Performance Assessment of Multivariate Control Systems

by

Timothy Spinner, B.S.

A Dissertation

In

Chemical Engineering

Submitted to the Graduate Faculty
of Texas Tech University in
in Partial Fulfillment of
the Requirements for
the Degree of

Doctor of Philosophy

Approved

Ragunathan Rengaswamy

Ronald Hedden

Rajesh Khare

Ismael Regis de Farias Jr.

Mark Sheridan
Dean of the Graduate School

December, 2014

ACKNOWLEDGEMENTS

First and foremost, I must thank my advisor, Dr. Raghunathan Rengaswamy, for his support, advice, and guidance during my doctoral studies at Texas Tech University. I would also like to thank Dr. Rajesh Khare, Dr. Ismael de Farias, Dr. Ronald Hedden, and Dr. Ram Iyer for their service on my doctoral committee, as well as their aid in my studies.

I owe a great deal of gratitude to Babji Srinivasan for his impact on my development as a researcher. I would like to express deep appreciation to the other group members whom I have had the pleasure to spend time with: Ulaganathan Nallasivam, Brian Bullecks, Jeevan Maddala, Parham Mobed, Vidyashankar Kuppuraj, Ravi Mandela, and Srinivasan Raman; all have been friends as much as colleagues. Special thanks to Qiuying Gu and Md Islam; together we have learned so much and shared many thought-provoking discussions.

Finally, I would like to dedicate this dissertation to my family. The love and encouragement of my parents, Dean and Susan Spinner; and my sisters Emily Hart, Rebecca Spinner, and Abigail Spinner has motivated me throughout the past five years.

TABLE OF CONTENTS

ABSTRACT

Modern process plants and refineries contain hundreds to thousands of loops under automatic controls. Poorly performing control loops cause off-spec product, waste energy, and cause maintenance and safety issues. Much of the process control engineer's time is spent finding the root cause of poor performance, meanwhile many underperforming loops are left alone for sake of insufficient knowledge and resources. The field of control loop performance assessment and diagnosis provides tools for control and plant engineers to identify poorly performing loops, diagnose root causes, and find corrective actions. Much as process controls automate much of the regulation and set point tracking functions within the process plant, these new tools aim to automate, as much as possible, plant monitoring and root cause diagnosis tasks.

Two of the main problems for poor plant performance are bad controller tuning and stiction in control valves. Concerning the former problem, most process controls in industrial environments use a proportional-integral (PI) algorithm. There exists a large body of literature on PI controller diagnosis and retuning, and one aim of this document is to classify and compare these methods to allow better understanding of their potential uses. Also herein, the use of the Hurst exponent as a tuning diagnosis measure is proposed, as well as retuning algorithms that can

incorporate a wide variety of existing tuning diagnosis measures in order to make retuning decisions. With respect to valve stiction, new theoretical results on the efficacy of Hammerstein model based valve stiction detection are presented. In addition, the Hammerstein model based stiction detection method is extended to the important case of interacting systems.

One key component of control loop performance assessment is benchmarking performance. When creating control loop performance benchmarking tools for multivariate control loops, there exists a large number of possible objectives to pursue. Early attempts focused on creating mathematical tools that could provide the theoretical minimum sum of variances of output variables. However, this objective is not always aligned with the goal of plant personnel, which is to increase economic benefit of the plant operation. Therefore, this thesis pursues economic benchmarking of multivariate control systems, which involves trading off variances of different process variables in order to achieve an optimal operating point. Earlier work in this area relied on non-convex programming with approximate solution methods. The current work instead performs a piecewise linear approximation of the non-convex constraints, allowing for a linear programming solution of the problem.

This document proposes tools that allow for an overall control loop assessment and diagnosis framework to be implemented. In the single-input single output case, this framework allows for completely data-based performance assessment and root cause diagnosis to be performed. For multivariate systems, the tools require more

system information than in the single-loop case, but with this cost comes the additional benefit of performance data directly related to the operation's economic objectives. In each case, the performance assessment and diagnosis use only non-invasive methods, so plant operation is not disrupted to perform the analysis. Adoption of these tools will allow for more efficient and better performing plant operations.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

OVERVIEW OF CONTROLLER PERFORMANCE ASSESSMENT

Process control refers to the automatic regulation and set-point tracking of chosen variables within a system that involves the processing chemicals or materials. This broad definition incorporates the use of process control techniques within the dissimilar fields of pulp and paper manufacturing, production of chemicals and chemicals products, petroleum refining, steel rolling, and electricity generation from fossil fuels. A process under automatic control typically produces products more safely, cheaply, reliably and with better precision than human operators can achieve alone. However, automated process controls are not without problems. Indeed, industrial studies have repeatedly shown that up to $2/3$ of control loops have performance issues [1, 2, 3]. These problems can lead to economic costs such as off-spec product that needs to be discarded or sent for reprocessing [4], energy costs due to actuator cycling, as well as safety issues.

To expand on the previous statement that $2/3$ of control loops have performance issues, a survey of $26,000$ industrial control loops had 22% rated as performing fairly, 10% rated as poorly performing, with 36% of controllers in open-loop, and only 32% rated at least acceptably [1]. ”Open-loop” refers to a situation where

the controller isn't performing any meaningful function, either because it has been switched off by an operator or else because the actuator output is saturated for an extended period of time. Malfunctions or degradation of control loop components such as sensors and actuators is very common during plant operation [5]. Improper tuning of controller parameters or selection of the wrong controller type is also responsible for many problems.

The field of controller performance assessment (CPA) seeks to aid plant personnel by providing noninvasive methods for detection of poorly performing control loops, diagnosis of the root cause for poor performance, and, where possible, providing solutions to correct identified problems. These actions are alternatively referred to as closed-loop monitoring and diagnosis (CLPM&D). The field shares goals with that of statistical process control, in that both concern the identification and reduction of excess variances in product quality. CPA is also closely aligned to the field of process fault detection, in that both seek to identify periods of abnormal performance and isolate the underlying causes. The main focus points of controller performance assessment are performance benchmarking, oscillation detection and diagnosis, and assessment of controller structure and tuning.

1.1 Review of Control Loop Assessment Techniques

The field of controller performance assessment has been rapidly growing since the appearance of the seminal work by Harris 25 years ago [6]. In that work, it was shown that a theoretical lower bound on the variance of the output variable of a single-input single-output (SISO) control loop could be constructed using only routine operating data and the time delay of the open-loop plant. This development led to the wide-spread use of the minimum variance index, which is the ratio of the theoretical minimum variance against the variance measured at the current time. Benchmarking loop performance with the minimum variance index allowed for problematic loops to be more easily identified in industrial processes [7], at which point diagnosis steps could be undertaken. As oscillations are one of the largest detriments to plant performance, techniques for automated detection were investigated [8, 9]. Control valve stiction and poor control loop tuning were identified as large contributors to the occurrence of oscillations and other performance problems, so methods to detect and correct these issues have also been a major priority for the control loop performance assessment field.

1.1.1 Performance Benchmarking

In the years following the initial publication on the minimum variance index, many new methods were introduced to extend this benchmark to different situa-

tions. For cases where the time delay is uncertain or where the user wants to define an easier to achieve benchmark, the use of the extended prediction horizon index has been proposed [10, 11]. A benchmark that used step disturbance rejection as an alternative performance goal was introduced [12]. It was strongly desired to extend the idea of benchmarking control loop performance to cases other than SISO loops having a general controller structure. Many works were concerned with performance benchmarking of multiple-input multiple-output (MIMO) systems. In MIMO processes, one or more of the manipulated variables affect more than one of the controlled outputs, which leads to interactions between separate control loops. Initial attempts to quantify the theoretical achievable output variance relied on the use of system interactor matrices [13, 14, 15]. The interactor is a relatively difficult to understand concept [16], and so Ko and Edgar introduced a method to calculate the same performance bound without calculating this matrix [17].

Several problems were identified in the use of minimum variance benchmarking. The first was the high requirement for open-loop plant information. Recently, for the assessment of SISO control loops, Srinivasan *et al.* [18] proposed the use of detrended fluctuation analysis for performance assessment, which does not require the open-loop time delay. For multivariate systems, Huang *et al.* proposed the use of the order of the interactor matrix instead of the full interactor matrix to quantify an approximate minimum variance benchmark [19]. This work also introduced the closed loop potential, which is a measure of the prediction error compared to a dead-

beat controller. Zhao *et al.* introduced the multi-step prediction error benchmark [20], wherein the impulse coefficients of the prediction error, normally employed for calculation of minimum variance benchmarks, are reweighted. This is done with accordance to the goal of some types of controllers to minimize the predicted error over a finite horizon.

Another problem frequently stated with the minimum variance benchmark is that the benchmark leads to too severe movements in the manipulated variables in an attempt to reduce the variance of output variables. Horch and Isaksson proposed a benchmark for SISO systems in which a single closed-loop pole is introduced for the decay of the output error after the time delay (similar to Lambda tuning methods) [21]. Huang and Shah proposed a similar concept allowing user specified dynamics for MIMO systems [22]. An important development was the introduction of linear quadratic Gaussian (LQG) benchmarking, wherein the trade-off of the variances of the manipulated and controlled variables is explicitly recognized [23]. In LQG benchmarking, the current loop performance is plotted against the Pareto optimal curve of input and output variables in order to judge whether the system lies close to some optimally controlled point.

Qin and Yu noted that multivariate minimum variance control also leads to minimum covariance between different output variables [24]. This led to the development of a data-based technique for performance assessment in which generalized eigenvector analysis of the covariance matrix was used to determine the worst di-

rections of performance compared to a historical benchmark [25]. After determining these directions, the contribution of the original set of outputs to each worst direction is identified and the highest contributing variables are assigned for further diagnosis [26]. On the other hand, Harris and Yu investigated the concept of degree-of-freedom analysis as a performance assessment tool [27]. The degree of freedom is defined as the difference between the total number of output variables and the number of manipulated variables that are not currently saturated. Positive degree of freedom allows the controller to regulate the output variables while simultaneously seeking to achieve other performance objectives, whereas negative degree of freedom means that some outputs cannot be maintained at their set points [27]. Gigi and Tangirala have investigated the quantification of interaction between control loops in decentralized control systems using frequency domain techniques [28, 29, 30]. Interaction is a fundamental factor determining the attainable performance by MIMO systems, although more work is required to make the connection clear.

1.1.2 Performance Assessment of Model Predictive Controllers

An area of increasing interest is the performance assessment of systems under model predictive control (MPC). An MPC controller determines each control move by solving an optimization problem, usually the sum over a finite horizon of a weighted combination of squared output errors and expended controller en-

ergy. The MPC formulation uses a model of the plant behavior to predict the effects of controller moves on the future plant behavior, and also allows the presence of constraints on manipulated and controlled variables. Hugo has stated that the previously introduced LQG benchmarks are unattainable for most industrial implementations of MPC due to the simplified random walk disturbance model they rely upon [31]. This disturbance model structure often inhibits performance, therefore, Julien *et al.* have proposed a benchmark where this inhibition is accounted for [32]. Schäfer and Cinar have proposed the use of a historical benchmark for MPC performance assessment, wherein the benchmark performance is chosen from a past operating period deemed to have been the best [33]. In alternative methods for MPC performance assessment, it has often been suggested to rerun the system in a simulation environment and to redesign the controller subject to the same operating point and constraints [34, 35, 36].

A particularly active area of MPC performance assessment is the automated identification of so-called model plant mismatch (MPM), a condition that occurs when the MPC optimizer uses a poor model of the real system behavior leading to inaccurate predictions. Loquasto and Seborg proposed the use of the pattern classification techniques such as neural networks or principle component analysis for detecting MPM [37, 38]. Badwe *et al.* proposed monitoring the correlation between model residuals and the manipulated variable signals to identify this problem [39], which is a similar idea to that proposed much earlier by [7] for loops under internal

model control. Badwe *et al.* also proposed the use of frequency domain techniques which can predict the effect of model plant mismatch on the economic performance [40]. Selvanathan and Tangirala also used a frequency domain technique that can separately identify mismatch in the model gain, time constant, or time delay [41]. Alternatively, Harrison and Qin proposed monitoring the autocorrelation of state estimate errors, which is also indicative of MPM [42].

1.1.3 Economic Performance Assessment

The concentration on reducing the variance of the output variable of the control loop sometimes leads to forgetting the primary objectives for doing so, which usually involve increasing operating profits. By reducing output variance, it is hoped to push the operating point closer to the constraint boundary as this is often where optimal economic performance is achieved [43]. However, it is usually the case that reducing the variance of one variable will cause another variable to increase in variance. Therefore, several works have treated the trade-off of variances of different control loop variables as constraints on achieving the maximum economic benefit. This problem is called minimum back-off operating point selection, as the operating point must be ”backed-off” from the operating constraint boundaries due to the normal variances encountered during dynamic operation [44, 45]. Peng *et al.* give an overview of many of the available formulations for this problem, finding problems in most of the existing formulations [46]. The formulation constructed

by Peng *et al.* results in a non-convex optimization problem which simultaneously allows for the determination of operating point and controller structure.

1.1.4 Non-Invasive Detection of Control Valve Stiction

Control valve stiction is one of the major root causes of plant-wide oscillations in the process industries [47]. This has led to a large number of works on stiction detection. The data-based Hammerstein model stiction detection algorithm of Srinivasan *et al.* [48] has been followed by many variants [49, 50, 51, 52, 53]. The advantage of this approach is that no *a priori* model is needed of the control loop, instead relying on the excitation of the loop variables by the presence of stiction in order to construct a model used within the stiction detection algorithm. It is noted that in some cases the Hammerstein model based approach fails to identify that stiction is present in loops where it occurs [54, 55]. Recently, Srinivasan *et al.*, have proposed an explanation for this failure [55]: if the controller bandwidth is larger than the process bandwidth, then the same frequency information will be present within both signals. Since the Hammerstein model based approach takes advantage of stiction excitation that is usually available in the plant output but not the controller signal, identification will fail in these cases. Srinivasan *et al.* have also proposed a reliability measure to quantify whether enough extra information is available in the plant output signal for the Hammerstein model stiction detection to work properly.

There are few methods available for isolation of stiction to a single valve within an interacting MIMO system. Haoli *et al.* proposed to sequentially change the gain across every loop of the interacting system [56]. In this method, the controller whose change in gain produces the largest shift in oscillation frequency is assumed to be connected to the valve with stiction. However, no theoretical backing is proposed, and it is unknown whether this method can have wide-spread application. Other than this, the only known solution to isolate stiction within an interacting system is to put the suspected loops under manual control and perform experimental tests. This type of manual diagnosis is labor intensive, disruptive to the process, and potentially unsafe [57].

1.1.5 PI and PID Controller Assessment

By far, the PID (proportional-integral-derivative) control law and particularly its PI (proportional-integral) implementation are unmatched in prevalence within the process industries, with one industrial survey reporting that 97% of low-level controllers had a PI or PID structure [1]. Meanwhile, close to 1000 research articles appear every year on the PI and PID control algorithms. As with the larger field of process control, there has been increasing interest in recent years to monitor PI/PID controller performance, diagnose problems, and take corrective action, in an automated fashion where possible. Therefore, an extensive review of the literature on PI controller performance diagnosis and automated retuning follows in Chapter 2.

1.2 Framework for Control Loop Performance Assessment and Diagnosis

Figs. 1.1 and 1.2 show proposed frameworks for controller performance assessment of single-input single-output and multiple-input multiple-output systems, respectively. The framework in Fig. 1.1 is a modification of the framework presented within [58] and is updated to reflect the state of the available techniques, while Fig. 1.2 is the extension of this framework to multivariate systems. As in previous frameworks for controller performance assessment and root cause diagnosis [5], a first step is looking for the presence of oscillations in control loop signals. Since oscillations are detrimental to control loop performance, it is assumed that the control engineer will want to diagnose and mitigate any oscillation sources before investigating other potential problems. Another reason for starting with oscillation detection is that the use of controller performance indices, such as the Hurst Index and the minimum variance index, are not recommended for use on oscillatory loops [59, 18].

As in the assessment tree for SISO systems in Fig. 1.1, if no oscillations are detected, then assessment of the loop can proceed according to one of several performance measures. Here, in Fig. 1.1, it is recommended to use the data-based measure, called the Hurst Index, as it requires no model information in its calculation [18]. Currently, in the process industries, model information is not available

for the large majority of control loops in a form that could be easily accessed by a large-scale automated assessment tool [60]. If the open-loop time delay is available, then the minimum variance index proposed by Harris can be used [6], and if a full plant model is available then either the generalized minimum variance benchmark can be used [61], or a benchmark specific to the particular controller type, such as the attainable PI performance benchmark [62], can be implemented. Other specialized benchmarks for single-loop control include ones for cascade control [63] or restricted structure control laws [64].

Whichever performance measure is chosen, the point is to classify the current loop performance as acceptable or problematic. If performance is not acceptable, then the diagnosis would move to the next level of the tree of Fig. 1.1. Single-input single-output control loops are very likely to use a proportional-integral control law. Therefore, it is useful to have a tool to diagnose problematic loops of this type, and so, Fig. 1.1 has includes a component for the diagnosis and retuning of this type of control loop.

For MIMO control loops, Fig. 1.2 shows that the first step is again to perform oscillation detection. In the case of no oscillations, the benchmarking proposed is the use of economic performance assessment. This is because a multivariate system, as the name implies, will have many variables, and the improvement in the quality of one variable will often be to the detriment of others, therefore this trade-off should occur in the context of increasing economic benefit. For oscillatory

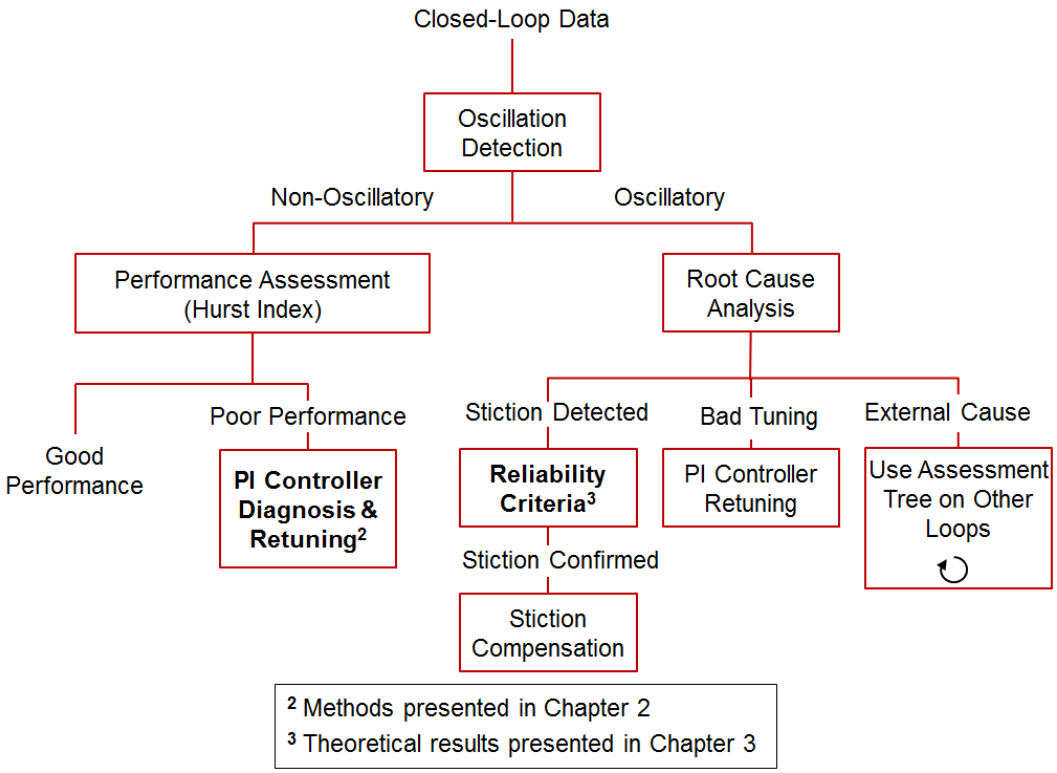

Figure 1.1: Performance assessment and root cause diagnosis tree for single-input single-output control systems

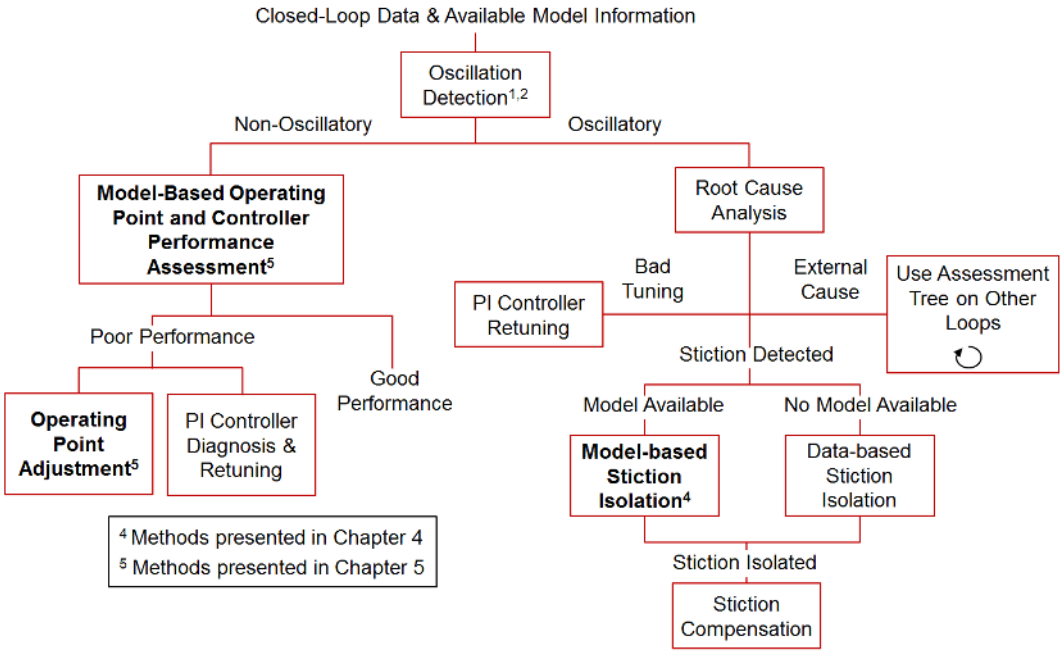

Figure 1.2: Performance assessment and root cause diagnosis tree for multiple-input multiple-output control systems

systems, stiction detection and isolation remains the main priority. Because other parts of the multivariate performance assessment procedure require a model, it is possible that a model will be available for use by the stiction detection algorithm. However, for instances when this is not the case, it is also proposed that data-based stiction detection in interacting systems should be pursued in the future.

1.3 Motivation

The control loop assessment trees of Figs. 1.1 and 1.2 provide a framework for determining control loop performance and isolating the cause of problems. However, several blocks on these trees need more investigation before the overall framework can be implemented:

- The literature on data-based tuning diagnosis and retuning of PI type controllers is fractured. An overall framework for viewing the multitude of different methods available for these purposes is required. Meanwhile, sluggish and aggressive tuning are commonly used terms with no commonly accepted definition. Automated retuning algorithms have been proposed that do not monitor the performance objective that they seek to improve [65]. Furthermore, stability of existing data-based retuning algorithms has not been explored.

- It has been noted that automated stiction detection algorithms fail for certain linear systems [54]. Recently, an explanation has been proposed along with an associated reliability measure for Hammerstein model based stiction detection [55]. However, theoretical backing for the proposed explanation has been lacking.

- Very little literature has appeared on isolation of stiction within interacting MIMO systems. Previous techniques rely on overly invasive methods such as making large changes in the controller gains several times throughout the system [56]. For SISO loops, the Hammerstein model based stiction detection technique requires only routine operating data and no invasive diagnostics [48]. Additionally, model information may be available for isolation of stiction within interacting systems.

- For economic performance assessment proposed in Fig. 1.2, existing techniques for creating a performance benchmark rely on the use of non-convex optimization problems, with computationally expensive branch and cut algorithms [46], or else the use of a two stage optimization process [66]. The optimization problem could be more efficiently solved by linear programming if piecewise linearization of the constraints is performed.

1.4 Objectives

Given the stated desire to fill gaps within the control loop assessment trees of Figs. 1.1 and 1.2, this thesis seeks to perform the following functions:

1. Present a unified classification for data-based loop tuning diagnosis and retuning techniques. Compare the class of data-based loop tuning diagnosis measures that rely on closed-loop operating data without artificial excitation or set-point changes. Propose the use of the Hurst exponent as an additional loop tuning diagnosis measure. Introduce a modification of data-based control loop retuning techniques for PI controllers in order to increase the performance of these algorithms. Highlight the issue of possible algorithm instability.

2. Provide theoretical justification for the proposed explanation of the failure of Hammerstein model based stiction detection within certain linear systems.

3. Extend Hammerstein model based stiction detection to MIMO systems. Allow for noninvasive isolation of detected stiction to a single valve within the interacting system.

4. Introduce a new method for the economic performance assessment of constrained multiple-input multiple-output systems. Transform the problem into

a simple form that can be solved through linear programming. Demonstrate its use for selection of the system operating point.

1.5 Organization

Fig. 1.3 provides an overview of the remainder of this document. Within this work, a variety of control loop performance assessment problems have been considered, and these can be categorized based on whether they apply to single-input single-output systems or for multivariate systems. Chapter 2 is concerned with the data-based performance assessment and retuning of single-input single-output control loops under PI control. After an in-depth review of existing techniques, the use of the Hurst Exponent as a loop tuning diagnosis measure is introduced, followed by the introduction of two new automated controller retuning algorithms. Simulation studies provide a comparison of several loop tuning diagnosis techniques and retuning algorithms. Chapter 3 introduces a theoretical result explaining the failure of stiction detection algorithms in certain linear SISO systems. Simulation examples demonstrate the effect of the theorized phenomenon in practical situations. Chapters 4 and 5 are concerned with performance assessment and improvement of multiple-input multiple-output control systems. First, an extension of Hammerstein-based stiction detection algorithm to the case of MIMO loops under decentralized control is proposed in Chapter 4. Simulation trials demonstrate the potential of this

technique. Following this, in Chapter 5, an algorithm for performance assessment and operating point selection for multivariate constrained control systems is introduced. Simulation examples illustrate the benefits of this new method. Finally, within Chapter 6, a summary of this work and its contributions is presented, and future work is proposed.

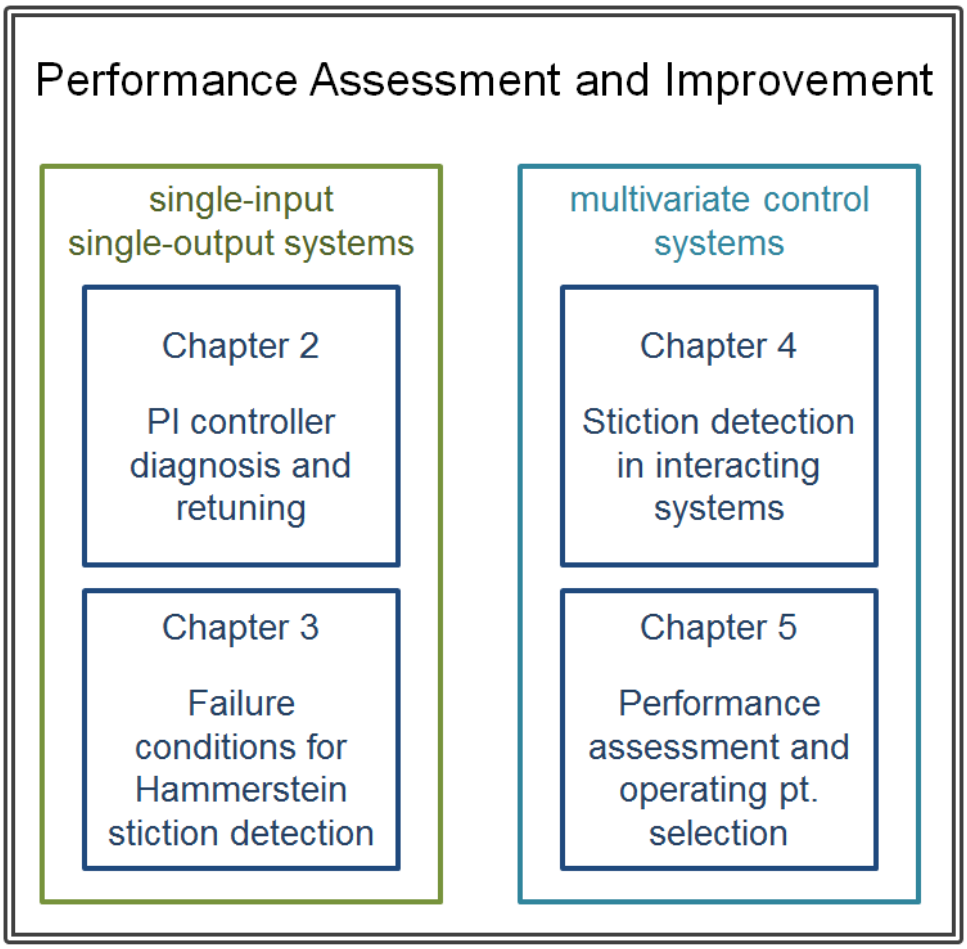

Figure 1.3: Outline of material contained in this document

Nomenclature

CLPM&D	closed-loop performance monitoring and diagnosis
CPA	controller performance assessment
LQG	linear quadratic Gaussian
MIMO	multiple-input multiple-output
MPC	model predictive control
MPM	model plant mismatch
PI	proportional-integral
PID	proportional-integral
SISO	single-input single-output

CHAPTER 2

AUTOMATED DIAGNOSIS AND RETUNING OF PROPORTIONAL-INTEGRAL (PI) CONTROLLERS

The aim of this chapter is to present a new look at the existing data-based and non-intrusive PI (proportional-integral) controller tuning assessment methods for SISO (single-input single-output) systems under regulatory control. Poorly tuned controllers are a major contributor to performance deterioration in process industries both directly and indirectly, as in the case of actuator cycling and eventual failure due to aggressive tuning. In this paper, an extensive review and classification of performance assessment and automated retuning algorithms, both classical and recent is provided. A subset of more recent algorithms that rely upon classification of poor tuning into the general categories of sluggish tuning and aggressive tuning are compared by their diagnostic performance. The Hurst exponent is introduced as a method for diagnosis of sluggish and aggressive control loop tuning. Also, a framework for more rigorous definitions than previously available of the terms "sluggish tuning" and "aggressive tuning"are provided herein. The performance of several tuning diagnosis methods are compared, and new algorithms for using these tuning diagnosis methods for iterative retuning of PI controllers are proposed and investigated using simulation studies. The results of these latter studies highlight

the possible problem of loop instability when retuning based upon the diagnoses provided by data-based measures. This chapter is excerpted from paper T. Spinner, B. Srinivasan, R. Rengaswamy. "Data-Based Automated Diagnosis and Iterative Retuning of Proportional-Integral (PI) Controllers. *Control Engineering Practice*, 29:23-41,2014".

2.1 Introduction

The typical process control engineer is responsible for several hundred or more loops. They must split their time between implementing new assets and maintaining existing controllers [1]. Perhaps as a result, surveys report that more than 60% of controllers provide less than acceptable levels of performance, leading to poor product quality and loss of production [1, 3, 2]. These industrial surveys indicate that at least 30% of control loops were increasing the variability of the process variable compared to the use of manual control, and another 36% of processes were in open loop. Among the major causes exists poor controller tuning, with sometimes one quarter of all loops never adjusted from the default controller parameters [2]. Lack of manpower and lack of tuning knowledge, combined with the time varying nature of process and disturbance behaviors [67] make the disappointing results unsurprising. The field of closed-loop performance monitoring and diagnosis (CLPM&D) seeks to provide tools to aid plant personnel in identifying poorly performing loops

and suggesting remedial action, which may include controller retuning.

CLPM&D is a maturing area with several excellent articles [68, 5, 69, 24, 70] and books [23, 65, 59] providing a general overview, with additional monographs available on more specialized topics such as valve stiction detection and diagnosis [71, 72]. Some of the major causes for poor control loop performance that CLPM&D techniques attempt to identify include oscillatory disturbances [73], sensor or actuator faults (such as in the case of valve stiction), or poor controller tuning. Interest in the field of CLPM&D has increased dramatically following the appearance of Harris's 1989 paper [6] on the minimum variance benchmarking of loop performance. Comparing against the theoretical minimum variance of the process variable provides control engineers a way to quantitatively assess the current performance of each loop. Since the original minimum variance benchmark only considered the performance limiting effects of time delay, subsequent works sought to include other limitations on loop performance such as interactions in multivariate systems [13, 15], right half plane zeros [74], but especially the effects of restricted controller structure [64]. Notably, several methods exist wherein process and disturbance model information for a given system is used in an attempt to define an upper bound on performance achievable by PI or PID control [62, 75, 76].

PID is the dominant controller algorithm within the process industries, with one survey reporting that 97% of controllers were of this type [1], while other references reporting that the actual implementation of these controllers is usually in PI form

[77, 78]. Although the seminal work of Ziegler and Nichols occurred over 70 years ago [79], research in the field of PID control is still experiencing a rapid growth in the number of publications [80], and a volume by ODwyer contains over 400 tuning correlations for PID controllers [81]. Still, proper tuning of these controllers for optimal performance is not always a priority that the responsible engineer has time for. Often these PI controllers are retuned only in the case that oscillations have been detected and thus remain sluggishly tuned otherwise due to lack of manpower or expertise [82]. Tuning in times of minimal disturbance can result in loops unable to properly attenuate common disturbances properly. Retuning is also necessary due to process changes or changes in operating regimes. Instrument wear and equipment fouling (e.g. [83]) cause the process dynamics to drift and time delays to increase. When applying PI or PID control to nonlinear processes, a change in operating regime should be an impetus for parameter retuning. Controller or process parameter changes within interacting loops will also necessitate retuning. All of these scenarios indicate a need for CLPM&D techniques to identify problematic loops in need of retuning.

