay) were used to estimate the interactor matrix (Huang *et al.*, 1997*a*). This is essentially tantamount to knowing the entire process model. Ko and Edgar (Ko and Edgar, 2001*a*) tried to eliminate the requirement of interactor matrix and developed an explicit "one-shot" solution for the MVC output expression. McNabb and Qin (McNabb and Qin, 2003) investigated the state space framework of MVC based performance monitoring. They derived a MTD matrix

by using extended state space formulation to substitute the interactor matrix. However, both methods still need a priori knowledge of the leading Markov parameters up to the delay order. To further loose the requirement of a priori process knowledge, Huang et al. (Huang *et al.*, 2005) presented an approach to estimate a suboptimal MIMO control performance benchmark from routine operating data with the information of the order of interactor matrix (OIM) rather than the full interactor. Therefore, such a benchmark is much easier to derive and more convenient to implement than the conventional MVC benchmark, although it does not provide the theoretical lower bound of achievable performance. Xia et al. (Xia *et al.*, 2006) recommended to use input-output (I/O) delay matrix to estimate the upper and lower bounds of the MIMO MVC performance with routine operating data. Compared to the interactor matrix, it is relatively easier to determine the time delays between each pair of input and output variables. The simplified computation procedures can then lead to an interval that includes the true MVC benchmark. It needs to be emphasized that the established performance benchmarks in both studies are significantly suboptimal to the MVC benchmark due to the reduced a priori knowledge. Therefore, none of these simplified approaches can result in the optimal performance bound like MVC benchmark.

The unrealistic requirement of a priori process knowledge and the complicated procedures of numerical estimation pose the challenging issues for the industrial applications of MVC based performance monitoring. However, the importance of MVC benchmark cannot be disregarded because it does

provide the absolute lower bound of the achievable performance. Therefore, it is highly desirable to derive the MVC benchmark by using more realistic process information and simpler computation algorithm. It is noticed that two special forms of interactor matrix, i.e. simple interactor and diagonal interactor, are much easier to derive with MIMO input-output time delay information, which can be obtained through either open-loop bump test or physical insights (Huang and Shah, 1999; Xia *et al.*, 2006; Huang *et al.*, 2006). Therefore, if the multivariate time delay structure can be factorized out in terms of diagonal matrix, the technical difficulty of estimating MIMO interactors would be avoided completely. In fact, the diagonal interactor matrix exists on the output-delay processes which have the uniform I/O time delays along each row vector of process transfer function matrix. Similarly, a right diagonal interactor matrix (Goodwin *et al.*, 2001) exists for the input-delay processes with the same delay terms along each column of transfer function. In this work, the existence condition and uniqueness of right diagonal interactor matrix are explored. Then the MVC benchmark using the right diagonal interactor is deducted. As a generalization of the special left or right interactor cases, the more complicated MIMO time delay structure may be characterized by the combination of left and right diagonal interactor matrices. Based on the combined left/right diagonal interactors, the computation procedures of MVC benchmark can thus be simplified using the similar principles as FCOR algorithm (Huang and Shah, 1999). More importantly, the unrealistic requirement of full plant model or leading Markov parameters is avoided and only

input-output delays are needed to construct the left/right diagonal interactors. In addition to MVC benchmark for performance monitoring, we also try to optimize the weighting matrix involved in LQ design to improve control performance. Eigenvalue decomposition on the output covariance matrix is implemented to find out the worst performance directions with the largest variance inflation. Then a suitable weighting matrix can be designed to rotate the original outputs to the eigendirections and the corresponding eigenvalues may be used to measure the relative importance of those directions. Through the iterative update of weighting matrix, the control performance in terms of output variance can be improved and pushed close to the MVC boundary.

The rest of the chapter is organized as follows. The preliminaries of MIMO interactor matrix and MVC benchmark are revisited in Section 4.2. The right diagonal interactor matrix is constructed for the special input-delay process in Section 4.3. The existence conditions and uniqueness of right diagonal interactor matrix are also discussed. The minimum variance output and control law are further derived with the right diagonal interactor in this section. The more general time delay structure is factorized into the combined left and right diagonal interactors in Section 4.4. The computation procedures for the corresponding MVC benchmark is then described. In Section 4.5, an EVD based adaptive selection strategy of weighting matrix is developed for LQ control design and performance improvement of MIMO processes. The results of a series of simulated examples are illustrated in Section 4.6. The chapter is summarized in Section 4.7.

4.2 Review of MIMO interactor matrix and MVC benchmark

Time delay is one of the most common causes that restrict the control performance. Relative to the scalar time delay in the univariate processes, the time delays of multivariate processes can be characterized by interactor matrix (Wolovich and Falb, 1976). Let a MIMO process be described by the following dynamic model

$$y_t = G(q)u_t + N(q)a_t \tag{4.1}$$

where $G(q)$ and $N(q)$ are process and disturbance transfer function matrices in the unit of backshift operator, respectively. u_t, y_t and a_t represent process input, output and disturbance vectors, where a_t is assumed to be normally distributed white noise with zero mean and covariance of Σ_a. In addition, both $G(q)$ and $N(q)$ are assumed to be proper and rational realizations. Then for the $n \times n$ transfer function matrix $G(q)$, there exists a non-singular and non-unique polynomial $n \times n$ matrix D such that $det(D) = q^r$ and

$$\lim_{q^{-1}\to 0} DG(q) = \lim_{q^{-1}\to 0} \widetilde{G}(q) = K \tag{4.2}$$

where $\widetilde{G}(q)$ is the delay-free transfer function matrix with only finite zeros, K is a full rank finite matrix and r denotes the number of infinite zeros of $G(q)$ (Goodwin and Sin, 1984; Huang and Shah, 1999). The interactor matrix D can be decomposed as

$$D = D_0 q^d + D_1 q^{d-1} + \cdots + D_{d-1} q \tag{4.3}$$

where d is the order of the interactor matrix and D_i $(i = 0, 1, \cdots, d-1)$ are the coefficient matrices. Based on the above definition, the interactor matrix D is obtained by factorizing out the multivariate time-delay structure from the process transfer function matrix $G(q)$ as follows

$$G(q) = D^{-1}\widetilde{G}(q) \tag{4.4}$$

The simplest form of interactor matrix is $D = q^d I$ for square transfer function matrix $G(q)$ and called *simple interactor matrix*. Although the simple interactor can be easily obtained by extracting the smallest I/O time delay term, it is rarely encountered in real MIMO processes. Another special case of interactor matrix is in the form of diagonal matrix as $D = diag(q^{d_1}, q^{d_2}, \cdots, q^{d_n})$. The time-delay structure of the output-delay process can be characterized by diagonal interactor matrix, which is obtained by extracting the common I/O delay terms along each row vector of process transfer function matrix. Not only restricted to output-delay case, some well-designed multivariate processes may also have the time delay structure in the form of diagonal interactor (Huang and Shah, 1999). In that situation, the smallest I/O delay terms along each row of process transfer function can be extracted as the diagonal entries of interactor. The general interactor may have different forms such as lower triangular, upper triangular, nilpotent and unitary matrices. The non-unique unitary interactor matrix satisfies the following condition

$$D^T(q^{-1})D(q) = I \tag{4.5}$$

The numerical algorithm for calculating the unitary interactor matrix can be found in Peng and Kinnaert (Peng and Kinnaert, 1992).

By minimizing the sum of the variances of the interactor-filtered output $\widetilde{y}_t = q^{-d}Dy_t$ as

$$min\ J = E(\widetilde{y}_t^T\widetilde{y}_t) = tr\big[cov(\widetilde{y}_t)\big] \tag{4.6}$$

the MIMO MVC law and benchmark can be derived (Huang *et al.*, 1997*b*; McNabb and Qin, 2003). For the unitary interactor D, the LQ objective function of the interactor-filtered output $\widetilde{y}_t$ is equivalent to the objective function of the original output y_t. Therefore, the actual control performance in current period can be evaluated by the following MVC based performance index

$$\eta = \frac{E(y_t^T y_t)_{\min}}{E(y_t^T y_t)} = \frac{E(\widetilde{y}_t^T\widetilde{y}_t)_{min}}{E(\widetilde{y}_t^T\widetilde{y}_t)} = \frac{tr[\mathrm{cov}(\widetilde{y}_{MV})]}{tr[\mathrm{cov}(\widetilde{y}_t)]} \tag{4.7}$$

where the value of η is always between 0 and 1 because MVC benchmark represents the theoretical lower bound of the achievable performance in terms of output variance.

4.3 MVC based MIMO control performance monitoring with right diagonal interactor

4.3.1 Introduction of right diagonal interactor matrix

Consider a $n \times n$ input-delay process transfer function matrix G, where the time delays for input variable 1 through n are $d_1', d_2', \ldots, d_n'$, respectively.

Then $G(q)$ can be expressed as

$$G(q) = \begin{bmatrix} \widetilde{G}_{11}q^{-d'_1} & \cdots & \widetilde{G}_{1n}q^{-d'_n} \\ \vdots & \ddots & \vdots \\ \widetilde{G}_{n1}q^{-d'_1} & \cdots & \widetilde{G}_{nn}q^{-d'_n} \end{bmatrix} \tag{4.8}$$

In this case, the time delay structure of the process transfer function matrix $G(q)$ cannot be formulated as a left diagonal interactor. However, it can be readily factorized as

$$G(q) = \begin{bmatrix} \widetilde{G}_{11} & \cdots & \widetilde{G}_{1n} \\ \vdots & \ddots & \vdots \\ \widetilde{G}_{n1} & \cdots & \widetilde{G}_{nn} \end{bmatrix} \begin{bmatrix} q^{-d'_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & q^{-d'_n} \end{bmatrix} = \widetilde{G}D_R^{-1} \tag{4.9}$$

where $D_R = diag(q^{d'_1}, q^{d'_2}, \cdots, q^{d'_n})$ is a diagonal matrix containing all the input delays. If $\lim_{q^{-1}\to 0} \widetilde{G}(q)$ is a full rank finite matrix, then D_R can be treated as a right interactor of $G(q)$. The reason that it is called "right" is because the delay-free transfer function matrix $\widetilde{G}(q)$ is post-multiplied by the time-delay structure matrix D_R instead of pre-multiplied by D_L. The general definition of right interactor matrix can be stated as follows (Goodwin *et al.*, 2001):

Definition 4.3.1. For any $n \times n$ proper and rational polynomial transfer function matrix $G(q)$, if a $n \times n$ non-singular polynomial matrix D_R satisfies

$$G(q) = \widetilde{G}(q)D_R^{-1} \tag{4.10}$$

and

$$\lim_{q^{-1}\to 0} \widetilde{G}(q) = \lim_{q^{-1}\to 0} G(q)D_R = K \tag{4.11}$$

with K denoting a full rank finite matrix and $det(D_R) = q^r$, then D_R is referred to as the right interactor matrix of $G(q)$.

Let $D_{I/O}$ be the I/O time delay matrix of $G(q)$ as

$$D_{I/O} = \begin{bmatrix} q^{-d_{1,1}} & \cdots & q^{-d_{1,n}} \\ \vdots & \ddots & \vdots \\ q^{-d_{n,1}} & \cdots & q^{-d_{n,n}} \end{bmatrix} \tag{4.12}$$

where $d_{i,j}$ $(1 \leq i, j \leq n)$ represents the time delay between the jth input and ith output. For any entry without connection between the jth input and ith output, set $d_{i,j} = \infty$ and $q^{-d_{i,j}} = 0$. The uniqueness of the right diagonal interactor matrix is stated in the following theorem.

Theorem 4.3.1. *If a right diagonal interactor matrix exists, then it can be factorized out from the I/O time delay matrix $D_{I/O}$ by extracting the minimum delay term along each column vector of $D_{I/O}$, i.e.*

$$D_R = diag(q^{d'_1}, q^{d'_2}, \cdots, q^{d'_n}) \tag{4.13}$$

with

$$d'_j = \min_i d_{ij} \tag{4.14}$$

the above diagonal right interactor matrix must be unique.

Proof. Suppose that the right diagonal interactor matrix is not unique and another realization is given by $D'_R = diag(q^{d''_1}, q^{d''_2}, \cdots, q^{d''_n})$, where $D'_R \neq D_R$. Therefore, at least one diagonal entry of D'_R is different from the corresponding entry of D_R. Without loss of generality, assume that $d''_j \neq d'_j$. Thus there exist only two scenarios:

(i). If $d''_j > d'_j = \min_i d_{ij}$, let $\min_i d_{ij} = d_{kj}$, then

$$\lim_{q^{-1} \to 0} \{G(q) D'_R\}_{kj} = \infty \tag{4.15}$$

which contradicts a necessary condition of right interactor matrix, i.e., $\lim_{q^{-1}\to 0} G(q)D'_R = K$ is a finite matrix.

(ii). If $d''_j < d'_j = \min_i d_{ij}$, then for all i we have $d''_j - d_{ij} < 0$, thus the jth column vector of $\lim_{q^{-1}\to 0} G(q)D'_R = K$ is zero and the condition that K is full rank does not hold.

Therefore, the assumption that $D'_R \neq D_R$ is overriden and the right diagonal interactor matrix must be unique. □

As an analogy, if a left diagonal interactor matrix exists, it is also unique and can be derived by exacting the minimum delay term along each row of I/O time delay matrix $D_{I/O}$ as $D_R = diag(q^{d_1}, q^{d_2}, \cdots, q^{d_n})$ with $d_i = \min_j d_{ij}$. To determine if a right diagonal interactor matrix exists for the process transfer function matrix $G(q)$, a sufficient condition can be established following the similar principles presented in Xia et al. (Xia *et al.*, 2006). Let the set of all $n \times n$ real matrices without any zero elements be

$$S(R) = \{R : R_{i,j} \neq 0,\ 1 \leq i, j \leq n\} \tag{4.16}$$

then the following theorem holds.

Theorem 4.3.2. *Consider the transfer function matrix $G(q)$ with its I/O time delay matrix $D_{I/O}$ and D_R given by Eqs.* (4.12) *and* (4.13)*, respectively. Define*

$$U = \lim_{q^{-1}\to 0} D_{I/O} D_R \tag{4.17}$$

If $U \odot R$ is always full rank for all $R \in S(R)$, where $\odot$ denotes the Hadamard (entry-wise) product of two matrices, then the right diagonal interactor matrix

exists.

Proof. The proof is straightforward and follows the definition of right interactor matrix. □

In addition to the I/O time delay matrix, if the leading $d_{I/O} = \max_{i,j} d_{i,j}$ Markov parameters are also known, then we have the following necessary and sufficient criterion to determine the existence of diagonal right interactor matrix.

Theorem 4.3.3. *Define the first-order I/O delay-free Markov parameter as*

$$G_0^f = \begin{bmatrix} g_{d_{1,1}}^{(1,1)} & \cdots & g_{d_{1,n}}^{(1,n)} \\ \vdots & \ddots & \vdots \\ g_{d_{n,1}}^{(n,1)} & \cdots & g_{d_{n,n}}^{(n,n)} \end{bmatrix} \tag{4.18}$$

where $g_{d_{i,j}}^{(i,j)}$ represents the $\{d_{i,j}+1\}$th-order Markov parameter of the transfer funcation entry $G_{i,j}(q)$ and the matrix D_R is given by Eq. (4.13)*. If $\lim_{q^{-1}\to 0}(G_0^f \odot D_{I/O})D_R$ is a full rank finite matrix, then the right diagonal interactor matrix exists.*

Proof. The proof simply follows the definition of right interactor matrix because

$$\lim_{q^{-1}\to 0}(G_0^f \odot D_{I/O})D_R = \lim_{q^{-1}\to 0} GD_R = \lim_{q^{-1}\to 0} \widetilde{G} \tag{4.19}$$

□

The first-order I/O delay-free Markov parameter G_0^f is essentially equivalent to the numerator gain matrix of the process transfer function $G(q)$. Hence

the above existence criterion of right diagonal interactor D_R can be checked as long as the numerator gain matrix is known, which should be easier to obtain than all of the leading $d_{I/O}$ Markov parameters. It needs to be pointed out that the above $d_{I/O}$ denotes the order of I/O delay matrix and could be larger than the order of interactor matrix d.

Example 1. Consider a 2×2 process transfer function matrix

$$G(q) = \begin{bmatrix} \frac{q^{-1}}{1-0.5q^{-1}} & \frac{q^{-5}}{1-0.9q^{-1}} \\ \frac{3q^{-2}}{1-0.4q^{-1}} & \frac{2q^{-3}}{1-0.3q^{-1}} \end{bmatrix}$$

the I/O delay matrix is

$$D_{I/O} = \begin{bmatrix} q^{-1} & q^{-5} \\ q^{-2} & q^{-3} \end{bmatrix}$$

and the first-order delay-free Markov parameter is

$$G_0^f = \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix}$$

By extracting minimum delay term along each column of $G(q)$, we have

$$D_R = \begin{bmatrix} q & 0 \\ 0 & q^3 \end{bmatrix}$$

Since

$$\lim_{q^{-1} \to 0} (G_0^f \odot D_{I/O}) D_R = \lim_{q^{-1} \to 0} \widetilde{G} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

is a full rank finite matrix, it can be concluded that the right diagonal interactor matrix D_R exists even though the process is not the special input-delay case.

If $G(q)$ is changed to

$$G(q) = \begin{bmatrix} \frac{q^{-1}}{1-0.5q^{-1}} & \frac{q^{-3}}{1-0.9q^{-1}} \\ \frac{3q^{-2}}{1-0.4q^{-1}} & \frac{2q^{-5}}{1-0.3q^{-1}} \end{bmatrix}$$

then

$$\lim_{q^{-1}\to 0}(G_0^f \odot D_{I/O})D_R = \lim_{q^{-1}\to 0}\widetilde{G} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

which is rank-deficient. Thus there exists no right diagonal interactor for this process and the factorized D_R is not a valid right interactor.

4.3.2 MIMO MVC law and benchmark under right diagonal interactor

For the class of processes with right diagonal interactors, e.g. input-delay process, we can further derive the MVC law and performance benchmark. Assume that the MIMO process described by Eq. (4.1) has a right diagonal interactor matrix $D_R = diag(q^{d'_1}, q^{d'_2}, \cdots, q^{d'_n})$, thus

$$\begin{aligned} y_t &= \widetilde{G}D_R^{-1}u_t + Na_t \\ &= \widetilde{G}\begin{bmatrix} u_1(t-d'_1) \\ u_2(t-d'_2) \\ \vdots \\ u_n(t-d'_n) \end{bmatrix} + Na_t \end{aligned} \tag{4.20}$$

Let $\widetilde{u}_t = D_R^{-1}u_t = \begin{bmatrix} u_1(t-d'_1) \\ u_2(t-d'_2) \\ \vdots \\ u_n(t-d'_n) \end{bmatrix}$, then Eq. (4.20) becomes

$$y_t = \widetilde{G}\widetilde{u}_t + ND_R \cdot D_R^{-1}a_t \tag{4.21}$$

Define $\widetilde{y}_t = q^{-d}y_t$, $\widetilde{N} = q^{-d}ND_R$ and $\widetilde{a}_t = D_R^{-1}a_t$, pre-multiplying q^{-d} on Eq. (4.21) yields

$$\widetilde{y}_t = q^{-d} \cdot \widetilde{G}\widetilde{u}_t + \widetilde{N} \cdot \widetilde{a}_t \tag{4.22}$$

It is obvious that the new system has a simple interactor matrix $q^{-d}I$, where $d = \max_{1\leq i\leq n} d_i$ is the order of the interactor matrix. Then based on Diophantine identity, $\widetilde{N}$ can be factorized as

$$\begin{aligned}\widetilde{N} &= \underbrace{F_0 + F_1 q^{-1} + \cdots + F_{d-1} q^{-(d-1)}}_{F} + Rq^{-d} \\ &= F + Rq^{-d} \end{aligned} \tag{4.23}$$

Substituting Eqs. (4.23) into (4.22) yields

$$\widetilde{y}_t = \widetilde{G}\widetilde{u}_{t-d} + F\widetilde{a}_t + R\widetilde{a}_{t-d} \tag{4.24}$$

Since $\widetilde{u}_{t-d}$ and $\widetilde{a}_{t-d}$ are uncorrelated with $\widetilde{a}_t$, then we have

$$\begin{aligned} tr\left[\mathrm{cov}(\widetilde{y}_t)\right] &= tr\left[\mathrm{cov}(\widetilde{G}\widetilde{u}_{t-d} + R\widetilde{a}_{t-d}) + \mathrm{cov}(F\widetilde{a}_t)\right] \\ &\geq tr\left[\mathrm{cov}(F\widetilde{a}_t)\right] = tr(FD_R^{-1}\Sigma_a D_R^{-1}F^T) \end{aligned} \tag{4.25}$$

MV control law is achieved when

$$\widetilde{G}\widetilde{u}_{t-d} + R\widetilde{a}_{t-d} = 0 \tag{4.26}$$

such that $\widetilde{y}_t = F\widetilde{a}_t$ which is also know as control invariant term (Huang and Shah, 1999). Recalling the fact that $\widetilde{G}$ is a delay-free and strictly proper transfer function matrix, we have the following relationship from Eq. (4.26):

$$\widetilde{u}_t = -\widetilde{G}^{-1}R\widetilde{a}_t \tag{4.27}$$

where the matrix inverse G^{-1} can be replaced by the pseudoinverse $G^{\dagger}$ for non-square MIMO processes. It follows that

$$u_t = -D_R\widetilde{G}^{-1}R\widetilde{a}_t \tag{4.28}$$

Inserting $\widetilde{y}_t = F\widetilde{a}_t$ into Eq. (4.28) yields

$$u_t = -D_R\widetilde{G}^{-1}R \cdot (q^d F)^{-1} y_t \tag{4.29}$$

Therefore, the MVC law under right diagonal interactor matrix is given by

$$Q|_{mv} = -D_R\widetilde{G}^{-1}R \cdot (q^d F)^{-1} \tag{4.30}$$

By comparing to the MVC law under left unitary interactor matrix, i.e.

$$Q|_{mv} = -\widetilde{G}^{-1}R \cdot (q^d F)^{-1} D_L \tag{4.31}$$

it can be found that the difference of the analytical expressions lies in pre-multiplying the right interactor matrix D_R or post-multiplying the left interactor matrix D_L. Under the established MVC law in Eq. (4.30), the minimum variance output for the original CV is

$$\begin{aligned} y_t|_{mv} = \quad & q^d \cdot \widetilde{y}_t|_{mv} = q^d \left(F_0\widetilde{a}_t + F_1\widetilde{a}_{t-1} + \cdots + F_{d-1}\widetilde{a}_{t-d+1}\right) \\ = \quad & q^d \left(F_0 D_R^{-1} a_t + F_1 D_R^{-1} a_{t-1} + \cdots + F_{d-1} D_R^{-1} a_{t-d+1}\right) \end{aligned} \tag{4.32}$$

Since

$$\begin{aligned} tr\left[\mathrm{cov}(\widetilde{y}_t)\right] &= tr\left[E\big(\widetilde{y}_t - E(\widetilde{y}_t)\big)\big(\widetilde{y}_t - E(\widetilde{y}_t)\big)^T\right] \\ &= tr\left[E(q^{-d}\big(y_t - E(y_t)\big)q^d\big(y_t - E(y_t)\big)^T)\right] \\ &= tr\left[\mathrm{cov}(y_t)\right] \end{aligned} \tag{4.33}$$

the variance of the original variables is equivalent to the variance of the filtered variables and the minimum variance under MVC law is given by

$$J_{MV} = tr\left[\mathrm{cov}(F\widetilde{a}_t)\right] = tr\left[\mathrm{cov}(FD_R^{-1}a_t)\right] = tr\left[FD_R^{-1}\Sigma_a D_R F^T\right] \tag{4.34}$$

For diagonal noise covariance Σ_a, Eq. (4.34) can be further simplified to

$$J_{MV} = tr\left(F\Sigma_a F^T\right) = tr\left(F_0\Sigma_a F_0^T + F_1\Sigma_a F_1^T + \cdots + F_{d-1}\Sigma_a F_{d-1}^T\right) \quad (4.35)$$

because $D_R^{-1}\Sigma_a D_R = \Sigma_a D_R^{-1} D_R = \Sigma_a$. It needs to be noted that, although the the factorized coefficients F_i may be different under the left and the right interactors, the sum of squares of F_i should be the same.

4.4 MVC based MIMO control performance monitoring with combined left/right diagonal interactors

4.4.1 Introduction of combined left/right diagonal interactors

The existence conditions of the right diagonal interactor matrix has been discussed in the previous section. Generally, neither the left nor right diagonal interactor matrix is guaranteed to exist and the simplified computation procedures based on a single diagonal interactor are applicable for some well-designed processes only. Therefore, an alternative strategy by using the combined left and right diagonal interactors is developed for more complicated processes.

For the process given in Eq. (4.1), the combined left/right diagonal interactors can be defined in the following statement

Definition 4.4.1. If the $n \times n$ process transfer function matrix $G(q)$ can be factorized as

$$G(q) = D_L^{-1} \cdot \widetilde{G}(q) \cdot D_R^{-1} \quad (4.36)$$

where $D_L = diag(q^{d_1}, q^{d_2}, \ldots, q^{d_n})$ and $D_R = diag(q^{d'_1}, q^{d'_2}, \ldots, q^{d'_n})$ are both diagonal matrices, and $\widetilde{G}$ satisfies

$$\lim_{q^{-1}\to 0} D_L G(q) D_R = \lim_{q^{-1}\to 0} \widetilde{G}(q) = K \tag{4.37}$$

with K being a full rank finite matrix, then D_L and D_R are called the combined left/right diagonal interactor matrices of $G(q)$.