Long before the formal advent of the CLPM&D field, there has existed a wide assortment of automated methods to assess and correct poor controller tuning, these belonging to the fields of automatic tuning and adaptive control [84, 85]. In fact, it is likely that many of the techniques discussed throughout this work were present in industry long before being documented in the literature. It is proposed that these

existing methods should be able to achieve improved outcomes if combined with other CLPM&D techniques. For example, many adaptive tuning methods would identify excessive oscillations as being associated with aggressive controller tuning [86], when in fact they could be present for a variety of reasons such as valve or sensor degradation other sources of hysteresis within the loop, or externally induced oscillations beyond the capable of loop to compensate. This is why it is recommended [5] to attempt detection of other specific types of malfunctions, for example by applying nonlinearity indices [87] or Hammerstein stiction detection methods [48], before assigning poor loop performance to controller tuning. In this way, controller tuning will not be inadvertently worsened when other problems are afflicting the loop.

CLPM&D based PI/PID tuning assessment techniques and the previously existing data-based automated tuning methods can be classified in a similar way. In one category, a model of the open loop process and possibly a model of the disturbance filter are used to calculate an upper bound on performance referred to as PI or PID achievable performance, and then the current performance is judged against this benchmark. As a result of the parameter optimization used to predict the best achievable performance, the optimal controller tunings are also acquired. However, the requirement of model information by these assessment techniques is not easily achieved. Process models are available for only a small minority of control loops [1]. Therefore, use of model-based techniques produces the need for either iden-

tification experiments or else the existence of specific conditions (excitation by set point changes) that may not be present in most loops. Thus the model information dependent techniques may be severely limited in applicability. Other types of PI/PID controller assessment and retuning require only operating data that includes responses to either step-type or stochastic disturbances, in order to diagnose and/or correct tuning problems. The comparative disadvantage of this type of assessment technique is that they cannot produce knowledge of the distance from the optimal set of tuning parameters, so that retuning with these techniques requires an iterative approach towards acceptable performance. However, the largest advantage of these data-based methods is that they can be implemented without the expense and intrusion of plant identification experiments.

2.1.1 Contributions of this Chapter

Data-based techniques for controller tuning assessment and correction have an important role to play in increasing process plant performance. This work presents review and discussion of several aspects concerning data-based diagnosis and retuning and new ideas and improvements to existing techniques are proposed. It should be noted that [65] gives a comprehensive overview of a wide range of topics within the CLPM&D framework and especially that Ch.13 of [65] includes the basis of several techniques improved by this work. In the following, we present:

1. A new categorization of the multitude of available loop tuning assessment and retuning techniques is proposed (Section 2.2).

2. A subset of the tuning assessment techniques is selected, concentrating on several diagnosis measures which categorize poor controller tuning as either sluggish or aggressive. First a description of each the selected diagnosis measures are provided (Section 2.3).

3. A novel use of the Hurst exponent as an additional tuning diagnosis measure will also be explored. Section 2.4 provides details of this schema.

4. Several definitions of the classifications "sluggish" and "aggressive" are examined, and a rigorous definition of these terms is proposed (Section 2.5).

5. Comparison studies rate the selected diagnosis measures based upon correct classification of PI controller parameters sets into the newly defined sluggish and aggressive categories. The diagnosis measures are calculated upon the responses of systems subjected to a step in load disturbance magnitude (Section 2.6).

6. Finally, use of the selected diagnosis measures within an iterative retuning algorithm is explored (Section 2.7).

Throughout, issues with and limitations of the use of data-based techniques are highlighted. The problem of stability during retuning and the insufficiency of

current and proposed techniques in this regard are stressed.

2.2 Classification of Tuning Assessment and Retuning Techniques

The idea of using automated performance monitoring and adaptive tuning of PID controller parameters has existed for many years. By 1950, Caldwell [88] had proposed an intricate mechanical design for adjusting the tuning knobs of a PID controller in order to reach Ziegler-Nichols [79] type tuning. In Caldwell's design, the integral and derivative gains were adjusted to be a set proportion and inverse proportion, respectively, of the closed-loop cycling period, while the proportional gain was further adjusted to prevent either continuous cycling or long aperiodic deviations. The act of continuously varying a controller's parameters may seem quite different from the task of monitoring control loop operating data and alerting operators when a poor condition is detected as a control loop tuning assessment tool might do. However, techniques for completing either duty can be similarly classified based upon the information they require from the loop under consideration along with the way in which they process this information.

Among automated tuning techniques, generally referred to as autotuning or adaptive algorithms, a first level of demarcation can be made based upon the method used for initiation of tuning [84, 89]; autotuning is defined as the case where the human user commands the system to undergo a single automated determination of new

controller parameters, whereas adaptive controllers continuously monitor the process and initiate tuning based upon their internal programmed logic. Among other identifiable differences, autotuning almost certainly involves some plant experiment to obtain the required information for tuning to occur, while adaptive control may or may not require plant experiments. Leva *et al.* [89] highlights a hybrid case, in which adaptive tuning is not performed continuously, but only upon fulfillment of some requirement, e.g. the control error exceeding some threshold. This is how most commercial applications of adaptive tuning behave in the process industries; Leva *et al.* and Jelali [65] refer to this type of implementation as autotuning, while strm and Wittenmark [84] refer to it as adaptive control. Here, we take the notation of strm and Wittenmark, so that controllers that even sporadically retune without user intervention are called adaptive. Adaptive controllers that retune upon detection of specific criteria use similar methods to CLPM&D tools that continuously monitor the process and then perform diagnostic functions when some certain conditions are detected.

For the remainder of this section, we will group PID adaptation methods together with PID controller assessment methods, and segregate this combined group of methods based upon some essential characteristics. The reasoning behind this is that several proposed CLPM&D methods incorporate a retuning function [65, 90], and several others provide estimated optimal controller parameters as a byproduct of their controller assessment procedure [62, 75]. Likewise many adaptive con-

trollers incorporate monitoring functions to detect when retuning is necessary and use diagnostic techniques to decide upon new controller parameters [86, 91]. Figure 2.1 shows the categorization of techniques for automated PID performance assessment and retuning. Techniques in Category I are classical adaptation techniques, with theory and practical aspects discussed in depth in [84] and [85] among other sources. These adaptive control algorithms can require intrusive methods such as relay feedback tests or the injection of pseudo-random binary sequences (PBRS) [69, 65], or more simply the switch to P-only control [92]. These techniques are distinguished from techniques in Category II, which do not require a priori models, change in feedback configurations, or artificial excitation of the loop in order to perform, with the possible exception of necessitating a single set-point change. This practical way of categorization of methods is based on the demands of many industrial users that there be no plant tests or upsets to the process under consideration during performance assessment and correction procedures [5]. Whether set-point response is a feature contained in normal operating data-sets varies , since in many loops of a process plant, set-point might never be changed [93]; however, there exist some applications where this data is readily available, and in other cases some of the existing methods relying on set-point responses may be modified to be made applicable to load disturbance responses, as in [94]. Whether a technique is meant for performance assessment or parameter retuning, it usually contains some diagnostic capability to determine the state of the loop.

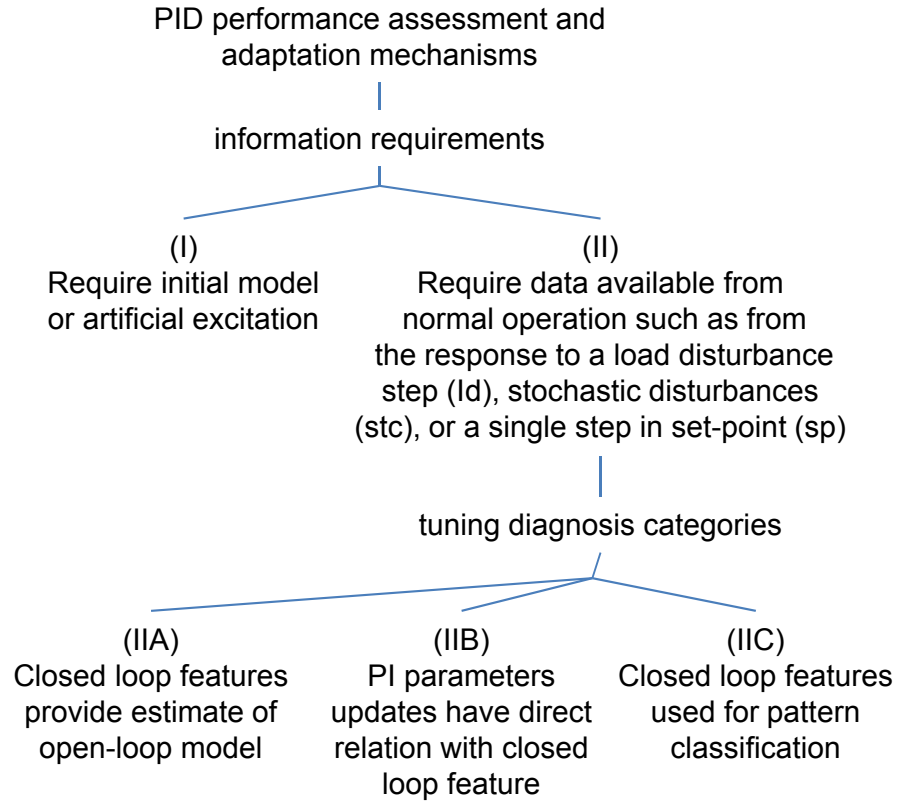

Figure 2.1: Classification of techniques for tuning assessment and retuning

It is shown in Figure 2.1that techniques in Category II are further divided into 3 subcategories based upon the way in which their functions are performed. Tables 1-3 provide a survey of techniques from categories IIA, IIB, and IIC, respectively. Within category IIA, features of the closed-loop response to a step change in either set-point or load disturbance magnitude are observed, and these features are combined with knowledge of the current controller settings to identify the parameters within simplified FOPTD (first order plus time delay) or SOPTD (second order plus time delay) open-loop model structures that are assumed for the plant under control. From the open-loop models, one of the many existing methods for PI/PID controller synthesis is used to generate a new set of tuning parameters according to the specified goal of the control loop. Notable in Category IIA is the work of Veronesi and Visioli, who applied a model simplification and parameter estimation technique to develop controller assessment and retuning techniques, first for self-regulatory loops with set-point changes [94], then later extended this method to the case of distributed-lag processes [95], processes with integrating plants [96], as well as cascade control loops [97]. Most recently they adapted their method for assessment and retuning of processes mostly subjected to load disturbances [94] and also presented methods for tuning of set-point filters when present.

Each of Table 2.1 - 2.3 contains a column describing the information required for performance assessment or retuning. This information requirement is of three types: response to a step change in set-point (sp), response to a step change in

load disturbance (ld), and response to stochastic disturbances (stoch). In Table 2.2, methods for adaptive tuning of PI and PID controllers are presented from Class IIB. These rely on identification of certain features of the closed-loop response (rise time, overshoot, period), and then setting the controller parameters as a prescribed function of the identified feature values. Several authors state that this procedure is similar to carrying out Ziegler-Nichols tuning, but in reality, it differs in several respects: (i) the features are identified on normal transient responses and not responses inflicted on the loop by experiments, and (ii) the responses may be decaying oscillatory in nature, rather than continuously cycling. Of the adaptive techniques of Category IIB, at least two have reached commercial success, these being the Foxborro EXACT controller which was installed on thousands of loops [84] and the PRAC (Pattern Recognition Adaptive Controller) which has been implemented in 500,000 digital controllers [86].

Table 2.3 describes members of Class IIC, methods which first perform categorization of the control loop tuning state before suggesting retuning action. The remainder of this work concentrates on this category of methods. From Table 2.3, it is clear that several of the methods involve classifying loop tuning into the categories "sluggish" and "aggressive". Rigorous definitions of these sets are absent from the literature, so Section 2.5 discusses a proposed classification scheme. The methods of classes IIB and IIC are the ideological descendants of the pattern recognition adaptive control methods pioneered by Bristol [121, 122]. He championed these

Table 2.1: Techniques for PID controller assessment and/or tuning from Category IIA

Reference	Type of Technique	Information Requirement	Model Estimation	Goal
Morilla *et al.* (2000)[98]	Adaptive PID retuning	sp	Estimates SOPTD (second-order plus time delay) closed-loop model, then uses knowledge of controller parameters to construct a FOPTD (first-order plus time delay) open-loop model	Achieve specified closed-loop relative damping
Bai and Zhang (2006)[99]	Adaptive PI retuning	sp/ld	Estimates discrete time FOPTD model using recursive least squares	λ - tuning
Qu and Zaheeruddin (2004) [100]	Adaptive PI retuning	sp	Estimates discrete time FOPTD model using recursive least squares. Needs user specification of time delay for accuracy.	Robust set-point tracking
Sung and Lee (1999) [101]	Adaptive PID retuning	sp	Use least-squares to identify high order open loop transfer function and then perform model reduction to obtain SOPTD model	Minimum ITAE
Veronesi and Visioli [102, 94, 95, 96, 97, 103]	Assessment and adaptive retuning of PID controllers	sp/ld	Identifies FOPTD or SOPTD plant models using current controller parameters and integrals of loop signals	Benchmarks loop IAE against IMC control
Yu *et al.* (2011) [104]	Assessment and adaptive retuning of PID controllers	sp	Identifies FOPTD or SOPTD plant models using current controller parameters and integrals of loop signals	Benchmark loop IAE against IMC control for various reference signals
Yu *et al.* (2012) [105]	Assessment and adaptive retuning of PID controllers	ld	Identifies FOPTD plant models using measured disturbance, current controller parameters and integrals of loop signals	Benchmark loop IAE, robustness, control effort, against DS-d control [106]

Table 2.2: Techniques for PID controller assessment and/or tuning from Category IIB

Reference	Type of Technique	Information Requirement	Parameter Tuning	Goal
Hensel *et al.* (2012a,2012b) [107, 108]	Adaptive PI retuning	sp	Steady state gain, overshoot	For remote sensors, utilizing level crossing sampling, minimize the objective function $J = Nsc \times ISE$ where Nsc is the number of samples transmitted
da Silva (1988) [109]	Adaptive PID retuning using expert system	sp/ld	Observed damping, overshoot, and period of decaying oscillatory response	Reach desired damping, overshoot, and period of decaying oscillatory response
Litt 1991 [110]	Adaptive PID retuning using expert system	sp	Observed damping, overshoot and overshoot ratio, rise time, and period of decaying oscillatory response	Reach desired damping, overshoot and overshoot ratio, rise time, and period of decaying oscillatory response
Seem 1998 and 2006 [91, 86]	Adaptive PI retuning	sp/ld	Extrema and slopes within responses to step changes in load or set-point	Minimize IAE
Bristol and Kraus 1984 [111], Kraus and Myron [112] (Foxboro EXACT)	Adaptive PID retuning	sp/ld	Observed damping, overshoot, and period of transient response	Desired damping, overshoot, and period

Table 2.3: Techniques for PID controller assessment and/or tuning from Category IIC

Reference	Type of Technique	Information Requirement	Categories	Categorization based upon	Goal
Zhou and Liu 1998 [113]	Adaptive PI retuning	sp	Long settling time with oscillations / acceptable settling time / long settling time with no oscillations	Spectral distribution of squared error	Low settling time and ISE
Hong *et al.* 1992 [114]	Adaptive PID retuning	sp	Five zones on the K Ti plane	Overshoot, decay ratio, settling time, and time delay	Minimum settling time
Seif 1992 [115]	Adaptive PI retuning	sp	Combinations of the 8 elementary patterns	Qualitative shape analysis of the process variable's step response	Not specified
Chia 2003 [67] (Controlsoft INTUNE)	Adaptive PID retuning	sp	Slow / medium / fast	Decay ratio, overshoot, settling time, oscillation symmetry	User-specified response speed
Salsbury 1999 [116]	Assessment of PI controller	sp/ld	Sluggish / no fault / oscillatory	Reversal index and settling time	Acceptable performance
Hgglund 1999 [82]	Sluggish loop diagnosis	ld	Sluggish / not sluggish	Idle index	Fast response with no overshoot

Reference	Type of Technique	Information Require-ment	Categories	Categorization based upon	Goal
Goradia 2005 [90]	Assessment and adaptive retuning of PI controllers	stoch	Sluggish / Optimal / Aggressive	Pattern matching of the estimated closed loop impulse response	Minimum mean squared error
Visioli 2005 [77]	Assessment of PI tuning	ld	For each of K and T: high / near optimal / low	Area, idle, output indices	Minimum IAE
Salsbury 2005 [117]	Assessment of PI tuning	ld	Sluggish / acceptable / aggressive	R-index	Critical damping
Howard and Cooper 2009, 2010 [118, 119, 120]	Assessment and adaptive retuning of PI controllers	stoch	Sluggish / well-tuned / aggressive	Relative damping index	User-defined damping factor
This paper (Section 4)	Assessment of PI tuning	stoch	Sluggish / aggressive	Hurst exponent	Minimum IAE

pattern recognition based approaches due to trouble with identification based adaptive controllers, which would sometimes cycle continuously due to poor model-plant agreement. This type of pattern recognition adaptation was also the basis for the aforementioned EXACT controller [111]. In Section 2.7, we present results that suggest that instability is also a concern for data-based retuning based on pattern classification.

Also notable within Tables 2.1, 2.2, 2.3 are several contributions [91, 86, 107, 108, 117, 116] from researchers in the field of HVAC (heating ventilation air-conditioning) controls. Like the process industries, this field is also cost sensitive, with insufficient manpower devoted to controller testing and tuning, and it relies heavily upon PI regulators [91]. HVAC systems are also nonlinear and subject to time varying system, loop interaction, and disturbance characteristics [91], complications which suggest adaptive tuning as necessary to maintain good control. The relevance of performance assessment and retuning to the field was indicated by a large survey of research into buildings controls systems issues, examining 40 studies which incorporated 450 control systems problems reported across more than 70 buildings, which suggested that software problems were responsible for about 1/3 of reported control system issues, and that regulator tuning was a significant contributor to this category [123, 124]. Moreover, the authors of the prior references indicated their suspicion that many system problems reported across other categories had originally been due to poor controller tuning, such as in the case

of actuator linkages being continuously cycled until mechanical failure. Within the presented diagnostic and retuning algorithms, the contributions from the HVAC field have carefully considered the handling of noise, the proper detection of transience, and the identification of problems beyond the controller's capability to correct [86, 91, 116, 117].

2.3 Review of Diagnosis Measures for Controller Tuning Assessment

This section reviews diagnosis measures available in the literature for assessing controller tuning. The usefulness of such techniques is by no means restricted to PI controllers; however, the remainder of this work concentrates upon this simple control structure both because it provides for tractable analysis and because it holds a dominant place in industrial applications [77, 78]. The techniques considered were developed for linear processes with relatively simple disturbances. Nonlinearities may affect interpretations of these measures by inflicting additional hysteresis in the response curves, necessitating the use of additional monitoring tools to detect the presence of sources of these confounding effects.

2.3.1 Controller Tuning Diagnosis Measures Which Require Step Changes in Load Disturbance

The methods which fall into this category give best results when the required controller output and/or process variable data satisfies the following conditions:

1. the data contains the system's response to a large step change in load disturbance so that the data represents a closed-loop step response;

2. when containing multiple step changes in load disturbance, the response to each are separable in time;

3. generally free of other disturbance types and noise.

To identify when conditions (1) and (2) are met from the recorded process measurement and controller output signals, an abrupt load detection procedure such as in [125] or [126] will be required. Suitable filtering is specific to each technique and discussed in the following subsections.

2.3.1.1 Idle Index

Hgglund [82] noted that often during a load disturbance response, sluggish tuning will yield positive correlation between the signs of the slopes in the controller output (OP) and process variable (PV) signals, and he proposed the idle index to detect this situation. Upon an isolated step response to a changing load disturbance,

define t_{pos} as the total time for which the increments of the process variable and controller output have the same sign and t_{neg} as the total time for which the increments have opposite signs. Then the Idle index (II) is defined as $= \frac{(t_{pos} - t_{neg})}{(t_{pos} + t_{neg})}$.

Extensive prefiltering of the data may be necessary in many practical situations. The range -0.4 < II < 0.4 was originally proposed [82] to indicate well-tuned loops, with values above 0.4 indicating sluggish tuning, and values below -0.4 providing no diagnosis as these could result from either well-tuned or aggressive systems. Later, Visioli [77] , among other requirements, specified that the idle index had to take on a value of less than -0.6 in order for the loop to be diagnosed as well-tuned. *Advantages*: This index was introduced solely for detection of sluggish systems, which makes it an attractive partner for one or more oscillation detection and diagnosis techniques. *Notes*: As a measure computed upon signal derivatives, this index is sensitive to the effects of noise, and its calculation requires isolated and denoised features from the closed loop load disturbance response in OP and PV to be provided. This can require a multistep procedure of load change detection, denoising, and exclusion of steady state data. These and various other implementation aspects are discussed in [127, 128].

2.3.1.2 Area Index

Visioli [77] found that aggressiveness or sluggishness of the loop's load disturbance response is characterized by the manner in which oscillation decays within

the OP (controller output) signal. For an isolated response to a step change in load disturbance, define A_1 as the area between the first two crossings of the final steady-state u_{ss} of the response of the OP and A_{tot} as the total area in the response of the OP after the first crossing of the final steady state value. Then, the index is calculated by $AI = \dfrac{A_1}{A_{tot}} = A_1 \left(\sum_{i=1}^{\infty} A_i \right)^{-1}$.

Figure 2.2a illustrates the definition of these quantities. If the response of the controller output signal does not present any crossing of its final steady state value, the index is simply set to unity. Calculation of the index requires isolated features from the closed loop load disturbance response within the OP signal to be provided, and additionally, it requires the load change response to be sufficiently abrupt in nature. This will require load change detection and abruptness quantification. Separating the effects of the load change from the effects of noise within the OP data also requires filtering or noise thresholding [77]. The range $0.3 < \text{AI} < 0.7$, among other requirements, was originally proposed [77] to indicate well-tuned loops, with lower values indicating aggressive tuning and higher values indicating sluggishness. Later, Jelali [129] proposed to extend the method to stochastic systems by applying the index to the estimated impulse response of the process output.

Advantages: As a measure calculated upon areas, it is relatively robust to noise.

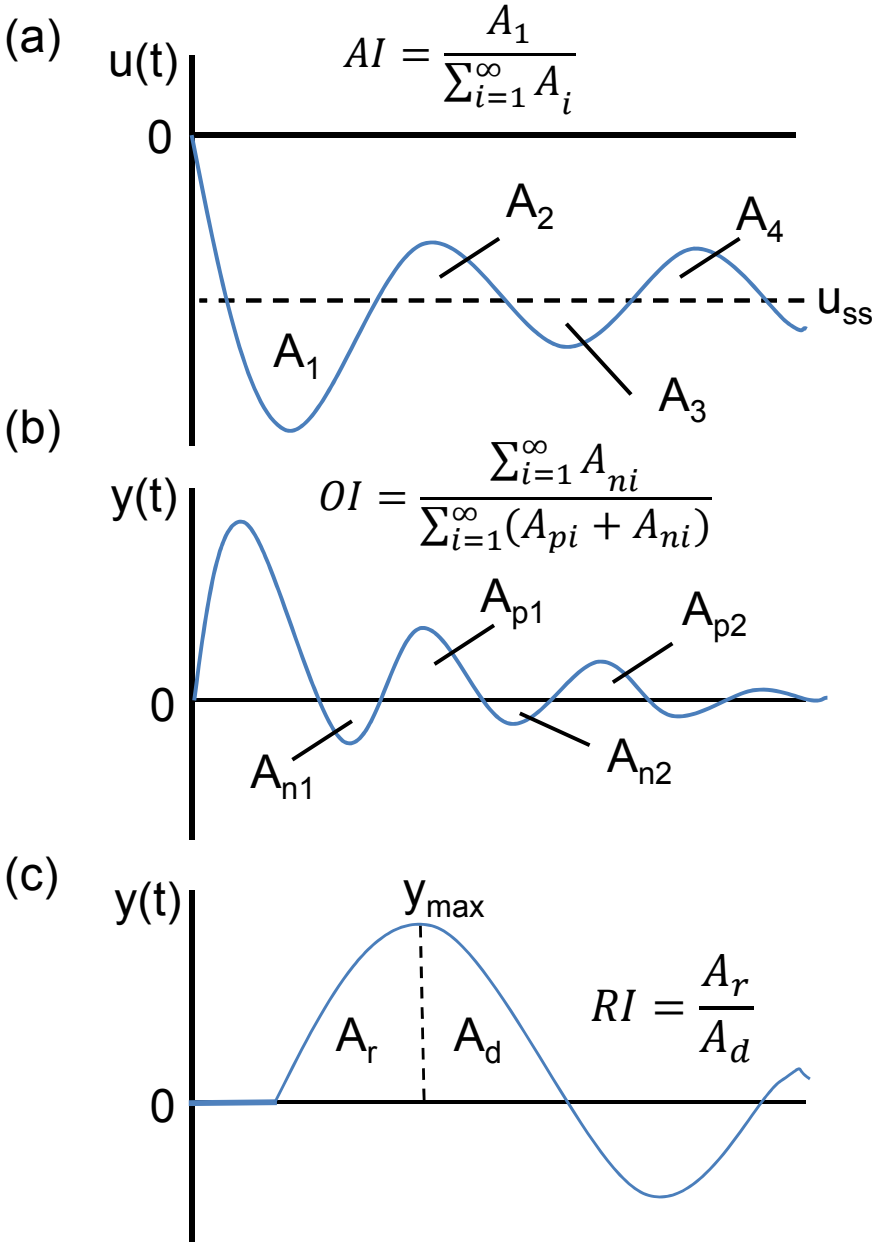

Figure 2.2: Calculation of the (a) the area index, (b) the oscillation index, and (c) the R-index.

2.3.1.3 Output Index

Visioli [77] also reported upon how overly high reset time could be diagnosed in oscillating loops by asymmetry of the oscillations about zero. This asymmetry occurs because there exists a dominant pole on the real axis within the closed loop transfer function. For an isolated response within the process variable (PV) data to a step change in load disturbance, define An as the integral of negative areas with respect to the final steady state value, and A_{tot} as the sum of all areas after the first zero crossing of the response. Then the Output Index is simply $OI = A_n/A_{tot}$ where $A_n = \sum_{i=1}^{\infty} A_{ni}$ and $A_{tot} = \sum_{i=1}^{\infty} A_{pi} + \sum_{i=1}^{\infty} A_{ni}$ with these quantities illustrated in Figure 2.2b. The index is only calculated in the case of an aggressive response being detected by other methods, because it requires the presence of a decaying oscillatory trend. Calculation of the index requires isolated features from the closed loop load disturbance response in PV to be provided. The beginning and ending of the response within the data need to be located with a load detection procedure and steady state detection, respectively. For loops already detected as being aggressive, Visioli [77] proposed that a value of OI < 0.35 indicates that the reset time is too high relative to its optimal setting. *Advantages*: As a measure calculated upon areas, it is relatively robust to noise. *Note*: This measure has only been proposed in conjunction with other diagnosis measures and possesses little utility on its own.

2.3.1.4 R-index

According to Salsbury [117], aggressiveness or sluggishness of the loop's load disturbance response can be characterized by shape of the initial response peak in the process output signal. For an isolated response within the PV to a step change in load disturbance magnitude, define A_r as the area between start of the disturbance response and its peak value, and define A_d as the area of between this peak value and the first zero crossing of the response. These quantities are illustrated in Figure 2.2c. Then, the R-index is calculated as $RI = \frac{A_r}{A_d}$. Calculation of the index requires isolated features from the closed loop load disturbance response in PV to be provided. Diagnosis criteria were not offered in the original work, but in the several examples presented, the values 0.7 and 0.85 were obtained from loops considered sluggish, values of 0.48 and 0.6 were obtained from systems considered well-tuned, and a value of 0.36 was obtained from data for an aggressively tuned system. *Advantages*: As with other measures calculated using areas, it is relatively robust to noise.

2.3.1.5 PI Controller Assessment using Multiple Diagnosis Measures

By combining several of the above diagnosis measures into one tuning assessment technique, Visioli [77] was able to give a more specific retuning recommenda-

tion for PI controllers than a simple sluggish or aggressive diagnosis could provide. Table 2.4 displays the detection criteria used for that purpose. This technique attempts to solve a very classical pattern recognition/classification problem: allow identification of systems from several categories by using the measurement of several features from each system(in this case, these are features of the closed loop response) [130]. *Advantages*: The combination of several features provides a more thorough diagnosis of incorrect controller tuning then the simple indication of sluggish/aggressive.

2.3.2 Controller Tuning Diagnosis Measures that can Handle Stochastic Systems

For the proper application of the techniques of Section 2.3.1 for tuning assessment, requires either: (i) the existence of a set of very specific conditions (i.e. a step change in load disturbance, the response to which can be isolated from other any other inputs to the closed-loop system), and/or (ii) copious data pre-processing in order to remove noise and steady state data. Many loops in the process industries may not regularly satisfy (i), and the use of the filtering procedures of (ii) may require tuning of noise thresholds and other parameters that can vary loop to loop and preclude ease of large scale automation. This subsection introduces techniques which are more robust to noise and whose performance may in fact increase with

noise exciting the closed loop response. The first technique makes use of a closed loop model fit upon the process variable (PV), while the second relies upon the autocorrelation function (ACF) of process variable.

2.3.2.1 Impulse Response Curve Method

Goradia *et al.* [90] suggested making sluggish and aggressive tuning diagnoses based upon pattern matching the system's estimated closed loop impulse response under disturbance to a set of 9 characteristic impulse response curves. The system's impulse response is estimated with an AR(20) model fit to the available process output data, from which a finite impulse response (FIR) model is obtained. Since the method relies on the estimated impulse response, data from routine operation (excluding actuator saturation, manual control efforts, or set-point changes) should be acceptable. It was originally proposed that the FIR coefficients would be visually matched to the characteristic response curves [90], though the use of neural networks for this purpose has also been demonstrated [129]. Characteristics of sluggish impulse response curves highlighted by the original authors included presence of offset and lack of zero crossings. For characterizing aggressive tuning, important quantities were the magnitude of overshoot and the number of discernible cycles of oscillation due to complex conjugate poles. *Advantages*: The use of the estimated impulse-response instead of raw OP or PV signals may allow for many of the diagnosis measures reviewed in Section 2.3.1 to have their use extended to purely

stochastic systems. *Note*: The method as first presented requires a human to perform the pattern matching. One can imagine many ways to automate this method, and it is an open problem to determine which choice of features of the impulse response can be used to achieve the most successful classification.

2.3.2.2 Relative Damping Index

It has been known for many years that the autocorrelation function is a valuable tool for control loop performance assessment [3, 79, 88, 127, 18]. Howard and Cooper [119] used the fact that if the closed-loop system can be properly described by linear model, then the autocorrelation function of the process variable data should theoretically have the same poles as the applicable linear model. The authors demonstrated this relation in a practical situation using boiler level data. A set of load disturbance response data was shown to be similar in shape to the ACF of data from a separate time period containing lower magnitude stochastic disturbances. Based on this useful property of the ACF, the authors further demonstrated sluggish/aggressive diagnoses based upon the damping coefficient, ζ, of a second order model fit to the estimated autocorrelation function of the system's closed loop output. The Relative Damping Index (RDI) is a rescaling of the damping coefficient with user defined bounds on aggressive and sluggish values, denoted by ζ_{agg} and ζ_{slug}, respectively. After the estimated ACF is obtained, the second order model is iteratively fit, where the number of ACF coefficients regressed upon varies between

successive iterations. We found that it was preferential to fit a discrete time model to the ACF first, which was then converted to the continuous time equivalent in order to obtain the damping coefficient. The discrete time model can better capture the behavior of the ACF since both can display a magnitude of one at lag zero, which is untrue of a continuous time model. An alternative solution was presented in [78], wherein a transformation of the ACF curve was used to make it take the appearance of an open loop step response, upon which the second order continuous time model could be properly fit. The damping coefficient was used to define the relative damping index, which is given by

$$RDI = \frac{\zeta - \zeta_{agg}}{\zeta_{slug} - \zeta} \tag{2.1}$$

Since the method index is calculated upon the estimated ACF, any PV data from routine operation (excluding actuator saturation, manual control efforts, or set-point changes) should be acceptable. The method has been demonstrated upon data containing both large load disturbance responses and stochastic noise. For self-regulating plants, the recommended criteria were that $\zeta < 0.6$ indicates aggressive tuning ($\zeta_{agg} = 0.6$) and $\zeta > 0.8$ indicates sluggishness ($\zeta_{slug} = 0.8$). Similar to the characteristic closed-loop impulse response patterns provided by Goradia *et al.* [90], the authors a set of autocorrelation curves were provided for visual determination of poor tuning. Advantages: By using the ACF of the process variable,

this technique can be applied to systems afflicted by mixtures of deterministic and stochastic disturbances. *Notes*: A suggested range of acceptable damping ratios is provided, however, it is unclear what condition the well-tuned specification indicates, whether it corresponds to minimization of a performance metric, acceptable bounds on some closed-loop response feature, or something else. Also, whether the diagnosis measure in this case (RDI) behaves as expected depends on how well a second order model can fit the ACF.

2.3.3 Remarks on Tuning Assessment Techniques

To assess the state of loop tuning, each of the above techniques identified some characteristic of a PV and/or OP signal, where these signals were from (i) an isolated load disturbance response, (ii) an estimated closed-loop impulse response, or (iii) an estimated autocorrelation function. For diagnosis measures defined upon one type of signal, it may be possible to extend their use to the other signal types (with a proper adjustment of detection criteria). One example [129] has already appeared in which the area index, originally designed for use on an isolated load disturbance response in the OP signal, was extended for use on stochastic data by calculating it for the estimated closed loop impulse of the PV signal instead. In Section 2.6, the performance of several of the tuning diagnosis measures reviewed here is examined. The assessment method relying upon pattern matching of the estimated closed-loop impulse response [90] is not considered further because of the

variety of different ways this technique could be automated.

2.3.4 Applications of Tuning Diagnosis Techniques

Kuehl and Horch [128] demonstrated the use of the idle index upon industrial flow loop and pressure data. It has also been reported that the idle index has been implemented in both ABB Industrial IT system [127] as well as a performance monitoring system at Eastman Chemical Company for 14,000 PID loops in 40 plants at nine sites worldwide [59]. Salsbury [117] gave results from several examples of using the R-index applied to monitor HVAC control systems. In some prior works, [90, 129] using the tuning assessment methods for iterative loop retuning towards an optimal value was proposed. Thus far, no documentation on the application of these techniques for automated retuning in an experimental or industrial context has appeared.

2.4 The Hurst Exponent Method for Loop Tuning Assessment

The previously introduced diagnosis measures rely upon the type of features present in the OP and PV signals, with the methods based upon responses to isolated step changes in load magnitude more suited to this type of disturbance and those relying upon the estimated impulse response or autocorrelation functions more suited to processes continuously excited by noise. Here we introduce a technique that always begins by processing the raw PV (process variable) data. To characterize

performance of control loops under PI control, we propose the use of the Hurst exponent, calculated by means of detrended fluctuation analysis (DFA) [3], to detect and quantify loop sluggishness or aggressiveness. In our previous work [18], a specific scaling of this quantity, referred to as the Hurst Index, was used to characterize overall loop performance for a wider class of systems. Therein, it was demonstrated that the Hurst Index could be used to estimate performance without any process knowledge, providing an index value that replicated trends in the widely used minimum variance index. Now, it is proposed to apply the unscaled Hurst exponent to diagnose specific types of poor performance in PI controlled loops, namely sluggish or aggressive loop tuning.

In the context of analyzing process signals, we rely on the Hurst exponent's ability to quantify the persistence of correlations or anti-correlations (such as present in an oscillatory signal) existing in a signal. More generally, the Hurst exponent is a measure applied to a time series to determine the self-similarity of the signal. A time series y_t is self-similar (contains sub-units which resemble the whole structure) if it holds that $y_t(k) \equiv a^{\alpha} y(k/a)$, where a represents the scaling factor along the x-axis (time axis), a^{α} the scaling factor along the y-axis, and exponent α is the self similarity parameter. Letting the scaling factors along the x and y axes be M_x and M_y, respectively, then the exponent is given by $\alpha = \dfrac{lnM_y}{lnM_x}$, or statistically, $\alpha = ln\sigma_y / ln\sigma_x$. The scaling exponent α is referred to as the (generalized) Hurst exponent, which can be obtained for a stationary signal mapping the time series

to a self similar process through integration. The method used here to calculate the Hurst exponent is detrended fluctuation analysis [131], which is regarded to be more reliable than several alternative methods [132]. A multitude of applications of the DFA procedure have been found in finance, earth science, and medicine (a large list of applications is collected in the introductions of [133, 134]. For example, the Hurst exponent calculated by DFA exhibits different values between the heart-beat signals of healthy individuals and sick patients [135].

The DFA procedure is presented in [48, 136] and can be performed on noise afflicted process loops without modification since the Hurst exponent calculation requires only routine operating data. Data containing set-point responses should be excluded from the computation though as these cause additional nonstationarity in the PV which will be misinterpreted if the exponent is computed without modification. Figure 2.3 summarizes the Hurst exponent calculation through DFA. Briefly, the PV data, of Panel (a) in Figure 2.3 is first numerically integrated resulting in a nonstationary signal as in Panel (b). This signal is then piecewise detrended by lines regressed over a data window of length L as seen in Panel (c). The root-mean-squared (rms) value of the resultant detrended data is termed the fluctuation, F(L), which is a function of the window length over which linear trends were regressed. The fluctuation calculation is repeated over a range of window lengths L. Finally, the Hurst exponent is calculated by fitting a linear trend between the logarithms of window size and fluctuation magnitude as shown in Panel (d) of Figure 2.3.

Mathematically, N samples of stationary PV signal y_t, which varies around its setpoint, are mapped to the detrended and integrated signal $Y_t = \sum_{k=1}^{N}(y_t - \bar{y})$, where $\bar{y}$ is the mean value. The integrated series is then segmented into W windows of length L. On each window $j = 1 \ldots W$, a linear trend is regressed according in order to the objective function

$$min_{a_j,b_j}\epsilon_j^2(L) = min_{a_j,b_j} \sum_{i=(j-1)L+1}^{jL} (Y_i - a_j i - b_j)^2 \tag{2.2}$$

and then this trend is subtracted. After this procedure is repeated for every window of length L, the detrended fluctuation on window length L can be calculated as $F(L) = \sqrt{\frac{1}{WL}\sum_{j=1}^{W}\epsilon_j^2(L)}$. The final scaling exponent reflects the growth in detrended fluctuations as the window size increases. This can be repeating the calculation over a range of L values, and obtaining the slope from a linear fit between L and $F(L)$. To avoid spurious artifacts occurring when the ratio of the window size and sample size becomes too large, it is recommended to keep $L \leq \frac{N}{4}$ [136].

For the purposes of sluggish and aggressive tuning diagnosis in this work, the window length parameter L of the DFA technique was varied from 100 to 400 samples at increments of 10. Based on the guideline of maximum window size of $\frac{N}{4}$, this translates to a total requirement of 1600 or more samples in a single record (with no set-point changes or other irregularities contained in the record). Our Matlab implementation of the Hurst exponent calculation takes less than 1 second to

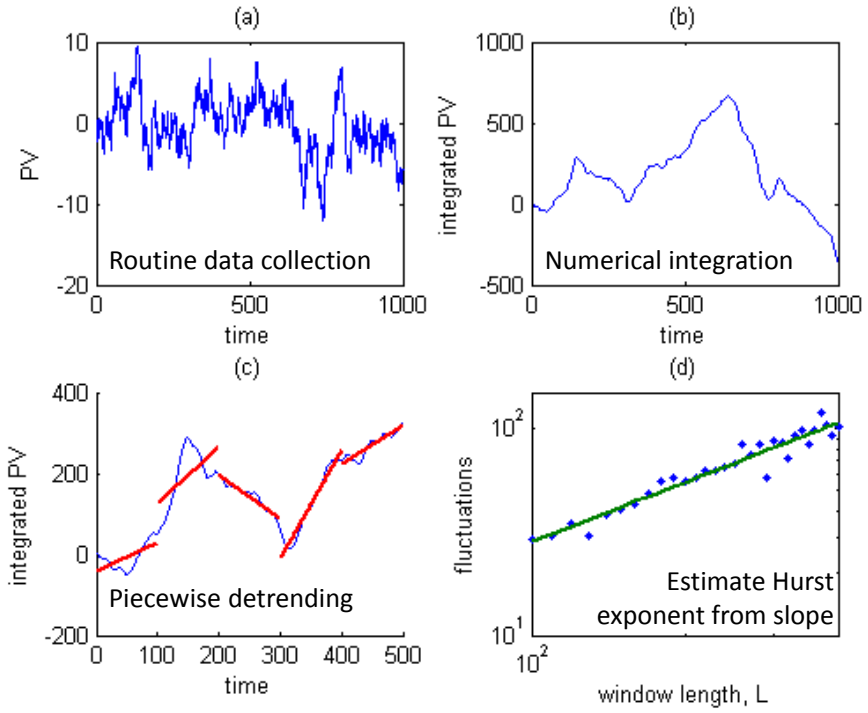

Figure 2.3: Illustration of the steps of Hurst exponent calculation.

run on a personal computer. The proposed use of the Hurst exponent as a tuning diagnosis measure is to combine it with other loop monitoring tools. Specifically, for an isolated control loop without interactions, these other diagnosis tools might include a nonlinearity or stiction detection algorithm to monitor equipment problems and a performance monitor such as the minimum variance index or Hurst index to indicate overall loop performance. If bad performance is indicated, and nonlinearity is not suspected to be the cause, the Hurst exponent can be used to diagnose a possible problem due to aggressive or sluggish (although the possibility of disturbances beyond the capability of the loop to correct would still remain in that scenario and false diagnoses would be possible). Table 2.5 shows the theoret-

ical Hurst exponent values of several basic signal types and the information that similar behavior within the PV signal of a control loop can yield about controller tuning. A signal appearing indistinguishable from white noise with no oscillations or extended deviations from zero is generally preferable to one that contains those features. Long drifts of the mean PV value indicate sluggish tuning and typically lead to high Hurst exponent values, while oscillations such as due to aggressive tuning will lead to lower Hurst exponent values. Over a broad range of loop types, the value of 0.5, indicative of white noise behavior is a useful standard to compare individual loops Hurst exponent values against. Higher values may indicate sluggish behavior, while lower values will indicate oscillatory behavior (which may be due to aggressive tuning). Within the Section 2.6 that follows, we make a comparison among the several diagnosis measures discussed so far.

2.5 Framework for the Evaluation of Controller Tuning

To proceed with the discussion of data-based controller tuning diagnosis methods, a framework from within which to understand the different existing diagnosis techniques is first presented. The methods for diagnosing poor controller tuning in the literature generally contain at least 2 components, these being: (i) a control loop performance metric, the optimization of which is the goal of controller retuning, and (ii) a data-based diagnosis measure to indicate in which direction to move the controller parameters in order to obtain a better value of the performance metric

Following these, a retuning technique to implement the suggestions of the diagnosis method is sometimes also considered. Items (i) and (ii) are discussed in subsections 2.5.1 and 2.5.2, respectively.

2.5.1 Control Loop Performance Metrics

In (i) above, the performance metric is often chosen to be either the integral of absolute error (IAE), integral of squared error (ISE), or the integral of time-weighted absolute error (ITAE), as reflected in Tables 2.1, 2.2, 2.3 of Section 2.2. Tuning to minimize ISE punishes large deviations from the set-point, but also produces aggressive action [138]. Shinskey [139] has suggested that the integral of absolute error has the closest relationship to economic considerations. Tuning to produce minimal ITAE gives the most conservative set of controller parameters, corresponding to the slowest response [138]. In the CLPM&D field, the metric chosen is often Harris's minimum variance index (MVI) [6], which is defined as

$$MVI = \frac{\sigma_y^2}{\sigma_{MV}^2} \tag{2.3}$$

where σ_y^2 is the variance of the process output measurement and σ_{MV}^2 is the theoretical minimum variance achievable for the given process with any linear controller. This index has a firm relation with ISE, one being the integrated square error, and σ_y^2 being the expected value of the square error when the process fluctuates

about its set-point. However, the open-loop plant time delay needed to compute the minimum variance index is likely to be unknown information and is not essential for the diagnosis and retuning algorithms. Therefore in this work, ISE is considered itself with the knowledge that the MVI could be recovered with the appropriate information, and that optimizing one metric should optimize the other too.

In the following, simulation results generated with Matlab Simulink are used to study the definitions of sluggish and aggressive tuning. Considering now only the case of systems under PI control, we introduce the symbols K and T to represent, respectively, controller gain and reset time, so that for continuous time, the PI controller Q has the form $Q(s) = K(s + \frac{1}{T})s^{-1}$. For illustration, a representative system with open-loop plant model $\frac{e^{-5s}}{10s+1}$ is chosen, and the closed-loop is subjected to a unit magnitude step disturbance at the plant input under various sets of PI controller parameters (K, T). Simulations are run for 2000 seconds at each parameter set. The following results seek to illustrate qualitative behavior of systems under PI control, and it is noted that the exact parameter values where certain features occur will vary from system to system. Figure 2.4a shows a plot of the K, T -plane with contours of IAE for the chosen system, along with the set of controller parameters providing global optimal IAE. In Figure 2.4b an additional boundary is defined to separate systems responding to the stepwise change in disturbance with overshoot in the PV signal from systems responding to the disturbance without overshoot. Lack of overshoot in the step response is considered necessary by some authors in order for the

control loop to be considered well-tuned [82], though opinions on this matter vary. In this case, reaching a constrained global optimum becomes the goal for retuning. In Figure 2.4b, it can be noted that for the chosen system, the constraint causes a minimal effect on the location and value of the optimum.

While most previous studies have pursued diagnoses of sluggish and aggressive tuning, Visioli proposed a method of detecting where the current parameters lie relative to the global optimum, as discussed in Section 2.3.1.5 and Reference [77]. The combination of diagnosis measures applied was used in an attempt to segment the K, T -plane into the categories shown in Figure 2.5. A small area of the plane surrounding the global optimum point is considered to be the region of acceptable tuning. The numerous classifications of Figure 2.5 make unnecessary the additional categorization of the loop behavior as sluggish or aggressive. The efficacy of this method is reviewed along with other proposed methods within Section 2.6.

2.5.2 Proposed Definitions for Sluggish and Aggressive

Now we consider the bulk of the literature on data-based tuning diagnosis, which concerns itself with the detection of sluggish and aggressive controller tuning. Within the literature, the definitions of what constitutes sluggish and aggressive tuning are almost as numerous as the techniques to detect these conditions. One way sluggish and aggressive tuning could be defined for PI controllers is by the corrective action suggested when each of these conditions are diagnosed. For

sluggish tuning, this corrective action is to increase the proportional gain and/or decrease the reset time, while for the case of aggressive tuning, it is suggested to decrease proportional gain and/or increase the reset time. Goradia *et al.* [90] stated that the performance metric, which was in their case the minimum variance index, also referred to as the CLPI (closed-loop performance index), behaved as follows: "*the CLPI, not surprisingly, takes a unimodal locus. Once the proper direction to improve the controller performance is determined, (i.e., to make the controller aggressive or detuned) one can proceed iteratively in that direction as long as CLPI continues increasing. After reaching the peak, CLPI starts to decrease even though we are moving in the same direction. The peak value of CLPI is the PI achievable performance.*" Figures 2.6a and 2.6b below show how the sluggish and aggressive regions are defined according to this idea, for the case where retuning of each tuning parameter is considered separately. As shown in each (a) and (b), adjusting a chosen controller parameter to move closer to its corresponding optimal curve will improve the value of the performance metric.

However, as shown in Figure 2.7a, when both parameters are considered together, there exist regions where the sluggish diagnoses for one parameter overlap with the aggressive diagnoses of the other parameter and vice versa. This behavior occurs because the curves $T_{opt}(K)$ and $K_{opt}(T)$ coincide at only a single point, this being the global optimum. As an example of the resulting contradictory behavior, at a given parameter set in Region IV, the IAE can be decreased either by decreasing

K or by decreasing T. So, if we follow the statements of Goradia *et al.*, parameter sets in this region would be considered aggressive when tuning gain and sluggish if we were tuning reset time. Obviously this definition of sluggish and aggressive is troublesome because our next step is to find a data-based diagnosis method to generally indicate sluggish or aggressive tuning, so we need well-defined regions to judge our measure's performance against. To resolve this, it is seen that even when leaving Regions III and IV in Figure 2.7a undefined, we can still classify most of the relevant portion of the K, T -plane into the categories of sluggish and aggressive as shown in Figure 2.7b. Here sluggish tuning corresponds to a condition where the performance metric could be improved by increasing gain and/or reducing reset time, while aggressive tuning is defined as a parameter set within the region where the performance metric can be improved through reduction of gain and/or increasing the reset time. At this point, the usefulness of a given data-based diagnosis measure can be judged according to whether it takes upon contrasting values between the sluggish and aggressive regions.

The definition of sluggish and aggressive against the one parameter optimal curves $T_{opt}(K)$ and $K_{opt}(T)$ captures a much larger portion of the K, T -plane compared to if we defined these two tuning classifications by judging the controller parameters relative to their global optimal values. That the classifications sluggish and aggressive would occupy a much reduced area in that case is demonstrated in Figure 2.8, which shows their definitions upon the stable region of parameter space

for a single system.

2.5.2.1 Definition of Sluggish and Aggressive with Consideration of Overshoot

Finally, let us revisit the case of a constraint placed upon overshoot within the load disturbance response. In Figure 2.4b a boundary line was plotted separating systems with and without overshoot in their disturbance responses. We could define sluggish tuning as parameter sets having lower gain and higher reset time than points upon this boundary and aggressive tuning as parameter sets having higher gain and lower reset time than points lying on the boundary line. Adding these definitions upon the optimal curves of Figure 2.7b, we can note from the resulting Figure 2.9 that:

1. Controller parameters in the sluggish region provide no overshoot in their disturbance step response but have worse performance than the constrained global optimum.

2. Controller parameters in the aggressive region contain overshoot and may or may not have a better value of the objective function than the constrained global optimum.

3. For the region of the K, T -plane shown, the constrained versions of the functions $T_{opt}(K)$ and $K_{opt}(T)$ have converged to a single curve (the overshoot

constraint line which separates sluggish and aggressive systems). The purpose of retuning sluggish systems is to approach the boundary line in order to increase performance. The purpose of retuning aggressive systems towards the boundary is to meet the constraint on overshoot.

2.5.2.2 Sluggish and Aggressive Definitions in the Literature

We now consider sluggish and aggressive definitions that have been used with data-based diagnosis measures in the literature. Jelali [129] largely followed the definitions of Goradia *et al.* [90] that were considered above in Figure 2.7. Hgglund [82] mentions that well-tuned loops have a fast response with no overshoot. This idea fits into the framework shown in Figure 2.9, which includes the presence of constraints on overshoot. Salsbury [117] proposed critical damping of the closed loop system as the optimal condition to aim for. Most closed loop systems will have too many poles for this goal to be achieved with the PI controller structure. However, the use of critical damping criteria does indicate that overshoot in the disturbance response is not considered acceptable to the author, and again we can view this detection method in the framework of Figure 2.9. Howard and Cooper [119] have defined sluggish, aggressive, and acceptable tuning based upon the value of their own diagnosis measure with user specified thresholds. Finally, as mentioned previously, Visioli [77] largely side-stepped the issue of sluggish and aggressive categorization by the use of a more thorough classification of controller tuning like

that shown in Figure 2.5.

2.5.2.3 Use of the Proposed Aggressive and Sluggish Classification framework

In the following sections, we consider the case of classifying tuning versus the optimal IAE of the process variable's load disturbance response (with no constraint on the amount of overshoot). This corresponds to the definitions of sluggish and aggressive of Figure 2.7b, discussed in Section 2.5.2.1.

2.6 Performance of Tuning Diagnosis Methods

In each of the works in which data-based controller tuning diagnosis methods were introduced or discussed, results for only several examples (at most, a few sets of tuning parameters for any single system) were shown. However, we find that the performance of the data-based diagnosis methods can vary considerably across different regions of the controller parameter space (the K, T -plane). By looking at several systems across a large portion of parameter sets likely to be encountered in practice, we hope to provide new results on these diagnosis methods performance.

2.6.1 Evaluation of Several Methods for Sluggish and Aggressive Detection

Four diagnosis measures from the literature for detecting sluggish or aggressive tuning, as well as the Hurst exponent based diagnosis technique introduced in Section 2.4, are evaluated on an example set of systems. The criteria for sluggish and aggressive classification by each method are provided in Table 2.6. These were chosen by a combination of guidelines from the literature and our experience using these methods on a set of training systems.

In order to obtain a training set of controller tuning, for each of the five continuous time systems in Table 2.7, the classifications of Figure 7b are produced for the case of a unit step change to the input of the corresponding disturbance filter. The systems in Table 2.7 were obtained from several different references on PI controller tuning or performance assessment [77, 98, 119, 140]. Notably, this set of systems contains both a variety of time delays and number of lags. This corresponds to a large range of values of the normalized dead-time parameter [141], η=(dead time)/(time constant), also referred to as the process's degree of difficulty [107]. The loop performance goal is taken to be minimum IAE of the disturbance step response, and the curves $T_{opt}(K)$ and $K_{opt}(T)$ were constructed accordingly. In this way, the classifications of the tuning parameter sets (K, T) for each of the five systems in Table 2.7 were created according to Figure 2.7b. For each of the five

Table 2.4: Visioli's proposed combination [77] of the area index, output index, and idle index to diagnose where current PI controller parameters are located relative to the set providing minimal IAE

	II < -0.6 (low)	-0.6 ≤ II ≤ 0 (med)	II > 0 (high)
AI < 0.35 (low)	K high*	T high	T high
0.35 ≤ AI ≤0.7 (med)	ok	K low, T low	K low, T high+
AI > 0.7 (high)	K low	K low, T low	K low, T high

* if OI <0.35, then T is also too high.
+ this combination is not provided for in Visioli's original paper and was given by [129]

Table 2.5: Theoretical Hurst exponent values of several signal types and what their appearance in process output data suggests about the state of control loop tuning

Signal type	Hurst exponent [136, 137]	Loop tuning interpretation
Pure sinusoidal oscillation	0	Aggressive controller is causing oscillations causing oscillations
White noise	0.5	Fast decay of loop response to disturbances indicates well-tuned loop
Brownian noise	1.5	Sluggish controller is failing to reject drifting disturbances

Table 2.6: Detection criteria

Diagnosis Method	Sluggish Condition	Aggressive Condition
Idle Index (II)	II > -0.6	II ≤ -0.6
R-Index (RI)	RI < 0.7	RI ≥ 0.7
Damping coefficient (ζ)	ζ ¿ 1	$\zeta \leq 1$
Hurst Exponent (HE)	HE > 0.5	HE ≤ 0.5
Area Index (AI)	AI > 0.35	AI ≤ 0.35

systems, 50 sets of tuning parameters were randomly chosen from within both the sluggish and aggressive classified portions of the (K, T) -plane. An additional 21 points were chosen from the regions classified as undefined in Figure 2.7b. At each set of tuning parameters, a closed loop simulation was run for a unit step change to the input of the disturbance filter. Then each of five diagnosis methods of Table 2.6 was applied to the resulting sampled process output and controller signals recorded once a second for 2000 samples. Simulations were conducted using Matlab Simulink.

The results for each diagnosis method are shown in Table 8, and when compared, it is seen that under the conditions studied the best choice for diagnosing aggressive and sluggish tuning versus the framework of Figure 2.7b is either the R-Index or the damping coefficient ζ , since they outperform the other methods in detecting both sluggish and aggressive systems. For each diagnosis measure, an approximately equal percentage of sluggish and aggressive systems are correctly detected (within a few percent). This indicates that changing the values of the detection criteria would have little positive effect on the overall performance. To see this fact, it is useful to consider each diagnosis measure acting as a simple classifier, with a threshold value separating sluggish and aggressive diagnoses. Changing the threshold can increase the correct diagnoses of one type of system (either sluggish or aggressive), but only at the expense of decreasing correct diagnoses of the other system type. Also, these results are for the case of a single isolated load distur-

bance. As discussed in Sections 2.3 & 2.4, the ability of each method to handle signals with other types of disturbances varies markedly. Notably, the Hurst exponent is expected to perform better in the case of stochastic disturbances.

It is notable that the training parameter sets were chosen from one of three categories (sluggish, aggressive, undefined), while the diagnosis methods only classified tuning into the two categories sluggish and aggressive; however, the reason for this are two-fold: (i) most of the tuning parameter sets in the undefined category are by no means well-tuned, and by putting these parameter sets into one of the categories "sluggish" or "aggressive", the retuning algorithms introduced in the next section can seek to achieve better tuning for these systems; (ii) Figure 2.10 Parts b-d show that there is no range of values for the considered diagnosis measures that can define a localized well-tuned region of the (K, T)-plane. That is because values of these diagnosis measures present open-contours, which contrast to the closed contours of the performance metric IAE (Part (a) of Figure 10). Therefore, we suggest, that the diagnosis measures whose plots are in Figure 2.10 (b-d), when used alone, are better suited for strictly making sluggish or aggressive diagnoses, since a category of well-tuned would inevitably contain systems far away from optimal performance when using these diagnosis measures.

For the Area Index, however, the use of a range of this diagnosis measure's values to define a well-tuned region on the K, T -plane makes more sense relative to the other methods. Figure 2.11 shows a plot of AI contours (solid lines) overlaid

upon the contours of IAE (dotted lines) for System 3. At the higher values of T in this plot, the contours have almost completely converged and the combined line separates systems having an AI value of unity from those having AI values of close to zero. At lower values of T, the contours of AI have opened up and now take on intermediate values. These intermediate values occur near the inner contour of IAE values which makes this diagnosis measure more attractive than the others to detect close to optimal tuning. In previous works, values of the AI between 0.35 and 0.7 were used to detect close to optimal tuning [77, 129]. Despite the advantageous behavior of the Area index at higher T values, from Figure 2.11, it is also seen that when moving from the point of optimal tuning towards the origin (low K and T values), the IAE increases in value, yet the contours of AI do not close. This means the index will give incorrect detections in this region if a well-tuned classification is defined. From the foregoing analysis, it is recommended to use the tuning diagnosis measures only to provide general diagnoses of sluggishness or aggressiveness, and not try to indicate well-tuned values, since these methods will often give this indication at far from optimal values of the performance objective. When used in this way, these methods will less satisfy the needs of continuous controller monitoring (because they will constantly indicate tuning is either sluggish or aggressive), but are instead more applicable for the iterative controller retuning methods introduced in Section 2.7.

2.6.2 Evaluation of Alternative Method for Tuning Diagnosis

As previously mentioned, Visioli [77] introduced a tuning diagnosis technique which combined three indices to qualitatively estimate the position of the controller's current parameter set versus that of the global optimal tuning set. Since the results of three separate indices are combined to make each diagnosis, as previously illustrated in Figure 2.5, many possible categories of diagnoses can result from the method. Figure 2.12 shows the results of this method applied to a grid of points on the K, T -plane. To aid in visualization of this method's performance, the suggested direction of retuning implied by the diagnosis at each point is depicted with an arrow. Diagonal arrows indicate that both tuning parameters have been diagnosed as away from their optimal values, while vertical and horizontal arrows indicate that the need for retuning only one of the parameters has been detected. The plots show large regions of the K, T -plane where the method works correctly for either one or both parameters. Nevertheless, a problem arises for both systems near a portion of the stability boundary (in the bottom center of each plot), where the method recommends decreasing T. The result occurs because even though the response is decaying oscillatory in this region, the idle index is in the medium range according to the criteria of Table 2.4. Depending on the magnitude of the step taken, following the diagnosis method's advice in this region will likely bring the control loop to an unstable setting, which of course is highly undesirable. The next section

will compare the performance of autotuning algorithms which utilize either: (1) one of the five diagnosis methods evaluated for sluggish and aggressive tuning detection in Section 2.6.1, or 2.6.2 the tuning classification method of Visioli.

2.7 Data-based retuning of PI controllers

Previous works in the field of CLPM&D have also introduced algorithms for automated retuning based upon model-free assessment of controller tuning [65, 90, 129]; the algorithms presented here contain additional logic to deal with cases where the diagnosis measures are incorrectly indicating the proper direction for adjustment of controller parameters. It is strongly noted that the tuning maps shown in the previous sections were for illustration only and are unavailable in practice unless the process and disturbance models are known, which is usually not the case (hence the use of data-based diagnosis measures). Therefore, in this section, we assume that the only information available is routine operating data which can be used to calculate the values of the tuning diagnosis measures. The simulation studies of this section are for the case of a single unit step in the magnitude of the disturbance variable, and it is assumed that the process output, controller output, and set-point data are available. The set-point data is solely used to exclude portions of the process variable (PV) and controller output (OP) recordings where set-point changes have an effect. The suitability of each detection method for other signal types (e.g. those driven by stochastic disturbances) was discussed in detail in Sections 2.3 &

2.4.

To retune controllers based upon the diagnosis methods discussed in Sections 3-6, two new algorithms were developed, whose purpose is to obtain better performance with regards to a metric such as IAE or ISE. One algorithm is for use with the user's choice of a single data-based diagnosis measure, and the second algorithm relies upon the diagnosis technique of Visioli [77] where the AI, II, and OI are used in complementary fashion. The first algorithm (an improvement of one proposed in [129]), shown in Figure 2.14, utilizes a user choice of one of the diagnosis measures appearing in Section 2.6. Considering the case where IAE is the performance metric, then for whichever diagnosis measure is chosen, the detection criteria of Table 2.6 are used to determine a diagnosis of either sluggish or aggressive tuning at the initial tuning parameter set (if a different performance metric is chosen then the associated detection criteria should be developed). Depending on the diagnosis at the original controller setting, a 20% change in the controller gain K is made. If the result is a decrease in IAE, a new diagnosis is made. If a switch in diagnosis occurs relative to the first one, the diagnosis measure is indicating that a minimum of IAE with respect to K was just crossed. This possible minimum is explored with smaller 10% steps in controller gain K. If the first change in gain led to an increase in IAE (a bad move occurred), the algorithm returns the system to the original parameter set.

The algorithm tunes one parameter K or T at a time, switching parameters af-

ter either an optimum point is detected or an IAE increase occured. The proposed method is supervised by the performance metric (in this case IAE), meaning that if this performance objective is worsened by moving in one direction, the algorithm will not continue in the same direction, regardless of the advice of the diagnosis measure. As previously suggested [90, 129], a basic step of 20% in the varied parameter was used in each iteration. Smaller steps of 10% were used to explore both the points providing minimum IAE in the varied parameter and also the boundaries between sluggish and aggressive diagnoses. To obtain faster convergence, larger step sizes should be considered in the case that the changes are carried out manually under human supervision.

The performance of the algorithm was explored with each of the diagnosis methods of Section 2.6 (Hurst exponent, idle index, area index, R-index, and damping coefficient ζ). Results were obtained for each choice of diagnosis method at each of the 250 starting sets of parameters. Two cases were attempted: (i) tuning both gain K and reset time T, and (ii) only retuning gain. The retuning was applied to noise free simulations of the system response to a unit step at the input of the disturbance filter. OP and PV signals were sampled at a rate of once per second during simulations of 2000 seconds. A set of 50 starting points for PI controller tuning on the K, T -plane were randomly generated for each of the 5 systems of Table 2.7. The starting points (K, T) were chosen to satisfy $IAE(K, T) < 4 \times IAE_{min}$, where IAE_{min} is the PI-achievable IAE for each system for the disturbance described, in order to

provide moderately (but not extremely) poor initial tuning. All simulations were performed using Matlab Simulink.

Table 2.9 contains one example of the use of the Hurst exponent along with the retuning algorithm of Figure 2.14 to search for better controller parameters. When a switch in the diagnosis occurs in Iteration 1, the algorithm neither continues in the original direction to check if a local minimum has been traversed, nor makes a full 20% step in the reverse direction as now indicated by the diagnosis measure. Instead the boundary between diagnosis classifications is explored with a 10% step. Even though a better value of the performance metric was reached in the first iteration, the retuning algorithm will not continue in the original direction to search for local minima when it is against the advice of the diagnosis measure, out of concern for stability. In this case the choice to reverse direction in fact provides us with the local minimum to within 10%. After reversing directions, the algorithm switches the tuning parameter to reset time T. In this case, when increasing the reset time according to the aggressive diagnosis, both the initial 20% move and a reduced 10% step in T cause an increase in IAE. Since no further improvement in IAE is occurring when making moves in either K or T according to the recommendations of the diagnosis measure, the algorithm terminates. The effects upon PV of the iterative retuning process detailed in Table 2.9 are illustrated in Figure 2.13. Although the results of the table indicate substantial improvement in performance, from the plot of PV it appears that the response is still oscillatory and more improvement should

be possible. This is a characteristic of the data-based retuning methods presented here, in that the significant improvement is usually achieved, but the algorithm still terminates far from the optimal controller parameter set due to the desires to limit the number of iterations and to make parameter changes only on the basis of support from the diagnosis measure and performance metric.

A second retuning algorithm is presented in Figure 2.15, wherein the determined direction of retuning relies upon the diagnosis criteria of Visioli [77]. During the initial step, either a 20% in a single controller parameter or 10% in each controller parameters is taken in accordance with the diagnosis criteria of Table 2.4. If this move resulted in an increase in IAE, the parameters were returned half way towards the original set. If at this smaller step, IAE was still increased compared to the original set, the search is terminated and the controller settings are returned to the optimal parameter set encountered thus far, against the advice of the diagnosis method (which is providing incorrect results). Thus this tuning method is supervised by IAE. The algorithm can also terminate due to a diagnosis of near optimal tuning according to the criteria of Table 2.6. One example application of algorithm of Figure 2.15 is presented in Table 2.10. It is noted that tuning according to the diagnosis criteria without considering changes in the performance metric, the parameters tend to cycle back and forth between differently classified regions leading to a large number of iterations and often times instability. Instead, as shown in Iteration 2 of Table 2.10, when the proposed algorithm with IAE supervision encounters

an indication that parameter changes need to be reversed, the controller is taken to the best parameter set encountered so far, and the algorithm is terminated. In this way, cycling between closely located sets of parameters does not occur.

Full results of all of the algorithms applied to the entire set of 250 retuning trials are displayed in Table 2.11. The results in the upper portion of the table refer to the algorithm of Figure 2.14 which was applied to each of the possible choices of diagnosis measures. One notable observation is that retuning of reset time in this first algorithm added marginal benefit; most of the improvement in performance was due to the initial retuning of gain alone. The algorithm terminated in a similar number of iterations for all five of the diagnosis measures tested, so no advantage of any technique is seen in this respect. Most notable from Table 2.11, it is seen that instability could be a major problem when retuning with several of these diagnosis measures. Specifically, the Hurst exponent, R-Index, and damping factor ζ all bring the systems to unstable parameter sets numerous times. While the Area and Idle indices did not have this problem in the trials studied, there is no guarantee that these methods will keep the loop at stable parameter sets for all systems. But due to the extreme problems encountered by the three other methods, it is recommended to only attempt this type of automated retuning with one of the more stable methods, either the AI or II. It is noted that the stability issue has not previously been discussed in the literature on these types of data-based controller diagnosis and retuning methods, and the issue needs more exploration. It is noted

the instability problem occurs when a set of parameters is diagnosed as sluggish (either correctly or incorrectly) and the resultant increase in gain or decrease in reset time moves the parameter set across the stability boundary. This implies that the problem could be mitigated by adjustment of the sluggish and aggressive detection criteria to shift the boundary of these classifications away from the stability line. However, even this solution could not guarantee to eliminate the problem, and it would be at the expense of algorithm performance.