The systematic diagram of the left/right interactors in closed-loop systems is given in Fig. 4.1. It needs to be pointed out that the combined left and right diagonal interactors may not be unique if they exist. There are two straightforward ways to factorize out the available left and right diagonal interactors: (i). first extract the minimum I/O time delay term $d_i = \min_j d_{ij}$ row by row from $G(q)$ to form a left diagonal interactor $D_L = diag\left\{q^{d_1}, q^{d_2}, \cdots, q^{d_n}\right\}$ and then extract the minimum I/O delay $d'_j = \min_i\{d_{ij} - d_i\}$ from $D_L G(q)$ column by column to obtain a right diagonal interactor $D_R = diag\left\{q^{d'_1}, q^{d'_2}, \cdots, q^{d'_n}\right\}$; (ii). reverse the order of row-column operations by first exacting the minimum delay of each column from $G(q)$ to generate D_R and then the minimum delay of each row from $G(q)D_R$ to get D_L. As an extension of Theorem 4.3.3, the existence of the combined left/right diagonal interactors can be checked by the following sufficient condition.

Theorem 4.4.1. *Given the first-order I/O delay-free Markov parameter G_0^f as defined in Theorem 4.3.3 and the matrices D_L and D_R obtained from the above two methods, if $\lim_{q^{-1}\to 0} D_L(G_0^f \odot D_{I/O})D_R$ is a full rank finite matrix, then the combined left/right diagonal interactors exist.*

Proof. The sufficiency of the condition simply follows Definition 4.4.1 because

$$\lim_{q^{-1}\to 0} D_L(G_0^f \odot D_{I/O})D_R = \lim_{q^{-1}\to 0} D_L G(q) D_R = \lim_{q^{-1}\to 0} \widetilde{G}(q) \tag{4.38}$$

The above condition is unnecessary because, unlike the single left or right diagonal interactor, the factorization of the combined left/right diagonal interactors is not unique. □

Example 2. Consider a 3×3 process transfer function matrix as

$$G(q) = \begin{bmatrix} \frac{2.5q^{-3}}{1-0.3q^{-1}} & \frac{1.74q^{-4}}{1-0.5q^{-1}} & \frac{q^{-5}}{1-0.2q^{-1}} \\ \frac{3q^{-2}}{1-0.4q^{-1}} & \frac{2q^{-5}}{1-0.8q^{-1}} & \frac{q^{-2}}{1-0.3q^{-1}} \\ \frac{0.8q^{-5}}{1-0.6q^{-1}} & \frac{1.3q^{-6}}{1-0.2q^{-1}} & \frac{1.5q^{-4}}{1-0.9q^{-1}} \end{bmatrix}$$

First we use row operation to factorize out the left diagonal interactor matrix

$$D_L = diag(q^3, q^2, q^4)$$

then carry out column operation on $D_L G(q)$ to figure out the right diagonal interactor

$$D_R = diag(1, q, 1)$$

Since the I/O delay matrix and the first-order delay-free Markov parameter are

$$D_{I/O} = \begin{bmatrix} q^{-3} & q^{-4} & q^{-5} \\ q^{-2} & q^{-5} & q^{-2} \\ q^{-5} & q^{-6} & q^{-4} \end{bmatrix}$$

and

$$G_o^f = \begin{bmatrix} 2.5 & 1.74 & 1 \\ 3 & 2 & 1 \\ 0.8 & 1.3 & 1.5 \end{bmatrix}$$

then we have

$$\lim_{q^{-1}\to 0} D_L(G_0^f \odot D_{I/O})D_R = \begin{bmatrix} 2.5 & 1.74 & 0 \\ 3 & 0 & 1 \\ 0 & 0 & 1.5 \end{bmatrix}$$

which is obviously a full rank finite matrix. According to the criterion stated in Theorem 4.4.1, the extracted D_L and D_R are valid left and right diagonal interactors. If we change the order of row-column operation, a right diagonal interactor matrix is first extracted as

$$D'_R = diag(q^2, q^4, q^2)$$

then the extracted left diagonal interactor from $G(q)D_R$ is

$$D'_L = diag(1, 1, q^2)$$

In this case,

$$\lim_{q^{-1}\to 0} D'_L(G_0^f \odot D_{I/O})D'_R = \begin{bmatrix} 0 & 1.74 & 0 \\ 3 & 0 & 1 \\ 0 & 1.3 & 1.5 \end{bmatrix}$$

which is still a full rank finite matrix and thus the existence criterion is also satisfied. It indicates that the factorization of the combined left/right diagonal interactors is not unique and there could exist more than one realization.

4.4.2 MIMO MVC law and benchmark under combined left/right diagonal interactors

Based on the combined left/right diagonal interactors, the MVC law and benchmark can be further derived. Consider the process model given in

Eq. (4.1) and substituting Eqs. (4.36) into (4.1) yields

$$D_L y_t = \widetilde{G} D_R^{-1} u_t + D_L N a_t \tag{4.39}$$

or

$$\begin{bmatrix} y_1(t+d_1) \\ y_2(t+d_2) \\ \vdots \\ y_n(t+d_n) \end{bmatrix} = \widetilde{G} \begin{bmatrix} u_1(t-d_1') \\ u_2(t-d_2') \\ \vdots \\ u_m(t-d_n') \end{bmatrix} + D_L N a_t \tag{4.40}$$

Define $\widetilde{y}_t = q^{-d} D_L y_t$, $\widetilde{N} = q^{-d} D_L N D_R$, $\widetilde{u}_t = D_R^{-1} u_t$ and $\widetilde{a}_t = D_R^{-1} a_t$, then we have

$$\widetilde{y}_t = (q^{-d} I) \cdot \widetilde{G} \widetilde{u}_t + \widetilde{N} \widetilde{a}_t \tag{4.41}$$

Obviously the above transformed system has a simple interactor matrix $q^{-d} I$. Therefore, it follows from Diophantine identity that

$$q^{-d} D_L N D_R = \widetilde{N} = F + q^{-d} R \tag{4.42}$$

Then, inserting Eqs. (4.42) into (4.41) yields

$$\widetilde{y}_t = \widetilde{G} \widetilde{u}_{t-d} + F \widetilde{a}_t + R \widetilde{a}_{t-d} \tag{4.43}$$

The fact that $\widetilde{u}_{t-d}$ and $\widetilde{a}_{t-d}$ are uncorrelated with $\widetilde{a}_t$ can lead to

$$tr\left[\mathrm{cov}(\widetilde{y}_t)\right] = tr\left[\mathrm{cov}(\widetilde{G} \widetilde{u}_{t-d} + R \widetilde{a}_{t-d})\right] + tr\left[\mathrm{cov}(F \widetilde{a}_t)\right] \tag{4.44}$$

thus

$$tr\left[\mathrm{cov}(\widetilde{y}_t)\right] \geq tr\left[\mathrm{cov}(F \widetilde{a}_t)\right] \tag{4.45}$$

Similar to the case of right diagonal interactor, MVC law is achieved when

$$\widetilde{G} \widetilde{u}_{t-d} + R \widetilde{a}_{t-d} = 0 \tag{4.46}$$

or

$$\widetilde{u}_t = -\widetilde{G}^{-1} R \widetilde{a}_t \tag{4.47}$$

Inserting $\widetilde{u}_t = D_R^{-1} u_t$, $\widetilde{y}_t = F\widetilde{a}_t$ and $\widetilde{y}_t = q^{-d} D_L y_t$ into the above equation leads to

$$u_t = -D_R \widetilde{G}^{-1} R (q^d F)^{-1} D_L \cdot y_t \tag{4.48}$$

Therefore, the MVC law can be expressed as

$$Q|_{mv} = -D_R \widetilde{G}^{-1} R (q^d F)^{-1} D_L \tag{4.49}$$

which includes both the left and right diagonal interactor matrices. The MVC law under the single left or right diagonal interactor can be treated as a special case of the above general form by replacing D_R or D_L with identity matrix. Furthermore, the MVC outputs of interactor-filtered and original variables are given by

$$\widetilde{y}_t|_{mv} = F\widetilde{a}_t = F_0 \widetilde{a}_t + F_1 \widetilde{a}_{t-1} + \cdots + F_{d-1} \widetilde{a}_{t-d+1} \tag{4.50}$$

and

$$y_t|_{mv} = q^d D_L^{-1} \cdot \widetilde{y}_t|_{mv} = D_L^{-1} \left(F_0 D_R^{-1} a_{t+d} + F_1 D_R^{-1} a_{t+d-1} + \cdots + F_{d-1} D_R^{-1} a_{t+1} \right) \tag{4.51}$$

respectively. Similar to the invariance of output variance under interactor filtering given in Subsection 4.3.2, the minimum variance of the interactor-

filtered variables is still equivalent to that of the original variables because

$$\begin{aligned} tr\left[\mathrm{cov}(\widetilde{y}_t)\right] &= trE\left[q^{-d}D_L\big(y_t - E(y_t)\big)\big(y_t - E(y_t)\big)^T D_L^T q^d\right] \\ &= trE\left[\big(y_t - E(y_t)\big)^T \underbrace{D_L^T D_L}_{I}\big(y_t - E(y_t)\big)\right] \\ &= tr\left[\mathrm{cov}(y_t)\right] \end{aligned} \tag{4.52}$$

Thus the minimum variance is given by

$$\begin{aligned} J_{MV} &= tr\left[\mathrm{cov}(FD_R^{-1}a_t)\right] \\ &= tr\left[FD_R^{-1}\Sigma_a D_R F^T\right] \end{aligned} \tag{4.53}$$

For diagonal Σ_a,

$$J_{MV} = tr\left(F\Sigma_a F^T\right) = tr\left(F_0\Sigma_a F_0^T + F_1\Sigma_a F_1^T + \cdots + F_{d-1}\Sigma_a F_{d-1}^T\right) \tag{4.54}$$

which is the same as the result deducted under the single right diagonal interactor.

Given closed-loop operating data and the combined left/right interactors D_L/D_R, F can be obtained by extracting the first d coefficients of the moving average model of the interactor filtered output $q^{-d}D_L y_t$ and post-multiplying them with D_R. Then the corresponding MVC benchmark can be estimated from Eq. (4.54).

4.5 MIMO control performance improvement using adaptive weighting selection

Minimum variance benchmark is essentially based on LQ objective function and a weighting matrix representing the relative importance of dif-

ferent variables is often desired. However, there is no systematic method to optimize the weighting matrix to improve the multivariable control performance. Within the LQ control design framework, the appropriate selection of weighting matrix usually depends on engineer's experience or trial-and-error tuning strategy. Therefore, it cannot guarantee the best weighting selection to lead to the optimal control performance.

In this section, a new eigenvalue decomposition based iterative approach is developed for the adaptive selection of weighting matrix. The weighted LQ design objective function is formulated as

$$\min \ J = y_t^T W y_t \tag{4.55}$$

where W denotes the output weighting matrix and satisfies $W > 0$. In actual control design, W is often selected in the form of diagonal matrix which is more feasible to tune. Nevertheless, such a diagonal W is probably a suboptimal selection with inferior control performance, especially for strong interacting MIMO systems. Due to the interactions among different CVs, a suboptimal diagonal weighting matrix is unable to efficiently reduce the output variance inflation along certain directions, as illustrated in Fig. 4.2. To address this issue, it is desirable to construct a non-diagonal weighting matrix with more emphasis on the major axis direction of the output covariance ellipse while less emphasis on the minor axis direction.

The basic idea of the adaptive design of weighting matrix is to first rotate the original output variables to the major and minor axis directions in

Fig. 4.2 and then assign the relative importance of each direction according to its magnitude of variance inflation. The major and minor axis directions essentially correspond to the principal component (PC) directions of the output data, which can be found through eigenvalue decomposition on the output covariance matrix as follows

$$cov(y^{(0)}) = P\Lambda P^T \tag{4.56}$$

where $y^{(0)}$ denote an initial period of operating data collected under an initial weighting matrix W_0. P is the eigenvector matrix corresponding to the PC directions and $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$ contains the eigenvalues, which measure the variance inflation along different directions. After the rotation of the original coordinates to the PC directions, the original output data are actually projected to the PC space as

$$t = P^T y_t \tag{4.57}$$

where t denote the projected scores of the output y_t. A non-diagonal weighting matrix W_1 can be designed to replace W_0 and rotate the outputs to the PC directions. Consequently, the objective function is updated to

$$\min \ J = y_t^T W_1 y_t = t^T t \tag{4.58}$$

Substituting Eqs. (4.57) into (4.58) can lead to the expression of W_1 as

$$W_1 = PP^T \tag{4.59}$$

If the magnitude of variance inflation along each PC direction is further employed to quantify the relative importance of the corresponding direction, the

objective function in Eq. (4.58) can be further modified to

$$\min \ J = \left(\Lambda^{\frac{1}{2}} t\right)^T \left(\Lambda^{\frac{1}{2}} t\right) = t^T \Lambda t \tag{4.60}$$

Accordingly, the weighting matrix is equivalent to

$$W_1 = P\Lambda P^T = cov(y^{(0)}) \tag{4.61}$$

Since $cov(y^{(0)})$ is positive definite, $W_1 > 0$ thus holds. The above equation indicates that the covariance matrix of the initial output data can be used as the updated weighting matrix and there is no need to calculate the principal components explicitly. The objective function in Eq. (4.60) can be rewritten as

$$\min \ J = \left(\Lambda^{\frac{1}{2}} P^T y_t\right)^T \left(\Lambda^{\frac{1}{2}} P^T y_t\right) = y'^T_t y'_t \tag{4.62}$$

where the updated control design objective is shown to be equivalent to the quadratic form of the output rotated and scaled by $\Lambda^{\frac{1}{2}} P^T$. In the implementation, the weighting matrix can be iteratively updated by generating a period of operating data $y^{(i)}$ under the weighting W_i and then replacing W_i with $W_{i+1} = cov(y^{(i)})$. The iterations are continued until there is no further significant improvement of control performance.

4.6 Simulation examples

4.6.1 Simulated example 1: Illustrative derivation of MVC benchmark

First let us consider a simple 2×2 process with the following plant and disturbance models

$$G(q) = \begin{bmatrix} q^{-1} & 0 \\ q^{-1} & q^{-2} \end{bmatrix}$$

and

$$N(q) = \begin{bmatrix} \frac{1}{1-q^{-1}} & 0 \\ 0 & \frac{1}{1-q^{-1}} \end{bmatrix}$$

where the plant has only input-output time delays without any other dynamics and the disturbance $a(t)$ is assumed to be Gaussian white noise with covariance $\Sigma_a = I$. The right diagonal interactor is

$$D_R = \begin{bmatrix} q & 0 \\ 0 & q^2 \end{bmatrix}$$

with the interactor order $d = 2$. Then

$$\widetilde{N} = q^{-2} N D_R = \begin{bmatrix} \frac{q^{-1}}{1-q^{-1}} & 0 \\ 0 & \frac{1}{1-q^{-1}} \end{bmatrix} = F + q^{-2} R$$

with the factorized matrix F as

$$F = \begin{bmatrix} q^{-1} & 0 \\ 0 & 1+q^{-1} \end{bmatrix}$$

Thus the minimum variance benchmark is given by

$$J_{MV} = tr(F \Sigma_a F^T) = tr(F_0 F_0^T + F_1 F_1^T) = 3$$

and the MVC output is

$$y_{t|mv} = q^2 F \widetilde{a}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1+q^{-1} \end{bmatrix} \cdot a_t$$

As a comparison, the MVC benchmark is also derived from the left unitary interactor, which is

$$D_L = \begin{bmatrix} -\frac{q}{\sqrt{2}} & -\frac{q}{\sqrt{2}} \\ -\frac{q^2}{\sqrt{2}} & \frac{q^2}{\sqrt{2}} \end{bmatrix}$$

In this case, F is factorized from $\widetilde{N} = q^{-2} D_L N$ and given by

$$F = \begin{bmatrix} -\frac{q^{-1}}{\sqrt{2}} & -\frac{q^{-1}}{\sqrt{2}} \\ -\frac{1+q^{-1}}{\sqrt{2}} & \frac{1+q^{-1}}{\sqrt{2}} \end{bmatrix}$$

which leads to the MVC benchmark as follows

$$J_{MV} = tr(F_0 F_0^T + F_1 F_1^T) = 3$$

It can be seen that the derived MVC benchmarks from the left unitary and the right diagonal interactors are exactly the same. The consistent results in this illustrative example verify the validity of the right diagonal interactor for MVC benchmark derivation.

4.6.2 Simulated example 2: MVC benchmark estimation under right diagonal interactor

This example has been investigated by Huang et al. (Huang *et al.*, 1997*b*) who constructed a unitary interactor matrix using the complicated numerical algorithm (Rogozinski *et al.*, 1987; Peng and Kinnaert, 1992) and then derived the MVC benchmark and performance indices. The 2×2 multivariate process used in this example is described by Eq. 4.1 with the following plant and disturbance models

$$G(q) = \begin{bmatrix} \frac{q^{-1}}{1-0.4q^{-1}} & \frac{K_{12}q^{-2}}{1-0.1q^{-1}} \\ \frac{0.3q^{-1}}{1-0.1q^{-1}} & \frac{q^{-2}}{1-0.8q^{-1}} \end{bmatrix}$$

and

$$N(q) = \begin{bmatrix} \frac{1}{1-0.5q^{-1}} & \frac{-0.6}{1-0.5q^{-1}} \\ \frac{0.5}{1-0.5q^{-1}} & \frac{1.0}{1-0.5q^{-1}} \end{bmatrix}$$

The disturbance $a(t)$ is a two-dimensional normally distributed white noise sequence with zero mean and covariance $\Sigma_a = 0.1I$. This MIMO process does not have a left diagonal interactor matrix and thus the tedious numerical procedures are needed to factorize out the multivariate time delay structure in terms of general interactor. However, a right diagonal interactor does exist since it is the special input-delay process. The following right diagonal interactor matrix is simply obtained by extracting the common I/O delay terms along each input

$$D_R = \begin{bmatrix} q & 0 \\ 0 & q^2 \end{bmatrix}$$

with the interactor order $d = 2$. Since

$$\lim_{q^{-1}\to 0} D_L G(q) D_R = \begin{bmatrix} 1 & K_{12} \\ 0.3 & 1 \end{bmatrix}$$

the existence criterion is always satisfied except for $K_{12} = 10/3$ when the above matrix is rank deficient. Then following the simplified procedures given in Section 4.3, we have

$$\widetilde{N} = q^{-d} N D_R = \begin{bmatrix} \frac{q^{-1}}{1-0.5q^{-1}} & \frac{-0.6}{1-0.5q^{-1}} \\ \frac{0.5q^{-1}}{1-0.5q^{-1}} & \frac{1.0}{1-0.5q^{-1}} \end{bmatrix} = F + q^{-d} R$$

where

$$F = \begin{bmatrix} q^{-1} & -0.6 - 0.3q^{-1} \\ 0.5q^{-1} & 1 + 0.5q^{-1} \end{bmatrix}$$

Thus the MVC output can be calculated as

$$y_t|_{mv} = q^d F \widetilde{a}_t = \begin{bmatrix} 1 & -0.6 - 0.3q^{-1} \\ 0.5 & 1 + 0.5q^{-1} \end{bmatrix} \cdot a_t$$

with the minimum variance index under MVC law as

$$J_{MV} = tr(F\Sigma_a F^T) = tr\left(F_0\Sigma_a F_0^T + F_1\Sigma_a F_1^T + \cdots + F_{d-1}\Sigma_a F_{d-1}^T\right) = 0.2950$$

For the purpose of comparison, the MVC output and the minimum variance from the unitary interactor are listed below

$$y_t|_{mv} = \begin{bmatrix} 1 - 0.02752q^{-1} & -0.6 - 0.1623q^{-1} \\ 0.5 + 0.09176q^{-1} & 1 + 0.5412q^{-1} \end{bmatrix} \cdot a_t$$

and

$$J_{MV} = tr(F\Sigma_a F^T) = 0.2938$$

Using a minimum variance controller, the simulated output signals from the unitary interactor and the right diagonal interactor are compared in Fig. 4.3. The actual variance indices of simulated MVC outputs are

$$tr\left(\mathrm{cov}(y_t|_{mv})\right) = 0.2885 \quad \textit{under unitary interactor}$$

and

$$tr\left(\mathrm{cov}(y_t|_{mv})\right) = 0.2894 \quad \textit{under right diagonal interactor}$$

The mean square errors of the simulated outputs between these two methods are $MSE_{y_1} = 1.965 \times 10^{-3}$ and $MSE_{y_2} = 1.053 \times 10^{-3}$, respectively. It can be observed that the theoretical and simulated MVC benchmarks under two kinds of interactors are highly consistent. The tiny difference is probably caused by the numerical errors from the estimation of unitary interactor D and moving average coefficients F_i.

When a multiloop minimum variance controller based on two single loops without interaction compensation

$$Q_c = \begin{bmatrix} \frac{0.5-0.2q^{-1}}{1-0.5q^{-1}} & 0 \\ 0 & \frac{0.25-0.2q^{-1}}{1-0.25q^{-2}} \end{bmatrix} \tag{4.63}$$

is implemented on the process as discussed by Huang et al. (Huang *et al.*, 1997*b*), the MVC benchmark based performance index η is calculated with varying numerator gain K_{12} in plant model. The results from two kinds of interactors are compared in Fig. 4.4. As K_{12} increases, the interaction becomes stronger and the control performance thus deteriorates because the interaction part is not compensated by the multiloop controller. It can be readily seen that the performance index values calculated from those two interactors are in well agreement. The comparison demonstrates that the right diagonal interactor based approach provides an easy and accurate estimate of MVC benchmark for MIMO control performance assessment with reduced process model information.

Next the plant model is changed to

$$G(q) = \begin{bmatrix} \frac{q^{-1}}{1-0.4q^{-1}} & \frac{K_{12}q^{-10}}{1-0.1q^{-1}} \\ \frac{0.3q^{-1}}{1-0.1q^{-1}} & \frac{q^{-10}}{1-0.8q^{-1}} \end{bmatrix}$$

with the delay on the second input increased from 2 to 10. The right diagonal interactor becomes

$$D_R = \begin{bmatrix} q & 0 \\ 0 & q^{10} \end{bmatrix}$$

where the order of the interactor is also increased to 10. Then

$$\widetilde{N} = F + q^{-10}R$$

with F being factorized as

$$F = \begin{bmatrix} q^{-9} & -0.6 - \frac{0.6}{2}q^{-1} - \frac{0.6}{2^2}q^{-2} - \cdots - \frac{0.6}{2^9}q^{-9} \\ 0.5q^{-9} & 1 + \frac{1}{2}q^{-1} + \frac{1}{2^2}q^{-2} + \cdots + \frac{1}{2^9}q^{-9} \end{bmatrix}$$

thus the MVC output is given by

$$y_t|_{mv} = \begin{bmatrix} 1 & -0.6 - \frac{0.6}{2}q^{-1} - \frac{0.6}{2^2}q^{-2} - \cdots - \frac{0.6}{2^9}q^{-9} \\ 0.5 & 1 + \frac{1}{2}q^{-1} + \frac{1}{2^2}q^{-2} + \cdots + \frac{1}{2^9}q^{-9} \end{bmatrix} \cdot a_t$$

and the minimum variance is

$$J_{MV} = tr(F\Sigma_a F^T) = 0.3063$$

As the comparison, the MVC output from the unitary interactor is given by

$$y_t|_{mv} = \begin{bmatrix} -0.9578q^9 & -0.2873 \\ -0.2873q^9 & 0.9578 \end{bmatrix} \cdot Fa_t$$

where

$$F = \begin{bmatrix} -1.1014q^{-9} & 0.2874q^{-9} \\ 0.1916 + \frac{0.1916}{2}q^{-1} + \cdots + \frac{0.1916}{2^9}q^{-9} & 1.1302 + \frac{1.1302}{2}q^{-1} + \cdots + \frac{1.1302}{2^9}q^{-9} \end{bmatrix}$$

and the corresponding minimum variance is

$$J_{MV} = tr(F\Sigma_a F^T) = 0.3049$$

As the delay order is raised, the best achievable performance under MVC law is degraded slightly due to the increased time-delay restriction. The simulated MVC output signals under two kinds of interactors are compared in Fig. 4.5. The actual variance indices of the simulated MVC outputs are

$$tr\left(\mathrm{cov}(y_t|_{mv})\right) = 0.2961 \quad \textit{under unitary interactor}$$

and

$$tr\left(\mathrm{cov}(y_t|_{mv})\right) = 0.2983 \quad \textit{under right diagonal interactor}$$

The mean square errors of the simulated outputs between those two interactors are $MSE_{y_1} = 2.5 \times 10^{-3}$ and $MSE_{y_1} = 1.2 \times 10^{-3}$ for two output channels, respectively. Apparently both the theoretical MVC benchmark and the simulated MVC outputs from two kinds of interactors match well with negligible numerical errors. The comparison of the computational results under the increased delay order further verifies the validity of the simplified solution to MVC benchmark by using the right diagonal interactor.

4.6.3 Simulated example 3: MVC benchmark estimation under combined left/right diagonal interactors

The example investigated by Xia et al. (Xia *et al.*, 2006) is reconsidered in this subsection. The plant and disturbance models of the 2×2 MIMO process are given below

$$G(q) = \begin{bmatrix} \frac{q^{-(d-1)}}{1-0.4q^{-1}} & \frac{0.7q^{-d}}{1-0.1q^{-1}} \\ \frac{0.3q^{-(d-2)}}{1-0.1q^{-1}} & \frac{q^{-(d-1)}}{1-0.8q^{-1}} \end{bmatrix}$$

and

$$N(q) = \begin{bmatrix} \frac{1}{1-0.5q^{-1}} & \frac{-0.6q^{-1}}{1-0.6q^{-1}} \\ \frac{0.5q^{-1}}{1-0.7q^{-1}} & \frac{1.0}{1-0.8q^{-1}} \end{bmatrix}$$

The disturbance $a(t)$ is set as two-dimensional normally distributed white noise sequences with zero mean and covariance of $\Sigma_a = 0.1I$.