Also, the lower section of Table 2.11 contains results from the second retuning algorithm of Figure 2.15, the IAE supervised algorithm which makes retuning decisions based on the criteria of Table 2.4. For contrast, the original version of this algorithm, proposed in [129], also for use along with the diagnosis method of [77], was implemented without IAE supervision. Lack of virtual supervision is ill-advised, since retuning can occur repeatedly in a given direction even though the performance metric to be minimized becomes worse at each step. As shown in Table 2.11, this leads to instability: 35 out of 250 trials were brought to unstable parameter sets for the algorithm without supervision, compared to only 2 trials in total for the IAE supervised method. Also, lack of supervision by the performance metric led to more than double the number of average iterations, for a small improvement in final IAE reached compared to the supervised algorithm. Using Table 11 to compare between the algorithms of Figures 2.14 & 2.15, it is concluded that tuning with the algorithm using the diagnosis criteria of Visioli provides the most benefit com-

pared to the other algorithm regardless of the choice of diagnosis measure. This result indicates the utility of combining the information available from several diagnosis measures. In the future, other combinations of diagnosis measures should be considered to see if further improvements with respect to performance and stability can be gained. Redundant information between different diagnosis measures could be investigated by use of singular valued decomposition, and support vector machines or other learning algorithms could then be applied to synthesize a final diagnosis based on the values of the set of diagnosis measures. Future work may also find that certain diagnosis techniques have better performance when dealing with data having certain disturbance types not considered here. For example, the Hurst exponent and ζ are both diagnosis measures designed for use with stochastic disturbances, as the underlying calculation of the detrended variance and the autocorrelation function, respectively, are statistical techniques for use with random data. Therefore, efficacy of these methods should improve for processes which are responding to disturbances that occur more frequently (having overlap in their response) and varied in direction and magnitude than the single step in load magnitude used in this study

Above all, the results of this section raise concerns with the stability of automated data-based techniques and more generally, retuning based upon heuristic measures of process behavior. It is well known that traditional adaptive systems require persistently exciting input in order to guarantee global exponential stability,

and it has been clearly demonstrated that lack of excitation can lead the adaptive controller to instability ([84] for example). The methods reviewed here, relying solely upon closed-loop features, are considered advantageous for industrial implementation because of their low requirements for experimentation and excitation. However, very little is known about these techniques with respect to stability. We have shown some very specific examples of how instability can be reached, but more rigorous mathematical foundations explaining stability of these methods are lacking both here and in the literature. As a practical matter, as proposed above, perhaps the combination of several diagnosis measures can also yield a better prediction of whether retuning in a considered direction will result in instability. It has been demonstrated here that the use of supervision by a performance metric is another way to mitigate the effects of erroneous diagnoses when performing heuristic data-based retuning.

2.8 Conclusions and Future Work

This work has examined several previously proposed data-based methods for PI controller tuning diagnosis and retuning. Most of the existing techniques for controller diagnosis classify poor tuning as sluggish or aggressive, but a rigorous definition of the meaning of these terms was previously absent. Here, we introduced a definition of the sluggish and aggressive categories upon the controller's parameter space that can be used to evaluate the performance of data-based con-

troller tuning diagnosis techniques. The Hurst exponent as a diagnosis measure for sluggish and aggressive loop tuning was introduced, and its performance was deemed comparable to the other available diagnosis measures. By examining behavior of the diagnosis measures across the relevant part of the K, T -plane, we gained insight into how these methods will perform under different tuning conditions. Finally, new adaptive algorithms for PI controller retuning were introduced that take advantage of the data-based diagnosis measures, and it was found that stability is a major concern that could limit the applicability of such techniques. Among the proposed automated retuning techniques based on data-based diagnosis measures, it was found that diagnosis based on the combined information from several measures provided superior performance to any single measure considered alone. This indicates that, in the spirit of using the values of multiple feature types for pattern recognition problems, future work in the area of data-based tuning diagnosis should concentrate on developing methods which synthesize information from several diagnostic measures.

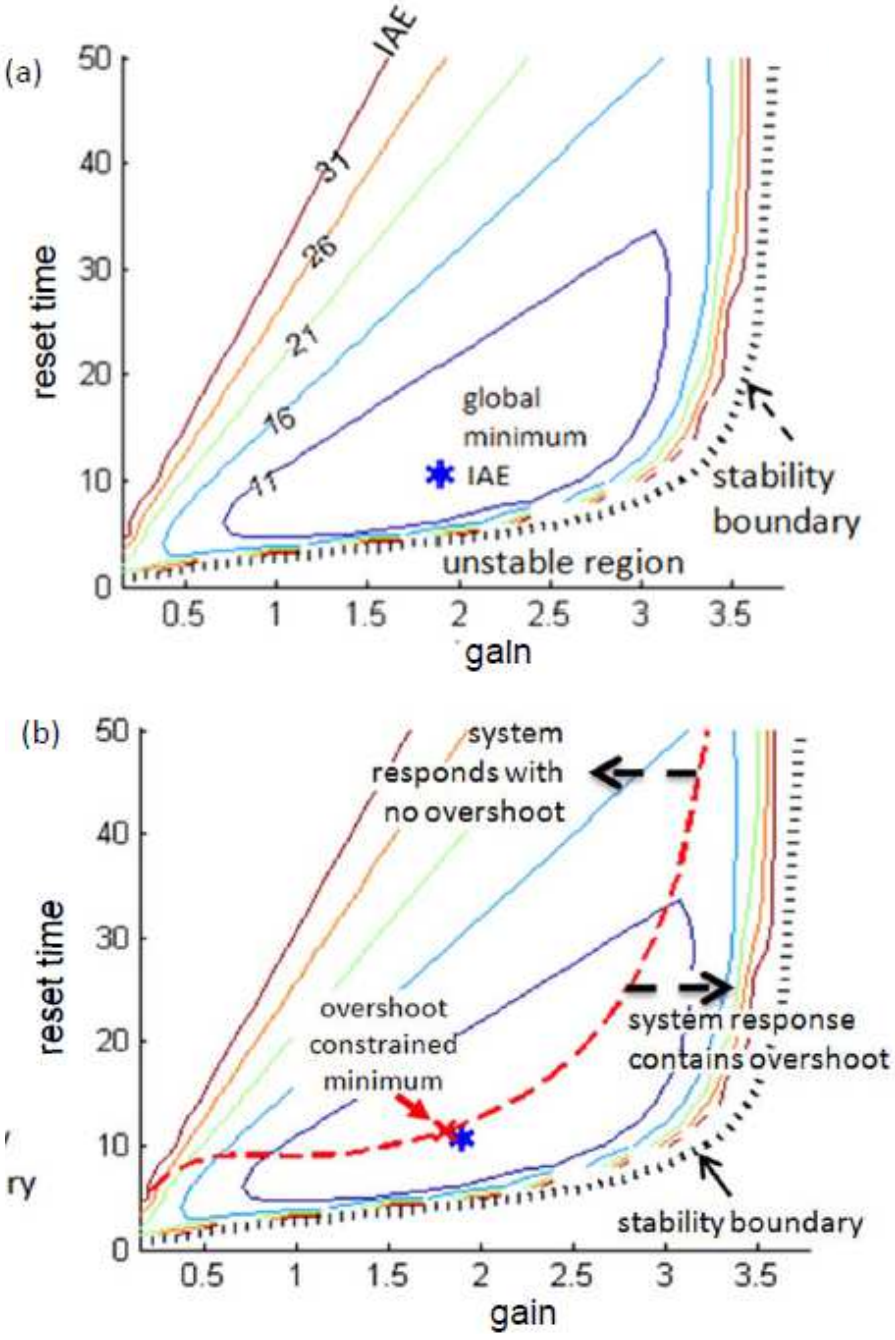

Figure 2.4: (a) Contours of IAE for a load disturbance response over the space of controller parameters. (b) The dashed line separates systems responding with overshoot from those which do not and contains the constrained global optimum (×)

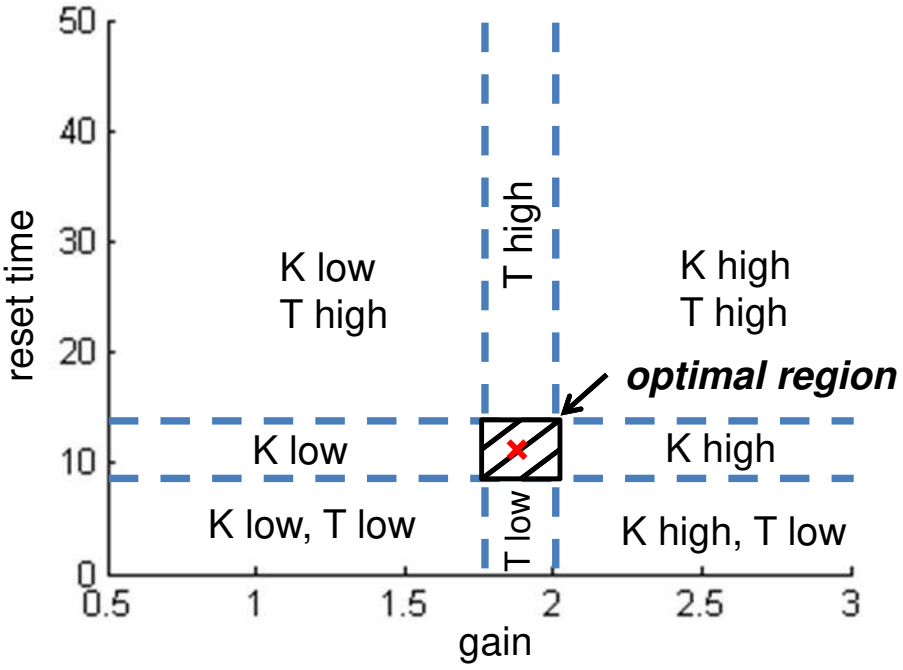

Figure 2.5: Classification of the current parameter values relative to the global optimal set (×).

Table 2.7: Systems used in simulation studies

Open Loop Plant	Disturbance Filter	Normalized dead-time
$\frac{e^{-5s}}{10s+1}$	$\frac{1}{10s+1}$	0.5
$\frac{1}{s+1^4}$	$\frac{1}{s+1^4}$	0
$\frac{e^{-3s}}{15s+15s+12s+1}$	$\frac{1}{17s+14s+1s+1}$	0.2
$\frac{e^{-5s}}{s+1^3}$	$\frac{1}{s+1^3}$	5
$\frac{e^{-10s}}{s+1s+2s+3}$	$\frac{1}{s+1s+2s+3}$	10

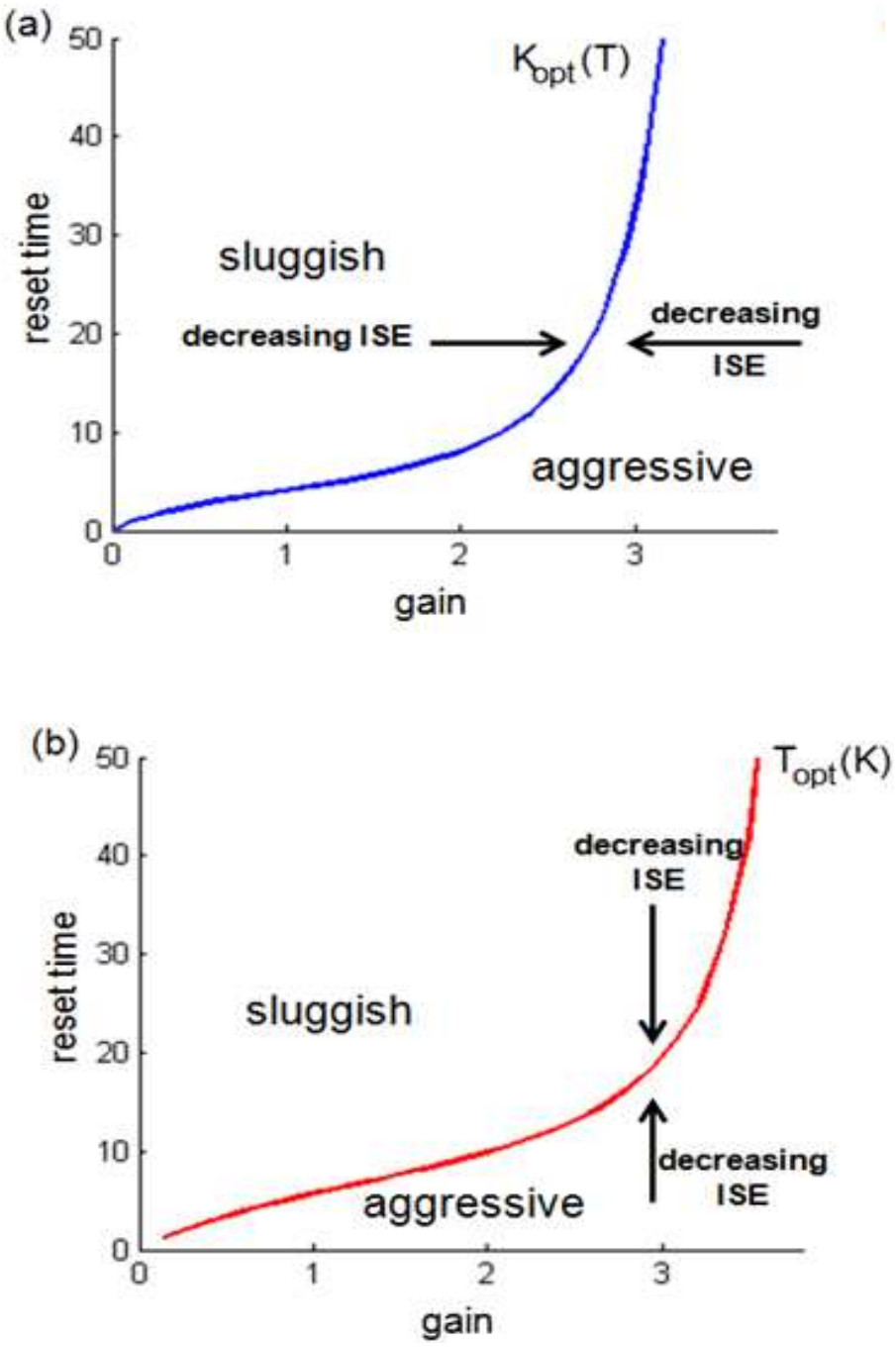

Figure 2.6: For System 1, definitions of sluggish & aggressive: (a) $K_{opt}(T)$, the optimal gain as a function of reset time and (b) $T_{opt}(K)$, the optimal reset time as a function of gain.

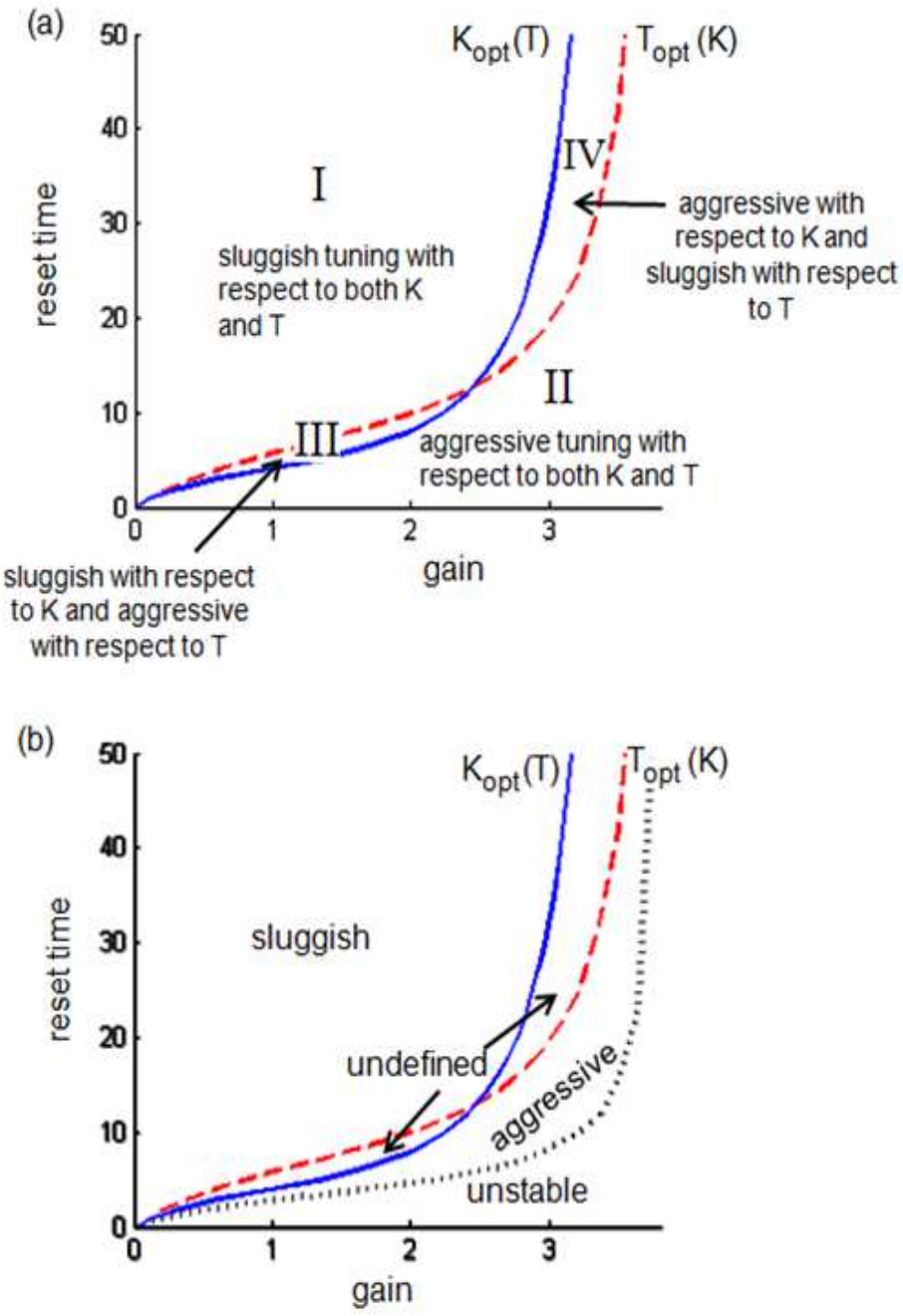

Figure 2.7: (a) Definitions of sluggish and aggressive resulting from overlapping the (a & b) plots of Figure 2.6. (b) Proposed segregation of the tuning plane into regions of sluggish, aggressive, and undefined controller parameters.

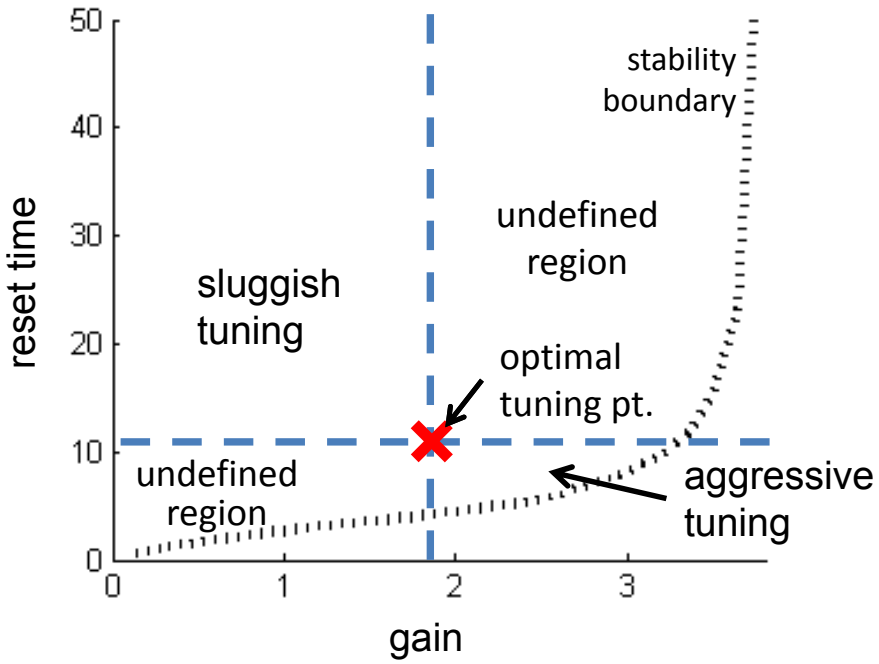

Figure 2.8: Sluggish and aggressive tuning definitions based upon parameter values relative to their values at the global optimum (×).

Table 2.8: Results of classification of 600 training sets of parameter pair (K, T) into categories sluggish and aggressive

diagnosis measures	sluggish systems	aggressive systems	systems in undefined regions	
	% correctly identified	% correctly identified	% identified as sluggish	% identified as aggressive
HE	85.6	81.2	19.0	81.0
II	80.4	75.2	61.0	39.1
AI	66.8	71.2	28.8	71.2
RI	93.6	97.2	14.3	85.7
ζ	92.4	95.6	24.8	75.2

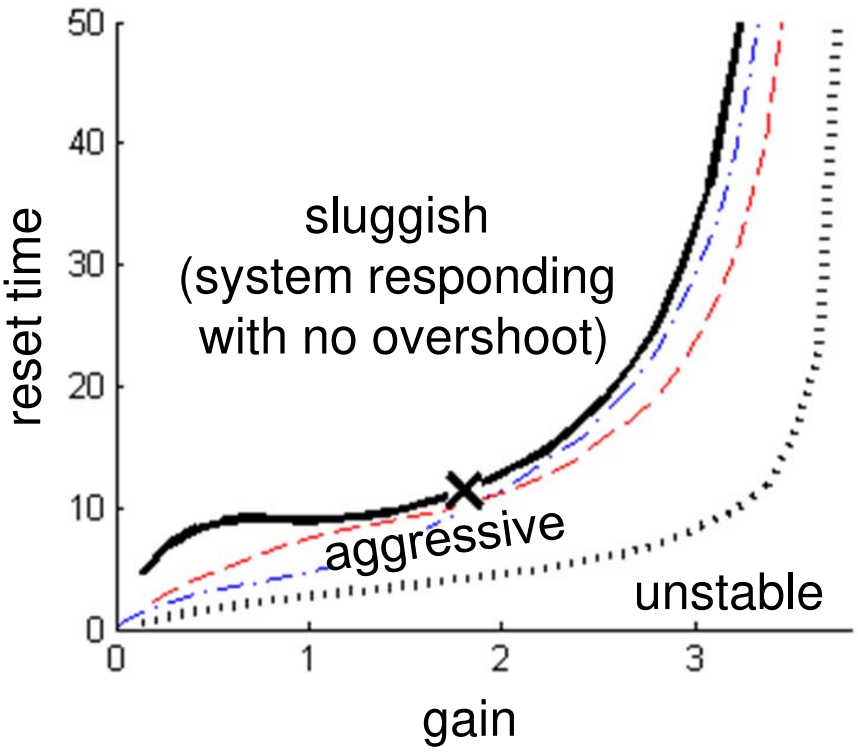

Figure 2.9: The proposed definitions of sluggish and aggressive regions of controller parameters under the presence of a constraint upon overshoot (solid line). Also shown are $T_{opt}(K)$ and $K_{opt}(T)$ (dashed and dash-dot lines, respectively), the constrained global optimum (x) and the stability boundary (dotted line).

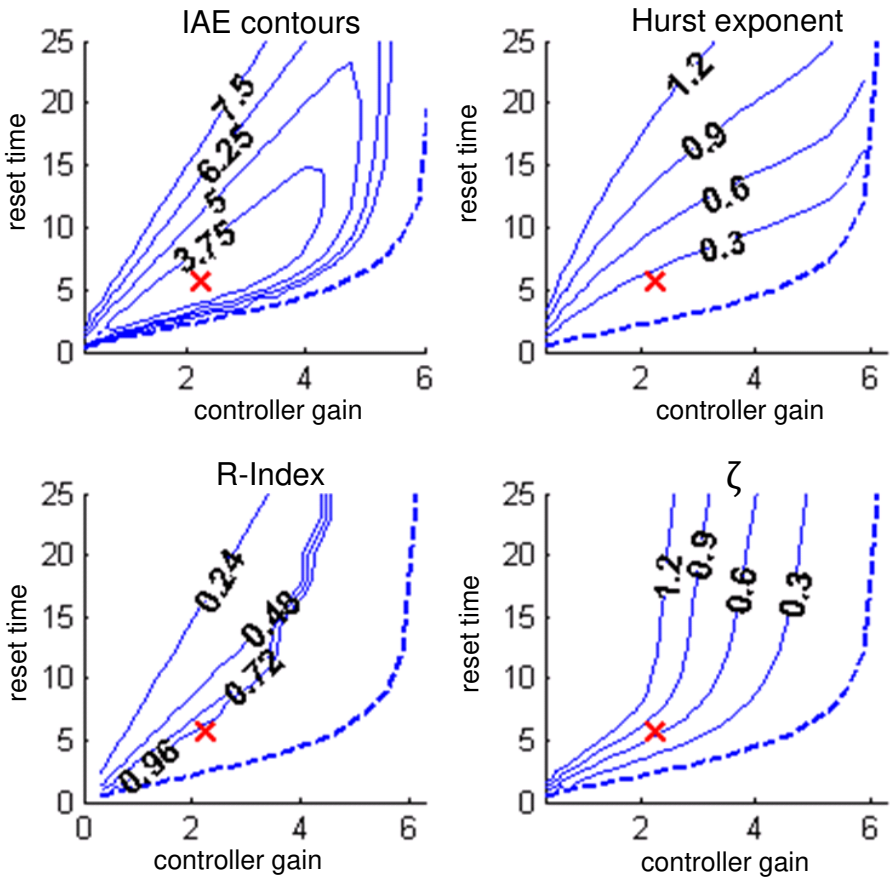

Figure 2.10: For System 5 subjected to the case of a unit load disturbance change, plots of integral of absolute error (IAE) and three different controller tuning diagnosis measures (Hurst exponent, R-index, ζ). In each plot the dotted line represents the stability boundary and the (x) is point where minimum IAE is obtained.

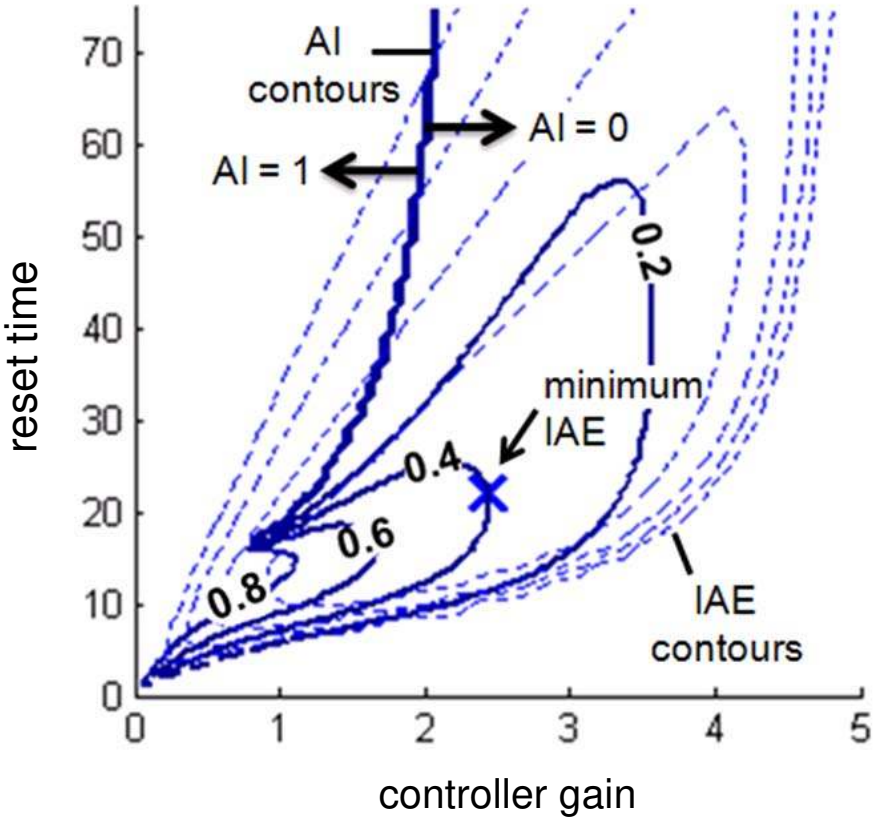

Figure 2.11: Contours of the area index overlaid upon contours of the integral of absolute error.

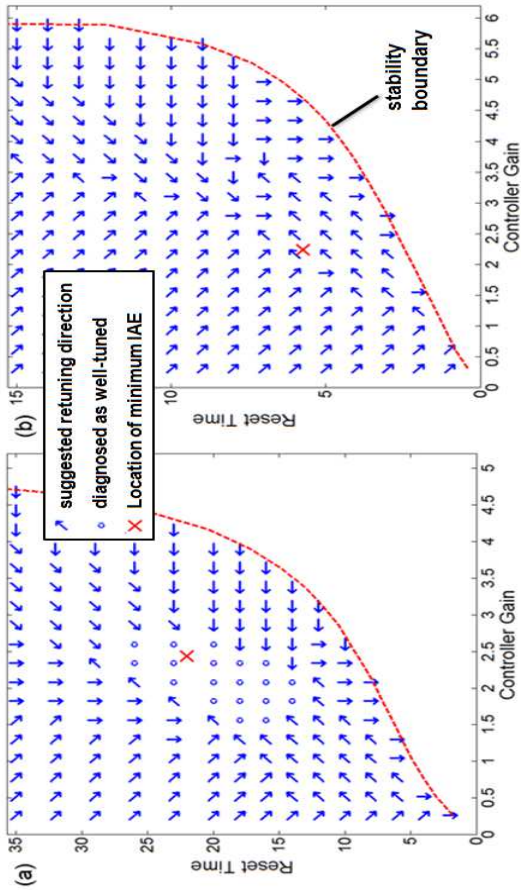

Figure 2.12: For (a) System 3 and (b) System 5 each subjected to the case of a unit load disturbance change, diagnosis of controller tuning using the method of Visioli [77]

Table 2.9: Demonstration of the retuning algorithm of Figure 2.14 for the case of isolated load disturbance changes afflicting System 2 with the Hurst exponent chosen as the method for diagnosis of sluggish or aggressive tuning

Iteration	K	T	IAE	HE	Comments
0	3.53	16.2	9.8	0.486	HE indicates aggressive turning, decrease gain 20%.
1	2.83	16.2	5.7	0.514	Diagnosis has changed, explore boundary at 10%.
2	3.18	16.2	5.3	0.499	Local minimum found within 10%. Now switch and increase T by 20% according to the aggressive diagnosis.
3	3.18	19.5	6.1	0.604	IAE has increased, try 10% step instead.
4	3.18	17.8	5.6	0.552	IAE is still increased, return to best parameter set.
5	3.18	16.2	5.3	0.499	Neither changing K nor T can give decreased IAE in indicated direction of retuning. Terminate.

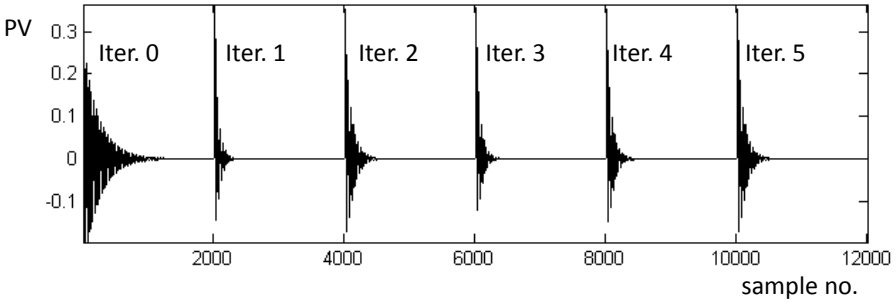

Figure 2.13: PV of System 2 corresponding to the results in Table 2.9

Table 2.10: Demonstration of the retuning method proposed for use with the diagnosis method of [77] for the case of isolated load disturbance changes afflicting System 5

Iteration	K	T	IAE	AI	II	OI	Comments
0	3.45	18.2	5.3	1.00	0.380		Increase gain and lower reset time according to AI, II.
1	3.80	16.4	4.3	1.00	-0.349		IAE has decreased, so continue retuning according to AI, II based diagnosis.
2	4.18	18.0	4.3	0.01	-0.787	0.002	Diagnosis method is indicating to reverse last move. Instead terminate at best parameter set reached.
3	3.80	16.4	4.3	1.00	-0.349		AI, II indicate to change turning, but the suggested move has previously been attempted. Terminate.

Table 2.11: Summary of retuning results for 5 systems retuned from 50 starting points

methods used	retuning using both parameters			K adjustment only		
	mean % change in IAE	mean iterations	# of trials instability reached	mean % change in IAE	mean iterations	# of trials instability reached
I. retuning based upon a single diagnosis measure	-20.5	7.2	38	-18.3	5.4	38
HE	-23.3	6.8	0	-23.2	5.8	0
II	-25.9	7.2	24	-24.4	5.6	24
RI	-18.1	6.7	0	-17.8	5.6	0
AI	-27.8	8.4	17	-25.0	5.6	17
Zeta						
II. turning based on Visioli's diagnosis method	-42.0	14.5	35			
unsupervised	-37.8	6.3	2			
with IAE supervision						
III. theoretically achievable	-47.5					

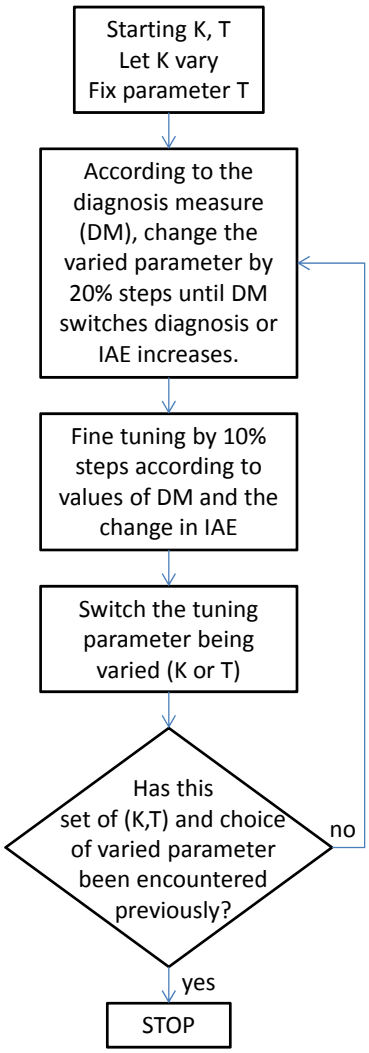

Figure 2.14: Proposed retuning algorithm using one of the sluggish/aggressive controller diagnosis methods (DM).

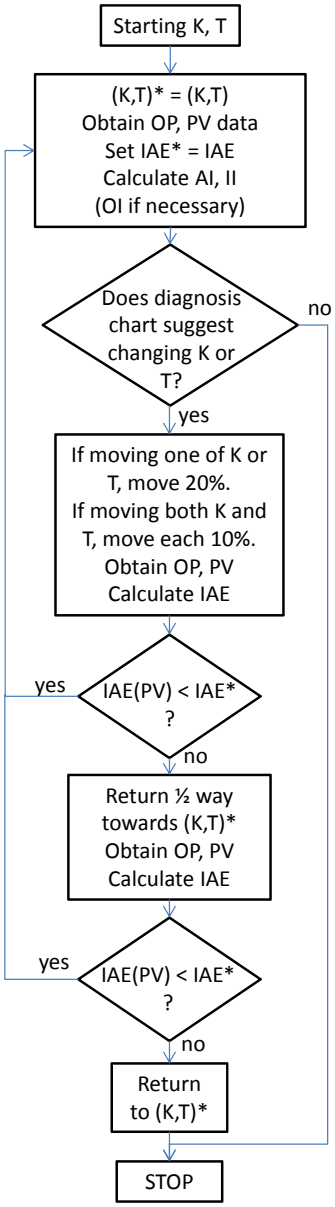

Figure 2.15: Proposed retuning algorithm using the controller diagnosis method of Visioli [77].

Nomenclature

ACF	autocorrelation function
AR	autoregressive
CLPM&D	closed-loop performance monitoring and diagnosis
CLPI	closed-loop performance index
DFA	detrended fluctuation analysis
FIR	finite impulse response
FOPTD	first order plus time delay
HVAC	heating ventilation airconditioning
IAE	integral of absolute error
ISE	integral of squared error
ITAE	integral of timeweighted absolute error
MVI	Harris's minimum variance index
OP	output
PI	proportional-integral
PID	proportional-integral-derivative
PRAC	pattern recognition adaptive controller
PV	process variable
RDI	relative damping index
SISO	single-input single-output
SOPTD	second order plus time delay

AI	area index
F	fluctuation
HE	Hurst Exponent
II	idle index
K	controller gain
K_{opt}(T)	optimal gain for a given integral time
L	rang of window length
M_x	scaling factors along the x axes
M_y	scaling factors along the y axes
N	number of samples
OI	output index
Q	PI controller
RI	R-index
W	number of windows
Y_t	detrended and integrated signal
T	reset time
$T_{opt}(K)$	optimal reset time for a given gain
a	the scaling factor along the x-axis
t_{pos}	the total time for which the increments of the process variable and controller output have the same sign
t_{neg}	the total time for which the increments of the process variable and controller output have the opposite sign
y_t	a time series
$\bar{y}$	mean value of y

α	the (generalized) Hurst exponent
ζ	damping coefficient
ζ_{agg}	the damping coefficient with user defined bounds on aggressive values
ζ_{slug}	the damping coefficient with user defined bounds on sluggish values
σ_y	variance of the process output measurement
σ_{MV}	theoretical minimum variance achievable for the given process with any linear controller
η	normalized dead-time parameter

CHAPTER 3

THEORETICAL CONDITIONS FOR THE DETECTION OF VALVE STICTION IN LINEAR SYSTEMS

Previous works introduced control valve stiction detection and quantification methods for closed-loop systems based on the identification of a Hammerstein element between the feedback controller and plant output signals. These techniques each rely upon the fact that the presence of valve stiction introduces nonlinearities in the closed-loop system, yet, no theoretical discussions have been presented which explain the conditions under which these methods will succeed or fail in properly detecting valve stiction. Therefore, the present work uses frequency domain analysis to provide a theoretical investigation of the identification of stiction in otherwise linear closed-loop systems. In this way, the failure of Hammerstein stiction detection techniques to positively identify valve stiction in certain systems in which it known to be present can be explained accordingly. This chapter is excerpted from paper: "T. Spinner, B. Srinivasan, R. Rengaswamy. On the detection of valve non-linearities in closed svstems. *Automatica* (under review)".

3.1 Introduction

Control valve stiction, static friction preventing proper movement, is one of the long-standing and common causes for oscillations in the process industries [142, 143], leading to poor controller performance, inferior quality products and larger rejection rates [144, 145]. As a result, much attention has been paid towards the development of techniques for detection and quantification of stiction in control valves. The presence of stiction in the control valve introduces non-linear behavior between the controller signal and manipulated variable [146, 143], and detection of these stiction induced nonlinearities forms the basis of several automated techniques to identify when oscillations are caused by stiction. These include shape based [147, 148], frequency based [149], and model based methods. Of the latter, Srinivasan *et al.* [150] first presented a Hammerstein model based identification technique for the diagnosis and quantification of valve stiction. The idea behind this approach is to fit a Hammerstein model (linear model & nonlinear element) between the controller output and process variable data. Later, several variants of this method were proposed by Jelali *et al.* [151], Lee *et al.* [152], Ivan *et al.* [153], Choudhury *et al.* [146] and Karra *et al.* [154] for stiction detection in linear closed-loop systems.

The Hammerstein model based approaches generally first filter the controller output data through their respective nonlinear valve models, which contain either

one or several parameters quantifying the amount of stiction present. An optimal ARMAX model is then found between this filtered controller output and the process variable, with the best linear model chosen on the basis of Akaike's Information Criterion [155]. This is step is repeated over a search of stiction parameter values to find the parameter (and associated best linear model) that minimizes the squared error between predicted and recorded process outputs. An optimal stiction parameter which is nonzero indicates the presence of valve stiction, while a parameter value of zero indicates its absence. While there have been several comparative studies of the various algorithms [59, 72] based on simulations, these studies do not provide any fundamental insights on the underlying reasons for the success or failure of these algorithms. In this work, frequency domain analysis is performed to identify certain conditions under which Hammerstein model based detection methods could succeed or fail in properly identifying the presence of control valve stiction. The approach proposed in this work is general and may also be used when analyzing other types of nonlinearities in control valves.

3.2 Motivation

Suppose valve stiction can be described by the following model:

$$v(t) = \Upsilon(\eta)u(t), \quad t = 1, 2, \ldots, N \tag{3.1}$$

where t denotes the observations of input and output at sampling instants $t = kT$ (sampling time $T = 1$) with input kept constant between the sampling instants. In the above equation, $u(t)$ is what the valve position would be in the absence of stiction and $v(t)$ is the actual valve position, which is the output of the nonlinear and possibly discontinuous operator $\Upsilon(\eta)$, where Υ has a known structure but depends on the unknown real vector of stiction parameters η. A special value of the parameter vector, $\eta = 0$ has the unique property that for all t, $\Upsilon(\eta = 0)u(t) = u(t)$. For all other values of η, this property does not hold. An estimate of η, denoted $\hat{\eta}$, results in estimated valve position $\hat{v}(t)$ from the relation

$$\hat{v}(t) = \Upsilon(\hat{\eta})u(t) \tag{3.2}$$

Now suppose that output measurement sequence $y(t)$ is generated by the output error process

$$y(t) = G(q)v(t) + e(t) \tag{3.3}$$

where $y(t)$ and $v(t)$ are as above, $e(t)$ is a sequence of independent samples from a white noise sequence, and $G(q)$ is a linear transfer function model. Here, q denotes the forward time shift operator. Also, suppose that $u(t)$ is related to $y(t)$ by the linear control law $u(t) = -K(q)y(t)$. Using an estimate of $\hat{\eta}$, known values of $u(t)$, and (3.2) to estimate the sequence $\hat{v}(t)$, an estimate of $y(t)$, denoted $\hat{y}(t)$, is obtained

from the following estimation equation

$$\hat{y}(t) = \hat{G}(q)\hat{v}(t) \tag{3.4}$$

where $\hat{G}$ is an estimate of $G(q)$ selected from the set $\mathcal{M}$. Then, the prediction error of $y(t)$ is

$$\begin{aligned} \epsilon(t) &= y(t) - \hat{y}(t) \\ &= G(q)v(t) - \hat{G}(q)\hat{v}(t) + e(t) \end{aligned} \tag{3.5}$$

Suppose the pair $(\hat{G}, \hat{\eta})$ is chosen to minimize the mean squared error of estimate $\hat{y}(t)$, which is given by

$$mse(\hat{y}(t)) = \frac{1}{N}\sum_{t=1}^{N} \epsilon^2(t). \tag{3.6}$$

In the following it will be demonstrated that in the presence of stiction ($\eta \neq 0$), under some conditions, the choice $\hat{\eta} = 0$ allows the quantity $mse(\hat{y}(t))$ to become minimum so that stiction is incorrectly identified as absent, and under other conditions, this cannot happen so that stiction will be correctly identified.

3.3 Main Result

In this section, Lemma 1 will provide frequency domain conditions for the minimization of $mse(\hat{y}(t))$, Theorem 1 will provide a sufficient condition for the correct

detection of the presence of stiction, and Theorem 2 will provide a sufficient condition for stiction to be incorrectly identified as absent. First, appropriate assumptions are introduced so that these developments may proceed.

Assumptions:

1. Signals $u(t)$ and $y(t)$ are known for all values of $t \in 1, 2, \ldots, N$ with t denoting the observations at sampling instant $t = kT$ (sampling Time $T = 1$). Further it is assumed that the input is kept constant during the sampling interval.

2. $G(q)$ is contained in $\mathcal{M}$ which is the search space for $\hat{G}(q)$ with q denoting the time shift operator.

3. $G(q)$ and all of the models in $\mathcal{M}$ are stable.

4. $G(q)$ and all of the models in $\mathcal{M}$ have at least one unit time delay meaning that both $G(q) = 0$ and $\hat{G}(q) = 0$ when $q = 0$.

5. $G(\omega_k)$, the frequency response of $G(q)$, has maximum magnitude $G_\infty = \max_{\omega_k} |G(\omega_k)|$ and all models $\hat{G} \in \mathcal{M}$ have the property $\max_{\omega_k} |\hat{G}(\omega_k)| \leq G_\infty$. By Assumption 3, $G_\infty < \infty$.

6. The Discrete Fourier Transform of signal $v(t)$, denoted $V(\omega_k)$, has a maximum bound $V_\infty = \max_{\omega_k} |V(\omega_k)| < \infty$.

7. Let $e(t)$ be a zero mean white noise process with independent samples, and

let its Discrete Fourier transform $E(\omega_k)$ have minimum and maximum bounds $E_{min} = \min_{\omega_k}|E(\omega_k)| > 0$ and $E_\infty = \max_{\omega_k}|E(\omega_k)| < \infty$, respectively.

8. The Discrete Fourier Transform of the control law satisfies the relation $U(\omega_k) = -K(\omega_k)(G(\omega_k)V(\omega_k) + E(\omega_k))$ exactly.

9. Assume that the controller output maintains some minimum power over all frequencies, so that $|U(\omega_k)| \geq U_{max}$, where $U_{max} = sup(|K(\omega_k)|)E_{min}$.

Lemma 1. Under the above assumptions, the minimization of $mse(\hat{y}(t))$ requires that $|G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)\hat{V}(\omega_k)| = 0$ for all $\omega_k = \dfrac{2\pi k}{N}$, $k = 1, 2, \ldots, N$, and this results in $mse(\hat{y}(t)) = var(e(t))$.

Proof. Signal $u(t)$ depends upon the signal $e(t)$ according to control law $u(t) = -K(q)y(t)$ and (3.3), and therefore $v(t)$ also has dependence by (3.1). Specifically sample u_t of signal $u(t)$ and therefore sample v_t of signal $v(t)$ depend upon samples $e_t, e_{t-1}, ..., e_0$ of signal $e(t)$. However, assuming plant models $G(q)$ and $\hat{G}(q)$ each contain at least one unit time delay, sample t of plant filtered signals $G(q)v(t)$ and $\hat{G}(q)v(t)$ are independent of sample e_t. Therefore,

$$\mathbb{E}[\epsilon^2(t)] = \mathbb{E}[\epsilon_r^2(t)] + var(e(t)) \tag{3.7}$$

where $\mathbb{E}$ is the expectation operator, and the quantity $\epsilon_r(t)$ is defined

$$\epsilon_r(t) = G(q)v(t) - \hat{G}(q)\hat{v}(t). \tag{3.8}$$

It is assumed that the sampled versions of each quantity also satisfy the same relationship, so that

$$mse(\hat{y}(t)) = \frac{1}{N}\sum_{t=1}^{N}\epsilon^2(t) = \frac{1}{N}\sum_{t=1}^{N}\epsilon_r^2(t) + var(e(t)) \tag{3.9}$$

and therefore the mean squared modelling error is divided into two independent components, $\frac{1}{N}\sum_{t=1}^{N}\epsilon_r^2(t)$ which can be affected by choice of $(\hat{G}, \hat{\eta})$, and $var(e(t))$ which cannot. Therefore the choice of $(\hat{G}, \hat{\eta})$ which minimizes MSE is the one that minimizes

$$\frac{1}{N}\sum_{t=1}^{N}\epsilon_r^2(t) = \frac{1}{N}\sum_{t=1}^{N}(G(q)v(t) - \hat{G}(q)\hat{v}(t))^2. \tag{3.10}$$

Let $\xi(\omega_k)$, $Y(\omega_k)$, $U(\omega_k)$, $V(\omega_k)$, $\hat{Y}(\omega_k)$, and $\hat{V}(\omega_k)$ represent the Discrete Fourier Transforms of signals $\epsilon_r(t)$, $y(t)$, $u(t)$, $v(t)$, $\hat{y}(t)$, and $\hat{v}(t)$, respectively. Also let $G(\omega_k)$, $\hat{G}(\omega_k)$, and $K(\omega_k)$ represent the frequency response functions of $G(q)$, $\hat{G}(q)$, and $K(q)$, respectively. By Parseval's theorem,

$$\sum_{t=1}^{N}\epsilon_r^2(t) = \sum_{k=1}^{N}|\xi(\omega_k)|^2 \tag{3.11}$$

where $\xi(\omega_k) = \xi(2\pi k/N) = \dfrac{1}{\sqrt{N}} \sum_{t=1}^{N} \epsilon_r(t) e^{-j2\pi kt/N}$ is the Discrete Fourier Transform (DFT) of $\epsilon_r(t)$. Taking the DFT of each side of (3.8) yields

$$\xi(\omega_k) = G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)\hat{V}(\omega_k) + R'(\omega_k) \tag{3.12}$$

where $R'(\omega_k)$ is an artifact of the Discrete Fourier Transform which can be neglected for large sample size N [156]. Then the problem of minimizing $mse(\hat{y}(t))$ over $(\hat{G}, \hat{\eta})$ is equivalent to minimizing the quantity $\sum_{k=1}^{N} \left| G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)\hat{V}(\omega_k) \right|^2$.

It has been assumed that the pair (G, η) is in the search space for $(\hat{G}, \hat{\eta})$, so that it is possible for $|G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)\hat{V}(\omega_k)|$ to become zero for all ω_k. This provides a minimum value of $mse(\hat{y}(t))$, $mse(\hat{y}(t)) = var(e(t))$. By Equations (9)-(12), if for any ω_k, $|G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)\hat{V}(\omega_k)| > 0$, then $mse(\hat{y}(t)) > var(e(t))$. Therefore, the choice of a pair $(\hat{G}, \hat{\eta})$ that is a minimizer for $mse(\hat{y}(t))$ requires $|G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)\hat{V}(\omega_k)| = 0$ for all ω_k, which results in $mse(\hat{y}(t)) = var(e(t))$.

Observation 1. When no stiction is included in the estimated model ($\hat{\eta} = 0$),

$$\begin{aligned}
\xi_r(\omega_k) &= \left| G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)U(\omega_k) \right| \\
&\geq |G(\omega_k)V(\omega_k)| - |\hat{G}(\omega_k)U(\omega_k)| \\
&\geq |G(\omega_k)V(\omega_k)| - |\hat{G}(\omega_k)|\,|U(\omega_k)| \\
&\geq |G(\omega_k)V(\omega_k)| - G_\infty\,|U(\omega_k)|
\end{aligned}$$

Therefore, $\xi_r(\omega_k) > 0$ if $G_\infty < \dfrac{|G(\omega_k)V(\omega_k)|}{|U(\omega_k)|}$.

Theorem 1. If there exists $\omega_C > 0$ such that for all $\omega_k > \omega_C$, $K(\omega_k) < \kappa = \dfrac{1}{2G_\infty}$, and there also exists a frequency $\omega_k > \omega_C$, $k = 1...N$, such that $|G(\omega_k)V(\omega_k)| > E_\infty$, and also $\eta > 0$, then the $\hat{\eta}$ minimizing $mse(\hat{y}(t))$ will require $\hat{\eta} > 0$.

Proof. Given $\omega_k > \omega_C$, by Assumption 8 and the existence of upper bound κ, $|U(\omega_k)| =$

$|K(\omega_k)\,(G(\omega_k)V(\omega_k) + E(\omega_k))| < \kappa(|G(\omega_k)V(\omega_k)| + E_\infty)$. Therefore,

$$\begin{aligned} \frac{|G(\omega_k)V(\omega_k)|}{|U(\omega_k)|} &> \frac{|G(\omega_k)V(\omega_k)|}{\kappa\,|G(\omega_k)V(\omega_k) + E(\omega_k)|} \\ &\geq \frac{|G(\omega_k)V(\omega_k)|}{\kappa\,(|G(\omega_k)V(\omega_k)| + E_\infty)} > \frac{1}{2\kappa} = G_\infty \end{aligned} \tag{3.13}$$

where the final inequality holds due to $|G(\omega_k)V(\omega_k)| > E_\infty$. Then by Observation 1, $\xi_r(\omega_k) > 0$ for at least one ω_k for the choice $\hat{\eta} = 0$. Since G and η are in the search space for $(\hat{G},\hat{\eta})$, it is possible to choose $(\hat{G},\hat{\eta})$ giving $\xi_r(\omega_k) = 0$ for all ω_k, and therefore any pair $(\hat{G},\hat{\eta} = 0)$ is not a minimizer for $mse(\hat{y}(t))$.

Remark 1. The above theorem states that under the present assumptions stiction will be correctly identified given the presence of (i) a cutoff frequency past which the magnitude of the controller's frequency response is sufficiently low and (ii) the presence of sufficient magnitude of plant-filtered stiction induced excitation $G(\omega_k)V(\omega_k)$ at frequencies greater than the controller cutoff frequency. This requires a sufficiently high plant cutoff frequency such that $G(\omega_k)$ has great enough

magnitude for this condition to be satisfied at some frequency $\omega_k > \omega_C$. Theorem 1 gives a sufficient condition for identification of stiction when it is present.

Remark 2. In prior works, the identified Hammerstein model used an ARMAX type model instead of an output error type model to model the linear component. In that case, the white noise filter included in the identified model could possibly be manipulated to capture the effects of stiction induced excitation at frequencies past the controller cutoff frequency. To prevent this from occurring in the prior works, the order of the ARMAX model was restrained by using the Akaike Information Criterion during linear model selection [155], which penalizes use of more model coefficients.

Observation 2. From discrete linear systems theory, if $\frac{|G(\omega_k)V(\omega_k)|}{|U(\omega_k)|} \leq G_\infty$ for all ω_k then there exists a $\hat{G} \in \mathcal{M}$ such that $\left|G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)U(\omega_k)\right| = 0$ for all ω_k.

Theorem 2. If there exists a $\omega_C > 0$ such that for all $\omega_k < \omega_C$, $K(\omega_k) > \kappa = \frac{V_\infty}{E_{min}}$, and if there also exists ω_G $(0 < \omega_G < \omega_C)$ such that for all $\omega_k > \omega_G$, $G(\omega_k) < \gamma = \frac{G_\infty E_{min} K_{min}}{V_\infty}$, where $K_{min} = \min_{\omega_k} |K(\omega_k)|$, then $\hat{\eta} = 0$ will always be a minimizer of $mse(\hat{y}(t))$ no matter the value of η.

Proof. By Assumption 9, $|U(\omega_k)| \geq |K(\omega_k)|E_{min}$ for all ω_k, and for $\omega < \omega_C$,

$|K(\omega_k)|E_{min} > \kappa E_{min}$. Also, in general, $|G(\omega_k)V(\omega_k)| \leq G_\infty V_\infty$, so that therefore

$$\begin{aligned} \frac{|G(\omega_k)V(\omega_k)|}{|U(\omega_k)|} &\leq \frac{G_\infty V_\infty}{|K(\omega_k)|E_{min}} \\ &< \frac{G_\infty V_\infty}{\kappa E_{min}} = G_\infty. \end{aligned} \tag{3.14}$$

Alternatively, when $\omega_k \geq \omega_C > \omega_G$, $|G(\omega_k)V(\omega_k)| \leq \gamma V_\infty$, and in general, $|U(\omega_k)|$ $\geq K_{min}E_{min}$, so that therefore

$$\frac{|G(\omega_k)V(\omega_k)|}{|U(\omega_k)|} \leq \frac{\gamma V_\infty}{K_{min}E_{min}} = G_\infty. \tag{3.15}$$

Therefore, since $\frac{|G(\omega_k)V(\omega_k)|}{|U(\omega_k)|} \leq G_\infty$ for all ω_k, then by Observation 2, there exists a $\hat{G} \in \mathcal{M}$ such that $\left|G(\omega_k)V(\omega_k) - \hat{G}(\omega_k)U(\omega_k)\right| = 0$ for all ω_k Therefore, $\hat{\eta} = 0$ is a minimizer for $mse(\hat{y}(t))$, and a model without stiction can be identified regardless of the value of the value of η.

Remark 3. The prior theorem states that under the existence of appropriate conditions, if plant cutoff frequency ω_G is less than the controller cutoff frequency ω_C (with appropriate definitions for each ω_C and ω_G), then a process model assuming no stiction ($\hat{\eta} = 0$) can always be identified that gives minimum mse, even in the presence of stiction. Theorem 2 provides a sufficient condition for incorrectly identifying the absence of stiction when it is present.

Remark 4. While there has been much debate in the literature on the effect of the

quality of the stiction model (one or two parameters) and its effect on stiction detection, theorem 2 shows that even in the case where the true process and stiction models are contained in the search space, under certain conditions, stiction will not be detected when present. While we have not shown that these are the only conditions (necessity) under which stiction detection could be compromised, establishing such conditions exist is of critical importance in further development of these techniques.

Remark 5. In [59], it can be seen that there are cases where Hammerstein based stiction detection fails whereas a shape based approach works. The shape based and other similar approaches can be viewed as imposing further constraints (frequency shaping) on the identified model that could help in improving stiction detection in some specific instances.

3.4 Simulation Examples

In this section, two simulation case study results are provided to illustrate Theorems 1 and 2 of the previous section. In the first case, corresponding to Theorem 1, the plant model is a discrete time transfer function $G(z) = \dfrac{0.957z^{-1}}{1 - 0.043z^{-1}}$ and the controller, $K(z) = \dfrac{0.5 - 0.45z^{-1}}{1 - z^{-1}}$, is of discrete PI form. The Bode magnitude plot of these two transfer functions is given in Figure 3.1 (a), which shows that the magnitude of the controller frequency response becomes lower than that of the plant by

10^{-2} Hz. Within Simulink, a one parameter valve stiction model [150] with $\eta = 2$ was added between the controller and plant blocks to simulate stiction. The closed loop was excited by both a unit step change in set point, as well as white noise of variance 2.5×10^{-3} added to the plant output. The simulation was conducted for 10^4 seconds with a sample time of 1 second. The resulting controller, plant, and valve signals ($u(t)$, $y(t)$, and $v(t)$) are shown in Figure 3.1 (b). In Figure 3.1 (c), the Discrete Fourier transform of each $u(t)$ and $y(t)$ are shown, as well as the extra information $Y_e(\omega_k)$ which is defined as $Y_e(\omega_k)= |G(\omega_k)V(\omega_k)| - G_\infty|U(\omega_k)|$. For each the case $\hat{\eta} = 2$ and $\hat{\eta} = 0$, ARX model identification with stability enforced was performed using the Matlab System Identification Toolbox. The G_∞ constraint was not explicitly enforced on the identified models, but enforcing stability led to this condition being satisfied over most of the frequency range of each estimated model. Fitting ARX models with up to 20 poles and 30 zeros, models for the $\hat{\eta} = \eta = 2$ case achieved $mse(\hat{y}(t))$ as low as 2.51×10^{-3}, while in the case of $\hat{\eta} = 0$, the lowest $mse(\hat{y}(t))$ was 0.13, meaning that, here, stiction is considered to be correctly identified. Additional coefficients may reduce the mse value further, however, due to the large amount of extra information present, it is unlikely models assuming $\hat{\eta} = 0$ will give lower mse than those using $\hat{\eta} = 2$ even if the G_∞ constraint is not strictly enforced.

In the second example, corresponding to the results of Theorem 2, a continuous

time plant $G(s) = \dfrac{3e^{-s}}{100s+1}$ and controller $K(s) = \dfrac{1}{s+1}$ were simulated for 10^3 seconds with a stiction parameter of $\eta = 0.4$. The variance of the white noise added to the plant output was 0.005, otherwise all other simulation conditions were the same as stated for the previous example. The Bode magnitude plots of Figure 3.1 (e) show that in this case the plant frequency response cuts out much earlier than that of the controller, and (f) of this figure displays the controller, plant, and valve signals from the response to a unit step change in set point. Figure 3.1 (g) contains the Discrete Fourier Transforms of each the controller and plant outputs. The extra information $Y_e(\omega_k)= |G(\omega_k)V(\omega_k)| - G_\infty|U(\omega_k)|$ is not shown, as it is zero at all frequencies for this system. Fitting ARX models with up to 20 poles and 30 zeros, the lowest mse achieved for the case of models using $\hat{\eta} = \eta = 0.4$ was 0.597, while for the case of models using $\hat{\eta} = 0$ a lower mse of 0.591 was obtained. Different results might be achieved by altering the maximum numbers of ARX coefficients, but in this case, using amounts of parameters similar to practical algorithms, no stiction was identified even though stiction was present.

3.5 Industrial Case Studies

In this section, the results of Section 3.3 are used to analyze stiction detection on industrial control loop datasets provided by Horch *et al.* [54]. Datasets for both an integrating level loop (tag name: LC011) and a non-integrating flow control loop

(tag name:FC525) are considered. The true cause for oscillations in these control loops was the presence of stiction in control valves (reported in [54]). The level control loop is an interesting case study since it has been reported in [54] that the model based segmentation approach for stiction detection of Stenman *et al.* [157] detected that there is no stiction in this loop. The controller and process outputs, $u(t)$ and $y(t)$, for the two loops are shown in Figure 3.2 (a) and (b). Hammerstein based approaches with both one parameter and two parameter stiction models are applied to each loop. The method using the one parameter stiction model provide estimated values of $\hat{\eta} = \hat{d} = 0$ (no stiction detected) for level loop and $\hat{\eta} = \hat{d} = 2.16$ (stiction detected) for the flow loop. Similar results were obtained by using a two parameter model [146]. For the level loop, the value of the two parameters are $\hat{\eta} = [\hat{S}, \hat{J}] = [0, 0]$ (indicating no stiction) while for the flow loop the values are $\hat{\eta} = [\hat{S}, \hat{J}] = [2, 0.5]$ (indicating presence of stiction).

The Discrete Fourier Transform of controller and process outputs for each loop are provided in Figure 3.2 (c) and (d). For the level loop there is no significant amplitude difference at high frequencies between the process and controller outputs. In other words, no extra amplitude information at frequencies beyond the controller cutoff frequency ω_c is found. However, for the flow loop data, there is a significant amplitude in process output at frequencies where the controller output magnitude is near zero. In this case, it is strongly suspected that $\omega_G > \omega_C$. For these industrial datasets, the plant and controller models are unknown so that extra information will

have to be analyzed differently than for the simulation examples. A threshold value of 1% of the maximum power value in the power spectrum of the process output is selected. Figure 3.2 (e) and (f) show the power of process output ($|Y(\omega)|^2$) at frequencies where the power spectrum of controller output ($|U(\omega)|^2$) is less than this threshold. From these plots it is evident that the necessary conditions for detection of the stiction nonlinearity by Hammerstein methods is not present in the integrating level loop, while for the flow loop, stiction detection is successful because of excitation that is present in the process output but absent in the controller signal.

3.6 Conclusions and Future Work

A frequency domain analysis of closed loop systems is performed to identify success and failure conditions for Hammerstein based stiction detection approaches in linear systems. In this work we demonstrate that under certain conditions stiction detection could be compromised even when there are no inherent limitations imposed on the identified process and stiction models *vis a vis* the true system. This important result could be used in the development of reliability measures for stiction detection algorithms. Simulation studies were used to illustrate the theoretical results. Case studies were presented to demonstrate the applicability of the proposed theoretical arguments in the industrial environment. Future work should also be directed towards theoretically studying the effect of both the process and stiction model mismatches on the Hammerstein model based stiction detection algorithms.

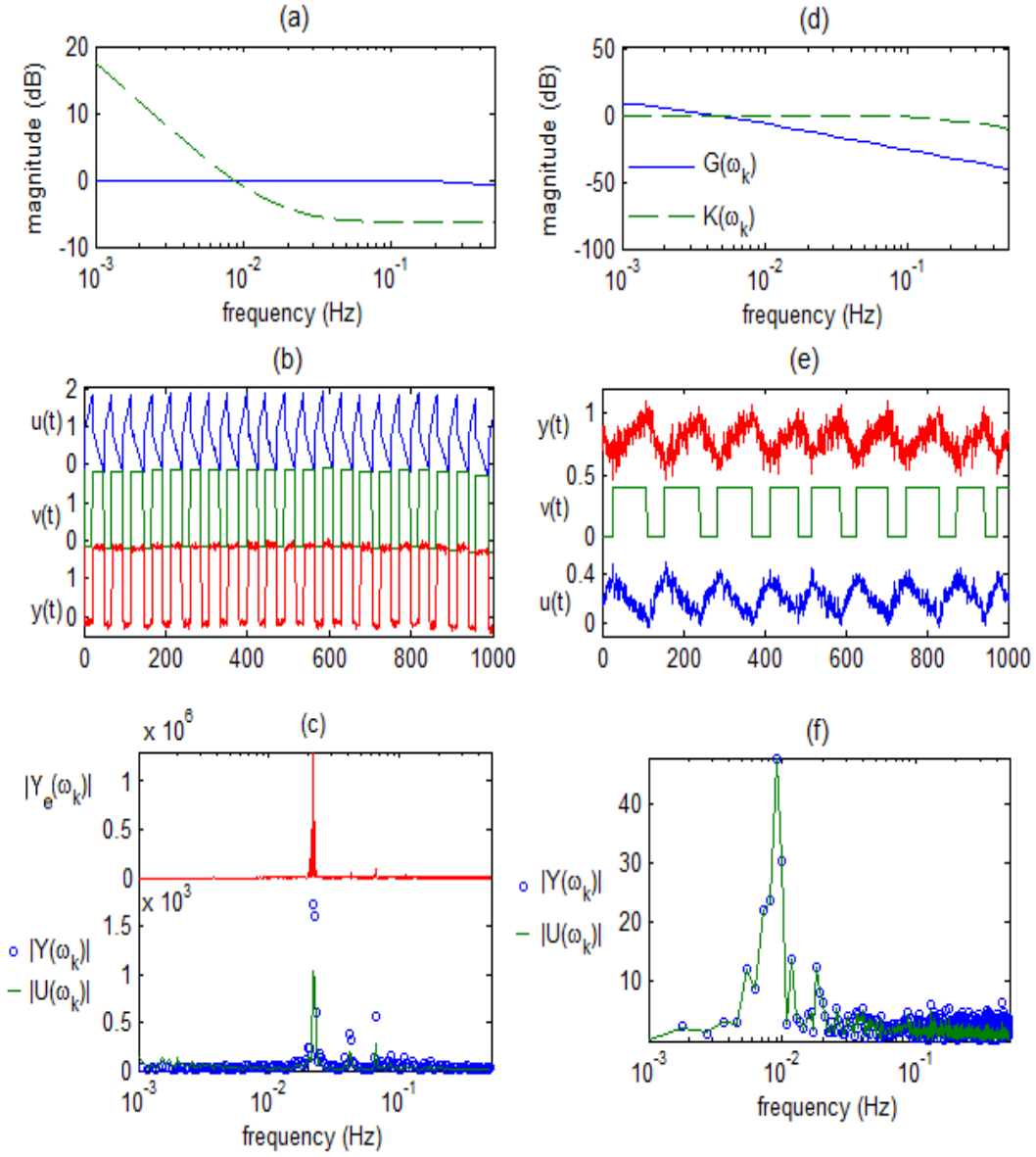

Figure 3.1: Simulation Ex. 1: (a) Bode magnitude plots for plant $G(\omega_k)$ and controller $K(\omega_k)$ for first the system. (b) Process output $y(t)$, controller output $u(t)$, and valve position $v(t)$ for the first system (c) DFT magnitudes of signals $u(t)$ and $y(t)$ and extra information $|Y_e(\omega_k)|$ for the first system. (d) Bode magnitude plots for plant $G(\omega_k)$ and controller $K(\omega_k)$ for the second system. (e) Process output $y(t)$, controller output $u(t)$, and valve position $v(t)$ for the second system (f) DFT magnitudes of signals $u(t)$ and $y(t)$ for the second system.