The above process transfer function matrix does not have a single left or right diagonal interactor because of the inherent rank deficiency caused by its

multivariate time-delay structure. For this process, however, there exist combined left/right diagonal interactors to characterize the time delay structure. The I/O delay matrix of this process is given by

$$D_{I/O} = \begin{bmatrix} q^{-(d-1)} & q^{-d} \\ q^{-(d-2)} & q^{-(d-1)} \end{bmatrix}$$

from which a pair of left and right diagonal interactors can be easily factorized as

$$D_L = \begin{bmatrix} q^{-(d-1)} & 0 \\ 0 & q^{-(d-2)} \end{bmatrix}$$

and

$$D_R = \begin{bmatrix} 1 & 0 \\ 0 & q \end{bmatrix}$$

The above left/right diagonal interactors are valid because

$$\lim_{q^{-1}\to 0} D_L G D_R = \lim_{q^{-1}\to 0} \widetilde{G} = \begin{bmatrix} 1 & 0.7 \\ 0.3 & 1 \end{bmatrix}$$

is a full rank finite matrix. Then following the procedures described in Section 4.4, the MIMO MVC benchmark can be calculated with any specified delay order d. Let $d = 3$, for instance, we can get

$$\widetilde{N} = q^{-d} D_L N D_R = \begin{bmatrix} \frac{q^{-1}}{1-0.5q^{-1}} & \frac{-0.6q^{-1}}{1-0.6q^{-1}} \\ \frac{0.5q^{-3}}{1-0.7q^{-1}} & \frac{q^{-1}}{1-0.8q^{-1}} \end{bmatrix} = F + q^{-d}R$$

with

$$F = F_0 + F_1 q^{-1} + F_2 q^{-2} = \begin{bmatrix} q^{-1} + 0.5q^{-2} & -0.6q^{-1} - 0.36q^{-2} \\ 0 & q^{-1} + 0.8q^{-2} \end{bmatrix}$$

It may be noticed that the interactor-filtered matrix $\widetilde{N}$ is actually delay-order invariant, although F changes as the order d increases. The MVC output is

then derived as

$$y_t|_{mv} = q^d D_L^{-1} F D_R^{-1} a_t = \begin{bmatrix} 1+0.5q^{-1} & -0.6q^{-1}-0.36q^{-2} \\ 0 & 1+0.8q^{-1} \end{bmatrix} \cdot a_t$$

From the unitary interactor

$$D = \begin{bmatrix} 0.9578q^2 & 0.2873q \\ -0.2873q^3 & 0.9578q^2 \end{bmatrix}$$

the MVC output can be computed as

$$y_t|_{mv} = \begin{bmatrix} 0.9578q^{-1}+0.4789q^{-2} & -0.2874q^{-2} \\ -0.2873-0.1437q^{-1}+0.4071q^{-2} & 1.1051q^{-1}+0.8546q^{-2} \end{bmatrix} \cdot a_t$$

The minimum variances from these two kinds of interactors are

$$J_{MV} = 0.345 \quad for\ unitary\ interactor$$

and

$$J_{MV} = 0.338 \quad for\ combined\ left/right\ diagonal\ interactor$$

In this case, it appears that the minimum variance from the combined left/right diagonal interactors is slightly smaller than that from the unitary interactor. The insignificant difference should come from the numerical errors of factorization of the moving average coefficients F_i. Actually, the numerical error disappears when the disturbance model is simplified to

$$N(q) = \begin{bmatrix} \frac{1}{1-q^{-1}} & 0 \\ 0 & \frac{1}{1-q^{-1}} \end{bmatrix}$$

where the minimum variance from both kinds of interactors is $J_{MV} = 0.4$. It indicates that the numerical error can be completely avoided if the disturbance model is simple enough. The simulated output signals under MVC law

from different interactors are plotted in Fig. 4.6. The mean square errors of the MVC outputs between those two methods are $MSE_{y_1} = 1.88 \times 10^{-2}$ and $MSE_{y_2} = 1.837 \times 10^{-2}$ for two controlled variables, respectively. In addition, the actual variances of the simulated MVC outputs from two kinds of interactors are given by

$$tr\left(\mathrm{cov}(y_t|_{mv})\right) = 0.3282 \quad for\ unitary\ interactor$$

and

$$tr\left(\mathrm{cov}(y_t|_{mv})\right) = 0.3207 \quad for\ combined\ left/right\ diagonal\ interactor$$

The comparison of the computational results indicates that the combined left/right diagonal interactors can be used to characterize the multivariate time-delay structure in a more straightforward way and result in the MVC benchmark with high accuracy.

If the MIMO multiloop controller in Eq. (4.63) is applied to the above process to close the loop, the minimum variance J_{MV} and the performance index η can be calculated with the varying delay order d, as shown in Fig. 4.7. It can be observed that the estimates of both indices from two kinds of interactors coincide well with each other. Under some delay orders, the J_{MV} estimates from the combined left/right diagonal interactor are a little smaller than that from the unitary interactor. This ignorable difference should be attributed to the numerical errors, which may arise from the estimation of the unitary interactor and the factorized F_i from $\widetilde{N}$. As the delay order is

increased, the increasing minimum variance implies that the theoretical lower bound is being pushed up by the added time-delay constraints. Meanwhile, the decreasing performance ratio index reveals that the performance of the multiloop controller is getting worse rapidly and there is bigger potential for performance improvement by retune the controller. This example shows that the combined left/right diagonal interactors are applicable for more general MIMO processes where the single left or right diagonal interactor is unavailable.

4.6.4 Simulated example 4: Adaptive weighting selection

In this subsection, a simulated example will be used to demonstrate the benefit of the proposed adaptive weighting selection strategy for multivariable control performance improvement. This example concerns a 2×2 multivariate process originally taken from Ko and Edgar (Ko and Edgar, 2001a) and has been modified on partial model parameters. The plant and disturbance models are given by

$$G(q) = \begin{bmatrix} \frac{-1.5q^{-3}}{1-0.659q^{-1}} & \frac{-0.167q^{-1}}{1-0.923q^{-1}} \\ \frac{-0.519q^{-5}}{1-0.784q^{-1}} & \frac{0.462q^{-3}+0.432q^{-4}}{1-0.874q^{-1}} \end{bmatrix}$$

and

$$N(q) = \begin{bmatrix} \frac{1-0.5q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} & \frac{-0.2q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} \\ \frac{-0.5q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} & \frac{1-0.8q^{-1}}{1-1.3q^{-1}+0.3q^{-2}} \end{bmatrix}$$

where the disturbance is assumed to be normally distributed white noise with zero mean and the covariance of $0.1I$. Suppose that there are no input/output constraints or plant/model mismatch. An unconstrained model predictive controller is designed starting with equal output weighting and zero input weight-

ing. The prediction and control horizons are set to 20 and 2, respectively. Under this controller setting, 2000 samples are collected initially for weighting adjustment. The covariance of the initial period of data can be calculated to update the output weighting matrix. Then the new controller is implemented to generate another 2000 samples for the further weighting update. Such procedures are iterated until there is no significant control performance improvement in terms of variance/covariance based indices.

In this example, the above procedures are repeated seven runs for the purpose of illustration and two performance indices, i.e. the trace and the determinant of output covariance matrix, are used to evaluate the magnitude of performance improvement. It needs to be emphasized that the trace of covariance is equivalent to the sum of variances for all output channels, while the determinant of covariance represents the volume of the output covariance ellipse as described in Chapter 2. Both indices can reflect the overall control performance of MIMO systems. By examining the evolution of performance indices in Figs. 4.8 and 4.9, one can find the similar trends of control performance as the output weighting matrix is iteratively updated. Both performance indices decrease rapidly after the first update of weighting matrix, which becomes

$$W_1 = \begin{bmatrix} 0.3207 & -0.1598 \\ -0.1598 & 0.5443 \end{bmatrix}$$

It means that the overall performance is improved dramatically due to the update of output weighting from the initial identity matrix to the above non-diagonal matrix. Such a non-diagonal weighting matrix points to the directions

with the largest variance inflation and put proportionally more control efforts on those directions to reduce the variability. It can be seen from Fig. 4.10 that the output covariance ellipse shrinks most significantly after the first update of weighting matrix. The volume of covariance ellipse appears to be compressed largely because the original coordinates are rotated to the orthogonal PC directions and thus the variation in those directions can be reduced more efficiently by control actions. More specifically, the variance in the first PC direction is reduced most because of the largest weighting assigned to this direction. Besides the overall performance, the univariate performance of both controlled variables after the first weighting update is also improved and pushed much closer to the corresponding univariate MVC bounds, as shown in Figs. 4.11 and 4.12. Notice that, different from the MIMO MVC benchmark, the above univariate MVC bounds are obtained by minimizing $var(y_1)$ and $var(y_2)$ independently. Owing to the interactions between those two CVs, the univariate bounds of y_1 and y_2 cannot be achieved simultaneously. The fact that the performance improvement of CV2 is more aggressive than that of CV1 may be explained by the orientation of PC1, which has a relatively smaller angle with CV2. The second update of weighting matrix

$$W_2 = \begin{bmatrix} 0.2552 & -0.0097 \\ -0.0097 & 0.2189 \end{bmatrix}$$

further promotes the overall performance with attenuated magnitude, as shown Figs. 4.8 and 4.9. The first weighting update compresses the covariance ellipse along the first PC direction and thus the second PC direction becomes the dominant one with the largest variance inflation. This explains why the

performance improvement from the second weighting update occurs primarily along the second PC direction in Fig. 4.10. As a result, it can be seen from Fig. 4.12 that the variance of CV1 is reduced more than that of CV2 after the second weighting update.

In addition to the first two updates of weighting matrix, the subsequent iterations seem not beneficial to the overall performance improvement because the covariance indices tend to be flat in Figs. 4.8 and 4.9. The reason is that, in this case, the multivariate control performance after the first two weighting updates has been close enough to the optimal boundary and there is no significant margin to further improve the performance. Therefore, in Figs. 4.10 and 4.12, the covariance ellipses and the univariate variances of two CVs almost coincide since the third update. The optimal iteration number in this example can thus be determined as $l = 2$ according to the performance trends of multiple iterations. In practice, this adaptive weighting selection method is more useful to find a suitable non-diagonal weighting matrix for strong interacting systems because the output covariance tends to be more elliptical.

4.7 Summary

In this chapter, simplified solutions to MVC benchmark based MIMO control performance monitoring and improvement are presented. In order to avoid the undesirable requirement of model information and the numerical complexity of calculating the general interactor, the special right interactor

in terms of a diagonal matrix is first adopted to derive the MVC benchmark. Some properties of right diagonal interactor, such as existence and uniqueness, are investigated and the simplified procedures to estimate the MVC benchmark using right diagonal interactor is established. To deal with processes with more complicated time-delay structure, the above idea is further generalized to construct the combined left/right diagonal interactors. Accordingly, the method of computing the MVC benchmark with the combined left/right diagonal interactors is also developed. It is shown that the single left or right diagonal interactor case can be unified into the general framework under the combined interactors. Although the existence of left/right diagonal interactors is not guaranteed, the proposed solution is still applicable for a wide class of MIMO processes. On the other hand, considering the fact that the underlying LQ objective of MIMO MVC benchmark is subjected to output weightings, an adaptive selection strategy of weighting matrix is designed to improve the control performance more efficiently. This method does not require a priori model information and is only based on the covariance of closed-loop output data. The simulation examples have demonstrated that the left/right diagonal interactor based approach can result in the accurate MVC benchmark with the alleviated model requirement and simplified computational procedures. In addition, the weighting selection strategy is illustrated to be capable of pushing the control performance close enough to the MVC boundary.

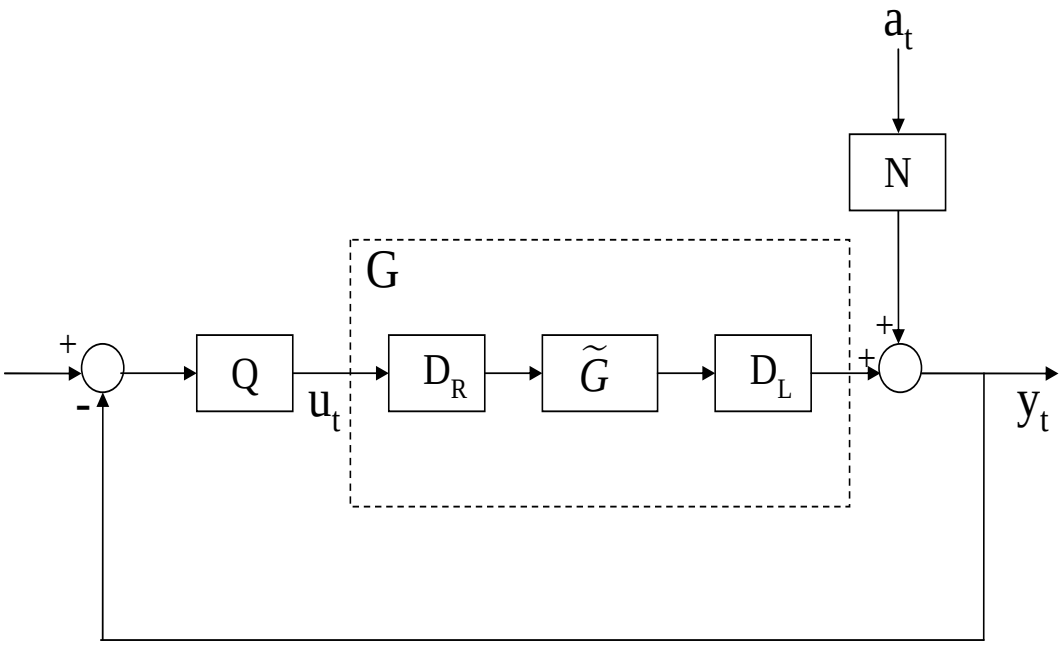

Figure 4.1: Systematic diagram of the combined left/right interactors in a closed-loop system

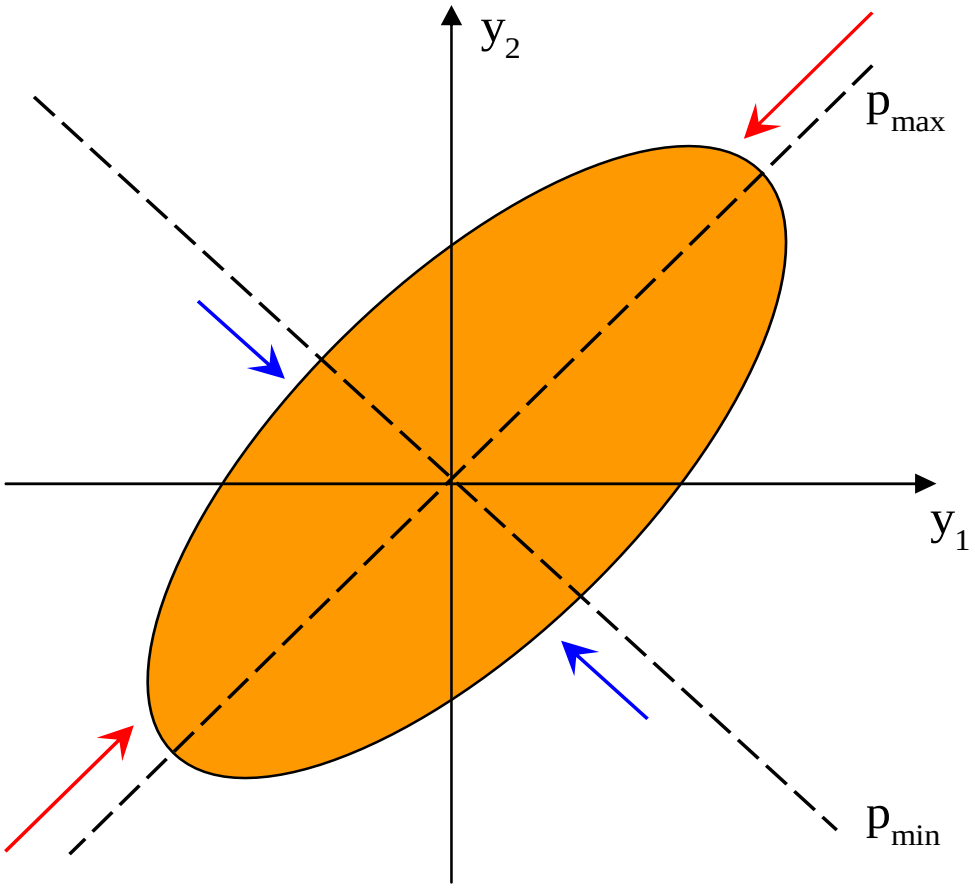

Figure 4.2: Illustration of weighting matrix design based on output variance inflation

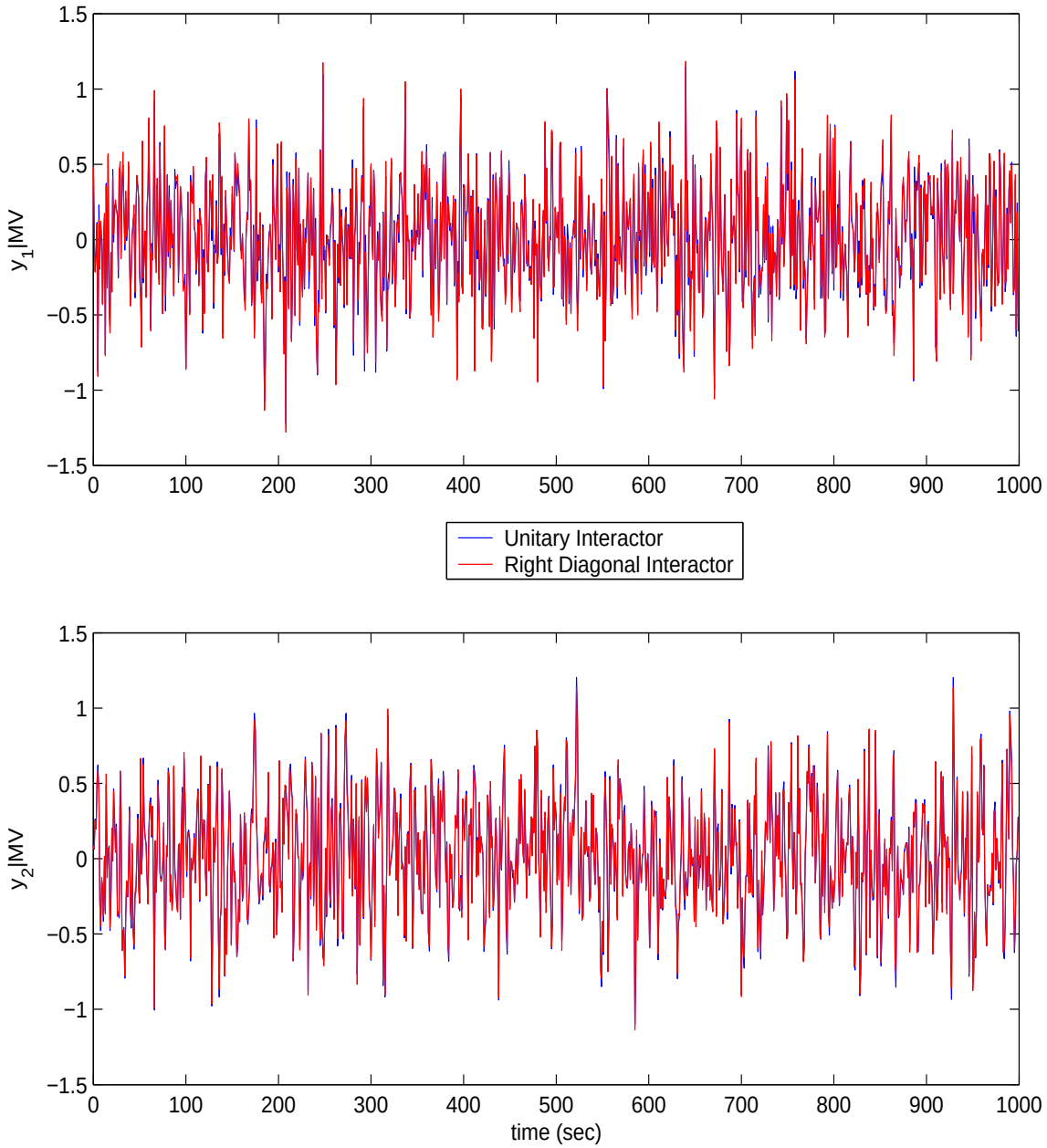

Figure 4.3: Comparison of simulated MVC output signals from unitary interactor and right diagonal interactor in simulated example 1

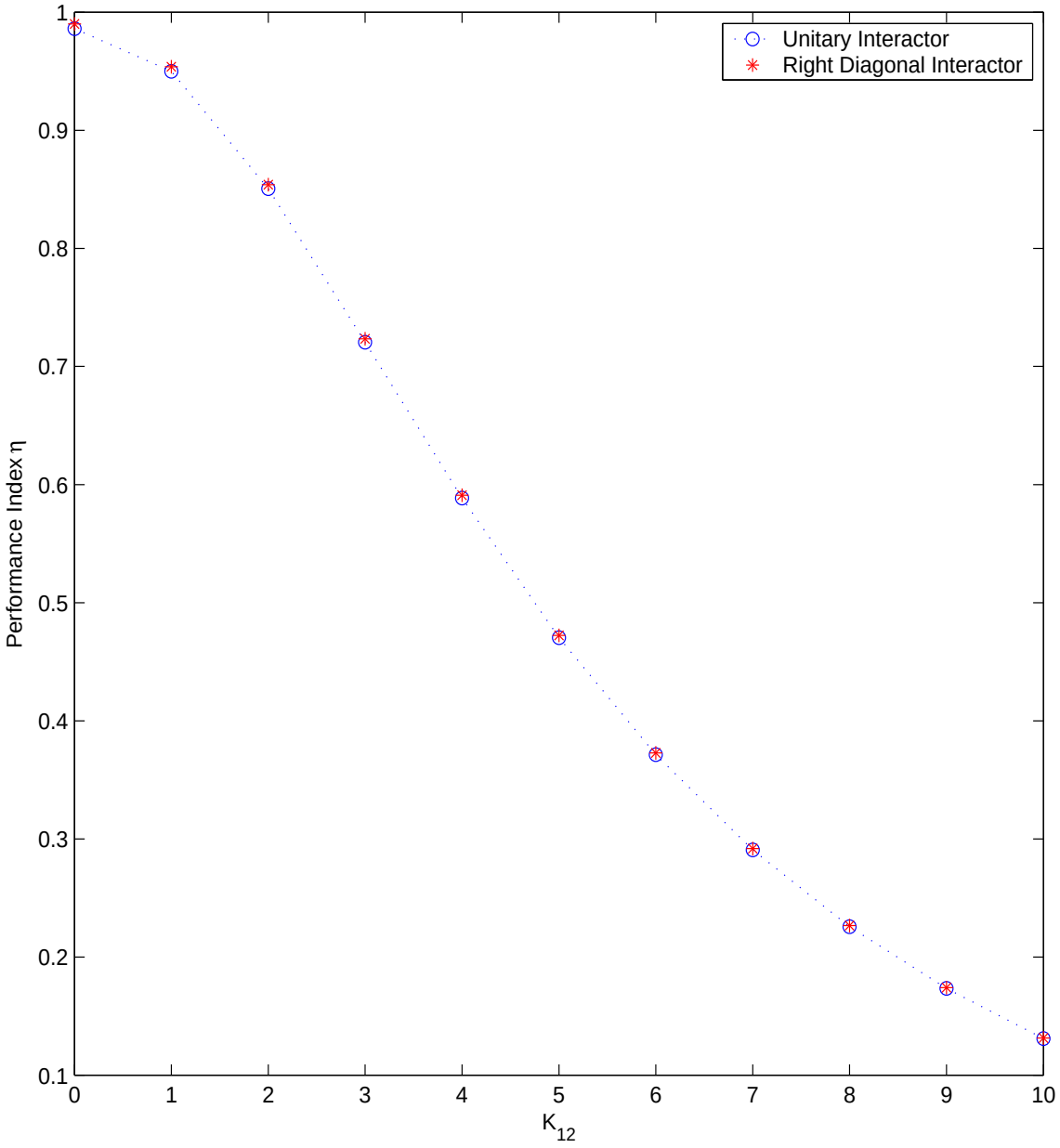

Figure 4.4: Comparison of MVC benchmark based performance index from unitary interactor and right diagonal interactor in simulated example 1