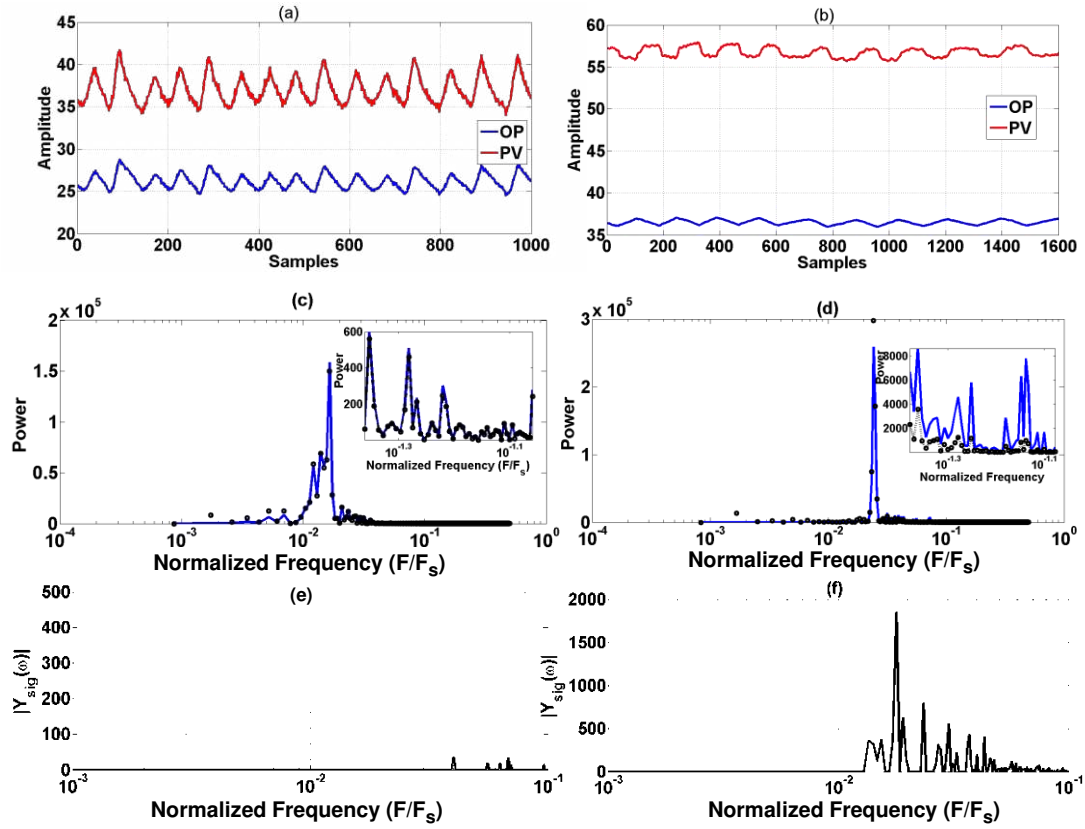

Figure 3.2: Simulation Ex. 2: (a) Process output $y(t)$ and controller output $u(t)$ from level loop (b) Process output $y(t)$ and controller output $u(t)$ from flow loop (c) DFT magnitudes of level loop signals $u(t)$ and $y(t)$ (d) DFT magnitudes of flow loop signals $u(t)$ and $y(t)$ (e) Significant power values in process output not found in controller signal for the level loop (f) Significant power values in process output not found in controller signal for the flow loop

Nomenclature

DFT	Discrete Fourier Transform
$ARMAX$	Autoregressivemoving-average model
$E(\omega_k)$	discrete fourier transform of signal $e(t)$
$G(q)$	a linear transfer function model
$G(\omega_k)$	the frequency response of $G(q)$
T	sampling time
$V(\omega_k)$	discrete fourier transform of signal $v(t)$
$e(t)$	sequence of independent samples from a white noise sequence
q	the forward time shift operator
t	sampling instants
$v(t)$	actual valve position
$\hat{v}(t)$	estimate valve position
$u(t)$	the valve position in the absence of stiction
$y(t)$	output measurement sequence
$\hat{y}(t)$	estimated output measurement sequence
$\mathbb{E}$	the expectation operator
Υ	the nonlinear and possibly discontinuous operator
$\epsilon_r(t)$	model output error
η	stiction parameter
$\hat{\eta}$	an estimate of η
$\xi(\omega_k)$	Discrete Fourier Transforms of signals $\epsilon_r(t)$
$\mathcal{M}$	set of models containing $G(q)$

CHAPTER 4

DETECTION OF STICTION IN INTERACTING SYSTEMS USING A HAMMERSTEIN MODEL APPROACH

Automated non-invasive diagnosis and localization of the root cause of oscillations in process plants is a widely held industry goal sought in order to stabilize product qualities and reduce equipment problems and energy costs. As stiction in control valves is one of the leading causes of oscillations in plant variables, detection and localization of valve stiction is a major part of any method for root cause diagnosis. Previous contributions have introduced a number of techniques that seek to achieve this objective, but most of the approaches presented are only applicable for the case of single-input single-output (SISO) processes. The current work seeks to extend one widely used approach, Hammerstein-model-based stiction detection, to the case of interacting plants. Because of the difficult nature of the problem, we first consider the case where a nominal linear plant model is known for the interacting system in question. With these approximate linear dynamics, we introduce an approach to identify in which loop the valve stiction is originating from, under the assumption that only a single valve in the interacting system is afflicted by stiction. The method efficacy is explored using simulation studies. The feasibility of extending the method to the case of unknown plant model via multivariate time-series

identification of the linear plant model is then briefly discussed. This chapter is excerpted from the conference paper: "T. Spinner, B. Srinivasan, R. Rengaswamy. Stiction Detection in Interacting Systems. *ADCONIP 2014*, Hiroshima, Japan".

4.1 Introduction

In the process industries, oscillations in the values of process variables can have immense economic costs due to equipment wear and increased product variability. Previous industrial surveys have indicated that a large percentage of control loops are affected by oscillations [158, 1]. A major cause for these oscillations is valve stiction (stiction being a term meaning static friction). The control valve is a weak link in the control loop, as it is often the only moving component in many processes [159]. Stiction induced oscillations caused by one control valve may propagate from control loop to control loop until a large portion of the entire plant is affected. Since large industrial processes may have hundreds to thousands of control loops, reliable and noninvasive automated stiction detection would serve as a valuable resource in order to locate the original source of oscillations so that the problem can be eliminated.

With this goal in mind, many valve stiction detection algorithms have appeared in the literature in recent years. Huang *et al.*[160] suggested that these methods could be classified into (i) descriptive statistic, (ii) pattern recognition, and (iii) model-based approaches. Alternatively, Babji *et al.* [73] have suggested to classify

stiction detection approaches into the categories of (i) shape-based, (ii) frequency-domain based, and (iii) model-based. In each case, the model-based methods referred to are those based on representing the process as a Hammerstein model. This technique was first introduced by [48], and many variants have been introduced since [49, 50, 51, 52, 53].

With few exceptions, the available techniques are focused on single-input single-output (SISO) processes. However, many industrial control loops are interacting, wherein the controlled or manipulated variables from one loop have direct impact on the controlled variables of one or more other loops. In this case, the nonlinear input-output behavior induced by stiction can appear in several loops simultaneously. Choudhury *et al.* [57] provided several methods for confirmation of valve stiction, two of which should be applicable to interacting systems, and these are (1) putting the controller for the suspected loop with stiction into manual control and see if the stiction induced limit cycles die down or (2) use valve positioner data and compare to the op (controller output) data to check for stiction. A third method presented in the same work was to change the controller gain and observe whether the oscillatory plant signals exhibited a change in frequency, with a change indicating the presence of valve stiction. Haoli *et al.* [56] have demonstrated that for interacting systems, this gain change method could fail to give the correct indication. A change in any controller gain within the interacting system could cause a shift in the oscillation frequency of the plant signals. Therefore, they proposed a modification

of this technique, wherein for an interacting system having several valves where the presence of stiction has been detected but not isolated to a particular valve, every controller gain in the interacting system was changed from its original setting, in order to see in which loop the controller gain change caused the largest magnitude change in oscillation frequency. The valve within the loop whose controller gain change caused the largest frequency shift is deemed to be the source of the stiction induced oscillations within the system.

All of the aforementioned methods that should correctly isolate valve stiction within interacting systems are fairly intrusive (unless a valve positioner was already installed). For instance, the modification of the gain change method presented in [56] requires 2 significant changes in gain for every control loop within the interacting system. Putting the loop in manual to complete the diagnosis is disruptive and possible unsafe [57], so this should be done as rarely as possible. An alternative to these methods is to use knowledge of the process topology to create a plant adjacency matrix, which can then be used to isolate the source of oscillation based upon the additional knowledge of which combination of process loops are exhibiting oscillations at the same frequency [161]. This method was designed in order to isolate the source of any type of oscillation whether due to poor tuning, external disturbance, or equipment malfunction or degradation, valve stiction included. However, in some instances, it is possible that the necessary topology information is not fully available or else possible that even with this information, the specific

fault and location cannot be resolved completely. Therefore, other techniques for resolving the location of valve stiction in interacting systems should still be useful.

In this work we will explore the localization of stiction detection on interacting systems using a Hammerstein-model-based detection approach wherein the linear plant model is assumed to be approximately known, and the nonlinear stiction element uses a one-parameter valve model during data-fitting. In the Section 2, the detection approach is described. In Section 3 several simulation examples are used to demonstrate the efficacy of the method as well as to highlight possible pitfalls. Finally, Section 4 contains discussion of the approach and conclusions are provided.

4.2 Stiction Detection Approach

This section introduces a stiction detection method for systems with internal interactions between control loops. First, some data-based valve stiction models are introduced which are useful for simulation and detection of stiction. Following this, a Hammerstein model-based stiction detection approach is proposed.

4.2.1 Valve Stiction Modeling

There are two types of valve stiction models, these being (i) physics-based models and (ii) data-driven models. In practical applications, a large number of the parameters required for physics based modelling have unknown value, and so data-driven models are used within stiction detection techniques. The phase plot pro-

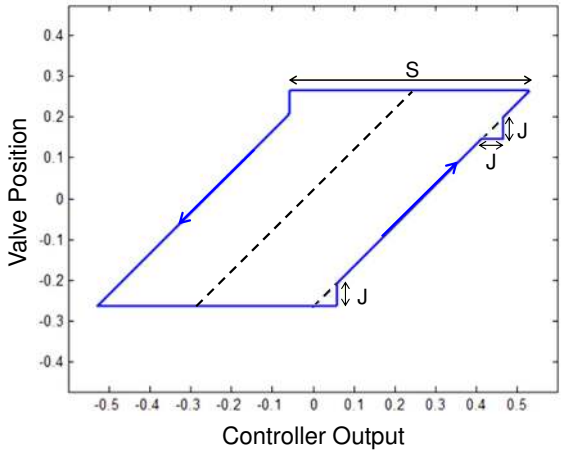

Figure 4.1: Idealized phase plot for a valve with stiction

duced by the data driven models of [162] and [144] is displayed in Fig. 4.1. These authors used two parameters to characterize a valve's stiction behavior, combined deadband and stickband, denoted S, and slip-jump J. A study by Garcia [163] comparing many types of physics based and data-driven valve models found that the model of [144] gave physically realistic input-output behavior under the simulations performed, and therefore this model is selected as the simulation model used in later sections.

For stiction detection however, a simpler data-driven model is used, this being the one-parameter model, which was used along with the original Hammerstein model-based stiction detection technique [48]. Here valve position $v(t)$ is related to its previous value $v(t-1)$ and the controller output $u(t)$ by the following simple expression,

$$v(t) = \begin{cases} u(t), & \text{if } |u(t) - v(t-1)| > d \\ v(t-1), & \text{otherwise} \end{cases}$$

where d is the parameter describing the magnitude of stiction present in the valve.

4.2.2 Hammerstein Model for Interacting Systems

A Hammerstein model is a block oriented model in which a nonlinear system is broken into two parts: a block containing a static nonlinearity, followed by a block containing linear dynamics. Several of the previous proposed methods for stiction detection rely upon this Hammerstein structure for stiction quantification, with the nonlinear data-driven stiction model contained in the first block, followed by a linear plant model. Then an iterative search is undertaken for identifying the parameters from each the linear and nonlinear blocks, with an outer loop for selecting the valve stiction parameters and an inner loop for identifying the linear dynamics.

The previously introduced methods pertained to SISO systems, which we now propose to extend to the case of interacting multiple-input multiple-output (MIMO) systems. Fig. 4.2 displays the assumed block form of the system under the Hammerstein assumption for the case of a 2×2 MIMO system. Controller outputs u_1 and u_2 enter the nonlinear element and are modified by data-driven valve models V_1

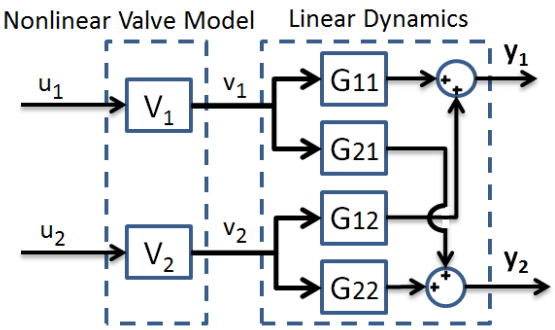

Figure 4.2: Example Hammerstein stiction model for a 2×2 system

and V_2 into valve positions v_1 and v_2. The linear dynamics then transform the valve position into measured outputs y_1 and y_2. Under the assumption of only one valve having stiction at a time, either V_1 or V_2 will be a nonlinear transformation, while the other simply acts as a pass-through of the controller signal.

Previous works used several types of models to identify the linear dynamics, including ARMAX (auto-regressive moving average with exogenous input) ([48], [53], [52]), extended ARMAX [51], and low-order model with time delay [50]. Possible multivariate extensions of these models include VARMAX (vector-ARMAX) or matrix transfer function models. In this work, we assume that a VARX (vector auto-regressive) approximation of the continuous-time multivariate system dynamics is available. Using the same one parameter valve stiction model as [48], we propose the following stiction location procedure for multivariate interacting systems:

1. Obtain a positive detection of stiction within the interacting system using one

of the previously existing techniques.

2. Assume the valve in Loop i has stiction and the other valves do not. Perform steps 2-6 for each $i = 1 \ldots n$ where n is the number of control loops in the system considered.

3. Perform a grid search over estimated stiction parameter $\hat{d}_i$ from 0 to $d_{i,max}$, where $d_{i,max}$ is defined $u_{i,max} - u_{i,min}$, while holding ($\hat{d}_j = 0$, $j \neq i$).

4. At each value of $\hat{d}_i$, transform u_i to linear plant input v_i using the one parameter valve model. Each of the other inputs u_j, $j \neq i$, enters the plant unaltered ($v_j = u_j$) since currently no stiction is assumed in the other valves.

5. Transform the plant inputs by the approximate VARX model to obtain plant output estimates $\hat{y}_k$, $k = 1 \ldots n$. Calculate the mse (mean-squared error) for each plant output, which is $mse(\hat{y}_k) = (y_k - \hat{y}_k)^2$ where y_k is the measurement signal of plant output k.

6. Calculate the MSE index, defined by $I_{mse}(\hat{d}_1, \ldots, \hat{d}_n) = \sum_{k=1}^{n} \frac{mse(\hat{y}_k)}{var(y_k)}$ for the current set $\hat{d}_i > 0$, $\hat{d}_j = 0$ ($\forall j \neq i$), where $var(y_k)$ is the variance of plant output signal k.

7. After looping through steps 2-6 for $i = 1 \ldots n$, select the valve most likely to contain stiction and the estimated severity based on the lowest value of $I_{mse}(\hat{d}_1, \ldots, \hat{d}_n)$. Since the search space consisted of having of only one nonzero

stiction parameter $\hat{d}_i$ at a time, the minimum will correspond to stiction in a single valve.

4.3 Simulation Results

To test the efficacy of the proposed method, simulation studies were carried out using Matlab and Simulink. For an example 2×2 MIMO system, we considered the distillation column model of [164]. The original continuous time transfer function model is used to simulate the plant, and two PI controllers were added. To simulate valve stiction, the model of [144] was used because of favorable characteristics of the model discussed in [163]. Within one valve model, the parameter set (S, J) was set to nonzero values to simulate stiction, while for the other loop, the valve parameters remained at $(S, J) = (0, 0)$. Simulations were run for 1000 seconds with sampling at a frequency of 1 Hz. In each case, a set-point change was used to excite the loop and no other external disturbances or nonlinearities were added.

Fig. 4.3 shows the controller and process outputs when the parameter sets $(S_1, J_1) = (0.6, 0.06)$ and $(S_2, J_2) = (0, 0)$ are used. Stiction induced oscillations are present in all plant signals, even though the valve in loop 2 had no stiction simulated. If considering the control loops separately, loop 1 takes on the classic appearance of a nonintegrating process under PI control that has stiction, with triangular wave controller output (u_1) and approximately rectangular wave process output (y_1). However, the effects of the valve nonlinearity will also appear in loop 2, and it can be

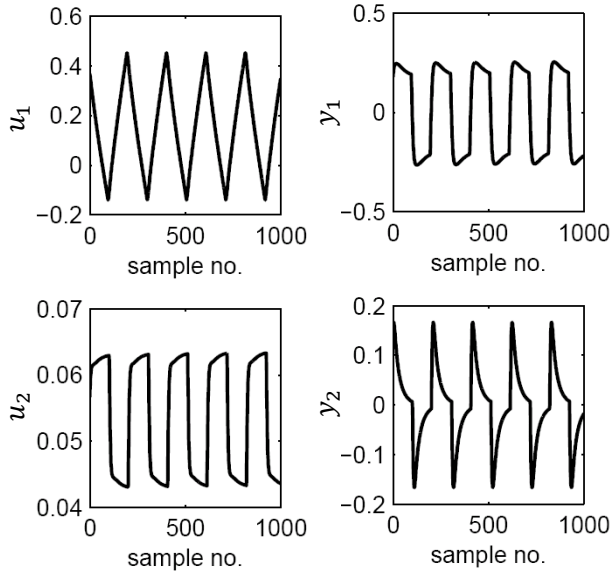

Figure 4.3: Simulated controller and process outputs for the Wood and Berry system under the case $(S_1, J_1) = (0.6, 0.06)$

shown that using SISO Hammerstein stiction detection on this loop will result in positive detection even though stiction is absent from the second valve.

An approximate linear model of the process was generated by taking a zero-order hold discretization of the original continuous time transfer function model to produce a discrete transfer function model, and then converting this into VARX form. Using the controller output data from Fig. 4.3, and assuming stiction sequentially in one loop at a time, predicted outputs $\hat{y}_1$ and $\hat{y}_2$ were generated for a range of stiction parameters on each valve.

Fig. 4.4 shows the mean-square error for each output with stiction assumed on each valve sequentially (one valve with stiction, the other without) during detection.

The pointed lines on each plot represent the mean square error (mse) resulting from assuming stiction in the first valve and no stiction in the second, while the circled lines provide the mse when stiction is assumed to be present only in the second valve. On the plots, each estimated stiction parameter ($\hat{d}_1$ in the case of the pointed line, $\hat{d}_2$ in the case of the circled line) is scaled by the span of the corresponding op signal so that the results can be presented together. For both estimated outputs $\hat{y}_1$ and $\hat{y}_2$, the minimum mean squared error is achieved when $\hat{d}_1 > 0.6$ ($\hat{d}_2 = 0$) (on the plots $\hat{d}_1 = 0.6$ corresponds to approximately 0.9 when scaled by span of the op). Since the minimum value of mse is obtained for each output when assuming stiction in valve 1 (pointed line), the method is correctly indicating stiction in this valve.

In this case, the results from each plot in Fig. 4.4 agree, and so the plot of I_{mse} in Fig. 4.5 reflects this, having a minimum value for $\hat{d}_1 \geq 0.6$ (≥ 0.9 on the plot after scaling by d_{max}) and $\hat{d}_2 = 0$. Therefore, it is concluded that stiction is correctly detected in the valve within loop 1 for this case. Interestingly, the indicated values of the stiction parameter ($\hat{d}_1 \geq 0.6$) correspond to the case where the valve is completely immobile for the duration of the predicted series. This result is probably due to different valve models being used during simulation and detection.

The simulation was again repeated, this time with stiction simulated in the valve in loop 2 with stiction parameter magnitudes of $(S_2, J_2) = (0.05, 0.05)$ ($(S_1, J_1) =$

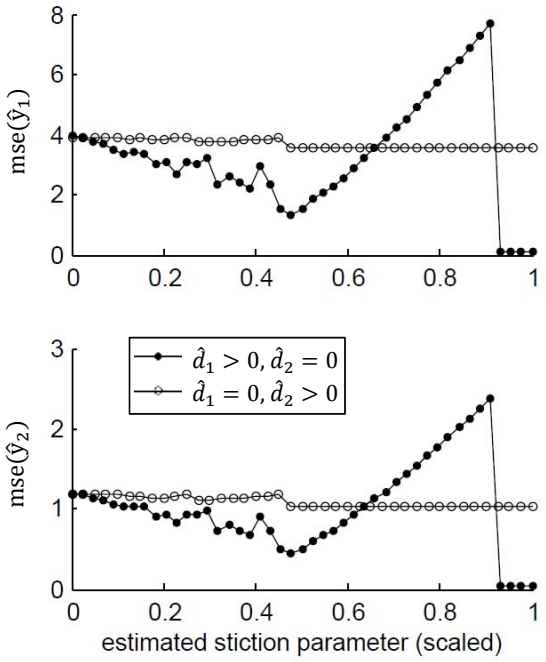

Figure 4.4: MSE of each predicted output for the Wood and Berry system under the case $(S_1, J_1) = (0.6, 0.06)$

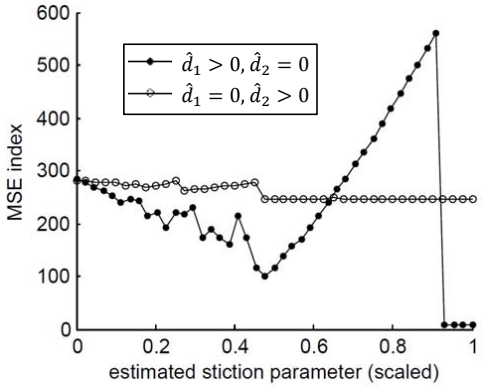

Figure 4.5: MSE index computed for the Wood and Berry system under the case $(S_1, J_1) = (0.6, 0.06)$

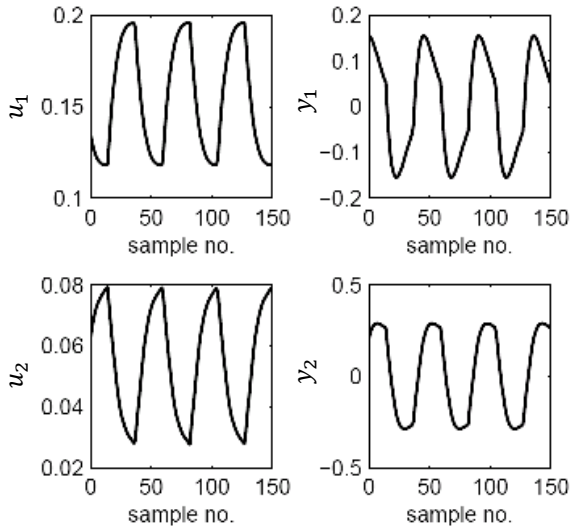

Figure 4.6: Simulated controller and process outputs for the Wood and Berry system under the case $(S_2, J_2) = (0.05, 0.05)$

$(0, 0)$). This is a special case where the deadband and stickband are equal, which provides input-output behavior quite similar to the one-parameter model used during identification. The data from this simulation is presented in Fig. 4.6, wherein each plant signal is vaguely sinusoidal. Again, the stiction detection proceeded according the method of the previous section. The mse for each output is plotted separately in Fig. 4.7. For each output, minimum mse was obtained for the parameter set ($\hat{d}_1 = 0$, $\hat{d}_2 = 0.048$) (on the plots, this value of $\hat{d}_2$ is scaled to approximately 0.37). Here, as could be expected in this special case, the result achieved gives $\hat{d}_2 \approx S = J$. The agreement of the results between each of the plots in Fig. 4.7 is also reflected in the MSE index in Fig. 4.8, wherein the same parameter set is identified.

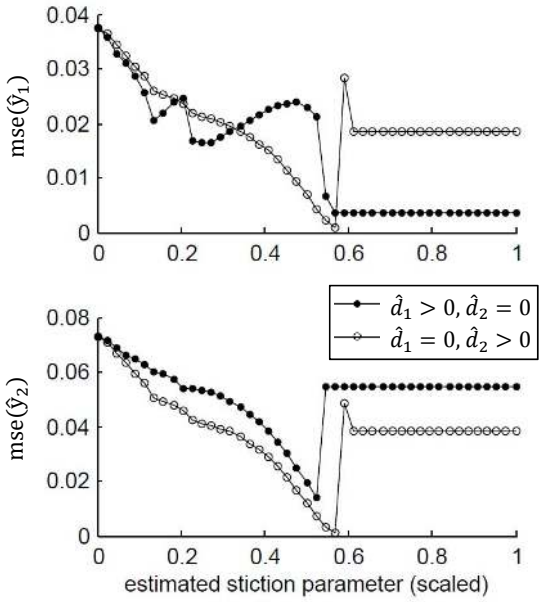

Figure 4.7: MSE of each predicted output for the Wood and Berry system under the case $(S_2, J_2) = (0.05, 0.05)$

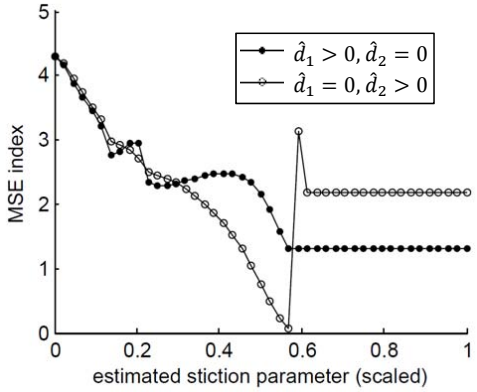

Figure 4.8: MSE index computed for the Wood and Berry system under the case $(S_2, J_2) = (0.05, 0.05)$

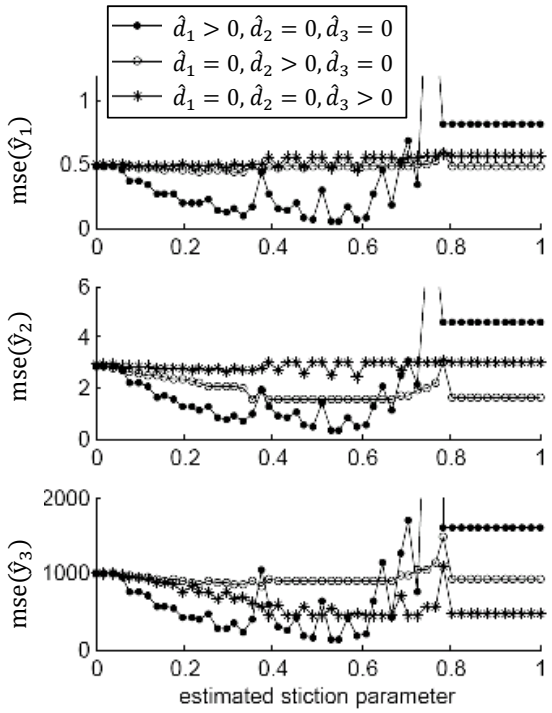

Figure 4.9: MSE of each predicted output for the Ogunnaike and Ray system under the case $(S_1, J_1) = (0.06, 0.04)$

A final simulation example uses a 3×3 MIMO transfer function simulation model obtained from [165]. Valve stiction was simulated in loop 1 using the same two parameter model as before and parameters ($S_1 = 0.06$, $J_1 = 0.04$). The simulated output was sampled at 1 Hz for 1200 seconds. Again, stiction detection occurs using the approach of the previous section, with an approximate model generated by taking a zero-order hold of the original transfer function model. Fig. 4.9 presents results from attempting to identify stiction on each valve. For each output, the lowest mse is obtained by assuming $(\hat{d}_1, \hat{d}_2, \hat{d}_3) = (0.54, 0, 0)$ (scaled values).

The MSE index was calculated, with the results in Fig. 4.10. A correct lo-

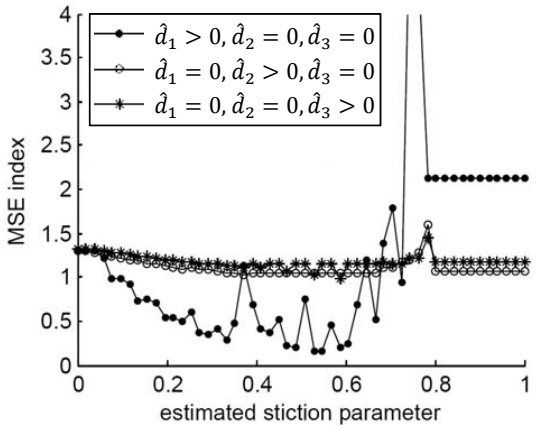

Figure 4.10: MSE index computed for the Ogunnaike and Ray system under the case $(S_1, J_1) = (0.06, 0.04)$

calization to the valve in loop 1 is provided with the minimum I_{MSE} reached at $(\hat{d}_1, \hat{d}_2, \hat{d}_3) = (0.54, 0, 0)$ (scaled by span of op). The MSE index is useful for selection of a final detection in cases where the results for each output considered separately do not agree, although this case was not found in any of the simulation results of this section.

4.4 Discussion and Conclusions

The results of the previous section indicate that performing localization of stiction detection in interacting multivariate systems using approximately known model information shows promise. The results of each detection were correct, with appreciable margin between correct and erroneous detection. These relatively convincing results in the case where an approximately known model is available, lead us to the

next question, which is the ability of this method to work when a model is not available and would have to be identified, in which case model structure (order) and time delays would have to be estimated too. The additional task of estimating multivariate linear dynamics from the closed-loop data will have unknown effect on the reliability of the method. The reliability could decrease if the parameter estimates converge far from true values, but the use of simultaneous linear model identification could also provide the flexibility to reduce errors due to plant or disturbance changes compared to the case of using previously determined dynamics. One possible way to decrease error may be to use a valve model for detection which is more able to capture the input-output behavior of the valve generating the data (in this study, the detection valve model is a one-parameter and the simulation model was a 2 parameter type). For practical applications, it would be necessary to know which valve model provides the most similar input-output characteristics as the sticky valve in the closed loop. There is some disagreement in the literature in this regard, with recent work [166] suggesting that a suitable valve model should have the valve reaching a stationary state during each sampling period, which is in contrast to the assumption in many previous works wherein the valve was assumed to stay in motion across sampling periods.

Extensive studies on industrial data-sets [72] have shown that, at best, the currently existing automated data-based techniques should be relied on as a screening mechanism in order to identify which subset of control valves should be selected

for more invasive tests in order to confirm valve stiction. This conclusion is because of the limited accuracy of the non-invasive techniques studied. This is still a valuable role for this type of technique to serve as it can greatly reduce the time and effort of plant personnel in locating the source of oscillations. In this role, it is essential that a stiction detection technique should have a low false negative rate, to avoid eliminating the sticky valve from consideration. Of course, reducing the number of false positives is also beneficial as it will reduce the number of invasive tests necessary in order to locate the valve with stiction that is the root cause of the oscillations. In a similar way, the currently proposed technique is not meant to provide a definite answer as to which valve contains stiction, but it can provide the most probable location for plant personnel to begin more invasive testing.

Future work can proceed in several directions, such as dropping the assumption that a nominal plant model is available for use. In the SISO case, it is known that the presence of stiction provides sufficient excitation for the closed-loop identification of plant models [48]. For the MIMO case, it would be necessary to fit a multivariate linear model, such as a VARX, VARMAX, or matrix transfer function model. Even in the SISO case, fitting linear dynamics is a computationally intensive procedure, which must be repeated at a large sampling of stiction parameters, so it is unknown if the multivariate case is feasible to be completed in a timely manner. A final challenge is industrial validation of the proposed approach, which will require obtaining data sets from the correct type of process (interacting) with

enough additional information (the condition of each control valve in the system, knowledge of disturbances, controller tunings, plant models) to properly test the diagnosis technique.

Nomenclature

d_i, $\hat{d}_i$	stiction parameter for valve i, and its estimate
I_{mse}	MSE index
J_i slip-jump parameter for valve i	
S_i	stickband parameter for valve i
u_i	controller output, loop i
v_i	valve position, loop i
V_i	valve transformation, loop i
y_i, $\hat{y}_i$	output variable of loop i, and its estimate

MIMO	multiple-input multiple-output
MSE	mean squared error
SISO	single-input single-output

CHAPTER 5

ECONOMIC PERFORMANCE ASSESSMENT OF MULTIVARIATE PROCESSES WITH CONSTRAINTS

Plant economic performance is most often related to the mean values of mean values of process variables; meanwhile most performance assessment techniques involve examining the variance and covariance of output variables. The combined approach is to trade-off between variances of different process variables in order to move the plant as close to the optimal economic operating point as operating constraints will allow. This problem is referred to as the minimum backed-off point operating point selection, and previous works have formulated it as a non-convex constrained optimization problem. The current work uses linearization of the relationship between the standard deviations of input and output variables in order to allow the problem to be solved as a small number of linear programs. This new approach is illustrated using simulation examples.

5.1 Introduction

Automated closed loop monitoring and diagnosis techniques for the process industries seek to aid plant personnel identify poorly performing control loops, find

root causes for problems, and perform remedial action. Identification of which control loops need attention is often achieved by comparing the current control loop performance against a historical, designed, or theoretical performance benchmark. Harris introduced the minimum variance index for single-input single-output (SISO) systems, with this measure comparing the variance of the measured variable against the benchmark of the theoretical achievable minimum variance [6]. Following the introduction of the minimum variance index, many other performance assessment techniques based on the variance and/or covariance of the output variable(s) have appeared in the literature, with techniques developed for both for SISO and interacting systems. The usefulness of these methods is often based on the assumption that reducing product variances will allow the operating point to be moved closer to the process constraints, and that this action will bring increased profits [167, 43]. While this assumption holds true for many cases, proper benchmarking of economic performance requires considering operating constraints along with price and cost information in addition to plant and disturbance model information.

Many performance assessment methods for interacting systems rely on a combination of open-loop plant information and closed-loop operating data. These include techniques developed by Huang *et al.* [168] and Harris [15] which are based on the use of the system interactor matrix. Ko and Edgar created a method capable of producing similar results using matrices of open-loop plant Markov coeffi-

cients [17]. Ko and Edgar also proposed techniques for determining the minimum variance benchmark for constrained systems [17], which requires a full plant and disturbance models. Each of the methods of Haung *et al.*, Harris, and Ko and Edgar determines a lower bound on a weighted combination of output variances. It has been frequently stated that minimum variance control of the outputs is not a common goal in industry, partly because this leads to excessive movement of manipulated variables. To simultaneously consider input and output variances, several techniques have been proposed, and these rely on full knowledge of the open-loop plant. Huang *et al.* proposed using the LQG (linear-quadratic Gaussian) tradeoff curve between a function of the input variances and a function of the output variances [23].

Another way to look at the input and output variances, providing a clearer connection to plant economics, is as limitations on the choice of the plant operating point. The standard deviations of plant variables determines how close that the actual plant operating point can come to the optimal operating point that lies upon the boundary formed by operating constraints. Finding the optimal backed-off point is called the minimum backed-off operating point selection [44, 45]. This usually involves transforming the original operating constraints on input and output values into a probabilistic form, wherein the constraint on each variable is written in terms of the standard deviation of that variable. Then the choice of operating point along with the input and output variance of each variable is selected within an

optimization problem, with a (usually linear) objective function representing plant economics.

The work of Chmielewski and Manthanwar [169] along with that of Peng *et al.* [46] formulated this as a problem with a linear objective function with convex and reverse convex constraints. They required the use of a branch and bound solution technique to solve for the optimal operating point. Nabil *et al.* formulated the problem in a similar manner, and used an iterative solution technique to deal with the non-convexity [66]. All of the aforementioned works use a state space model of the plant. Zhao *et al.* considered a more restricted problem in which not all of the input and output variances were free to vary [170]. Their formulation was valid for any model type provided that the relation between input and output standard deviations could be provided.

The current work provides economic performance assessment and operating point selection for a class of multivariate systems. It uses a multivariate finite impulse response model of the plant and disturbances. A collection of Pareto optimal standard deviations is then generated. A piecewise linearization of the Pareto optimal space is then used to generate linear inequality constraints and linear relationships for the standard deviations. The optimal economic operating performance can then be determined by solving a small sequence of linear programs. Section 2 contains the formulation of the operating point and variance tradeoff problem to be solved. Section 3 contains the proposed method for solving this problem. Section

4 contains a simulation example to demonstrate the proposed approach. Finally, conclusions and ideas for future extensions are provided.

5.2 Preliminaries

This chapter will rely on several techniques from the field of machine learning including k-means clustering, principle component analysis, and support vector machines. Therefore, this section provides a brief overview of each of these three methods.

5.2.1 k-Means Clustering

Clustering is the task of grouping a collection of unlabeled (unclassified) data points into several distinct categories. Clustering techniques are widely applied across the fields of machine learning, data-mining, and pattern recognition. One popular method to achieve the desired data-partitioning is the use of k-means clustering, which is also variously referred to as the isodata algorithm or Lloyd's algorithm [130], which divides the data into k sets, S_i, $i = 1 \dots k$, each with centroid m_i. For every data point $x^j \in \mathbb{R}^N$, $j = 1 \dots n$ (N is the dimension of the data, n is the number of data points), through assignment to a particular set S_i, k-means aims to minimize the within cluster sum of squares (WCSS), defined as the squared Euclidean distance of the point to centroid of its assigned cluster. The algorithm comprises iterating a two-step procedure which consist of:

1. Set membership assignment based on minimum WCSS,

$$S_i = \{x^j \| x^j - m_i\|^2 \leq \|x^j - m_p\|^2, \forall p = 1 \ldots k\}. \tag{5.1}$$

2. Recalculation of the centroid of each set,

$$m_i = \frac{1}{|S_i|} \sum_{x^j \in S_i} x^j, i = 1 \ldots k \tag{5.2}$$

where $|S_i|$ is the number of elements assigned to cluster i.

In Matlab, the seeding of the initial centroid positions is performed randomly by default [171]. The algorithm has converged to a local minimum when the classification of each individual point does not change between consecutive iterations. Many variants of the k-means algorithm exist [172], including those which seek to provide a global minimum WCSS to every data point.

5.2.2 Support Vector Machines

Within machine learning applications, one common objective is to separate data from two or more sets through the creation of one or more decision boundaries. Support vector machines (SVMs) are one popular and powerful tool for achieving this goal. In order to create the decision boundary between two types of data, first SVM must be provided with labelled training data from the two classes. If it is

possible that a single hyperplane is able to separate the training data from the two separate classes, then the data is referred to as linearly separable. If an SVM is applied in this case, the resulting decision hyperplane (boundary) will maximize its distance from each set, being placed in what can be intuitively thought of as the middle of margin between the two sets of training data [130]. The closest points to the decision boundary are referred to as the support vectors.

Often times, it is impossible to separate the training data by a single hyperplane boundary, and therefore the data is only nonlinearly separable. Nonlinear separation can possibly be achieved through the combination of support vector machines with the so-called *kernel trick* in order to nonlinearly transform the the training data [173]. However, it is commonly desired to still use a linear boundary, and instead minimize some measure of classification error. In this case, soft-margin SVM is the preferred approach, wherein the distance from the decision boundary of each incorrectly classified point is added to the objective function to be minimized [174]. In the current work, linear soft-margin support vector machines are used to provide boundaries on a set of linear surfaces.

5.2.3 Principle Component Analysis

Principle component analysis (PCA) is a statistical technique that is used for both identifying patterns in data and dimension-reduction of data. In the pattern-recognition framework, the data considered is a set of feature vectors. A feature

vector can be described as either an N-dimensional set of features or as a data point in N-dimensional space.

To begin the PCA procedure [175, 176], the data are first placed in an $N \times n$ matrix, X, where the rows correspond to different features and the columns correspond to different data points (with each point corresponding to, for example, a separate observation of a process or a different object in a collection, depending on the problem in question). The mean across each row is collected in $N \times 1$ vector μ, which is subtracted from each column to provide mean-subtracted matrix $\bar{X}$. The covariance matrix, C, of the data contained in X, is then provided by

$$C = \frac{1}{n-1}\bar{X}\bar{X}^T.$$

Assuming C is positive definite, its eigendecomposition can be written

$$C = VDV^T$$

where D is an $N \times N$ diagonal matrix whose entries are the eigenvalues of C, ordered from largest to smallest, while V is an $N \times N$ orthogonal matrix whose rows are the eigenvectors of C, ordered according to their corresponding eigenvalues. The largest eigenvalue corresponds to the direction of maximum variation in the original data. If the eigenvalues are scaled so that they sum to unity, the largest scaled

eigenvalue is the fraction of the total variation captured by the first eigenvector, the value of the second eigenvalue is the additional variation captured by the second eigenvector, and so forth. The direction of maximum variation is referred to the first principle component, the direction capturing the maximum amount of residual variation is called the second principle component, and so on.

During data reduction, components of less significance are ignored and thereby the final data set will have lesser dimension. For example, to reduce the feature dimension from N dimensions to L dimensions, the first L columns of the eigenvectors matrix V are used to form $N \times L$ matrix W. To project a given $N \times 1$ feature-vector or data-point, x, to L dimensions, the formula

$$\tilde{x} = W^T(x - \mu)$$

where $L \times 1$ vector $\tilde{x}$ is the projection of x onto the first L eigenvectors of C. To obtain the point corresponding to $\tilde{x}$ in the original N-dimensional space, the relation

$$\hat{x} = W\tilde{x} + \mu = WW^T(x - \mu) + \mu$$

is used. The point $\hat{x}$ lies on the hyperplane formed by the first L eigenvectors of C.

For a given value of the reduced dimension, L, PCA provides the linear projection giving the lowest mean-squared-error compared to the original data set. There-

fore, PCA can be used to provide the hyperplane of a given dimension that provides the best-fit linear relationship between a set of data points. In the current work, PCA is used to simultaneously lower the data dimension and provide piecewise linearization of the set of data.

5.3 Problem Formulation

In this section, an optimization problem is formulated for choosing the operating point for a certain class of constrained linear systems subjected to stochastic disturbances. A key component of the method, which has also been applied in previous works, is the use of a probabilistic formulation using both hard and soft constraints, a soft constraint referring to one that is satisfied at some confidence level. The formulation uses the assumption of a normal distribution of the process variables. For example, given the pair of lower and upper bound soft constraint vectors ($\underline{y}$, $\bar{y}$) on plant output y, if the constraints are permitted to be violated at up to 5% of the time, the constraint can be written as $\underline{y} + z_y\sigma_y \leq y^* \leq \bar{y} - z_y\sigma_y$ where $z_y = 1.96$ is the standard deviation multiplier, y^* is the mean of y, and σ_y is the standard deviation of y. Standard deviation multiplier z_y can be adjusted based on the level of tolerance for constraint violations on each variable. In the case of plant inputs u, the upper and lower bound constraint vectors $\underline{y}$ and $\bar{y}$ are often hard constraints due to physical limitations, such as the opening range on a valve. In this case, a similar inequality can be written, $\underline{u} + z_u\sigma_u \leq u^* \leq \bar{u} - z_u\sigma_u$, where the standard deviation multiplier of

the inputs is set to $z_u = 3$. This will cause the controller to respect the constraint at least 99.7% of the time. Then the occurrences of actuator saturation should be rare, and have very little effect on process performance.

The multivariate model form assumed for this work is that of a discrete finite impulse response model

$$(y_t - y_o) = H(z^{-1})(u_t - u_o) + N(z^{-1})a_t \tag{5.3}$$

where y_t is the vector of output values at time step t, u_t is the vector of input values, a_t is the vector of random disturbance impulses from a multivariate normal distribution, (y_o, u_o) is the nominal operating point around which the linear model is valid, z^{-1} is the backshift operator, $H(z^{-1}) = \sum_{i=1}^{N} H_i z^{-i}$ is the finite impulse model for the plant, and $N(z^{-1}) = \sum_{i=1}^{N} N_i z^{-i}$ is the finite impulse model of the disturbance filter, with N the order of the finite impulse response models.

Assuming a linear dependence of plant output revenues and plant input costs on the output and input values, then the plant economic optimization should proceed

according to:

$$
\begin{aligned}
max \qquad J &= c_y y^* - c_u u^* && (eqn.5.4a)\\
sub\,ject\,to: \quad y^* &= y_o + \Delta y && (eqn.5.4b)\\
u^* &= u_o + \Delta u && (eqn.5.4c)\\
\Delta y &= K\Delta u && (eqn.5.4d)\\
\underline{y} + z_y\sigma_y &\leq y^* \leq \bar{y} - z_y\sigma_y && (eqn.5.4e)\\
\underline{u} + z_u\sigma_u &\leq u^* \leq \bar{u} - z_u\sigma_u && (eqn.5.4f)\\
(y_t - y_o) &= H(z^{-1})(u_t - u_o) + N(z^{-1})a_t && (eqn.5.4g)
\end{aligned}
\tag{5.4}
$$

where (y^*, u^*) is the chosen steady-state operating point, Δy and Δu are vectors of the deviations of the chosen operating point from the nominal point (y_o, u_o), K is the steady state gain matrix, c_y is the economic benefit coefficient vector, and c_u is the cost coefficient vector. The following section details how this problem can be solved as a series of linear programs.

5.4 Solution Method

The problem of Eqn. 5.4 above is not yet in a solvable format. This section details how a relationship between the entries in the vector of output standard deviations σ_y and the entries in the vector of standard deviations of the manipulated variables, σ_u, can be generated. Following this, a piecewise linear relationship between the entries in these vectors can be created by a novel method. This allows the

problem of Eqn. 5.4 to be solved by solving a small set of linear programs.

5.4.1 Generating Pareto Optimal Standard Deviations

The objective function J of Eqn 5.4 is linear and each of the constraints 2a-2f is linear, however, in Eqn 5.4, it is not explicit how the set of output and input variances (σ_y, σ_u) are to be generated. A formula included within the work of Ko & Edgar [17] allows for the generation of optimal variances σ_y^2 and σ_u^2 for a given weighting of the each individual input and output variable, in the case that the process model is of the form of Eqn 5.4g. Therefore, this method can be used to generate a collection of points $(\sigma_y, \sigma_u)_i$ where each point i corresponds to a different set of weighting coefficients.

Before proceeding with the method of Ko & Edgar, one notable modification is that the current technique seeks to simultaneously minimize a weighted sums of variances of both inputs and outputs, instead of only output variances as in the previous method. To proceed, the plant and disturbance models must be augmented with the plant inputs treated as additional outputs. The augmented model takes the form

$$\widetilde{y_t} = \begin{bmatrix} (y_t - y_o) \\ (u_t - u_o) \end{bmatrix} = \begin{bmatrix} H(z^{-1}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} (u_t - u_o) + \begin{bmatrix} N(z^{-1}) \\ \mathbf{0} \end{bmatrix} a_t = \widetilde{H}(z^{-1})(u_t - u_o) + \widetilde{N}(z^{-1}) a_t \tag{5.5}$$

where $\widetilde{H}(z^{-1})$ is the augmented plant filter, and $\widetilde{N}(z^{-1})$ is the augmented disturbance filter.

Now returning to the method of Ko & Edgar, from the MIMO plant filter impulse coefficients $\widetilde{H}_i$ that make up plant model $\widetilde{H}(z^{-1})$, a dynamic matrix can be defined:

$$G_{N+1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \widetilde{H}_1 & 0 & 0 & 0 \\ \vdots & \ddots & 0 & 0 \\ \widetilde{H}_N & \cdots & \widetilde{H}_1 & 0 \\ 0 & \widetilde{H}_N & \cdots & \widetilde{H}_1 \\ 0 & 0 & \widetilde{H}_N & \cdots \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & \widetilde{H}_N \end{bmatrix} \tag{5.6}$$

Also, matrix $D_{N+1} = \begin{bmatrix} \tilde{N}_0 \\ \vdots \\ \tilde{N}_N \end{bmatrix}$ can be constructed from disturbance filter impulse coefficients $\tilde{N}_i$ describing the total disturbance effect resulting from a single white noise vector impulse a_0. Therefore, the effect of a_0 on augmented output $\tilde{y}_t$ can be written

$$\begin{bmatrix} \tilde{y}_0 \\ \vdots \\ \tilde{y}_N \end{bmatrix} = G_{N+1} \begin{bmatrix} (u_0 - u_0) \\ \vdots \\ (u_N - u_0) \end{bmatrix} + D_{N+1} a_0 \tag{5.7}$$

Suppose it is desired to achieve the minimum of a weighted sum of input and output variances $\sigma_Y^2 = \sum_{i=1}^{p} w_{y,i}\sigma_{y,i}^2 + \sum_{j=1}^{n} w_{u,j}\sigma_{u,j}^2$ through proper manipulations of $(u_t - u_o)$. First construct weighting matrix $Q = diag\left(\begin{bmatrix} w_{y,1} & \dots & w_{y,p} & w_{u,1} & \dots & w_{u,n} \end{bmatrix}\right)$ and also Q_{N+1}, which contains $N+1$ repetitions of Q on its block diagnonal. The augmented closed-loop output corresponding to minimum σ_Y^2 is given by $(\tilde{y}_t)_{opt} = \sum_{i=0}^{2N} Y_i a_{t-i}$ where Y_i are the filter coefficients calculated from the least squares projection:

$$\begin{bmatrix} Y_0 \\ \vdots \\ Y_N \end{bmatrix} = [\boldsymbol{I} - G_{N+1}(G_{N+1}^T Q_{N+1} G_{N+1})^{+} G_{N+1}^T Q_{N+1}] D_{N+1} \tag{5.8}$$

The output covariance matrix $\Sigma_{\tilde{y}}$ corresponding to the minimized σ_Y^2 is computed through $\Sigma_{\tilde{y}} = \sum_{i=0}^{2N} Y_i \Sigma_a Y_i^T$. This final result is due to the statistical independence of any terms within white-noise sequence a_t, which provides $cov(a_{t-i}, a_{t-j}) = \mathbf{0}$ [23]. Finally, the vectors σ_y and σ_u can be obtained by taking the square root of the diagonal elements of $\Sigma_{\tilde{y}}$. To create the space of Pareto optimal (σ_y, σ_u), the elements of weighting matrix Q should each be varied from 10^{-4} to 10^4.

5.4.2 Linearization of (σ_y, σ_u) Space

The Pareto optimal standard deviations will lie within a space over $(n_u + n_y)$ dimensions, where n_u is the number of inputs and n_y the number of outputs. To create a piecewise linear approximation of this set, the first requirement is to segregate the original set into a finite number of subsets. Then for the second step, on each subset, a linear approximation is performed. The methods chosen for these steps in this work are (i) k-means clustering and (ii) principle component analysis (PCA). Since k-means clustering has a random component, steps (i) and (ii) are repeated in parallel 5-10 times in order to find a segregation and projection providing minimum total projection error.

The hyperplanes resulting from PCA will not be bounded. As a first step in bounding these planes, it is proposed to segregate the subsets determined by k-means via planes determined by support vector machines. Further bounding can be performed by fitting a bounding polytope [177]. In the 2-input 2-output case, a

simple procedure is demonstrated in the next section. Finally, from the linearization, the constraint Eqn 5.4g is turned into a pair of matrix equality and inequality constraints that describe each approximating hyperplane and its bounds, such as

$$\begin{aligned} A^i \begin{bmatrix} \sigma_y \\ \sigma_u \end{bmatrix} &\leq b^i \\ & (eqn.2g') \\ C^i \begin{bmatrix} \sigma_y \\ \sigma_u \end{bmatrix} &= d^i \end{aligned} \tag{5.9}$$

where A^i, b^i, C^i, d^i are coefficient matrices corresponding to the i^{th} linearized cluster, $i = 1 \ldots k$, where k is the number of clusters used. An overview of the proposed piecewise linearization procedure is given in Fig. 5.1. A detailed example, illustrating how this procedure is used follows in the next section.

5.5 Simulation Example

The system studied is a discrete approximation of the Wood & Berry column model [164] having plant model

$$T(z) = \begin{bmatrix} \frac{0.744z^{-2}}{1-0.9419z^{-1}} & \frac{-0.8789z^{-4}}{1-0.9535z^{-1}} \\ \frac{-1.302z^{-4}}{1-0.9329z^{-1}} & \frac{-0.5786z^{-8}}{1-0.9123z^{-1}} \end{bmatrix} \tag{5.10}$$

with steady state gain, $K = \begin{bmatrix} 12.8 & -18.9 \\ 6.60 & -19.4 \end{bmatrix}$, output noise filter

$$N(z) = \begin{bmatrix} \frac{0.09}{(1-0.7z^{-1})(1-0.7z^{-1})} & \frac{0.2z^{-1}}{1-0.8z^{-1}} \\ \frac{-0.2z^{-1}}{1-0.8z^{-1}} & \frac{-0.09}{(1-0.7z^{-1})(1-0.7z^{-1})} \end{bmatrix} \tag{5.11}$$

and noise covariance $\Sigma_a = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$. The proposed method requires transformation of the plant and disturbance models, and so each discrete transfer function was approximated by a finite impulse response model of length $N = 100$.

The nominal operating point and input and output constraints were:

$$\begin{bmatrix} y_{1,o} \\ y_{2,o} \\ u_{1,o} \\ u_{2,o} \end{bmatrix} = \begin{bmatrix} 90 \\ 90 \\ 55 \\ 55 \end{bmatrix} \quad \left(\underline{y}, \bar{y}\right) = \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 100 \\ 100 \end{bmatrix}\right) \left(\underline{u}, \bar{u}\right) = \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 60 \\ 60 \end{bmatrix}\right).$$

The objective function was:

$$\max_{\Delta y, \Delta u, \sigma_u, \sigma_y} J = c_y y^* - c_u u^* = \begin{bmatrix} 5 & 5 \end{bmatrix} \begin{bmatrix} y_1^* \\ y_2^* \end{bmatrix} - \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} u_1^* \\ u_2^* \end{bmatrix}$$

The Ko & Edgar method is used to minimize the weighted sum: $w_{y,1}\sigma_{y,1}^2 + w_{y,2}\sigma_{y,2}^2 + w_{u,1}\sigma_{u,1}^2 + w_{u,2}\sigma_{u,2}^2$, providing the Pareto optimal standard deviations for

each particular set of chosen weights, where each of the weights $w_{y,1}, w_{y,2}, w_{u,1}, w_{u,2}$ were varied in the range 10^{-4} to 10^{4}. This resulted in about 35000 points in $(\sigma_{u,1}, \sigma_{u,2}, \sigma_{y,1}, \sigma_{y,2})$ space. Two subspaces of the resulting 4-dimensional space are shown in Fig. 5.2.

In order to reduce the space of Pareto optimal standard deviations to a piecewise linear form, the procedure of Fig. 5.1 is used. Data is clustered using k-means (using $k = 5$), then each cluster is reduced to 2-dimensions using principle component analysis (PCA). This was repeated on the original data-set 10 times, since the k-means has a random component, and it was desired to minimize the final residual of the projected data. For the final set of selected clusters, the first 2 covariance eigenvectors determined via PCA captured about 96 % of the data variance. The data was projected onto the PCA-determined planes associated with each cluster, and then boundaries separating the clusters were determined using support vector machines. Each cluster had $(k - 1) = 4$ boundaries determined via use of support vector machines (SVM) to separate it from the other data clusters. The Pareto optimal standard deviation points divided into the five clusters, as well as the lower dimension planes determined by PCA, are shown in Fig. 5.5. Each plane has already been restricted by the SVM determined boundaries. However, it is clear that each plane extends significantly outside of the region that the data covers.

To eliminate the extensions outside the region of the data, each cluster of data is projected into 2-dimensional space via the previously determined PCA eigenvec-

tors, and within this space, simple bounding lines are fit. Then the 2-d boundary lines are transformed back into planes in 3-dimensional space using the inverse of the PCA projection. Fig. 5.5 shows boundary lines in the projected 2-dimensional space for an example projected data cluster. Two lines are extended from each the left-most, right-most, bottom-most, and top-most projected data points, creating a polytope containing all of the projected points.

Through the previous steps, the original $(\sigma_{u,1}, \sigma_{u,2}, \sigma_{y,1}, \sigma_{y,2})$ space could then be approximated in the form of five piecewise linear and bounded spaces. Each space was described by a best-fit plane formed via PCA, boundaries separating the space from the other data clusters (determined through SVM), and the additional linear boundaries created to contain the space. The plane and boundaries were added to the optimization problem as linear constraints, and with the objective and other constraints already in linear form, the problem could be easily solved using linear programming. Therefore, the optimization was carried out once for every piecewise linear space, then the results for each were compared to pick a final optimal operating point. The best case was found to be

$$\begin{bmatrix} y_1^* \\ y_2^* \\ u_1^* \\ u_2^* \end{bmatrix} = \begin{bmatrix} 94.3 \\ 92.6 \\ 55.3 \\ 55.0 \end{bmatrix} \quad \& \quad \begin{bmatrix} \sigma_{y,1} \\ \sigma_{y,2} \\ \sigma_{u,1} \\ \sigma_{u,2} \end{bmatrix} = \begin{bmatrix} 2.84 \\ 3.72 \\ 1.57 \\ 1.30 \end{bmatrix} \tag{5.12}$$

Now, this second matrix of standard deviations refers to a point that exists on the linear piecewise approximation of the space of Pareto optimal standard deviations, and it might be questioned whether such a point is really obtainable. To provide validity to the result, the data set generated by the Ko & Edgar method was searched for a similar point, and a very close one was found, $(\sigma_{y,1}, \sigma_{y,2}, \sigma_{u,1}, \sigma_{u,2}) = (2.83, 3.73, 1.56, 1.30)$. From this estimate of the optimal operating point, the current economic performance can be assessed by a ratio measure showing percent of economic potential that has been lost due to the current choice of operating point, or $\gamma = \frac{J_{opt} - J_{current}}{J_{opt}} \times 100$. If we considered the initial operating point (u_o, y_o) versus the optimal point for instance, we would get $\gamma = \frac{613.8 - 580.0}{613.8} \times 100 = 5.8\%$ lost profit relative to the optimal case.

Considering the case traditionally referred to as "minimum variance" or "singular LQ" where the weights $w_{y,1} = w_{y,2} = 1$ and $w_{u,1} = w_{u,2} = 0$, the standard deviations take on the values $\begin{bmatrix} \sigma_{y,1} \\ \sigma_{y,2} \\ \sigma_{u,1} \\ \sigma_{u,2} \end{bmatrix} = \begin{bmatrix} 2.53 \\ 3.72 \\ 3.79 \\ 2.09 \end{bmatrix}$. Comparing this result to the optimal case above shows that the system is close to minimum variance on variable y_2 in the optimal case. Because of the problem formulation (initial operating point and constraints), there is apparently no operating point where the minimum variance relation can hold. This is an important result showing the benefit of the current

analysis. The traditional benchmark of multivariate minimum variance is unachievable by this system due to constraints. Since some limiting constraints refer to physical limits such as the amount a valve can open, they must be considered in order to provide an accurate benchmark for system performance. Additionally, even if minimum variance were achievable, it might not provide the optimal economic operation.

The piecewise linearization component is the most complex component of the proposed performance assessment procedure. Although the procedure is currently automated, using the user-input of plant and disturbance models to generate a piecewise linearized set of spaces to represent the space of Pareto optimal standard deviations, it is also encouraged for the user to examine the quality of the piecewise linear approximation. If the original space is too skewed by the linearization procedure, it is suggested to perhaps adjust the number of iterations that the clustering/PCA loop runs or else change the number of piecewise linear spaces used in the procedure. Continuing work is examining the use of generalized PCA in order to streamline the process of approximating the optimal space with piecewise hyperplanes.

5.6 Conclusion

A novel approach for economic performance assessment of constrained multivariate systems has been proposed. The proposed method requires plant, disturbance, and constraint information, but returns an estimate of both the optimal

achievable economic performance as well as the minimum backed-off operating point that can provide this performance level.

Future work will explore the use of generalized principle component analysis to streamline the linearization and dimension reduction of the Pareto optimal standard deviations. Also, analytical approaches towards the piecewise linearization of the relationship between Pareto optimal standard deviations will also be attempted.

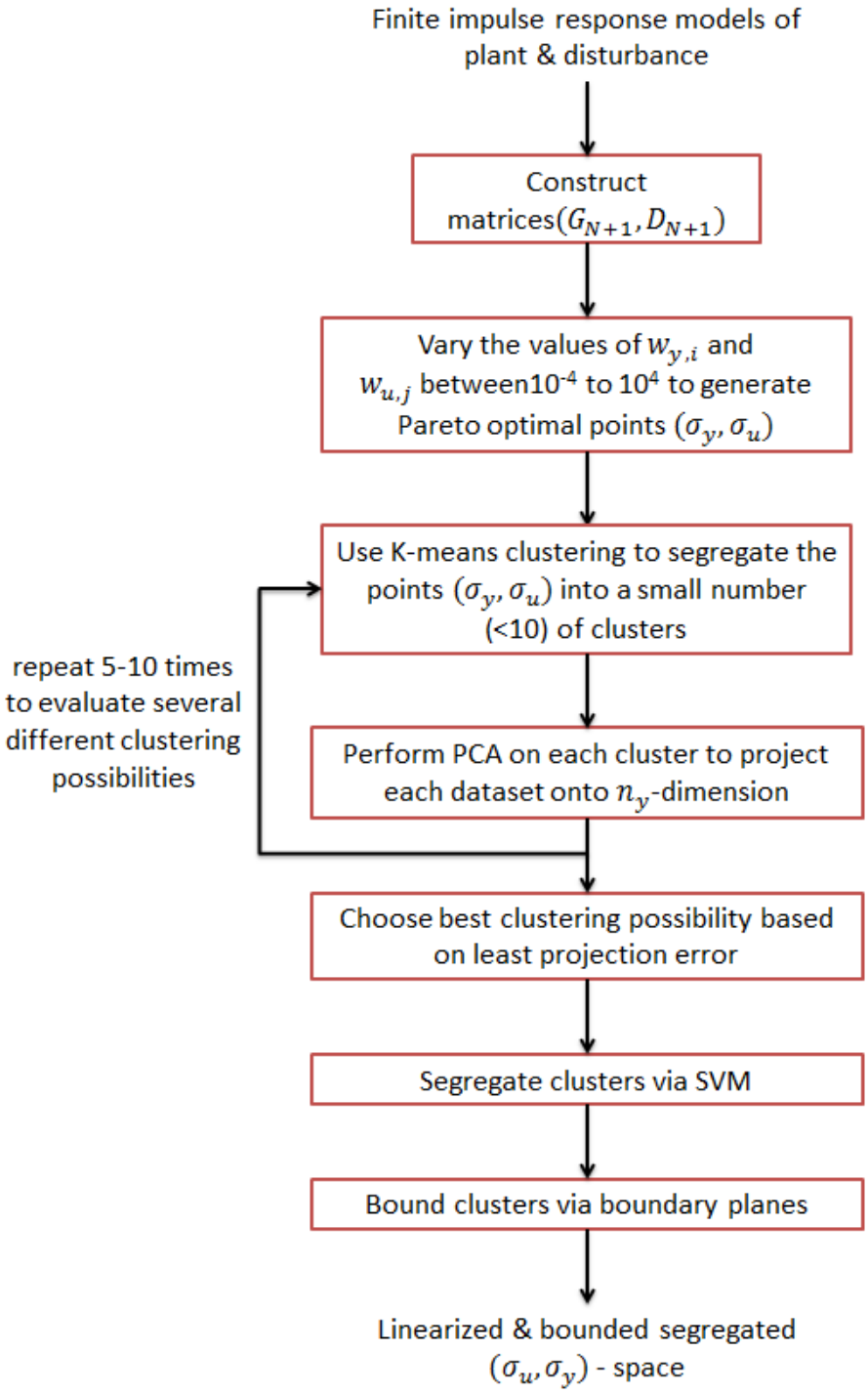

Figure 5.1: Pareto space linearization procedure.

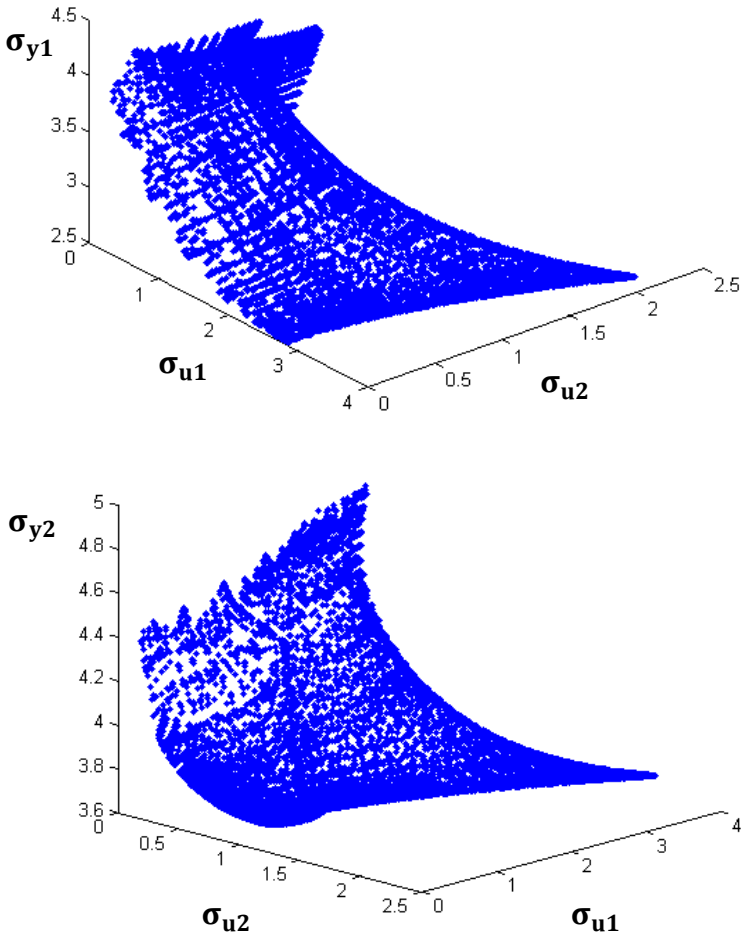

Figure 5.2: Subspaces of the $(\sigma_{u,1}, \sigma_{u,2}, \sigma_{y,1}, \sigma_{y,2})$ space.