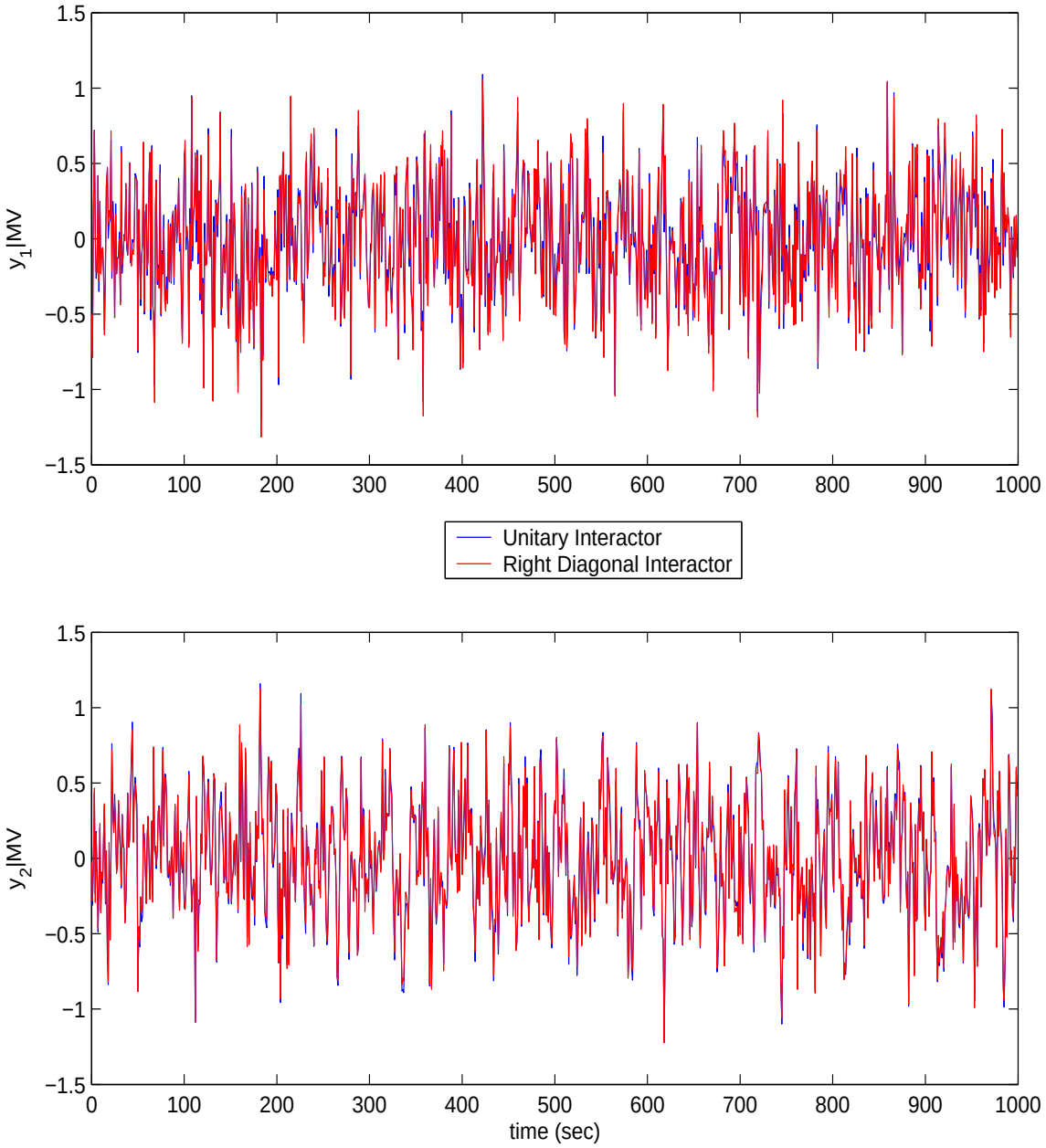

Figure 4.5: Comparison of simulated MVC output signals from unitary interactor and right diagonal interactor in simulated example 1 with increased delay order

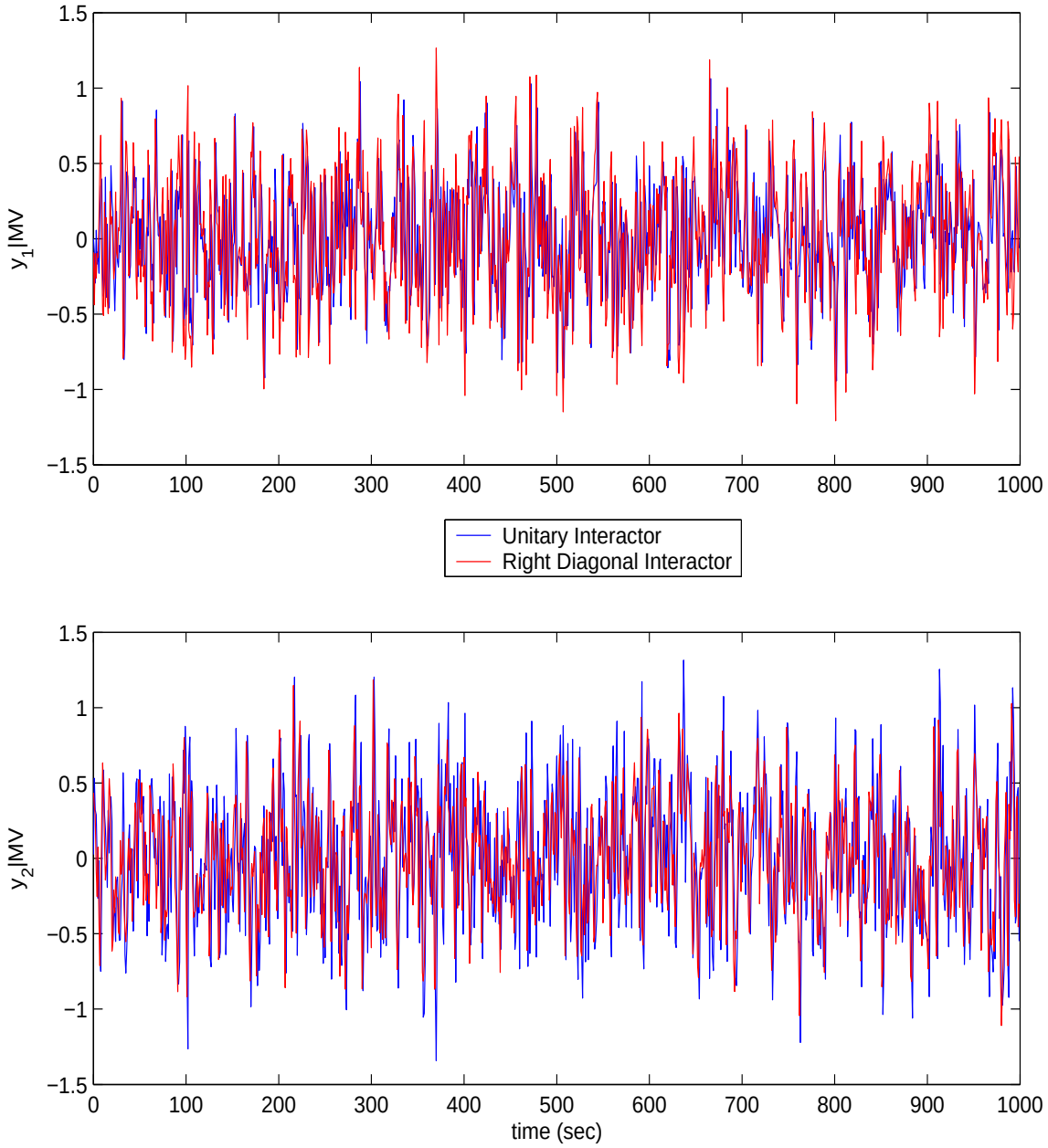

Figure 4.6: Comparison of simulated MVC output signals from unitary interactor and combined left/right diagonal interactors in simulated example 2

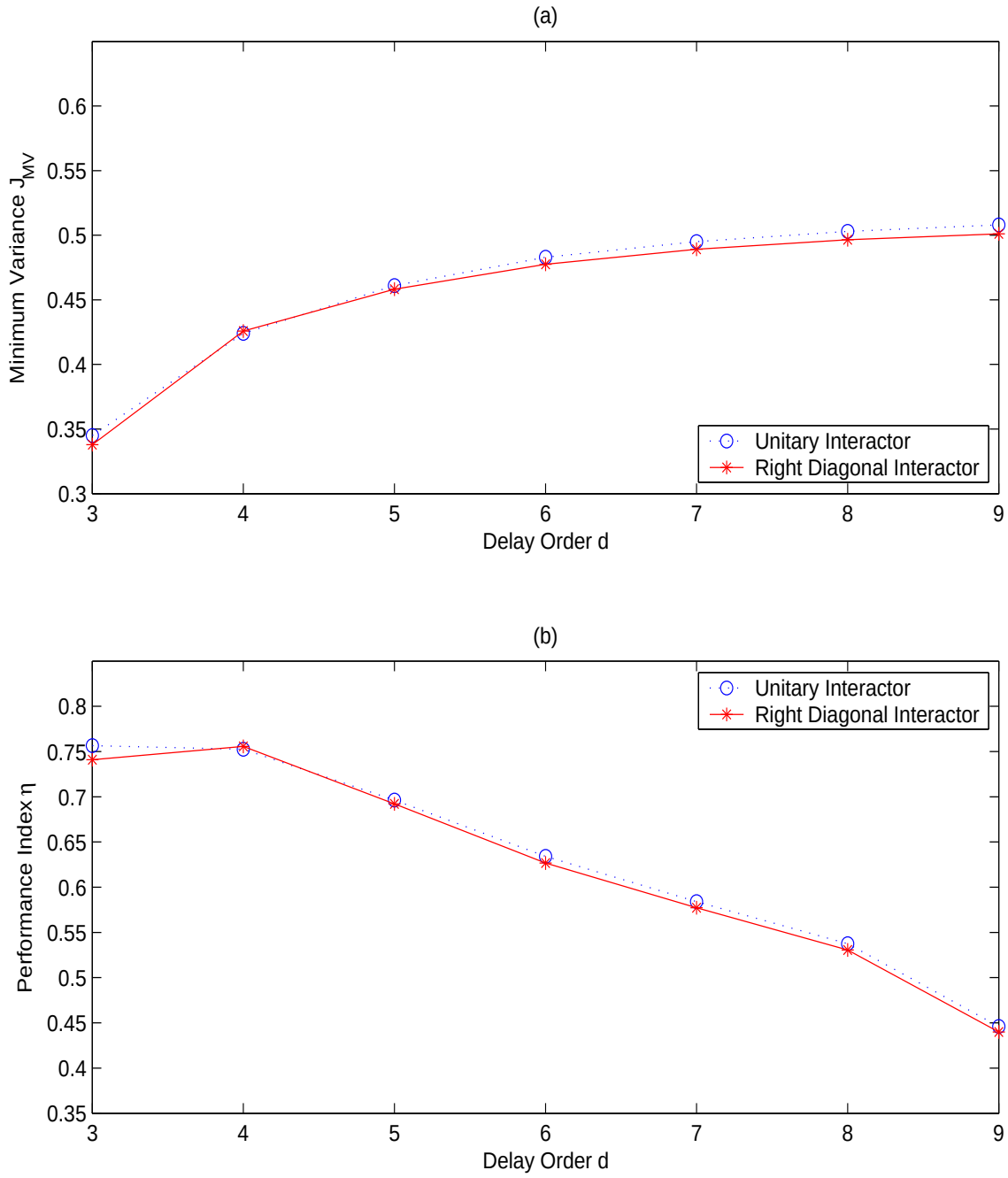

Figure 4.7: Estimates of (a) minimum variance J_{MV} and (b) performance index η from unitary interactor and combined left/right diagonal interactors in simulated example 2

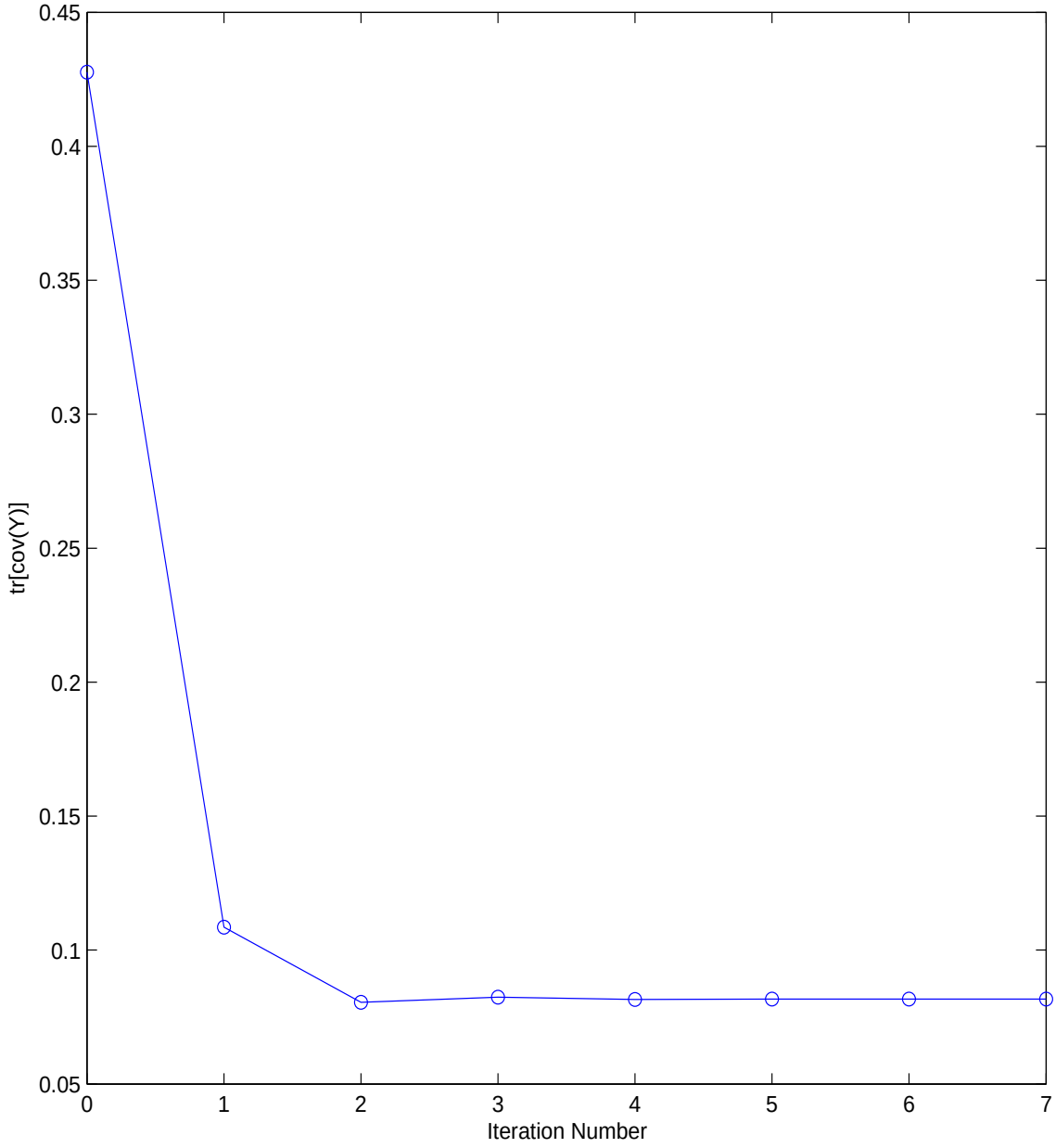

Figure 4.8: Performance trend in terms of trace of output covariance with iterative weighting update in simulated example 3

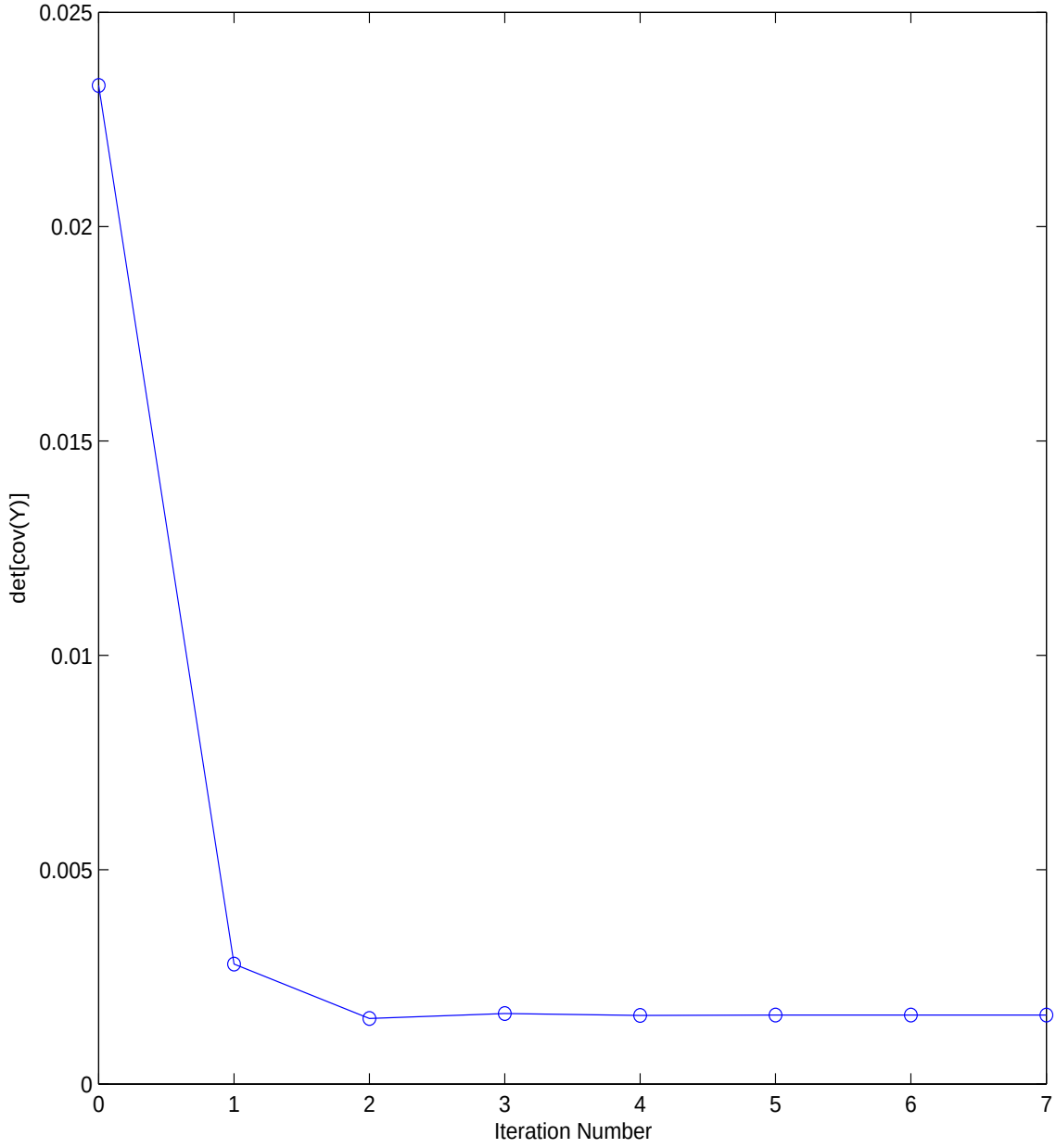

Figure 4.9: Performance trend in terms of determinant of output covariance with iterative weighting update in simulated example 3

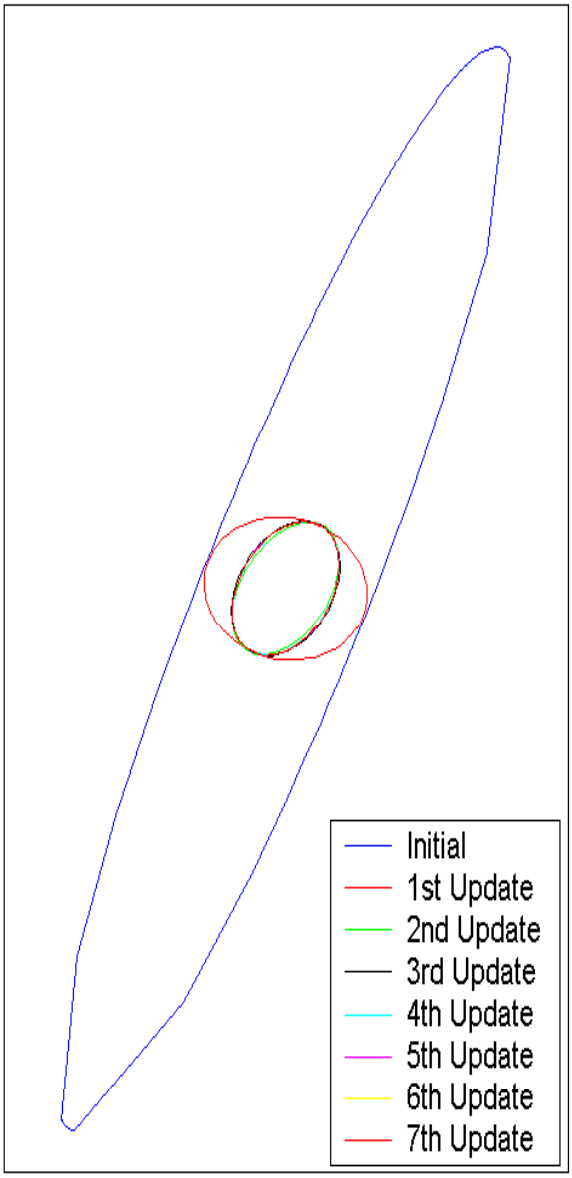

Figure 4.10: Comparison of output covariance ellipses with iterative weighting update in simulated example 3

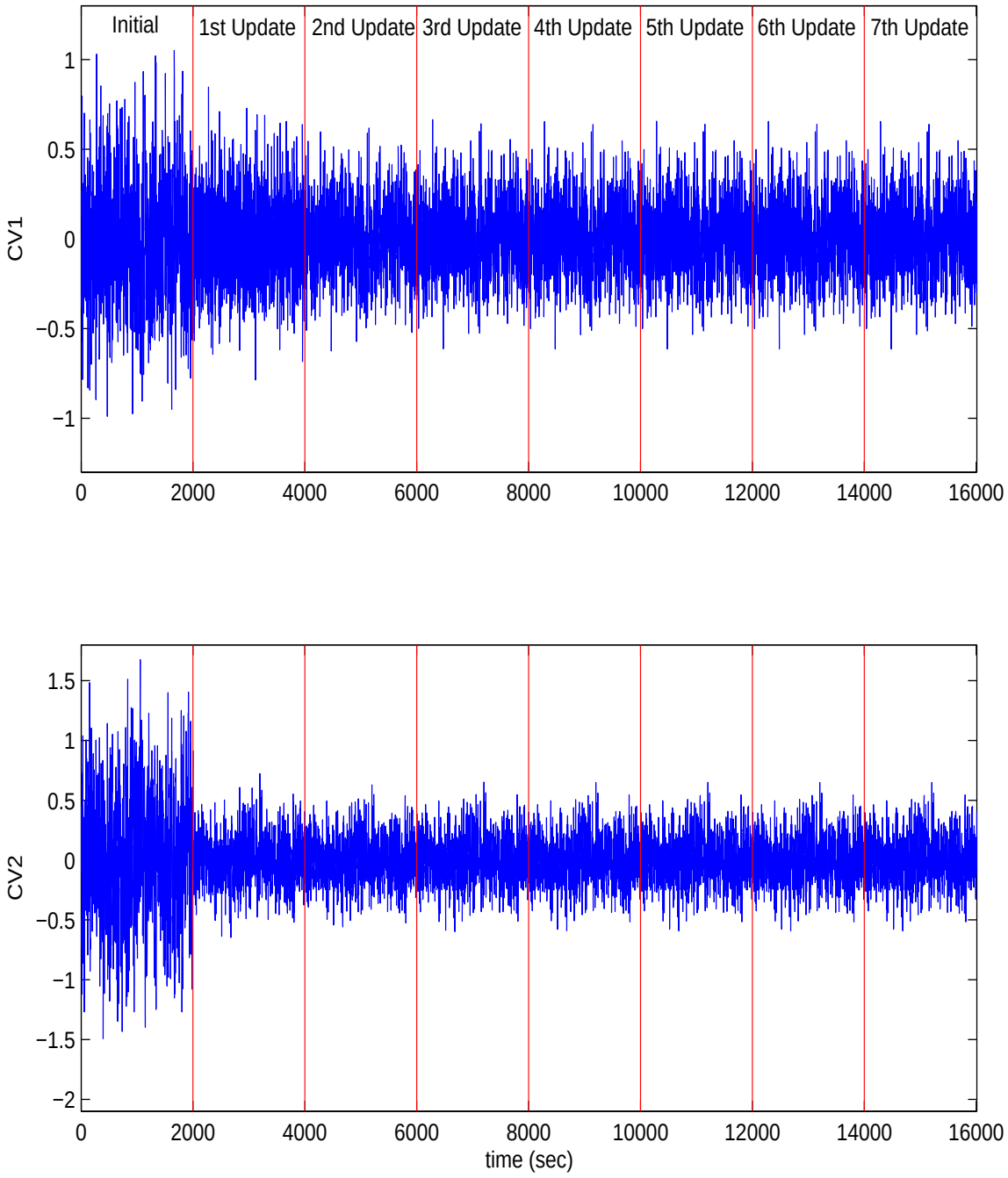

Figure 4.11: Comparison of output signals with iterative weighting update in simulated example 3

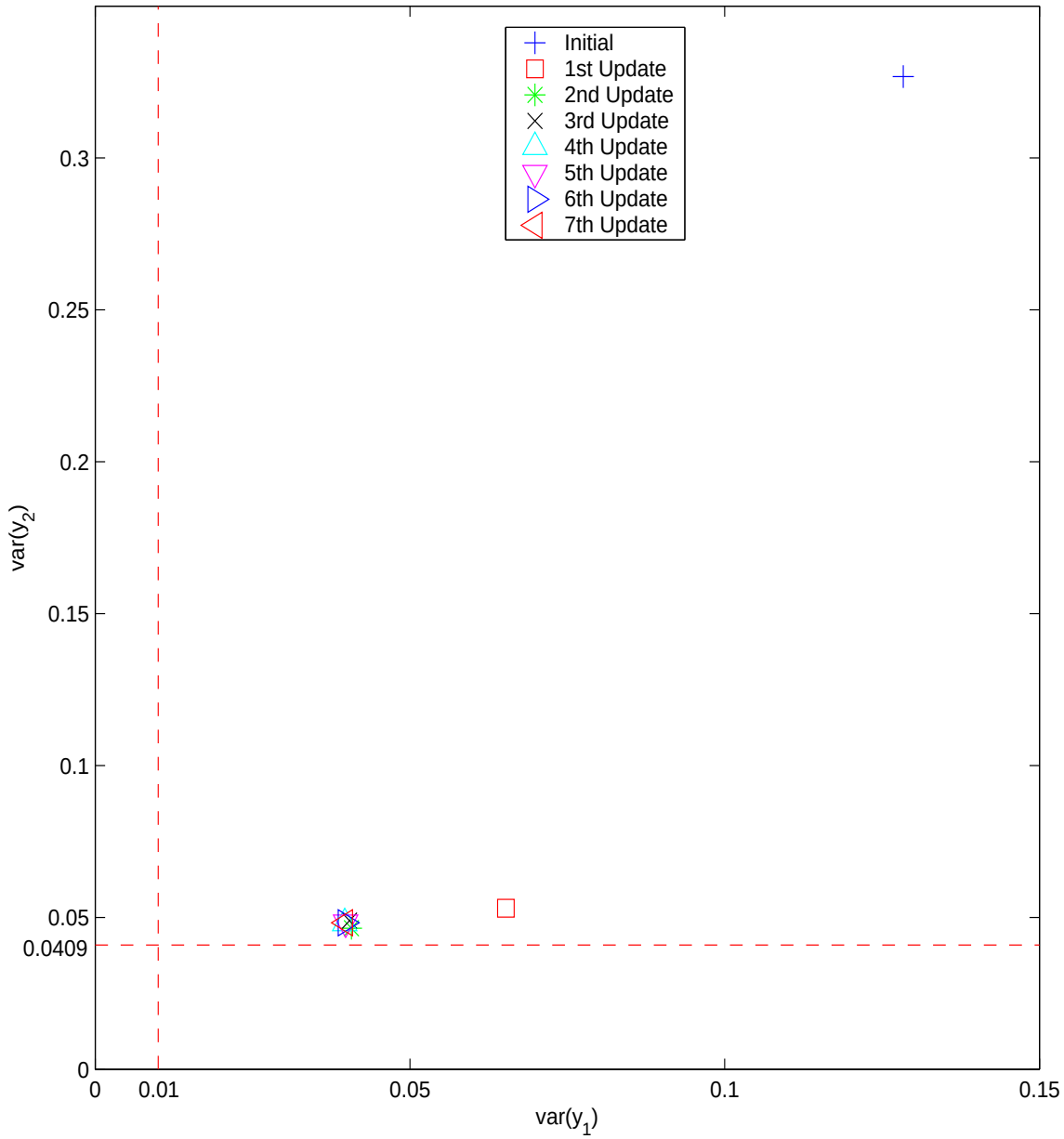

Figure 4.12: Performance comparison in terms of single output variance with lower bounds under different iterations of weighting updates in simulated example 3

Chapter 5

Variance component analysis based fault diagnosis of multi-layer overlay lithography processes

5.1 Introduction

Lithography process is one of the most critical steps in semiconductor manufacturing and largely responsible for driving improvement in the design of device circuits (Bode *et al.*, 2004). During the fabrication, the lithography process is utilized repeatedly to make each layer properly sized and aligned with its adjoining layer to create a functional device. Hence overlay constitutes one of the key metrics related to the lithography processes. In multi-stepper lithography processes, the displacement error of an exposed image field relative to the previous field is a major source of process faults. There are a number of physical reasons leading to overlay misalignment errors, such as reticle distortion errors, exposure tool image field distortion, alignment errors, tracking errors, and wafer distortion errors from processing (Levinson, 1999).

The major difficulty for overlay process monitoring is due to the large amount of required information and complicated variation propagation throughout the multi-stage processes. The quality information flow of the products in multistage manufacturing systems and the complex interaction between the

process faults and the product quality characteristics make the issue very challenging (Ding *et al.*, 2002; Zhou *et al.*, 2003). For instance, the effects of certain process faults on the product quality in the intermediate stage may propagate along the multi-layer lithography processes and different types of faults probably have similar manifestation on the final product quality. In addition, the absence of full in situ metrology means that only partial intermediate and final metrological patterns are available. Due to all of these reasons, some powerful process monitoring techniques such as PCA, PLS and FDA are unable to deal with the multi-stage overlay lithography processes.

Aimed at this objective, a variance component analysis based approached is applied to solve the challenging multi-layer overlay process monitoring problem. First a state-space model for overlay processes is derived from the fundamental physical principles between the photolithography steppers and the metrology sensors in Section 5.2. Then a general input-output linear formulation is developed from the state-space realization and the VCA technique is employed to estimate the mean and variance components in different layers in Section 5.3. A hypothesis testing method is adopted to detect the significant faults in multi-stepper overlay processes and the estimated mean/variance components can be further used to diagnose the magnitude and orientation of misalignment errors. In Section 5.4, the above procedures are examined in a series of simulated examples and the computational results are presented to verify the validity and effectiveness of this diagnosis method for overlay lithography process monitoring. The chapter is summarized in Section 5.5.

5.2 Modeling of Multi-stage Overlay Lithography Processes

Fundamental overlay lithography process model has been proposed to describe the quantitative relationship between the photolithography steppers and the metrology sensors (Brink *et al.*, 1988). The model concerning the overlay accuracy can be divided into two aspects based on the sources of errors. One part of the model is focused on the intrafield sources produced by the fitting problems between the light source filter lens and mask within the same image field. The other part is associated with the interfiled sources of errors resulting from the relative positions between mask and wafer. The overlay misalignment error sources are illustrated in Fig. 5.1 (Bode, 2001). The mathematical model of intrafield error sources can be expressed as follows

$$\Delta e_r^x = T_{rx} + (M_r + M_a)x_r - (R_r + R_a)y_r + O(x_r^2, y_r^2) \tag{5.1}$$

and

$$\Delta e_r^y = T_{ry} + (M_r - M_a)y_r + (R_r - R_a)x_r + O(x_r^2, y_r^2) \tag{5.2}$$

where M_r is the reticle magnification and M_a denotes asymmetric magnification. R_r represents the reticle rotation and R_a is asymmetric reticle rotation. T_{rx} and T_{ry} are the translation error components in the x and y directions, respectively. x_r and y_r are the die coordinates with the center of die as the origin. $o(x_r^2, y_r^2)$ stands for the second or higher order terms.

Similarly, the interfiled model may be formulated as

$$\Delta e_g^x = T_{gx} + S_x x_g - (R_g + R_n)y_g + o(y_g^2) \tag{5.3}$$

and

$$\Delta e_g^y = T_{gy} + S_y y_g + R_g x_g + o(x_g^2) \tag{5.4}$$

where S_x and S_y are grid scales. R_g and R_n are grid orthogonal and non-orthogonal rotation. x_g and y_g represent the grid coordinates with the center of wafer as the origin. $o(x_g^2)$ and $o(y_g^2)$ are the second or higher order terms.

If the second or higher order terms are ignored, then the above model for each layer can be approximated as linear model with two parts. The first part is intrafield model as follows

$$\Delta e_r^x = T_{rx} + (M_r + M_a)x_r - (R_r + R_a)y_r \tag{5.5}$$

and

$$\Delta e_r^y = T_{ry} + (M_r - M_a)y_r + (R_r - R_a)x_r \tag{5.6}$$

The other part is interfiled model given by

$$\Delta e_g^x = T_{gx} + S_x x_g - (R_g + R_n)y_g \tag{5.7}$$

and

$$\Delta e_g^y = T_{gy} + S_y y_g + R_g x_g \tag{5.8}$$

In every stage, let the state variable be

$$x(k) = \begin{bmatrix} x_g(k) \\ x_r(k) \\ y_g(k) \\ y_r(k) \end{bmatrix} \tag{5.9}$$

where $x(k)$ represents the intrafield and interfiled coordinate information of each layer.

Furthermore, the system input $u(k)$ and output measurement $y(k)$ are defined as

$$u(k) = \begin{bmatrix} \Delta e_g^x(k) \\ \Delta e_r^x(k) \\ \Delta e_g^y(k) \\ \Delta e_r^y(k) \end{bmatrix} \tag{5.10}$$

and

$$y(k) = \begin{bmatrix} \Delta e^x(k) \\ \Delta e^y(k) \end{bmatrix} = \begin{bmatrix} \Delta e_g^x(k) \\ \Delta e_g^y(k) \end{bmatrix} + \begin{bmatrix} \Delta e_r^x(k) \\ \Delta e_r^y(k) \end{bmatrix} \tag{5.11}$$

where the input variable $u(k)$ and the output $y(k)$ denote the potential process faults and the metrology errors of products in each layer, respectively. Assume that the components of $u(k)$ are independent of each other. The general linear state space model can be derived to characterize the overlay lithography process in the mode of layer by layer

$$x(k+1) = A(k)x(k) + B(k)u(k) + w(k) \tag{5.12}$$

and

$$y(k) = C(k)x(k) + v(k) \tag{5.13}$$

where $w(k)$ and $v(k)$ are process noise and measurement noise, respectively. The N-layer overlay process is illustrated in Fig. 5.2.

In the overlay process model, if the system noise terms are considered small enough and negligible, the state space realization may be simplified to the following form

$$x(k+1) = A(k)x(k) + B(k)u(k) \tag{5.14}$$

and

$$y(k) = C(k)x(k) \tag{5.15}$$

then we have

$$\begin{aligned} y(k) &= \begin{bmatrix} \Delta e_g^x(k) \\ \Delta e_g^y(k) \end{bmatrix} + \begin{bmatrix} \Delta e_r^x(k) \\ \Delta e_r^y(k) \end{bmatrix} \\ &= \begin{bmatrix} S_x & M_r + M_a & -(R_a + R_n) & -(R_r + R_a) \\ R_g & R_r - R_a & S_y & M_r - M_a \end{bmatrix} \cdot \begin{bmatrix} x_g(k) \\ x_r(k) \\ y_g(k) \\ y_r(k) \end{bmatrix} \\ &= C(k) \cdot x(k) \end{aligned} \tag{5.16}$$

with

$$C(k) = \begin{bmatrix} S_x & M_r + M_a & -(R_a + R_n) & -(R_r + R_a) \\ R_g & R_r - R_a & S_y & M_r - M_a \end{bmatrix} \tag{5.17}$$

and the translational terms are ignored due to the normalization of data.

Further, the state update can be written as

$$x(k+1) = \begin{bmatrix} x_g(k) \\ x_r(k) \\ y_g(k) \\ y_r(k) \end{bmatrix} + \begin{bmatrix} \Delta e_g^x(k) \\ \Delta e_r^x(k) \\ \Delta e_g^y(k) \\ \Delta e_r^y(k) \end{bmatrix} = x(k) + u(k) \tag{5.18}$$

such that $A(k) = I_{4\times4}$ and $B(k) = I_{4\times4}$.

5.3 Root-cause identification based on mean and variance component estimation

5.3.1 Variance component analysis

First define the state transition matrix ϕ and input-output transition matrix γ as

$$\Phi(k,i) = \begin{cases} A_{k-1}A_{k-2}\cdots A_i & (i<k) \\ I_{n_x} & (i=k) \end{cases}$$

and

$$\gamma(k,i) = \begin{cases} C(k)\Phi(k,i)B(i) & (i \le k) \\ 0 & (i>k) \end{cases}$$

Without loss of generality, set the initial state $x_0 = 0$. The state-space model in Section 2 is then transformed into a linear input-output model as follows (McCulloch and Searle, 2001; Zhou *et al.*, 2004)

$$y(k) = \sum_{i=1}^{k} \gamma(k,i)u(i) \tag{5.19}$$

Given M samples, let $Y_j^T = \left[y_j^T(1), y_j^T(2), \cdots, y_j^T(N)\right]$ be the output measurements of the jth sample along all the N layers. Stacking the M samples along all the N layers yields

$Y^T = \left[Y_1^T \cdots Y_j^T \cdots Y_M^T\right]$,

$\Gamma^T = \left[\underbrace{\gamma^T \cdots \gamma^T \cdots \gamma^T}_{M}\right]$,

and

$Z = block \quad diag(\underbrace{\gamma, \cdots, \gamma, \cdots \gamma}_{M})$

The stochastic fault-quality model for the M samples in all layers is then written as (Zhou *et al.*, 2004)

$$Y = \Gamma U + Z\widetilde{u} \tag{5.20}$$

where U and $\tilde{u}$ represent the fixed and random effects of $u(i)$, respectively. The above model is also termed as a general mixed linear model.

Based on the above model, the mean and covariance are given by

$$\mu_Y = E(Y) = \Gamma U \tag{5.21}$$

and

$$\Sigma_Y = \mathrm{cov}(Y) = F_1\sigma_{u_1}^2 + \cdots + F_P\sigma_{u_P}^2 \tag{5.22}$$

where P denotes the number of potential faults along all layers and $F_i = \sum_{k=1}^{M} Z_{:p(k)} {Z_{:p(k)}}^T$ with $p(k) = (k-1) \times P + i$. Since there are four potential faults in each layer of overlay processes, then we have $P = 4N$.

Using MINQUE algorithm (Rao and Kleffe, 1988; Searle *et al.*, 1992; McCulloch and Searle, 2001), the variance components can be estimated by solving the following linear equations

$$S^T\widehat{\Sigma} = q \tag{5.23}$$

where $q^T = [q_1, \cdots, q_j, \cdots, q_P]$ with $q_j = Y^T(M\Sigma_0 M)^- F_j (M\Sigma_0 M)^- Y$, $M = I - \Gamma(\Gamma^T\Gamma)^-\Gamma^T$, and $[S_{ij}] = tr(F_i(M\Sigma M)^- F_j(M\Sigma M)^-)$. Here A^- denotes the Moore-Penrose pseudoinverse of a matrix A and Σ_0 is the initial selection of variance components. From the above variance component estimates, the least-square estimator of mean component U is given by

$$\widehat{U} = (\Gamma^T\Sigma^{-1}\Gamma)^-\Gamma^T\Sigma^{-1}Y \tag{5.24}$$

5.3.2 Hypothesis testing for root cause identification

As is known, normal process variations cannot be completely avoided in the multi-layer overlay lithography processes. In this case, a design tolerance threshold ϵ_i is needed to determine if there is significant bias/variance error occurring in the *ith* layer. The one-sided hypothesis test on the variance components is established as follows (Zhou *et al.*, 2004; Seifert, 1993; Khuri *et al.*, 1998)

$$H_0 : \sigma_i^2 \leq \epsilon_i \tag{5.25}$$

vs

$$H_a : \sigma_i^2 > \epsilon_i \tag{5.26}$$

where $i = 1, 2, \ldots, P$. The threshold ϵ_i is chosen based on process knowledge or technical specifications. Similarly, the source of bias error may be identified by the two-sided hypothesis testing on the mean components as

$$H_0 : \mu_i = 0 \tag{5.27}$$

vs

$$H_a : \mu_i \neq 0 \tag{5.28}$$

Thus, the MINQUE based test statistics for mean and variance components μ_i and σ_i^2 can be constructed as

$$T_0 = \frac{\mu_i}{\sqrt{[(\Gamma^T \Sigma^{-1} \Gamma)^-]_{ii}}} \quad for\ mean\ components \tag{5.29}$$

and

$$T_1 = \frac{\sigma_i^2 - \epsilon_i}{\sqrt{[S^{-1}]_{ii}}} \quad for\ variance\ components \tag{5.30}$$

Under normal approximation (Kleffe and Seifert, 1988; Zhou *et al.*, 2004) and the confidence level $(1-a)\times 100\%$, the statistical decision rules are stated as follows:

Reject H_0 in Eq. (5.25) if $|T_0| > \Phi^{-1}(1-\alpha/2)$

and

Reject H_0 in Eq. (5.27) if $T_1 > \Phi^{-1}(1-\alpha)$.

If the null hypothesis H_0 in Eqs. (5.25) or (5.27) is rejected, then we may conclude that significant variance or bias error is present in the corresponding layer.

5.4 Applications to multi-layer overlay lithography processes

In this section, variance component analysis method is applied to detect and diagnose the misalignment errors in the multi-layer overlay lithography processes. As summarized in Table 5.1, a number of simulated examples are designed to test the effectiveness of the VCA approach for overlay process monitoring.

5.4.1 Simulated example 1: three-layer 36×4 overlay process

First, a three-layer overlay process with 36 sample points per field and 4 sample fields per wafer is simulated. The schematic diagram of the process is displayed in Fig. 5.3 (a) and (b). The physical parameter values used in this simulated example are listed in Table 5.2 (Lin and Wu, 1999). The state

matrices are

$$A(k) = B(k) = I_{4\times 4}$$

and

$$C(k) = \begin{bmatrix} 100 & -8240 & 0 & -100 \\ 100 & 100 & 100 & 153 \end{bmatrix}$$

In the overlay process simulation, the normal operational condition is assumed for the $1st$ and $3rd$ layers based on the original physical model. However, two scenarios of misalignment errors are designed and applied to the $2nd$ layer. In the first scenario, a positive bias error of magnitude 6 along both X and Y directions is added with normal variance of $\sigma^2 = 0.1$. Due to the error propagation along the multiple layers, the metrology data in the third layer will be driven away from the targets by the fault in the second layer. The test statistics for the mean and variance components are shown in Fig. 5.4. As described in Section 5.2, each layer involves four components representing the intrafield and interfield error information along both X and Y directions. It can be easily seen that the test statistics for both mean and variance components in Layer II, i.e. components 5 through 8, exceed the 95% confidence limits. Hence it indicates that there exist significant bias and variance errors in the second layer. For the components in Layers I and III, however, the test statistics for mean and variance components are all below the confidence limit lines. Therefore, it is concluded that no bias or variance error occurs in these two layers. To find out more about the bias error in the second layer, the mean component estimates are plotted in Fig. 5.5. The bias error is inferred to be along the positive X and Y directions because of the upward bars of the

estimated mean components in the second layer. In addition, the estimated mean components in Layer II are around 6, which implies the magnitude of bias error is 6. The diagnosis results based on mean component estimation are consistent with the simulation design. The test statistics and the confidence limits can be used to determine if there is an active fault happening in any layer. Then the mean component estimates provide diagnostic information on the misalignment errors, such as the orientation and magnitude.

In the second scenario, a variance error with inflated variance of $\sigma^2 = 1.5$ is generated and mixed with the bias error in the second layer. The test statistics and the corresponding confidence limits for the mean and variance components are shown in Fig. 5.6. It can be observed that both T0 and T1 statistics of components 5 through 8 are over the confidence limit, which is the statistical indication of both bias and variance errors in Layer II. For the components in Layers I and III, on the contrary, the corresponding test statistics are below the confidence limits and thus no significant bias/variance errors in those two layers. The MINQUE estimates of mean and variance components are further shown in Fig. 5.7. The mean component estimates in Layer II indicate that the bias error is positive with the magnitude around 6. Meanwhile, it can be inferred from the estimated variance components that the variance error is about 1.5 in Layer II. The computational results demonstrate that the mixed bias and variance errors can be diagnosed from the mean and variance components, respectively. Through the estimation and hypothesis testing procedures, the misalignment error sources in any individual layer can

be detected and diagnosed by using the metrology data.

5.4.2 Simulated example 2: three-layer 49×4 overlay process

In this simulated example, the three-layer overlay process is configured with 49 sample points per field and 4 fields per wafer. The physical parameter values for the fundamental model used in this example are given in Table 5.3 (Lin and Wu, 1999). The state matrices can be calculated from the model parameters as follows

$$A(k) = B(k) = I_{4\times 4}$$

and

$$C(k) = \begin{bmatrix} 100 & 100 & 0 & -100 \\ 100 & 100 & 100 & 154 \end{bmatrix}$$

In the simulation, a negative bias error with magnitude of 8 and a variance error with variance of $\sigma^2 = 1.0$ in both X and Y directions are generated and added to the $2nd$ layer. The hypothesis testing results are depicted in Fig. 5.8. The bias and variance errors can be readily detected because both T0 and T1 statistics for the components 5 through 8 in Layer II are larger than the corresponding 95% confidence limits. Thus it is desirable to further check the MINQUE estimates of mean and variance components to diagnose the error sources. As shown in Fig. 5.9, the mean component estimates in the second layer are close to 8, which is the same as the designed magnitude of the bias error in the simulation. Meanwhile, the downward bars of mean components in Layer II indicate that the bias error is along the negative X and Y directions

which is also consistent with the designed fault. The variance component estimates, on the other hand, reveal that the variance of the error in Layer II is around 1.0, which is the induced variance level. Not only the bias and variance errors mixed in the second layer can be isolated and detected correctly, but the qualitative and quantitative characteristics of the error sources are also diagnosed from the mean/variance component estimates.

5.4.3 Simulated example 3: six-layer 49×4 overlay process

In the third simulated example, a more complicated six-layer 49×4 overlay process with the same physical parameter values as given in Table 5.3 is used and multiple error sources occurring in different layers are designed. Two scenarios are tested with multiple errors in different layers. In the first scenario, two bias errors with magnitudes of 6 and 8 along the positive X and Y directions are added to the $2nd$ and $5th$ layers, respectively. As shown in Fig. 5.10, the test statistics of mean components 5 through 8 in Layer II and 17 through 20 in Layer V exceed the confidence limit. Thus the $2nd$ and $5th$ layers are inferred to contain significant bias errors. On the other hand, the fact that the test statistics of all variance components are below the corresponding confidence limit implies no significant variance errors in any layer of the overlay process. The mean component estimates are given in Fig. 5.11. The upward bars of the mean components in Layers II and V indicate that both bias errors are in positive directions. Furthermore, the magnitudes of the bars are around 6 and 8 in Layers II and V, respectively, which coincide

with the simulation design.

The second scenario is devised in the way of positive (magnitude of 6) and negative (magnitude of 8) bias errors in Layers II and V, respectively. Similarly, the error sources in these two layers can be detected according to the test statistics in Fig. 5.12. The opposite directions of the bias errors are readily seen from the upward and downward bars in the estimated mean components in Fig. 5.13. In addition, the estimated values of the mean components in Layers II and V are around 6 and 8, respectively, and reflect the magnitudes of misalignment errors approximately.

5.4.4 Simulated example 4: three-layer 49×4 overlay process with one-way error

In the previous examples, the bias or variance errors are applied along two ways of both X and Y directions. To test the capability of the method for diagnosing one-way fault along a single direction, two scenarios are designed in a three-layer 49×4 overlay process with the model parameters given in Table 5.3. In the first scenario, a positive bias error of magnitude 8 along X direction is applied to the second layer. The second scenario is to insert a negative error of the same magnitude along Y direction to layer II. The hypothesis test statistics of these two test scenarios are shown in Figs. 5.14 and 5.15, respectively. In Fig. 5.14, only the mean components 5 and 6 exceed the confidence limit and all variance components are below the confidence limit. The $5th$ and $6th$ components correspond to the X-way misalignment error in the second layer.

Likewise, the Y-way bias error in Layer II for the second scenario is detected from Fig. 5.15 as the corresponding mean components 7 and 8 are beyond the confidence limit. The estimated mean components in Figs. 5.16 and 5.17 can further provide the orientation and magnitude information of the detected bias errors. In the first scenario, the error is along positive X direction because the mean component bars are upward. Similarly, the downward mean component bars indicate negative Y-way error in the second scenario. Furthermore, the magnitudes of the relevant mean components are around 8 and consistent with the simulation design in both scenarios. The simulation results in this example show the capability of the presented method to isolate and diagnose the misalignment error along a single direction in multi-layer overlay processes.

5.5 Summary

In this chapter, a variance component analysis approach for the fault detection and diagnosis of multi-layer overlay lithography processes is presented. A multi-stage state space model for the overlay processes is derived from the physical principles and further formulated into the general mixed linear model. Then the MINQUE algorithm is used to estimate the mean and variance components in different layers. Subsequently, a statistical hypothesis testing strategy is applied to examine the significance of each mean/variance component and detect the error sources in different layers. Results on a series of simulated examples demonstrate that different types of errors in multiple layers can be isolated and detected correctly by using test statistics of

mean/variance components. Furthermore, the orientation and magnitude of the errors can be diagnosed effectively from the estimated mean/variance components. This approach is shown to be a promising technique for root-cause identification of multi-layer overlay lithography processes in semiconductor manufacturing.

Table 5.1: Summary of simulated examples for multi-layer overlay process monitoring

		Bias Error	Variance Error
Example 1	Scenario 1	Layer II (+6)	N/A
	Scenario 2	Layer II (+6)	Layer II (1.5)
Example 2		Layer II (-8)	Layer II (1.0)
Example 3	Scenario 1	Layer II (+6) & Layer V (+8)	N/A
	Scenario 2	Layer II (+6) & Layer V (-8)	N/A
Example 4	Scenario 1	Layer II (+8, X-way)	N/A
	Scenario 2	Layer II (-8, Y-way)	N/A

Table 5.2: Physical model parameter values of simulated 36×4 overlay lithography process (All of the parameter values are in unit of 100 nm)

	Interfield						Intrafield					
	T_{gx}	T_{gy}	S_x	S_y	R_n	R_g	T_{rx}	T_{ry}	M_r	M_a	R_r	R_a
Value	2739	3426	100	100	0	100	0	0	-4043.5	-4196.5	100	0

Table 5.3: Physical model parameter values of simulated 49×4 overlay lithography process (All of the parameter values are in unit of 100 nm)

	Interfield						Intrafield					
	T_{gx}	T_{gy}	S_x	S_y	R_n	R_g	T_{rx}	T_{ry}	M_r	M_a	R_r	R_a
Value	2472	3091	100	100	0	100	0	0	127	-27	100	0

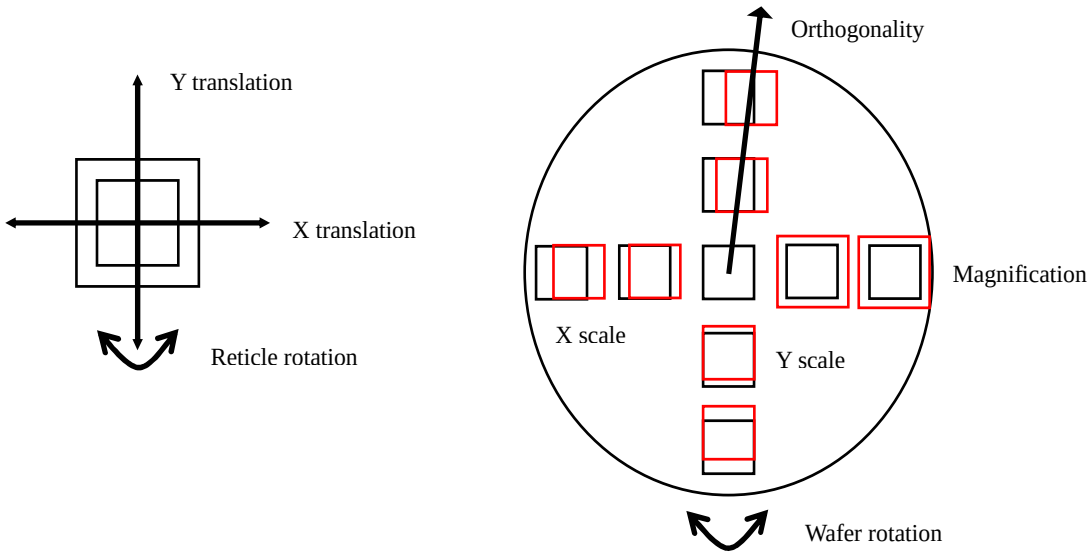

Figure 5.1: Illustration of overlay misalignment error sources

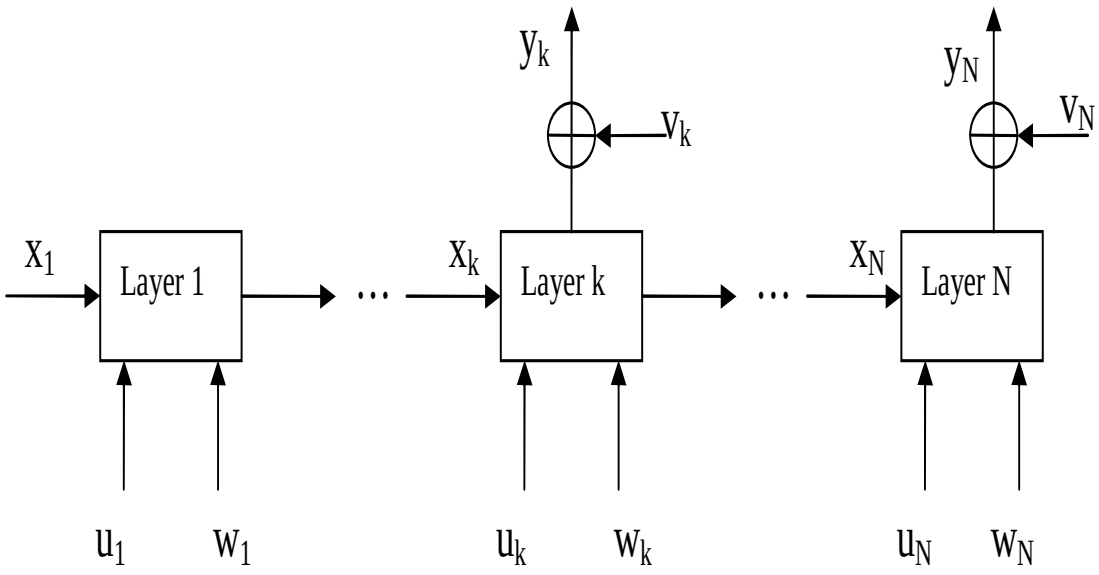

Figure 5.2: Systematic diagram of the multi-layer overlay lithography process

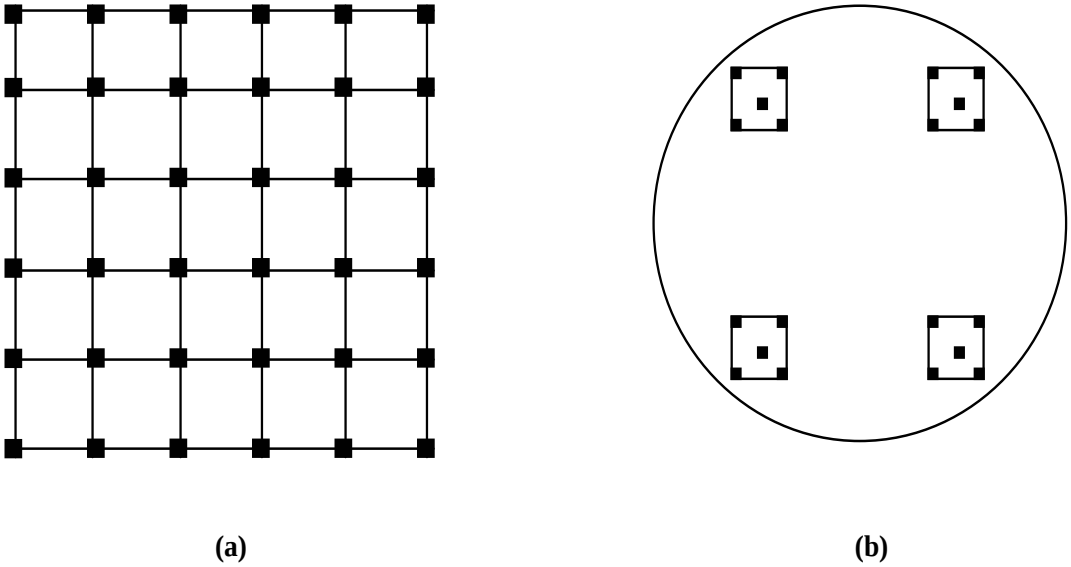

Figure 5.3: The schematic diagram of the 36×4 overlay lithography process with (a) 36 sample points in a field and (b) 4 fields in a wafer

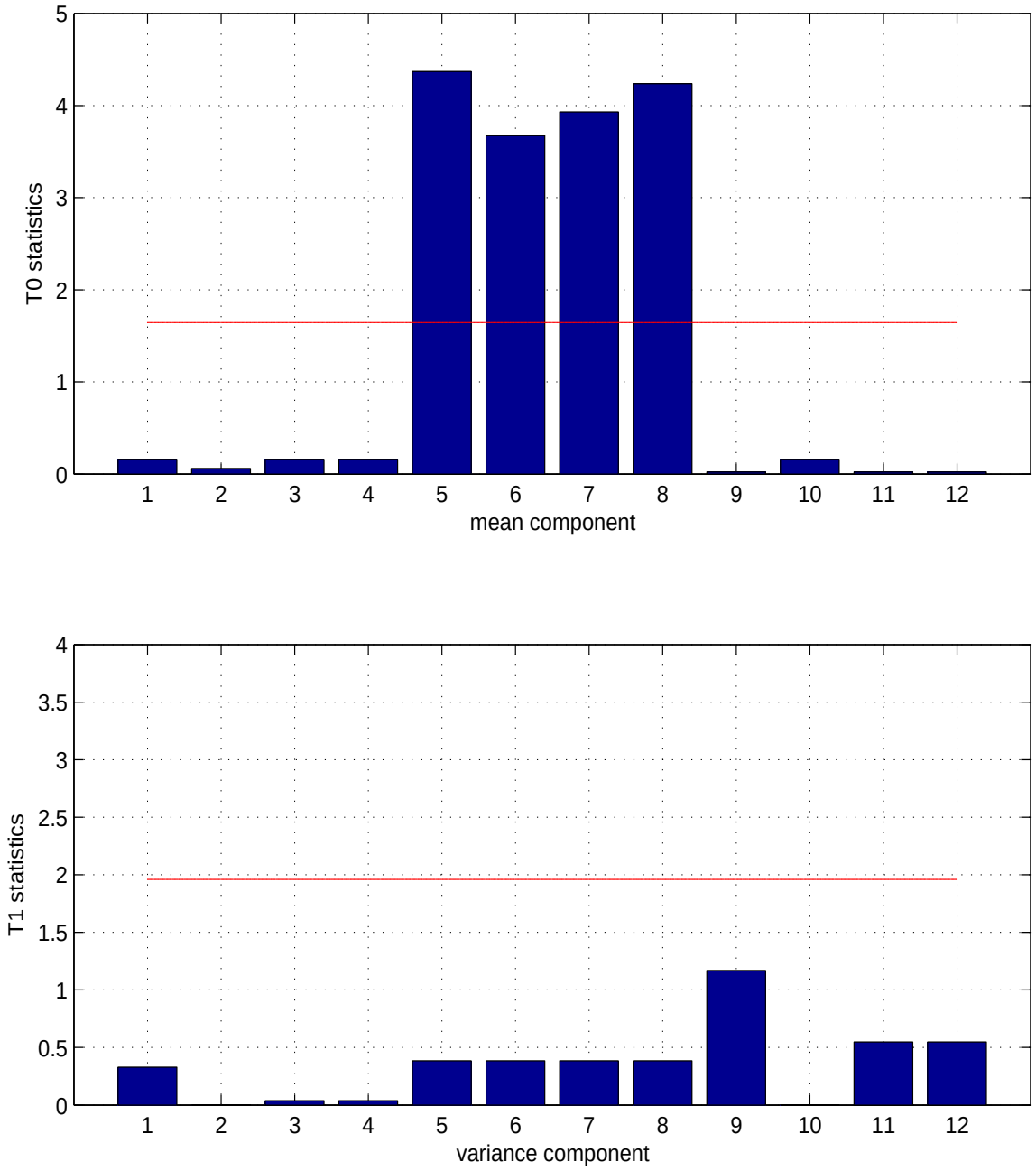

Figure 5.4: Simulated example 1: Test statistics for mean and variance components of the 3-layer 36×4 overlay process with only bias error in the $2nd$ layer

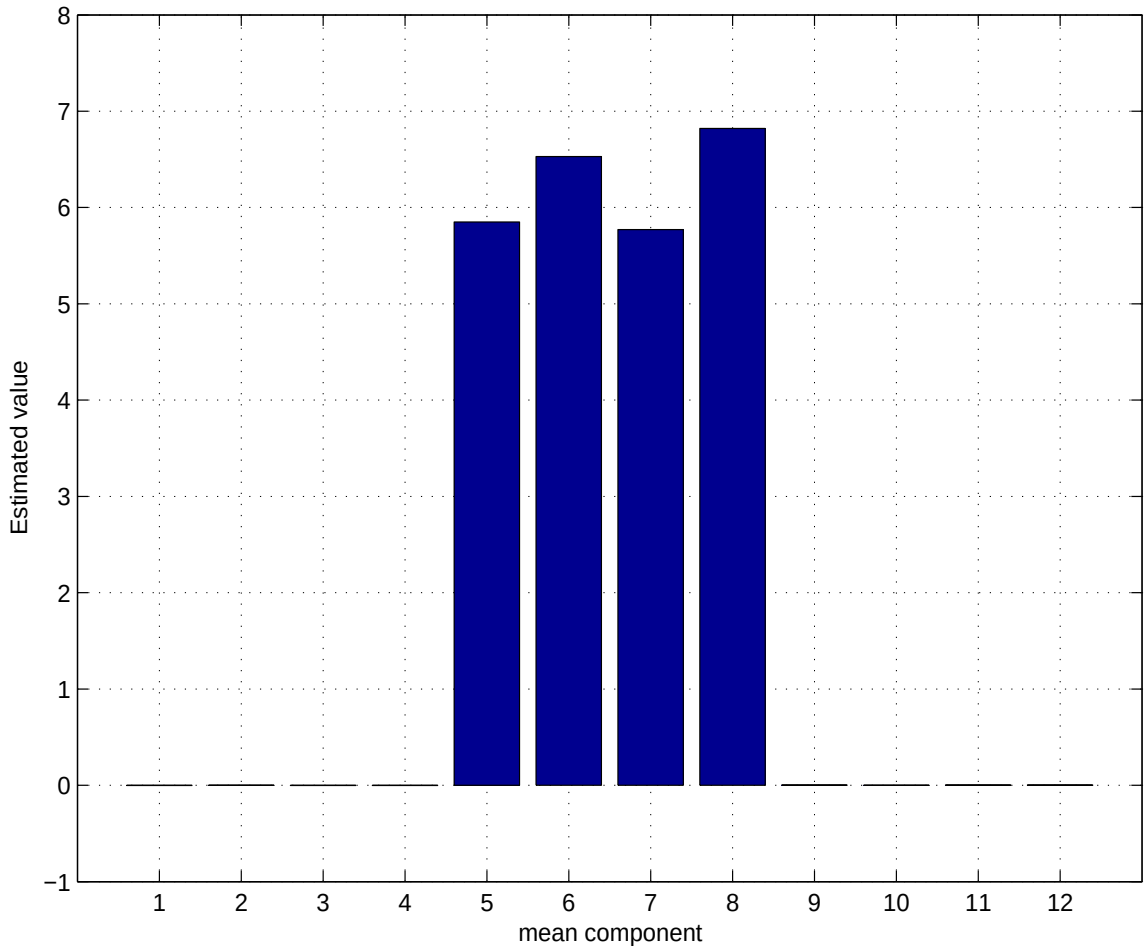

Figure 5.5: Simulated example 1: Mean component estimation of the 3-layer 36×4 overlay process with only bias error in the $2nd$ layer

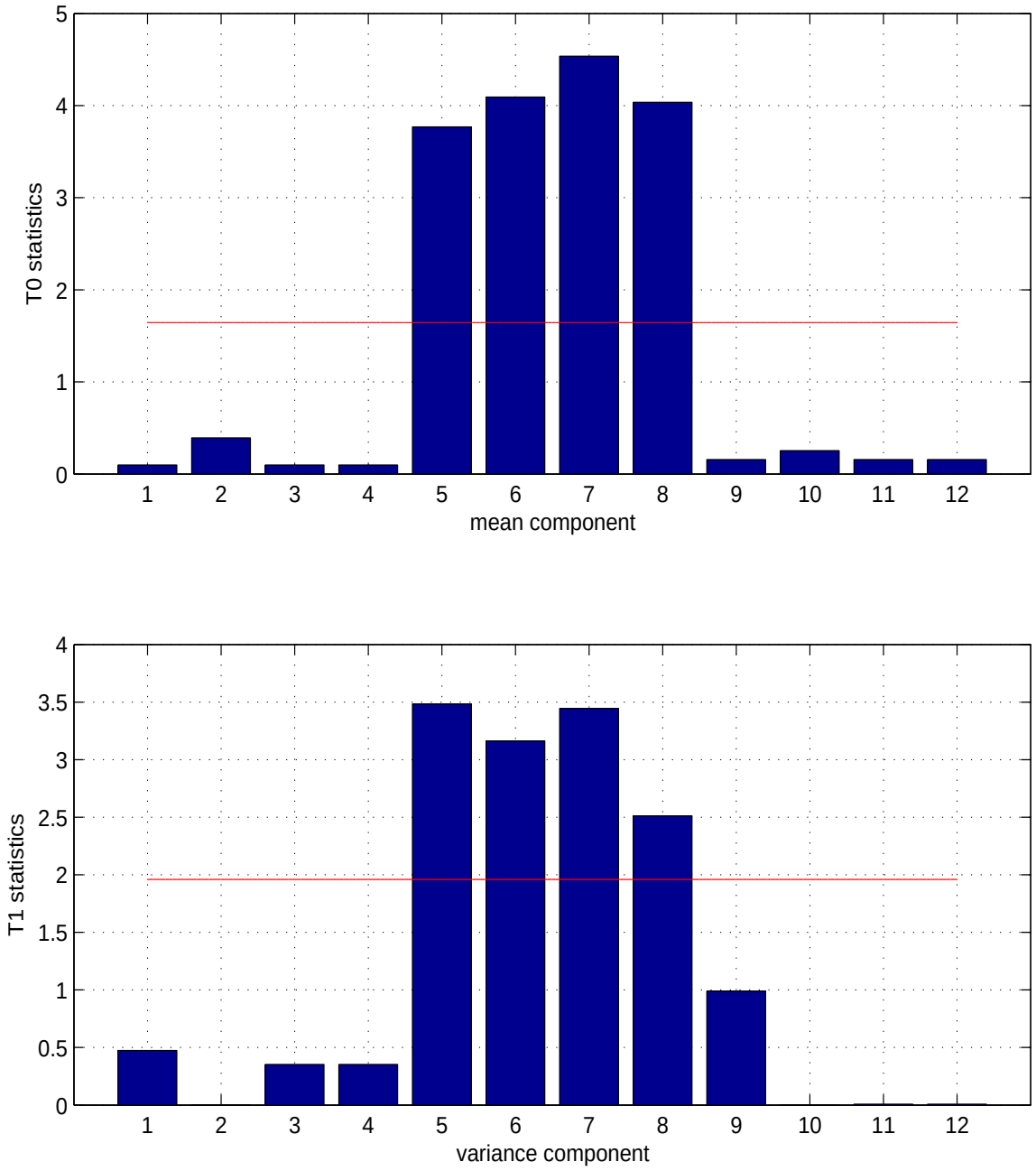

Figure 5.6: Simulated example 1: Test statistics for mean and variance components of the 3-layer 36×4 overlay process with both bias and variance errors in the $2nd$ layer

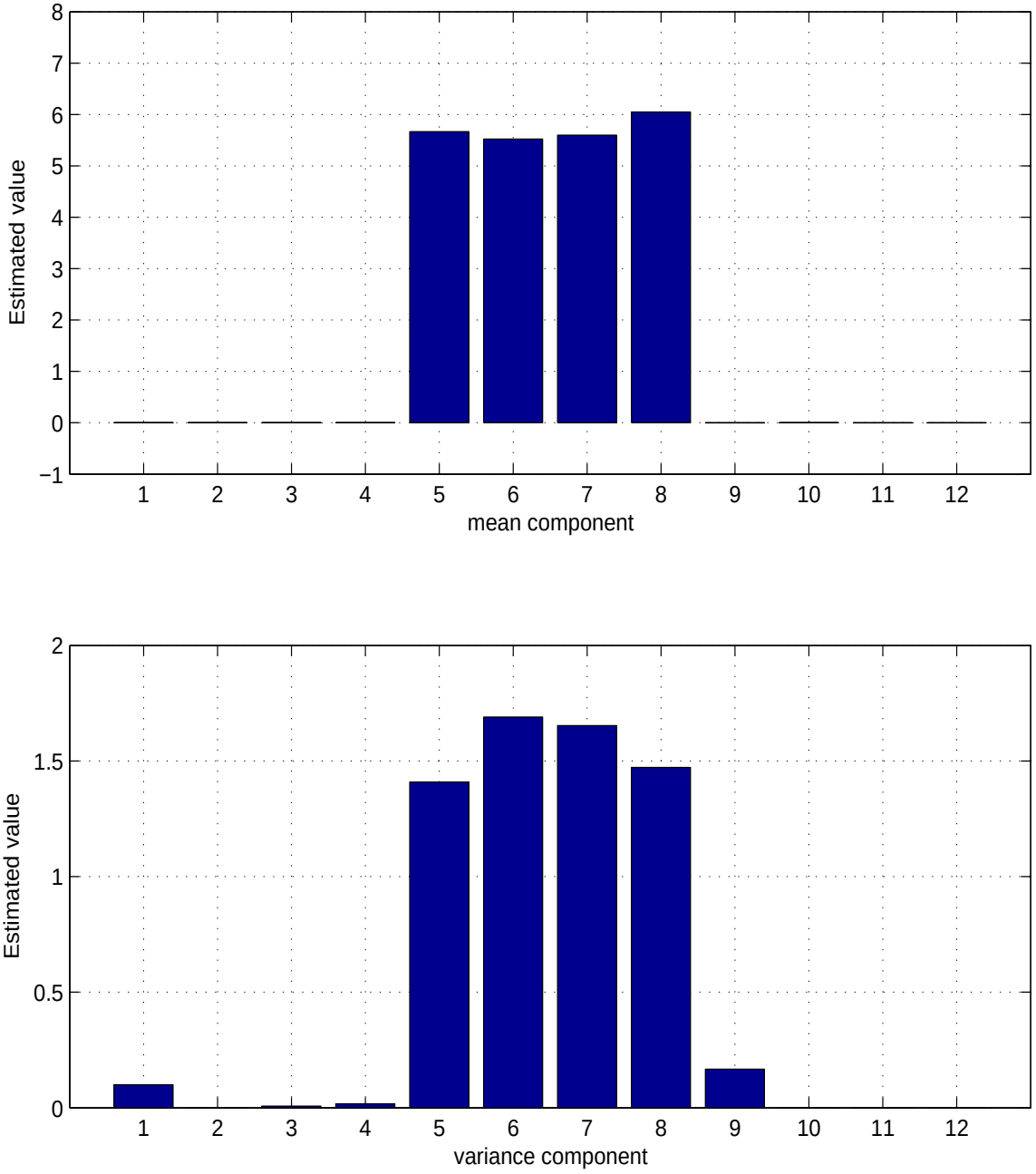

Figure 5.7: Simulated example 1: Mean and variance component estimation of the 3-layer 36×4 overlay process with both bias and variance errors in the $2nd$ layer

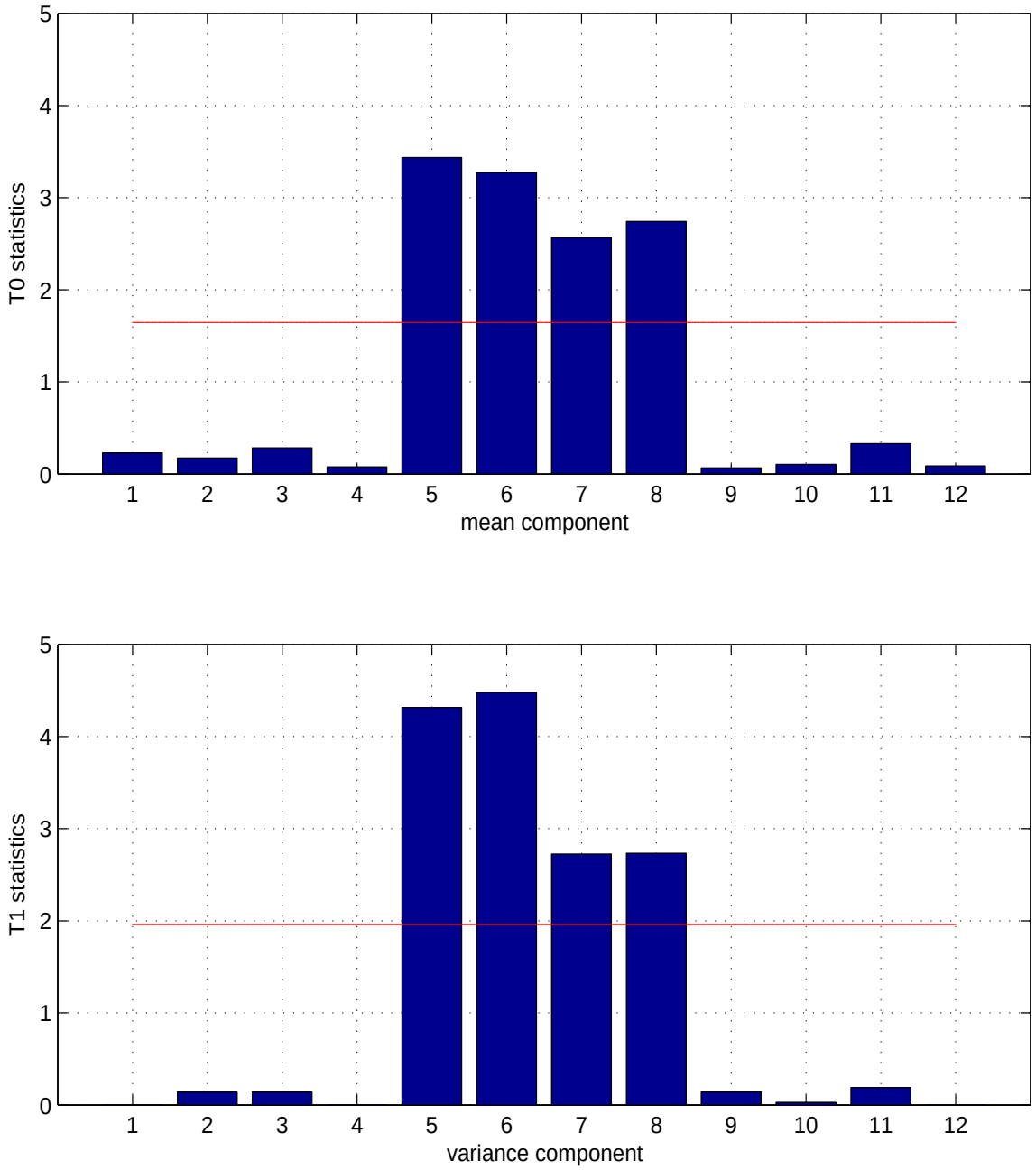

Figure 5.8: Simulated example 2: Test statistics for mean and variance components of the 3-layer 49×4 overlay process with both bias and variance errors in the $2nd$ layer

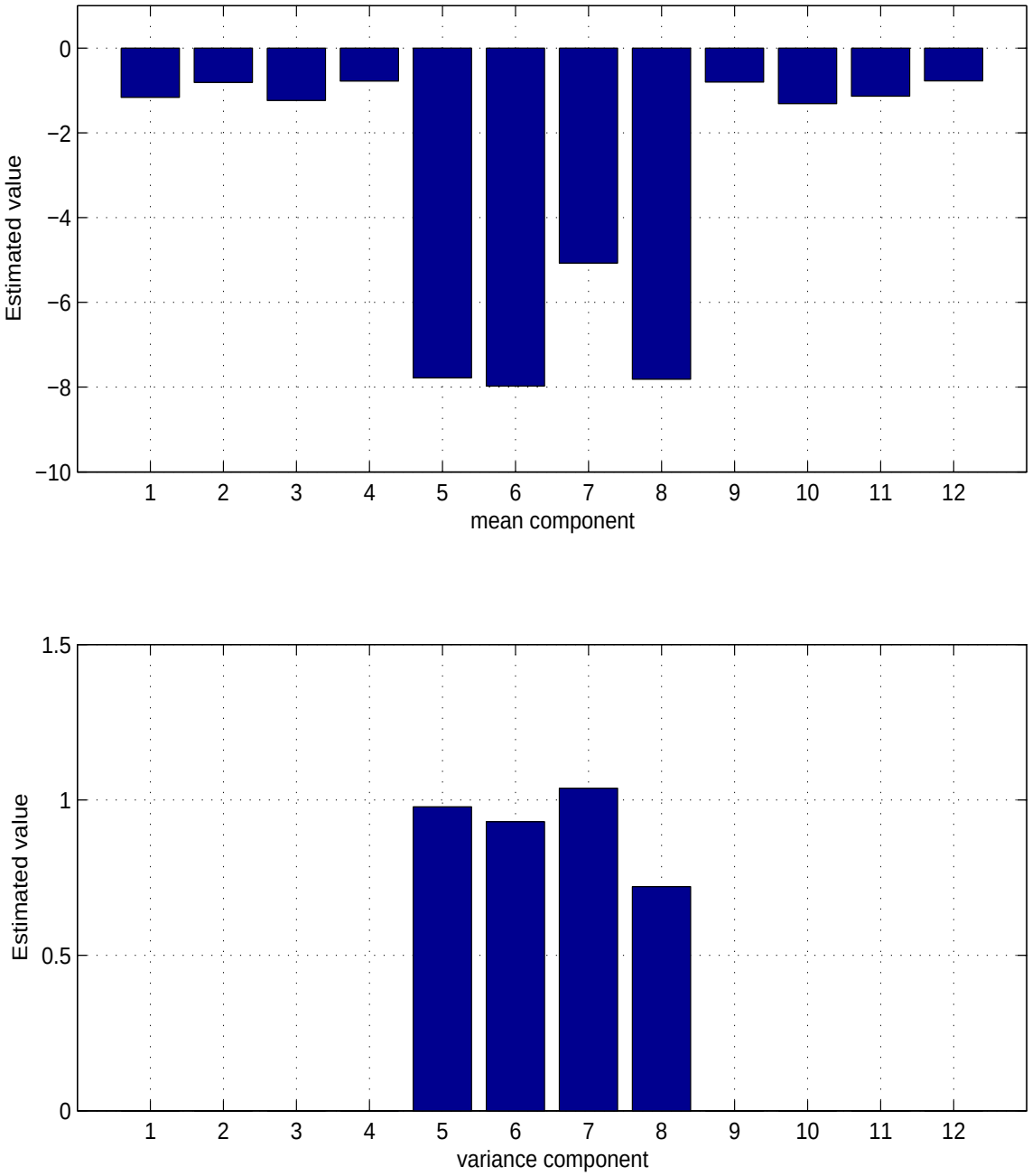

Figure 5.9: Simulated example 2: Mean and variance component estimation of the 3-layer 49×4 overlay process with both bias and variance errors in the $2nd$ layer

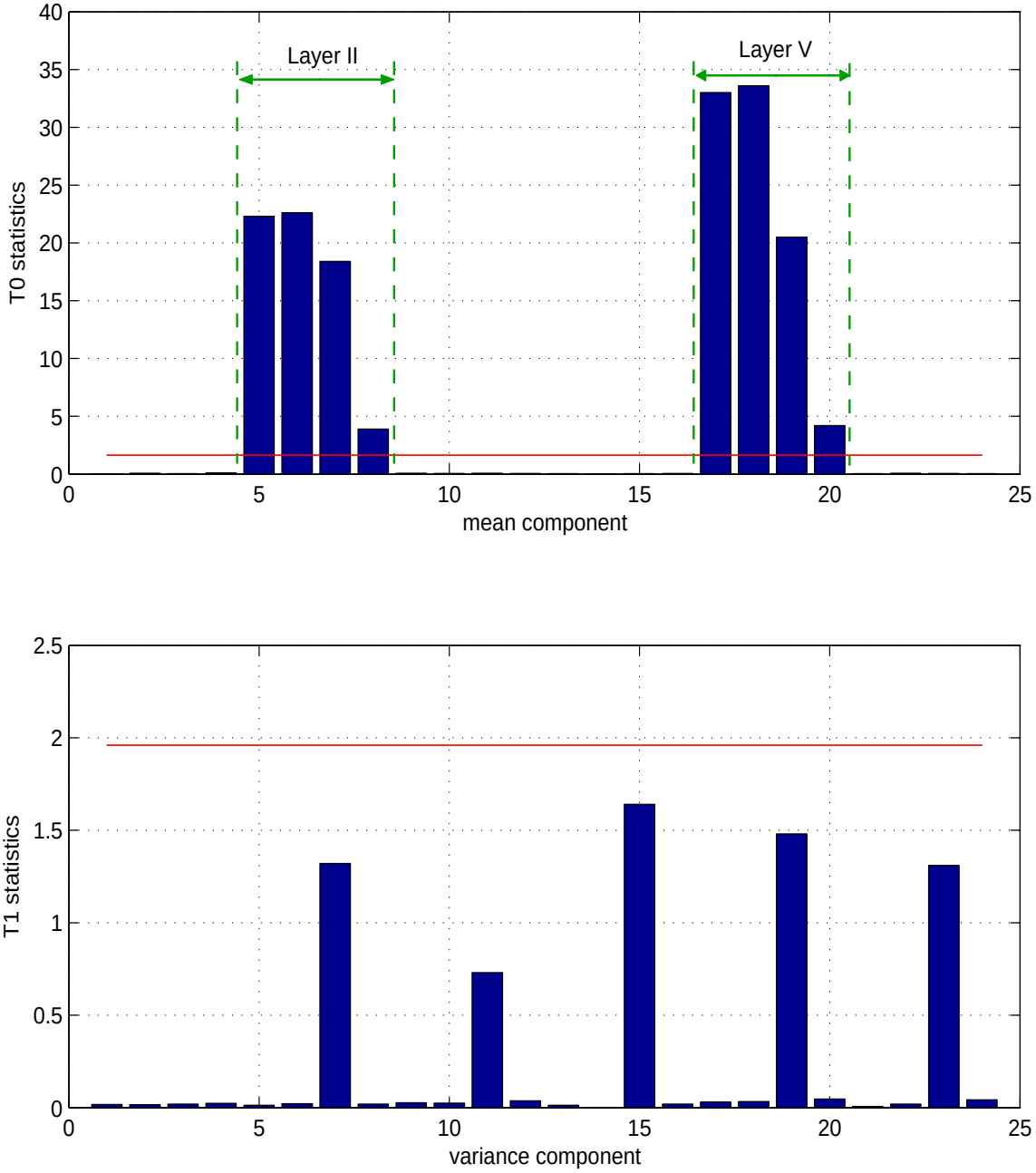

Figure 5.10: Simulated example 3: Test statistics for mean and variance components of the 6-layer 49×4 overlay process with both positive bias errors in the $2nd$ and $5th$ layers

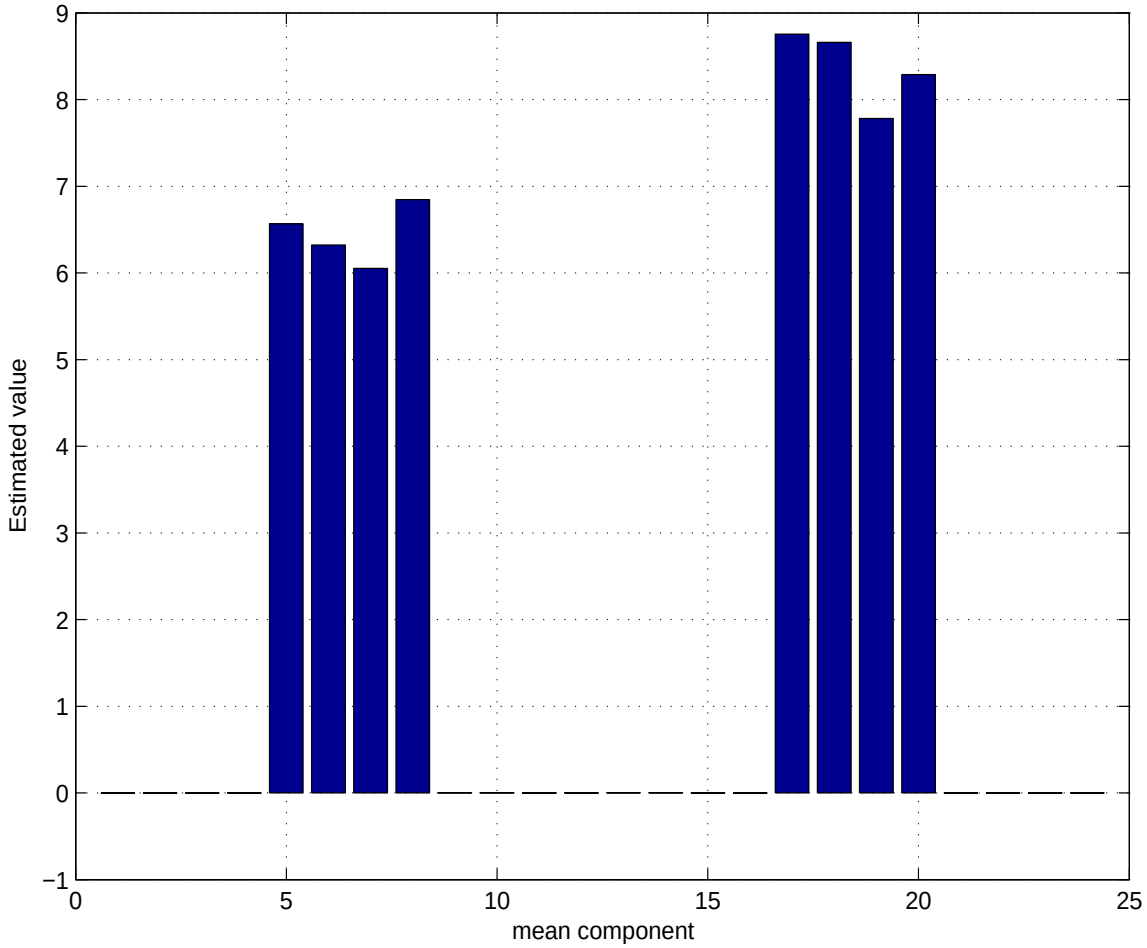

Figure 5.11: Simulated example 3: Mean component estimation of the 6-layer 49×4 overlay process with both positive bias errors in the $2nd$ and $5th$ layers

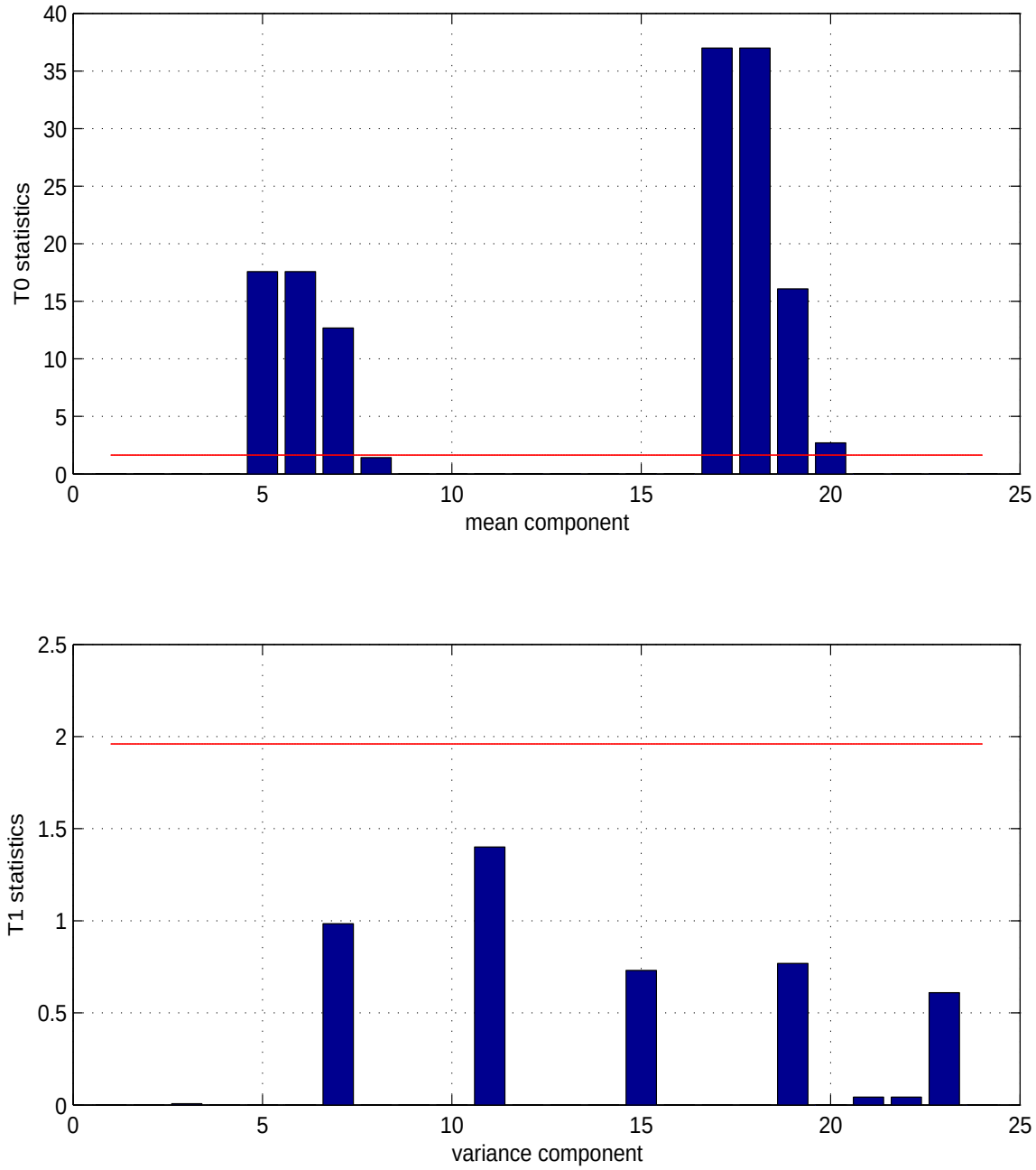

Figure 5.12: Simulated example 3: Test statistics for mean and variance components of the 6-layer 49×4 overlay process with positive and negative bias errors in the $2nd$ and $5th$ layers, respectively

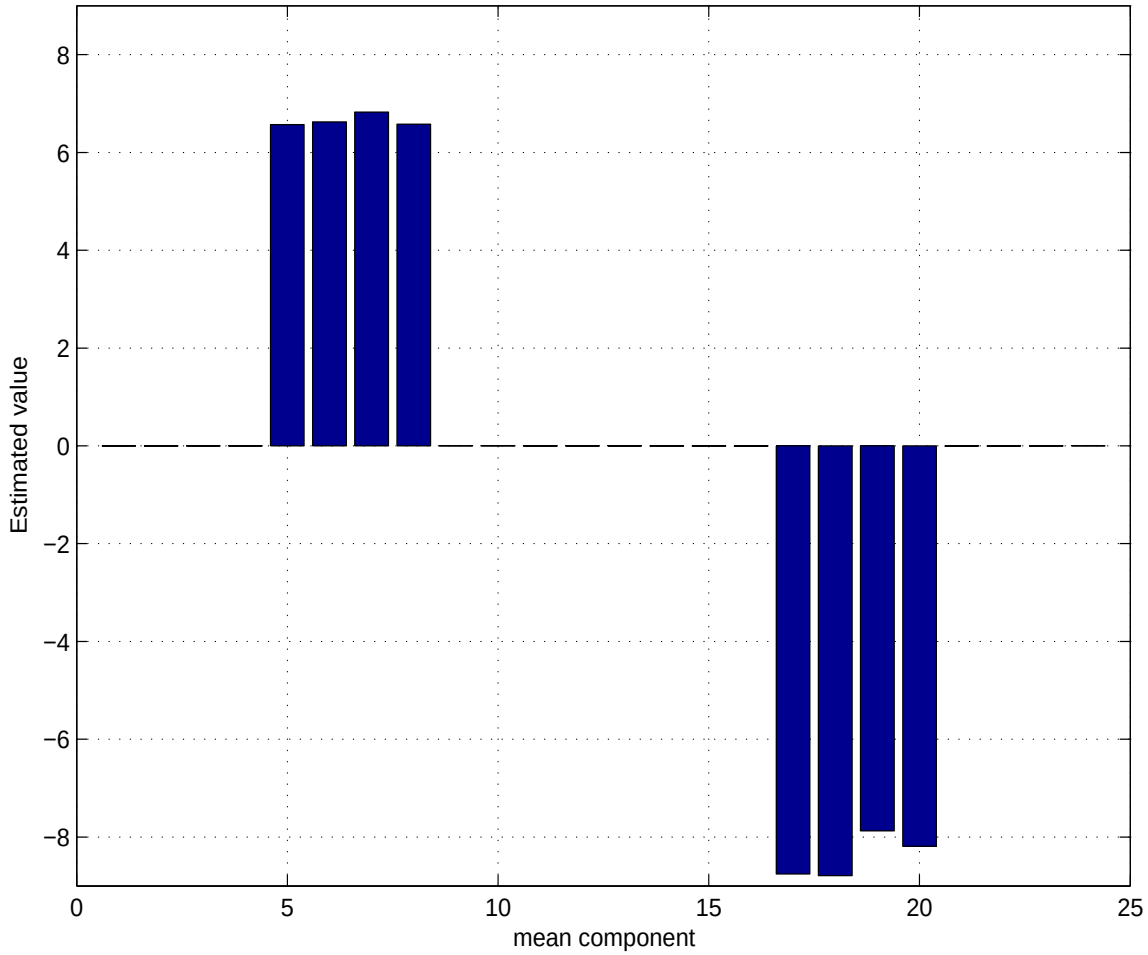

Figure 5.13: Simulated example 3: Mean component estimation of the 6-layer 49×4 overlay process with positive and negative bias errors in the $2nd$ and $5th$ layers, respectively

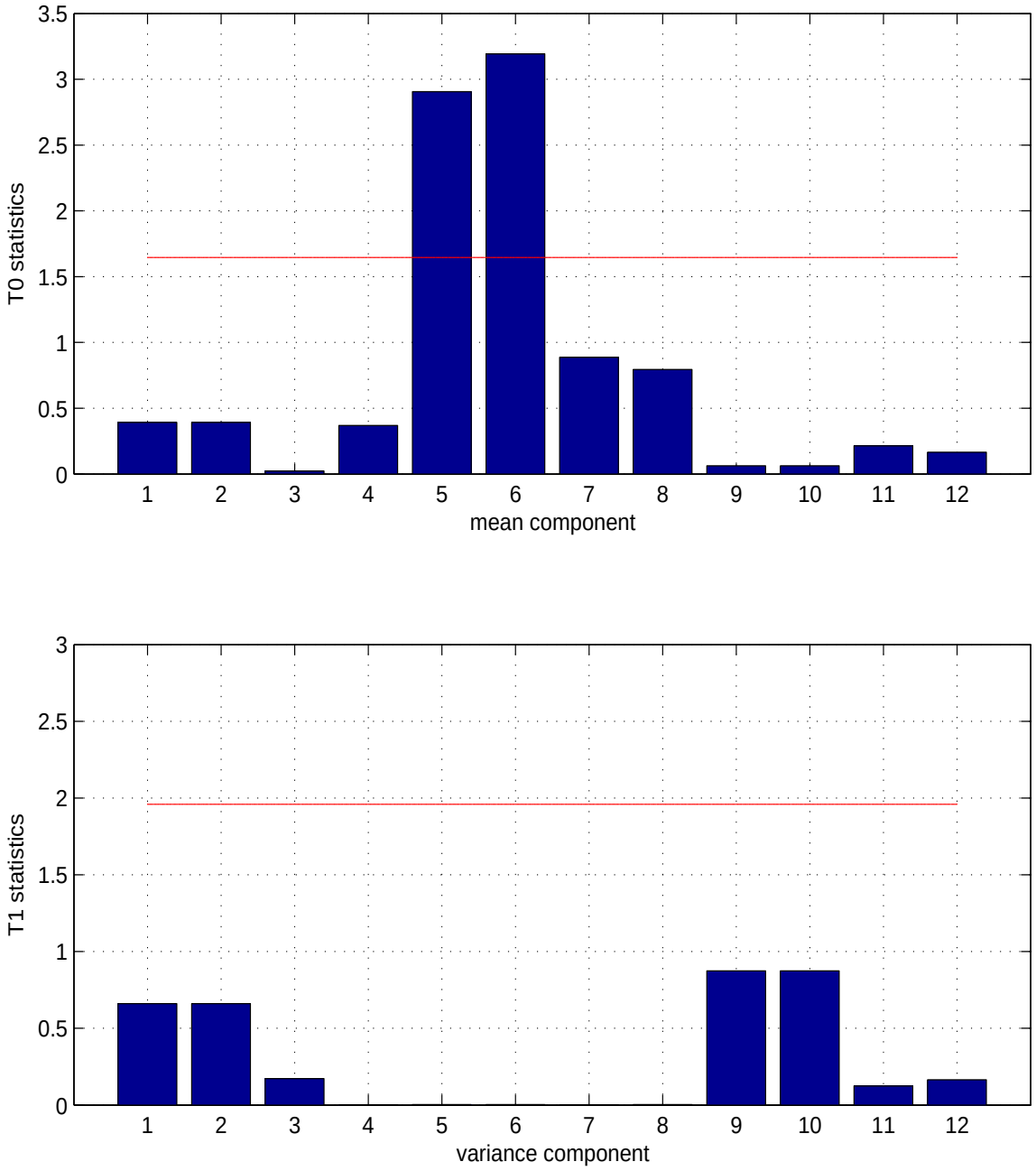

Figure 5.14: Simulated example 4: Test statistics for mean and variance components of the 3-layer 49×4 overlay process with only X-way bias error in the $2nd$ layer

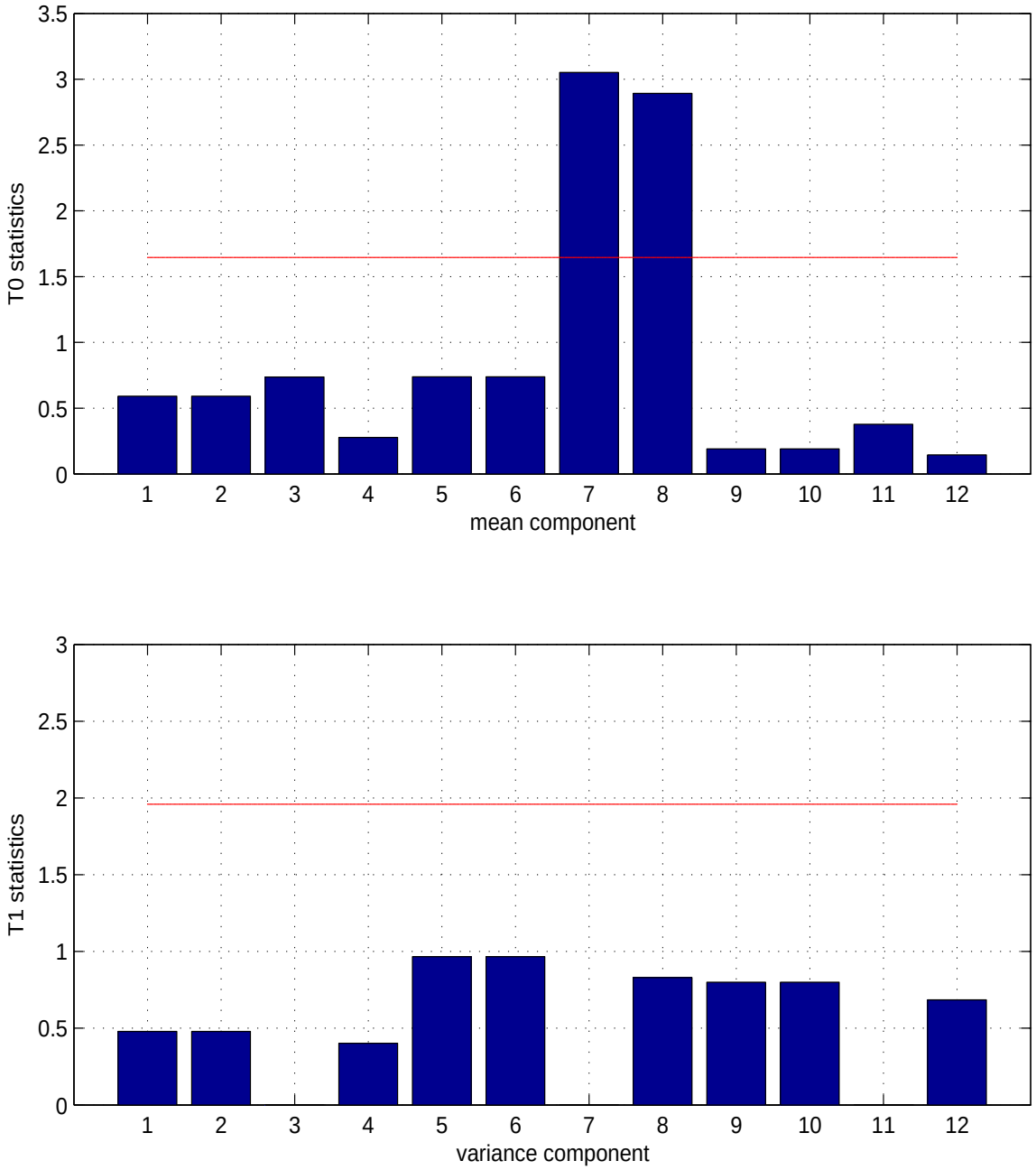

Figure 5.15: Simulated example 4: Test statistics for mean and variance components of the 3-layer 49×4 overlay process with only Y-way bias error in the $2nd$ layer

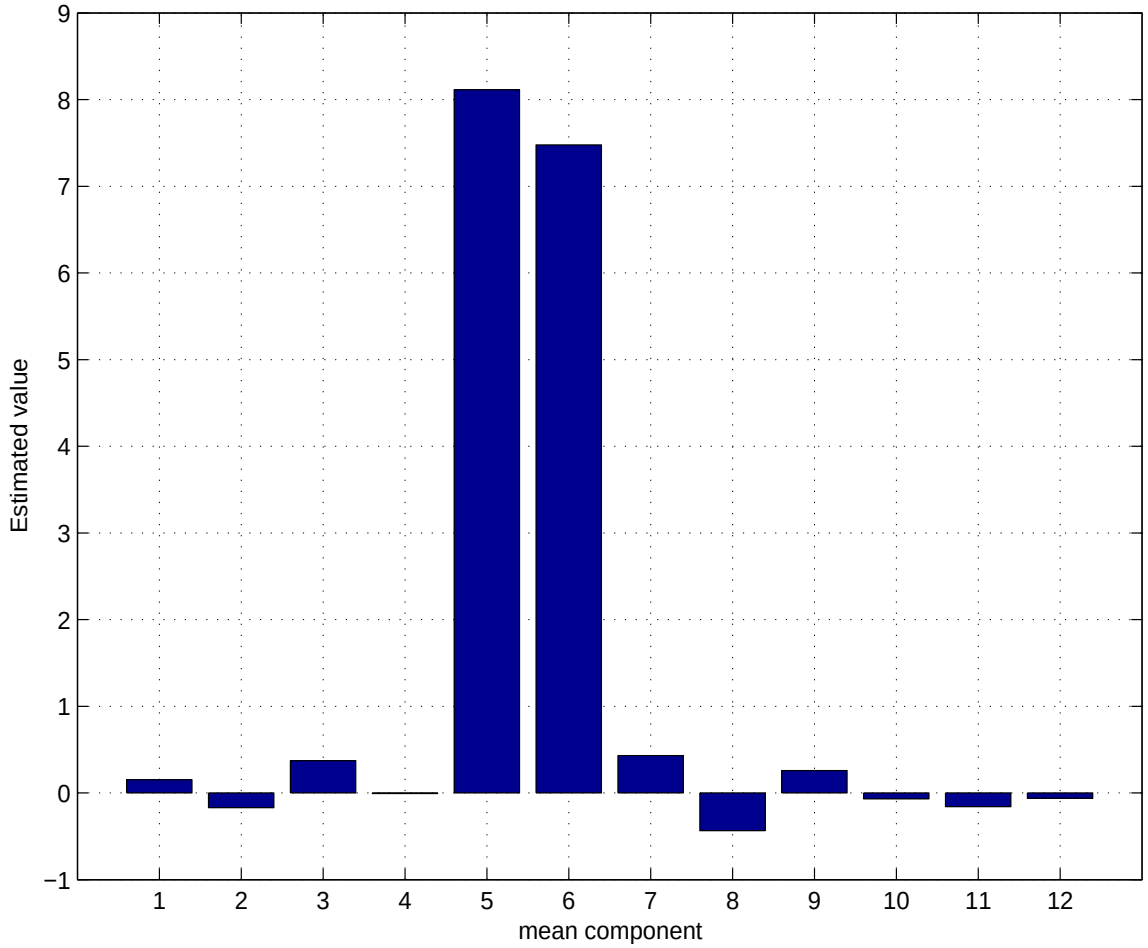

Figure 5.16: Simulated example 4: Mean component estimation of the 3-layer 49×4 overlay process with only X-way bias error in the $2nd$ layer

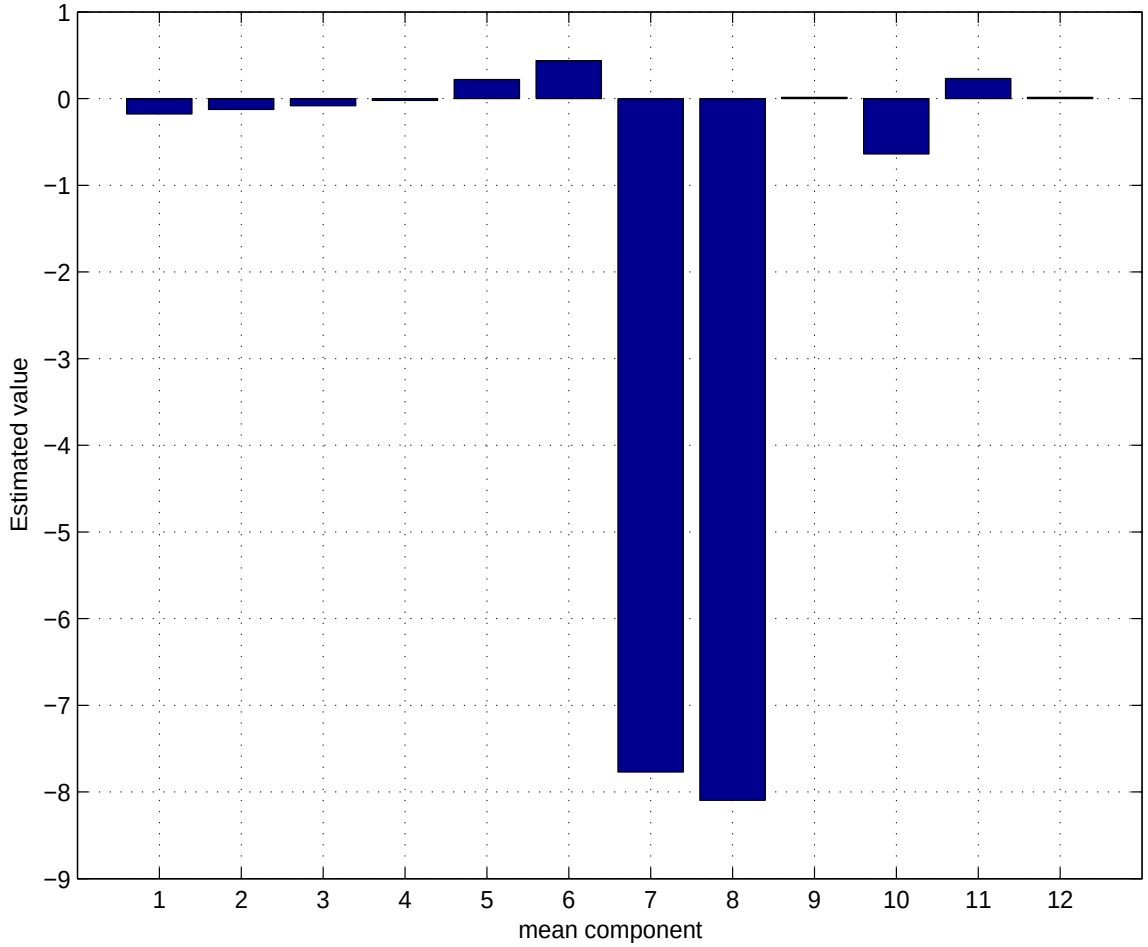

Figure 5.17: Simulated example 4: Mean component estimation of the 3-layer 49×4 overlay process with only Y-way bias error in the $2nd$ layer

Chapter 6

Conclusions and Future Directions

6.1 Conclusions

This research presents some innovative solutions to the challenging issues in control performance monitoring and process monitoring areas. The major contributions are summarized as follows:

1. A data-driven benchmark method, which uses historical operation data with desirable performance, is proposed for MIMO control performance monitoring. This benchmark is specified by users based on their process experience and can be easily implemented in industrial plants to avoid the model requirements of interactor-based or model-based MVC and LQG benchmarks.

2. Within data-driven covariance monitoring framework, generalized eigenvalue analysis is used to find the directions with the largest or smallest variance inflation ratio. Then statistical inference methods are developed to determine the directions or subspace with significantly worse or better control performance relative to the benchmark. Covariance based indices are also defined to assess the extent of performance degradation/improvement within the respective subspace. The generalized

eigenvalues as well as these performance indices are shown to be invariant to data scaling.

3. Following the data-driven benchmark, two types of multivariate contribution methods are developed for control performance diagnosis within the worse or better performance subsapce. In the loading based contribution approach, bootstrap resampling technique is employed to estimate the distribution and statistics of eigenvector loadings and the confidence intervals is then derived. For the angle based contribution method, the cosine of the angle between each loop/variable is defined as contribution index and further proven to be equivalent to canonical correlation. The asymptotical statistics of canonical correlation is then used to deduct the confidence limit of the angle based contribution. Both methods are shown to be capable of diagnosing controlled variables with degraded or improved control performance.

4. A simplified solutions to MVC benchmark for control performance monitoring is established. Analogous to the conventional left interactor, a right diagonal interactor matrix is first adopted to derive the MVC benchmark from time delays only without the knowledge of process model. To accommodate a wider class of MIMO processes, both left and right diagonal interactors are further integrated to characterize the more complicated time delay structure. Two straightforward means of factorizing out the left/right diagonal interactors from only I/O time delays

are provided. Subsequently, the estimation method of MVC benchmark using the combined left/right diagonal interactors is also developed.

5. To optimize the weighting selection in the underlying LQ objective of MVC benchmark, an iterative strategy has been designed. The eigenvalue analysis is carried out on the output covariance, and a non-diagonal weighting matrix can then be constructed by rotating the original outputs to the eigendirections with the eigenvalues measuring the relative importance. Such weighting matrix is shown to be equivalent to the output covariance matrix of the closed-loop operating data and the weighting update procedures can be iterated to push the performance close to the MVC bound.

6. The challenging problem of multi-layer overlay lithography process monitoring is investigated and the variance component analysis approach is used to deal with the complicated error propagation along the multiple steppers. A multi-stage state space model to characterize the overlay processes is derived from the fundamental principles and further formulated into the general mixed linear model. Then a hypothesis testing procedure is applied to detect the active faults mixed in different layers. Further, the mean and variance components representing the deterministic and stochastic effects are estimated to diagnose the magnitude and orientation of misalignment errors.

6.2 Future Directions

Based on the research work that we have accomplished, some interesting directions are pointed out for the further enhancement of process and control performance monitoring framework.

1. In current data-driven performance monitoring approach, only controlled variables are considered. The manipulated variables, however, also deserve analysis to detect the input constraint change or input saturation which may have significant impact on control performance.

2. The means to isolate the influence between the external disturbances and disturbance dynamics change on performance variation should be investigated and the subspace projection based methods may be potentially useful to solve this problem.

3. The effect of time delay structure of MIMO processes on the separability of the combined left/right diagonal interactors is worth further research effort to find out simpler existence criterion.

4. The adaptive weighting selection is currently focused on output weighting matrix and the similar idea may be extended to input weighting for performance improvement.

5. Generalization of the VCA based fault diagnosis approach to handle more complicated error sources is important for semiconductor manufacturing.

Bibliography

Åström, K. J. (1976). State of the art and needs in process identification. In: *AIChE Symposium Series*. pp. 184–194.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, Hoboken, NJ. Third Edition.

Åström, K. J. (1970). *Introduction to Stochastic Control Theory*. Academic Press. San Diego, California.

Åström, K. J. and B. Wittenmark (1997). *Computer-controlled systems: Theory and design*. Prentice Hall Inc, New Jersey.

Bialkowski, W. L. (1993). Dreams versus reality: a review from both sides of the gap. *Pulp and Paper - Canada* **94**, 19–27.

Bode, C. A. (2001). Run-to-run control of overlay and linewidth in semiconductor manufacturing. PhD thesis. The University of Texas at Austin.

Bode, C. A., B. S. Ko and T. F. Edgar (2004). Run-to-run control and performance monitoring of overlay in semiconductor manufacturing. *Control Eng. Practice* **12**, 893–900.

Box, G. E. P. and G. M. Jenkins (1970). *Time series analysis: Forecasting and control*. Holden-Day, San Francisco, CA.

Brink, M. A., C. G. M. Mol and R. A. George (1988). Matching performance for multiple wafer steppers using an advanced metrology procedure. *SPIE Integrated Circuit Metrology, Inspection and Process Control II* **921**, 180–197.

Chiang, L.H., E.L. Russell and R.D. Braatz (2000). Fault diagnosis and fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics Intell. Lab. Syst.* **50**, 243–252.

Chiang, L.H., E.L. Russell and R.D. Braatz (2001). *Fault Detection and Diagnosis in Industrial Systems*. Advanced Textbooks in Control and Signal Processing. Springer-Verlag. London, Great Britain.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, New York, NY.

Desborough, L. and T. J. Harris (1992). Performance assessment measures for univariate feedback control. *Can. J. Chem. Eng.* **70**, 1186.

Desborough, L. and T. J. Harris (1993). Performance assessment measures for univariate feedforward/feedback control. *Can. J. Chem. Eng.* **71**, 605–616.

Ding, Y., J. Shi and D. Ceglarek (2002). Diagnosability analysis of multi-station manufacturing processes. *ASME J. Dynamics Systems, Measurement and Control* **124**, 1–13.

Dugard, L., G.C. Goodwin and X. Xianya (1984). The role of the interactor matrix in multivariable stochastic adaptive control. *Automatica* **20**(5), 701–709.

Dunia, R., S. J. Qin, T. F. Edgar and T. J. McAvoy (1996). Identification of faulty sensors using principal component analysis. *AIChE J.* **42**, 2797–2812.

Ender, D. (1993). Process control performance: not as good as you think. *Control Engineering* **9**, 180–190.

Ender, D. (1999). Common cause of control loop cycling. *Control Engineering* **46**, 78–78.

Goodwin, G. C. and K. S. Sin (1984). *Adaptive Filtering Prediction and Control*. Prentice-Hall. Englewood Cliffs, New Jersey.

Goodwin, G. C., S. F. Graebe and M. E. Salgado (2001). *Control System Design*. Prentice Hall, Upper Saddle River, New Jersey.

Grimble, M. J. (2002). Controller performance benchmarking and tuning using generalized minimum variance control. *Automatica,* **38**, 2111–2119.

Harris, T. J. (1989). Assessment of control loop performance. *Can. J. Chem. Eng.* **67**(10), 856–861.

Harris, T. J. (2004). Statistical properties of quadratic-type performance indices. *J. Proc. Cont.* **14**, 899–914.

Harris, T. J. and C. T. Seppala (2002*a*). Increasing customer value of industrial control performance monitoring - honeywell's experience. In: *Chemical Process Control - CPC VI*. CACHE. Tuscon, Arizona. pp. 169–189.

Harris, T. J. and C. T. Seppala (2002*b*). Recent developments in controller performance monitoring and assessment techniques. In: *Chemical Process Control - CPC VI*. CACHE. Tuscon, Arizona. pp. 208–222.

Harris, T. J., C. T. Seppala and L. D. Desborough (1999). A review of performance monitoring and assessment technique for univariate and multivariate control systems. *Journal of Process Control* **9**, 1–17.

Harris, T. J., F. Boudreau and J.F. MacGregor (1996). Performance assessment of multivariable feedback controllers. *Automatica* **32**(11), 1505–1518.

He, Q. P., S. J. Qin and J. Wang (2005). A new fault diagnosis method using fault directions in fisher discriminant analysis. *AIChE J.* **51**, 555–571.

Hoo, K. A., M. J. Piovoso, P. D. Schnelle and D. A. Rowan (2003). Process and controller performance monitoring: overview with industrial applications. *Int. J. Adaptive Control and Signal Processing* **17**, 635–662.

Horch, A. and A. J. Isaksson (1999). A modified index for control performance assessment. *J. Proc. Cont.* **9**, 475–483.

Huang, B. (1997). Multivariate Statistical Methods For Control Loop Performance Assessment. PhD thesis. University of Alberta.

Huang, B. and S. L. Shah (1998). Practical issues in multivariable feedback control performance assessment. *Journal of Process Control* **8**, 421–430.

Huang, B. and S.L. Shah (1999). *Performance Assessment of Control Loops: Theory and Applications*. Advances in Industrial Control. Springer-Verlag. London, Great Britain.

Huang, B., S. X. Ding and N. Thornhill (2005). Practical solutions to multivariate feedback control performance assessment problem: reduced a priori knowledge of interactor matrices. *J. Proc. Cont.* **15**, 573–583.

Huang, B., S. X. Ding and N. Thornhill (2006). Alternative solutions to multi-variate control performance assessment problems. *J. Proc. Cont.* **16**, 457–471.

Huang, B., S.L. Shah and H. Fujii (1997*a*). The unitary interactor matrix and its estimation using closed-loop data. *J. Proc. Cont.* **7**(3), 195–207.

Huang, B., S.L. Shah and K.Y. Kwok (1997*b*). Good, bad or optimal? performance assessment of MIMO processes. *Automatica* **33**(6), 1175–1183.

Jelali, M. (2006). An overview of control performance assessment technology and industrial applications. *Control Eng. Practice* **14**, 441–466.

Kano, M., B. Hasebe, S. Hashimoto and I. Ohno (2003). Monitoring independent components for fault detection. *AIChE J.* **49**, 969–976.

Khuri, A. I., T. Mathew and B. K. Sinha (1998). *Statistical tests for linear mixed models*. John Wiley and Sons, New York.

Kleffe, J. and B. Seifert (1988). On the role of MINQUE in testing of hypothesis under linear mixed models. *Communication in Statistics - Theory and Methods* **17**, 1287–1309.

Ko, B. S. and T. F. Edgar (2000). Performance assessment of cascade control loops. *AIChE J.* **46**, 281–291.

Ko, B. S. and T. F. Edgar (2001*a*). Performance assessment of constrained model predictive control systems. *AIChE J.,* **47**, 1363–1371.

Ko, B. S. and T. F. Edgar (2001*b*). Performance assessment of multivariable feedback control systems. *Automatica* **37**, 899–905.

Kourti, T. and J. F. MacGregor (1996). Multivariate SPC methods for process and product monitoring. *J. Qual. Tech.* **28**, 409–428.

Kozub, D. and C. E. Garcia (1993). Monitoring and diagnosis of automated controllers in the chemical process industries. In: *AIChE Annual Meeting*. St. Louis, MO.

Kozub, D.J. (1996). Controller performance monitoring and diagnosis: experiences and challenges. In: *Fifth Int. Conf. on Chemical Process Control* (J.C. Kantor, C.E. Garcia and B.C. Carnahan, Eds.). AIChE and CACHE. Tahoe, CA. pp. 83–96.

Kulkarni, A., V. K. Jayaraman and B. D. Kulkarni (2005). Knowledge incorporated support vector machines to detect faults in tennessee eastman process. *Computer Chem. Eng.* **29**, 2128–2133.

Lee, J., S. J. Qin and I. Lee (2006*a*). Fault detection and diagnosis based on modified independent component analysis. *AIChE J.* **52**, 3501–3514.

Lee, Y. H., H. D. Jin and C. H (2006*b*). On-line process state classification for adaptive monitoring. *Ind. Eng. Chem. Res.* **45**, 3095–3107.

Levinson, H. J. (1999). *Lithography process control*. SPIE Optical Engineering Press, Bellingham, WA.

Lin, Z. C. and W. J. Wu (1999). Multiple linear regression analysis of the overlay accuracy model. *IEEE Trans. Semiconduct. Manuf.* **12**, 229–237.

Lynch, C.B. and G.A. Dumont (1996). Control loop performance monitoring. *IEEE Trans. Cont. Sys. Tech.* **4**(2), 185–192.

MacGregor, J. F., C. Jaeckle, C. Kiparissides and M. Koutoudi (1994). Process monitoring and diagnosis by multiblock PLS methods. *AIChE J.* **40**, 826–828.

Martin, E. B. and A. J. Morris (1996). Non-parametric confidence bounds for process performance monitoring charts. *J. Proc. Cont.* **6**, 349–358.

McCulloch, C. E. and S. R. Searle (2001). *Generalized, linear and mixed models*. John Wiley and Sons, New York.

McNabb, C. A. and S. J. Qin (2003). Projection based MIMO control performance monitoring – I. Covariance monitoring in state space. *J. Proc. Cont.* **13**(8), 739–757.

McNabb, C. A. and S. J. Qin (2005*a*). Fault diagnosis in the control invariant subspace of closed-loop systems. *Ind. Eng. Chem. Res.* **44**, 2359–2368.

McNabb, C. A. and S. J. Qin (2005*b*). Projection based MIMO control performance monitoring – II. Measured disturbances and setpoint changes. *J. Proc. Cont.* **15**, 89–102.

Mehranbod, N., M. Sorough and C. Panjapornpon (2005). A method of sensor fault detection and identification. *J. Proc. Cont.* **15**, 321–339.

Mehranbod, N., M. Sorough, M. Piovoso and B. A. Ogunnaike (2003). Probabilistic model for sensor fault detection and identification. *AIChE J.* **49**, 1787–1802.

Paulonis, M. A. and J. W. Cox (2003). A practical approach for large-scale controller performance assessment, diagnosis, and improvement. *J. Proc. Cont.* **13**, 155–168.

Peng, Youbin and Michel Kinnaert (1992). Explicit solution to the singular LQ regulation problem. *IEEE Trans. Auto. Cont.* **37**(5), 633–636.

Politis, D. N. (1998). Computer-intensive methods in statistical analysis. *IEEE Sig. Proc. Mag.* **1**, 39–55.

Qin, S. J. (1998). Recursive PLS algorithms for adaptive data modeling. *Comput. Chem. Eng.* **23**, 503–514.

Qin, S. J. (2003). Statistical process monitoring: Basics and beyond. *J. Chemomotrics* **17**, 480–502.

Qin, S. J. and J. Yu (2007). Recent developments in multivariable controller performance monitoring. *J. Proc. Cont.* **17**, 221–227.

Rao, C. R. and J. Kleffe (1988). *Estimation of variance components and applications*. North-Holland, Amsterdam, The Netherlands.

Rinehart, N. and F. Jury (1997). How control valves impact process optimization. *Hydrocarbon Processing* **76**, 53–60.

Rogozinski, M., A. Paplinski and M. Gibbard (1987). An algorithm for calculation of a nilpotent interactor matrix for linear multivariable systems. *IEEE Trans. Auto. Cont.,* **32**, 234–237.

Searle, S. R., G. Casella and C. E. McCulloch (1992). *Variance components*. John Wiley and Sons, New York.

Seem, J. E. (1998). A new pattern recognition adaptive controller with application to hvac systems. *Automatica,* **34**, 969–982.

Seem, J. E. (2006). An improved pattern recognition adaptive controller. In: *Proceedings of the 2006 American Control Conference*. Minneapolis, MN.

Seifert, B. (1993). Root cause estimation and statistical testing for qualify improvement of multistage manufacturing processes. *J. Statistical Planning and Inference* **36**, 253–268.

Shah, S.L., R. Patwardhan and B. Huang (2002). Multivariate controller performance assessment: methods, applications and challenges. In: *Chemical Process Control - CPC VI*. CACHE. Tuscon, Arizona. pp. 190–207.

Stanfelj, N., T. E. Marlin and J. F. MacGregor (1993). Monitoring and diagnosis of process control performance: The single-loop case. *Ind. Eng. Chem. Res.* **67**(10), 856–861.

Tong, H. and C. M. Crowe (1995). Detection of gross errors in data reconciliation by principal component analysis. *AIChE J.* **41**, 1712–1722.

Tyler, M.L. and M. Morari (1996). Performance monitoring of control systems using maximum likelihood methods. *Automatica* **32**(8), 1145–1162.

Wolovich, W.A. and P.L. Falb (1976). Invariants and canonical forms under dynamic compensation. *SIAM J. Cont. Opt.* **14**(6), 996–1008.

Xia, H., P. Majecki, A. Ordys and M. Grimble (2006). Performance assessment of MIMO systems based on I/O delay information. *J. Proc. Cont.* **16**, 373–383.

Zhou, S., Y. Chen and J. Shi (2004). Root cause estimation and statistical testing for qualify improvement of multistage manufacturing processes. *IEEE Trans. Automation Sci. Eng.* **1**, 73–83.

Zhou, S., Y. Ding, Y. Chen and J. Shi (2003). Diagnosability study of multistage manufacturing processes based on linear mixed-effects models. *Technometrics* **45**, 312–325.

Vita

Jie Yu, the only son of Zonglai Yu and Wenxia Yang, was born in Tongling, Anhui, P.R. China on November 20, 1977. He received his high school diploma in June, 1996 from Tongling high school in Tongling, Anhui. Mr. Yu entered Zhejiang University, Hangzhou in the fall of 1996, where he received the Bachelor of Science degree in Bioengineering in 2000 and the Master of Science in Engineering degree in Biochemical Engineering in 2003. Mr. Yu entered graduate school at the University of Texas at Austin in August, 2003, and was admitted to Ph.D. Candidacy at the University of Texas at Austin in February, 2006.

Permanent address: 1600 Wickersham Lane Apt 1085
Austin, Texas 78741

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.