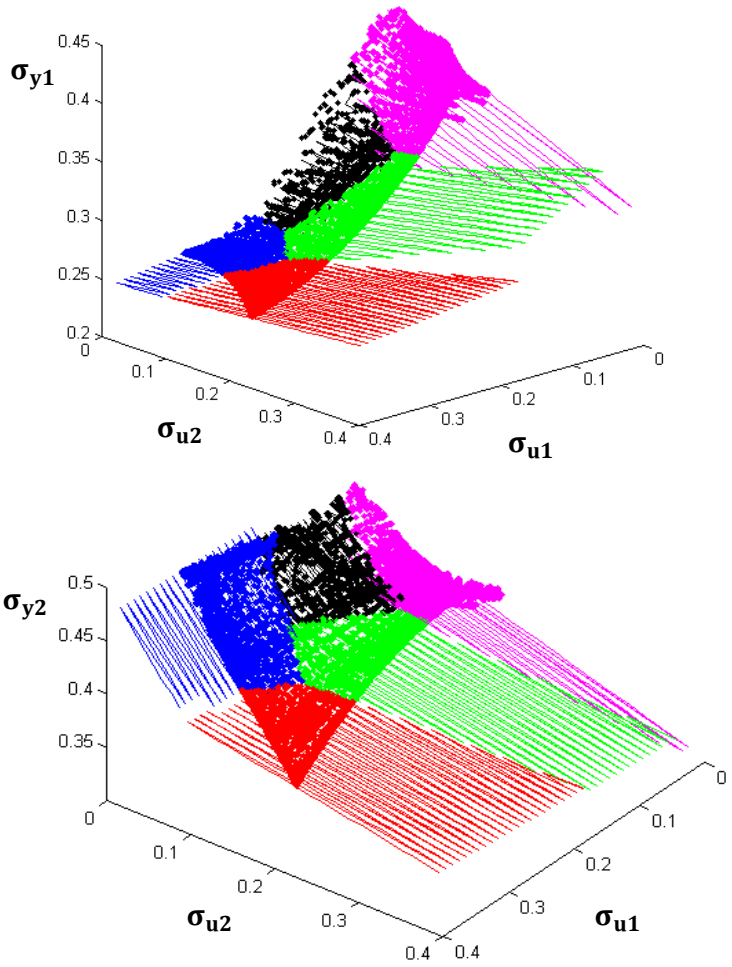

Figure 5.3: Subspaces of the Pareto optimal data showing clusters and the PCA-determined planes associated with each cluster.

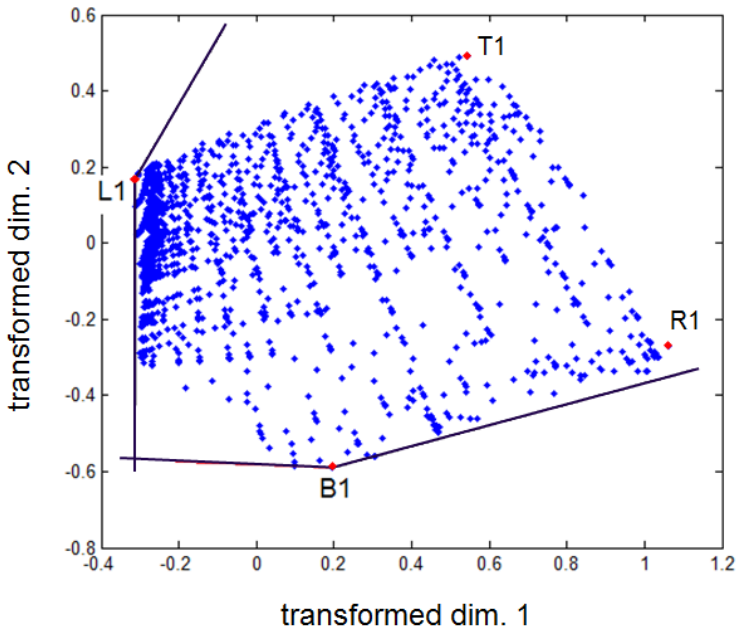

Figure 5.4: Boundary lines through the left-most (L1) and bottom-most (B1) points. Not shown are boundaries created using the top-most (T1) and right-most (R1) points.

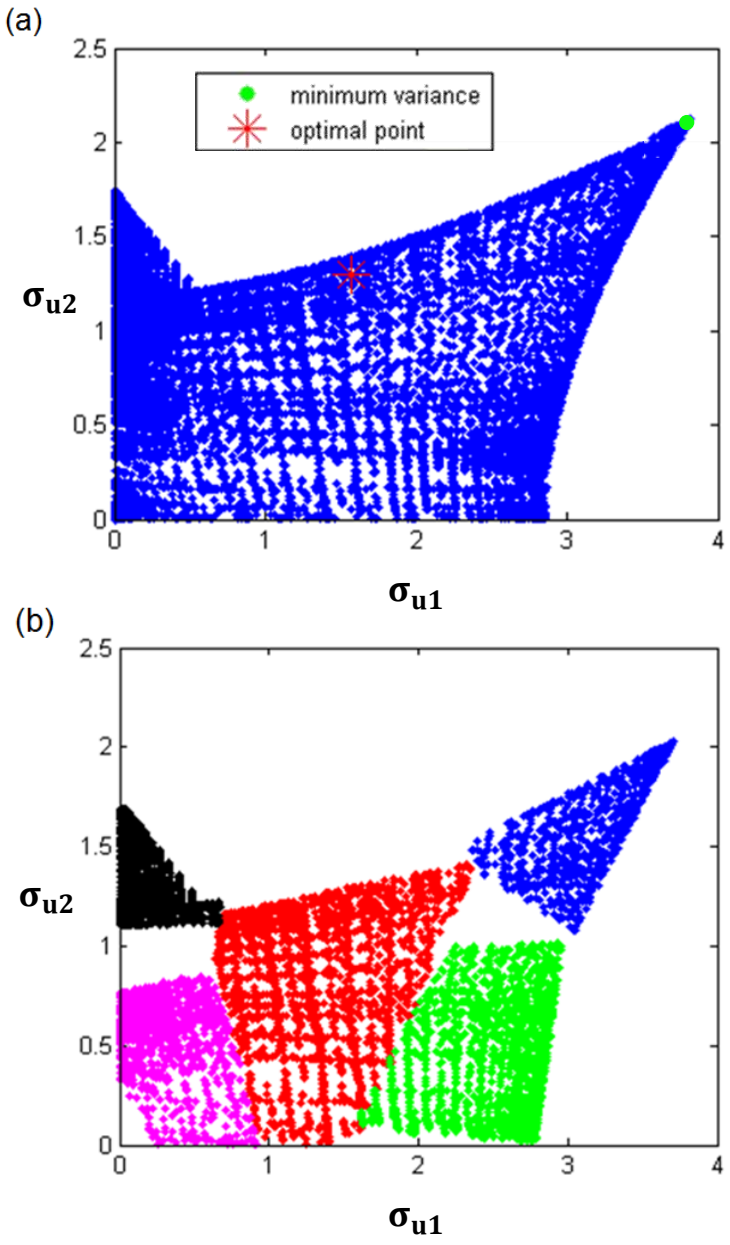

Figure 5.5: (a) $(\sigma_{u,1}, \sigma_{u,2})$ coordinates of the Pareto optimal points on the $(\sigma_{u,1}, \sigma_{u,2}, \sigma_{y,1}, \sigma_{y,2})$. (b) $(\sigma_{u,1}, \sigma_{u,2})$ coordinates after linearization of the previous data.

Nomenclature

a_0	single white noise vector impulse
A	coefficient matrices
C	coefficient matrices
a_t	the vector of random disturbance impulses from a multivariate normal distribution
c_u	cost coefficient vector
c_y	economic benefit coefficient vector
H	the finite impulse function
$\tilde{H}(z^{-1})$	augmented plant filter
$\tilde{H}_i$	augmented plant filter impulse coefficient
J	objective function
n_u	the number of inputs
n_y	the number of outputs
N	the order of the finite impulse response models
$\tilde{N}_i$	disturbance filter impulse coefficient
$\tilde{N}(z^{-1})$	the augmented disturbance filter
Q	weighting matrix
u	plant input
u^*	steady-state operating point of plant input
u_o	nominal input operating point
$w_{u,i}$	weight factor of input variance
$w_{y,i}$	weight factor of output variance
y	plant output
y^*	steady-state operating point of plant output
y_o	nominal output operating point
$\underline{y}$	lower bound on plant output
$\overline{y}$	upper bound on plant output
Y_i	the filter coefficients calculated from the least squares projection
z^{-1}	backshift operator
z_u	the standard deviation multiplier of the inputs
z_y	the standard deviation multiplier
y^*	the mean of the plant output y

Δu	vectors of the deviations of the chosen operating point from the nominal point u_o
Δy	vectors of the deviations of the chosen operating point from the nominal point y_o
σ_y	the standard deviation of the plant output y
σ_u	the standard deviation of the plant input u
Σ	output covariance matrix
γ	percent of potential economic potential lost due to the current choice of operating point

PCA	principle component analysis
SVM	support vector machines
SISO	single-input single-output
MIMO	multi-input multi-output
LQG	linear-quadratic Gaussian

CHAPTER 6

SUMMARY, CONTRIBUTIONS, AND FUTURE WORK

6.1 Summary

The noninvasive controller performance assessment and diagnosis techniques presented in this document offer a promising set of tools for plant personnel to identify poorly performing control loops, diagnose root causes of poor performance, and seek remedial action. Chapter 1 presents two sets of performance assessment and diagnosis trees that allow for an overall framework to be applied loop by loop in a large process plant, and the techniques of Chapters 2-5 contribute to filling in several nodes within this framework.

It is important for researchers and practitioners to be aware of the pitfalls of methods presented in the literature. In Chapter 2, it is shown that several techniques for diagnosing sluggish or aggressive tuning, techniques both present in the literature and newly introduced, often provide incorrect diagnoses. This is not unexpected; any simple classifier is bound to give erroneous results sometimes. However, it is noted that few examples of wrong classifications were presented in literature, and automated retuning algorithms have appeared in prior works that take no account of the possibility for wrong classifications. It was shown herein how such

oversights can lead to poor performance and control loop instability.

The work of Chapter 3 presents a similar theme. Hammerstein model based stiction detection sometimes gives wrong results, therefore we should seek to understand under which conditions this method fails. Srinivasan *et al.* presented a stiction diagnosis reliability measure that can be run in parallel to the stiction detection algorithm [55]. The results of Chapter 3 provide more support for the use of this measure. Chapter 4 proposed one way to bring noninvasive stiction detection to interacting systems. As this is one of the initial efforts in this area, the uncertainties and limitations that accompany the method are still waiting to be explored.

Chapters 4 and 5 attempt to contribute some new techniques to field of multivariate control system performance assessment and diagnosis. For multivariate systems, it's all about the interactions. In the case of stiction detection, interactions can cause confounding results if SISO control loop techniques are simply applied to each loop in the multivariate system. For performance assessment, interactions can either aid or inhibit loop performance compared to a SISO point of view, although it is usual for the interactions to be detrimental if trying to apply decentralized control in a loop by loop fashion. But the main suggestion for performance assessment of multivariate systems is to choose the right objective to examine. In SISO systems, it is often the case that if you can reduce the variance of the (single) output variable, then you can push the operating point closer to the constraint boundary, and higher profit may follow. As more variables and more objectives are added, the number

of tradeoffs increase. The way that the problem was handled in this work was to directly optimize the economic objective, and any other information was viewed as constraints on this goal. Many works in the performance benchmarking literature have focused on the interactor matrix, the weighted interactor, or other solutions to the singular LQ problem, but as of yet these techniques have not been strongly connected to the actual objectives of process plants, which are often economic in nature with some other requirements (operability, safety, reliability) mixed in.

6.2 Contributions

The main contributions of this work include:

1. Rigorous definitions of sluggish and aggressive tuning were provided. The use of the Hurst exponent as a tuning diagnosis measure was proposed. New frameworks for viewing the efficacy of tuning diagnosis measures and retuning algorithms were introduced. Tuning algorithms were proposed that can incorporate a variety of tuning diagnosis measures in their operation. Potential instability of data-based retuning algorithms was demonstrated.

2. Theoretical results were introduced providing the failure condition for Hammerstein model based stiction detection algorithms, providing an explanation for results previously seen in the literature. These results lend support to

a previously proposed reliability measure for model-based stiction detection algorithms.

3. A method to adapt the Hammerstein model based stiction detection algorithm to the isolation of stiction to a particular valve within interacting multivariate systems was proposed.

4. A new algorithm for economic performance assessment was proposed. Unlike existing methods, the new algorithm transforms the economic benchmarking and operating point selection problem into a linear problem, allowing for efficient solution by linear programming.

6.3 Future Work

Future work can proceed in several directions:

1. The multivariate methods proposed in this document rely heavily on open-loop plant model information. This could be a hindrance to wide-scale adoption in industry, as models are not maintained for a majority of process loops. Therefore, data-based performance assessment and diagnosis tools should be continued to be explored for multivariate systems. This is most likely not an easy problem to solve, for example:

    - Attempts to extend the Hurst Index to multivariate systems either in

a loop by loop fashion or else through the use of the detrended cross correlation coefficient has so far yielded no conclusive results.

- Data-based Hammerstein based stiction detection may produce an unwieldily large space of model structure parameters if straightforward extension to multivariate systems is pursued.

However, the economic benefit of these control loop assessment and diagnosis techniques increases if they become easier to implement on more loops. Therefore, data-based assessment of multivariate loops should still be pursued.

2. The theoretical results on Hammerstein stiction detection presented herein are valid for single-input single-output loops only. The extension of these results to multivariate systems could be very interesting, as several sets of plant and controller dynamics will determine whether the presence of stiction is able to be isolated to a particular loop. Most importantly, it is suggested that the stiction reliability measure previously proposed for SISO systems should be extended to multivariate systems.

3. For multivariate systems, PI/PID control is commonly implemented in industry within a decentralized control framework. Very little literature on the assessment of these systems is currently available. Tuning of these systems is difficult as the change in parameters for one controller will affect all interact-

ing loops. Also, the controller designer must be concerned with robust stability in the event of sensor, actuator, or controller failures. These difficulties should not be seen as a drawback to the problem, but instead an opportunity for control loop assessment tools to provide assistance to plant personnel in increasing performance from their control systems.

Nomenclature

LQ	Linear - Quadratic
PI	Proportional-Integral
PID	Proportional-Integral-Derivative
SISO	Single-Input Single-Output

BIBLIOGRAPHY

[1] L. Desborough and R. Miller. Increasing customer value of industrial control performance monitoring-honeywell's experience. In *Proceedings of Chemical Process Control VI*, pages 172–192, 2002.

[2] D.B. Ender. Process control performance: Not as good as you think. *Control Engineering*, 40(10):180–190, 1993.

[3] W.L. Bialkowski. Dreams versus reality: a view from both sides of the gap. *Pulp & Paper Canada*, 94(11):19–27, 1993.

[4] Y. Zhang, D. Monder, and J.F. Forbes. Real-time optimization under parametric uncertainty: a probability constrained approach. *Journal of Process Control*, 12(3):373–389, 2002.

[5] M. Jelali. An overview of control performance assessment technology and industrial applications. *Control Engineering Practice*, 14(5):441–466, 2006.

[6] T. J. Harris. Assessment of control loop performance. *The Canadian Journal of Chemical Engineering*, 67(5):856–861, 1989.

[7] N. Stanfelj, T.E. Marlin, and J.F. MacGregor. Monitoring and diagnosing process control performance: the single loop case. *Industrial & Engineering Chemistry Research*, 32(2):301–314, 1993.

[8] T. Hägglund. A control-loop performance monitor. *Control Engineering Practice*, 3(11):1543–1551, 1995.

[9] N.F. Thornhill and T. Hägglund. Detection and diagnosis of oscillation in control loops. *Control Engineering Practice*, 5(10):1343–1354, 1997.

[10] L. Desborough and T. Harris. Performance assessment measures for univariate feedback control. *The Canadian Journal of Chemical Engineering*, 70(6):1186–1197, 1992.

[11] A. Ingimundarson and T. Hägglund. Closed-loop performance monitoring using loop tuning. *Journal of Process Control*, 15(2):127–133, 2005.

[12] P.-G. Eriksson and A.J. Isaksson. Some aspects of control loop performance monitoring. In *Proceedings of the Third IEEE Conference on Control Applications*, pages 1029–1034, 1994.

[13] B. Huang, S. L. Shah, and E. K. Kwok. On-line control performance monitoring of mimo processes. In *Proceedings of the American Control Conference*, volume 2, pages 1250–1254, 1995.

[14] B. Huang and S.L. Shah. The role of the unitary interactor matrix in the explicit solution of the singular lq output feedback control problem. *Automatica*, 33(11):2071–2075, 1997.

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

[16] T.J. Harris. Interpretations of multivariate performance assessment indices. *Journal of Process Control*, 19(4):701–710, 2009.

[17] B.-S. Ko and T. F. Edgar. Performance assessment of multivariable feedback control systems. *Automatica*, 37(6):899–905, 2001.

[18] B. Srinivasan, T. Spinner, and R. Rengaswamy. Control loop performance assessment using detrended fluctuation analysis (dfa). *Automatica*, 48(7):1359–1363, 2012.

[19] B. Huang, S. X. Ding, and N.F. Thornhill. Practical solutions to multivariate feedback control performance assessment problem: reduced a priori knowledge of interactor matrices. *Journal of Process Control*, 15(5):573–583, 2005.

[20] Y. Zhao, Y. Gu, H. Su, and B. Huang. Extended prediction error approach for mpc performance monitoring and industrial applications. In *Proceedings of the 17th IFAC World Congress*, pages 6–11, 2008.

[21] A. Horch and A.J. Isaksson. A modified index for control performance assessment. *Journal of Process Control*, 9(6):475–483, 1999.

[22] B. Huang and S.L. Shah. Practical issues in multivariable feedback control performance assessment. *Journal of Process Control*, 8(5):421–430, 1998.

[23] B. Huang and S. L. Shah. *Performance assessment of control loops: theory and applications*. Springer, Berlin, 1999.

[24] S.J. Qin and J. Yu. Recent developments in multivariable controller performance monitoring. *Journal of Process Control*, 17(3):221–227, 2007.

[25] J. Yu and S.J. Qin. Statistical mimo controller performance monitoring. part i: Data-driven covariance benchmark. *Journal of Process Control*, 18(3):277–296, 2008.

[26] J. Yu and S.J. Qin. Statistical mimo controller performance monitoring. part ii: Performance diagnosis. *Journal of Process Control*, 18(3):297–319, 2008.

[27] T.J. Harris and W. Yu. Analysis of multivariable controllers using degree of freedom data. *International Journal of Adaptive Control and Signal Processing*, 17(7-9):569–588, 2003.

[28] S. Gigi and A.K. Tangirala. Quantification of directed influences in multivariate systems by time-series modeling. In *Proceedings of the International Conference on Control, Automation, Communication and Energy Conservation (INCACEC)*, pages 1–6, 2009.

[29] S. Gigi and A.K. Tangirala. Quantitative analysis of directional strengths in jointly stationary linear multivariate processes. *Biological cybernetics*, 103(2):119–133, 2010.

[30] S. Gigi and A.K. Tangirala. Quantification of interaction in multiloop control systems using directed spectral decomposition. *Automatica*, 49(5):1174–1183, 2013.

[31] A. Hugo. Performance assessment of dmc controllers. In *Proceedings of the American Control Conference*, volume 4, pages 2640–2641, 1999.

[32] R.H. Julien, M.W. Foley, and W.R. Cluett. Performance assessment using a model predictive control benchmark. *Journal of Process Control*, 14(4):441–456, 2004.

[33] J. Schäfer and A. Cinar. Multivariable mpc system performance assessment, monitoring, and diagnosis. *Journal of Process Control*, 14(2):113–129, 2004.

[34] O.A.Z. Sotomayor and D. Odloak. Performance assessment of model predictive control systems. In *Proceedings of the 2006 International Symposium on Advanced Control of Chemical Processes*, volume 6, pages 2–5, 2006.

[35] B.-S. Ko and T. F. Edgar. Performance assessment of constrained model predictive control systems. *AIChE Journal*, 47(6):1363–1371, 2001.

[36] Q. Zhang and S. Li. Enhanced performance assessment of subspace model-based predictive controller with parameters tuning. *The Canadian Journal of Chemical Engineering*, 85(4):537–548, 2007.

[37] F. Loquasto III and D.E. Seborg. Model predictive controller monitoring based on pattern classification and pca. In *Proceedings of the American Control Conference*, volume 3, pages 1968–1973, 2003.

[38] F. Loquasto and D.E. Seborg. Monitoring model predictive control systems using pattern classification and neural networks. *Industrial & Engineering Chemistry Research*, 42(20):4689–4701, 2003.

[39] A.S. Badwe, R.D. Gudi, R.S. Patwardhan, S.L. Shah, and S.C. Patwardhan. Detection of model-plant mismatch in mpc applications. *Journal of Process Control*, 19(8):1305–1313, 2009.

[40] A.S. Badwe, R.S. Patwardhan, S.L. Shah, S.C. Patwardha, and R.D. Gudi. Quantifying the impact of model-plant mismatch on controller performance. *Journal of Process Control*, 20(4):408–425, 2010.

[41] S. Selvanathan and A.K. Tangirala. Diagnosis of poor control loop performance due to model-plant mismatch. *Industrial & Engineering Chemistry Research*, 49(9):4210–4229, 2010.

[42] C.A. Harrison and S.J. Qin. Discriminating between disturbance and process model mismatch in model predictive control. *Journal of Process Control*, 19(10):1610–1616, 2009.

[43] C. Pryor. Autocovariance and power spectrum analysis derive new information from process data. *Control Engineering*, 29(11):103–106, 1982.

[44] L. Narraway, J. D. Perkins, and G. Barton. Interaction between process design and process control: economic analysis of process dynamics. *Journal of Process Control*, 1(5):243–250, 1991.

[45] L. T. Narraway and J. D. Perkins. Selection of process control structure based on linear dynamic economics. *Industrial & Engineering Chemistry Research*, 32(11):2681–2692, 1993.

[46] J.-K. Peng, A. M. Manthanwar, and D. J. Chmielewski. On the tuning of predictive controllers: The minimum back-off operating point selection problem. *Industrial & Engineering Chemistry Research*, 44(20):7814–7822, 2005.

[47] M.A.A.S. Choudhury, S.L. Shah, and N.F. Thornhill. *Diagnosis of process nonlinearities and valve stiction: data driven approaches*. Springer, Berlin, 2008.

[48] R. Srinivasan, R. Rengaswamy, S. Narasimhan, and R. Miller. Control loop performance assessment. 2. hammerstein model approach for stiction diagnosis. *Industrial & Engineering Chemistry Research*, 44(17):6719–6728, 2005.

[49] M. A. A. S.. Choudhury, M. Jain, and S.L. Shah. Stiction–definition, modelling, detection and quantification. *Journal of Process Control*, 18(3):232–243, 2008.

[50] K.H. Lee, Z.Ren, and B. Huang. Novel closed-loop stiction detection and quantification method via system identification. In *Proceedings of the 4th International Symposium on Advanced control of industrial processes ADCONIP*, pages 4–7, 2008.

[51] S. Karra and M.N. Karim. Comprehensive methodology for detection and diagnosis of oscillatory control loops. *Control Engineering Practice*, 17(8):939–956, 2009.

[52] M. Jelali. Estimation of valve stiction in control loops using separable least-squares and global search algorithms. *Journal of Process Control*, 18(7):632–642, 2008.

[53] L. Zhi X. Ivan and S. Lakshminarayanan. A new unified approach to valve stiction quantification and compensation. *Industrial & Engineering Chemistry Research*, 48(7):3474–3483, 2009.

[54] A. Horch. Oscillation diagnosis in control loops stiction and other causes. In *Proceedings of the American Control Conference*, 2006.

[55] B. Srinivasan, T. Spinner, and R. Rengaswamy. A reliability measure for model based stiction detection approaches. In *Proceedings of the 8th IFAC Symposium on the Advanced Control of Industrial Processes*, pages 750–755, 2012.

[56] Yu Haoli, S Lakshminarayanan, and Vinay Kariwala. Confirmation of control valve stiction in interacting systems. *The Canadian Journal of Chemical Engineering*, 87(4):632–636, 2009.

[57] M. A. A. S. Choudhury, V. Kariwala, N.F. Thornhill, H. Douke, S.L. Shah, H. Takada, and J.F. Forbes. Detection and diagnosis of plant-wide oscillations. *The Canadian Journal of Chemical Engineering*, 85(2):208–219, 2007.

[58] B. Srinivasan. *Control Loop Performance Assessment and Diagnosis*. Ph.D. Thesis, Texas Tech University, Lubbock, TX, 2012.

[59] A.W. Ordys, D. Uduehi, and M.A. Johnson. *Process control performance assessment: from theory to implementation*. Springer, Verlag, 2007.

[60] N.F. Thornhill and B. Huang. Management and monitoring of process assets. In *Proceedings of the 4th International Symposium on Advanced control of industrial processes ADCONIP*, pages 4–7, 2008.

[61] M.J. Grimble. Controller performance benchmarking and tuning using generalised minimum variance control. *Automatica*, 38(12):2111–2119, 2002.

[62] B. S. Ko and T. Edgar. Assessment of achievable pi control performance for linear processes with dead time. In *Proceedings of the American Control Conference*, volume 3, pages 1548–1552, 1998.

[63] B.-S. Ko and T.F. Edgar. Performance assessment of cascade control loops. *AIChE Journal*, 46(2):281–291, 2000.

[64] M.J. Grimble. Restricted structure controller tuning and performance assessment. *IEE Proceedings-Control Theory and Applications*, 149(1):8–16, 2002.

[65] M. Jelali. *Control Performance Management in Industrial Automation*. Springer, London, 2013.

[66] M. Nabil, S. Narasimhan, and S. Skogestad. Economic back-off selection based on optimal multivariable controller. In *Proceedings of the 8th IFAC International Symposium on Advanced Control of Chemical Processes*, pages 792–797, 2012.

[67] T.-L. Chia and L. Irving. Expert-based adaptive control. In *Techniques for Adaptive Control*, pages 203–232. Butterworth-Heinemann, 2002.

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

[69] S.J. Qin. Control performance monitoringa review and assessment. *Computers and Chemical Engineering*, 23(2):173–186, 1998.

[70] F. Qi Y. Shardt, Y. Zhao, K. Lee, X. Yu, B. Huang, and S. Shah. Determining the state of a process control system: Current trends and future challenges. *The Canadian Journal of Chemical Engineering*, 90(2):217–245, 2012.

[71] M. A. A. S.. Choudhury, S.L. Shah, N.F. Thornhill, and D. S. Shook. Automatic detection and quantification of stiction in control valves. *Control Engineering Practice*, 14(12):1395–1412, 2006.

[72] M. Jelali and B. Huang. *Detection and diagnosis of stiction in control loops*. Springer, Berlin, 2010.

[73] S. Babji, U. Nallasivam, and R. Rengaswamy. Root cause analysis of linear closed-loop oscillatory chemical process systems. *Industrial & Engineering Chemistry Research*, 51(42):13712–13731, 2012.

[74] M. L. Tyler and M. Morari. Performance assessment for unstable and nonminimum-phase systems. In *Preprints of the IFAC workshop on on-line fault detection supervision chemical process industries*, 1995.

[75] B. S. Ko and T. F. Edgar. Pid control performance assessment: The single-loop case. *AIChE Journal*, 50(6):1211–1218, 2004.

[76] A. Y. Sendjaja and V. Kariwala. Achievable pid performance using sums of squares programming. *Journal of Process Control*, 19(6):1061–1065, 2009.

[77] A. Visioli. Method for proportional-integral controller tuning assessment. *Industrial & Engineering Chemistry Research*, 45(8):2741–2747, 2006.

[78] K. J. Åström and T. Hägglund. Pid controllers: theory, design, and tuning. *Instrument Society of America, Research Triangle Park, NC*, 1995.

[79] J. Ziegler and N. Nichols. Optimum settings for automatic controllers. *Transactions of the ASME*, 64(11), 1942.

[80] Q.-G. Wang, Z. Ye, W.-J. Cai, and C.-C. Hang. *PID control for multivariable processes*, volume 373. Springer, 2008.

[81] A. O'Dwyer. *Handbook of PI and PID controller tuning rules*, volume 2. World Scientific, 2006.

[82] T. Hägglund. Automatic detection of sluggish control loops. *Control Engineering Practice*, 7(12):1505–1511, 1999.

[83] S. Matsumura, K. Ogata, S. Fujii, and H. Shioya. Adaptive control for the steam temperature of thermal power plants. In *Proceedings of the IEEE International Conference on Control Applications*, volume 2, pages 1105–1109, 1998.

[84] K. J. Åström and B. Wittenmark. *Adaptive control*. Courier Dover Publications, 2013.

[85] R. Isermann, D. Matko, and K.-H. Lachmann. *Adaptive control systems*. Prentice-Hall, Inc., 1992.

[86] J. Seem. An improved pattern recognition adaptive controller. In *Proceedings of IFAC Workshop on Energy Saving Control in Plants and Buildings*, pages 117–122, 2006.

[87] S. Choudhury, S. L. Shah, and N. F. Thornhill. Diagnosis of poor control-loop performance using higher-order statistics. *Automatica*, 40(10):1719–1728, 2004.

[88] W. I. Caldwell. Control system with automatic response adjustment, 1950. US Patent 2,517,081.

[89] A. Leva, C. Cox, and A. Ruano. *Hands-on PID autotuning: a guide to better utilisation*. IFAC, 2002.

[90] D. Goradia, S. Lakshminarayanan, and G. Rangaiah. Attainment of pi achievable performance for linear siso processes with deadtime by iterative tuning. *The Canadian Journal of Chemical Engineering*, 83(4):723–736, 2005.

[91] J. E. Seem. A new pattern recognition adaptive controller with application to hvac systems. *Automatica*, 34(8):969–982, 1998.

[92] M. Shamsuzzoha and S. Skogestad. The setpoint overshoot method: A simple and fast closed-loop approach for pid tuning. *Journal of Process Control*, 20(10):1220–1234, 2010.

[93] B. Srinivasan, U. Nallasivam, and R. Rengaswamy. Diagnosis of root cause for oscillations in closed-loop chemical process systems. In *Preprints of the 18th IFAC World Congress*, volume 18, pages 13145–13150, 2011.

[94] M. Veronesi and A. Visioli. Performance assessment and retuning of pid controllers. *Industrial & Engineering Chemistry Research*, 48(5):2616–2623, 2009.

[95] M. Veronesi and A. Visioli. A pid automatic tuning method for distributed-lag processes. In *Proceedings of the 7th IFAC International Symposium on Advanced Control of Chemical Processes*, volume 7, pages 762–767, 2009.

[96] M. Veronesi and A. Visioli. Performance assessment and retuning of pid controllers for integral processes. *Journal of Process Control*, 20(3):261–269, 2010.

[97] M. Veronesi and A. Visioli. Simultaneous closed-loop automatic tuning method for cascade controllers. *IET Control Theory & Applications*, 5(2):263–270, 2011.

[98] F. Morilla, A. Gonzáles, and N. Duro. Auto-tuning pid controllers in terms of relative damping. In *Preprints of IFAC Workshop PID00*, pages 161–166, 2000.

[99] J. Bai and X. Zhang. A new adaptive pi controller and its application in hvac systems. *Energy Conversion and Management*, 48(4):1043–1054, 2007.

[100] G. Qu and M. Zaheeruddin. Real-time tuning of pi controllers in hvac systems. *International journal of energy research*, 28(15):1313–1327, 2004.

[101] S. W. Sung and I.-B. Lee. On-line process identification and pid controller autotuning. *Korean Journal of Chemical Engineering*, 16(1):45–55, 1999.

[102] M. Veronesi and A. Visioli. Performance assessment and retuning of pid controllers for load disturbance rejection. In *Preprints of IFAC conference on advances in PID control*, pages 530–535, 2012.

[103] M. Veronesi and A. Visioli. An industrial application of a performance assessment and retuning technique for pi controllers. *ISA transactions*, 49(2):244–248, 2010.

[104] Z. Yu, J. Wang, B. Huang, and Z. Bi. Performance assessment of pid control loops subject to setpoint changes. *Journal of Process Control*, 21(8):1164–1171, 2011.

[105] Z. Yu and J. Wang. Assessment of proportional–integral control loop performance for input load disturbance rejection. *Industrial & Engineering Chemistry Research*, 51(36):11744–11752, 2012.

[106] D. Chen and D. E. Seborg. Pi/pid controller design based on direct synthesis and disturbance rejection. *Industrial & Engineering Chemistry Research*, 41(19):4807–4822, 2002.

[107] B. Hensel, J. Ploennigs, V. Vasyutynskyy, and K. Kabitzsch. A simple pi controller tuning rule for sensor energy efficiency with level-crossing sampling. In *Proceedings of the 9th International Multi-Conference on Systems, Signals and Devices (SSD)*, pages 1–6, 2012.

[108] B. Hensel, V. Vasyutynskyy, J. Ploennigs, and K. Kabitzsch. An adaptive pi controller for room temperature control with level-crossing sampling. In *Proceedings of the UKACC International Conference on Control*, pages 197–204, 2012.

[109] M. A. da Silva, F. Gomide, and W. Amaral. A rule based procedure for self-tuning pid controllers. In *Proceedings of the 27th IEEE Conference on Decision and Control*, pages 1947–1951, 1988.

[110] J. Litt. An expert system to perform on-line controller tuning. *Control Systems, IEEE*, 11(3):18–23, 1991.

[111] E. Bristol and T. Kraus. Life with pattern adaptation. In *Proceedings of the American Control Conference*, pages 888–893, 1984.

[112] T. W. Kraus and T. J. Myron. Self-tuning pid controller uses pattern recognition approach. *Control Engineering*, 31(6):106–111, 1984.

[113] Q. Zhou and M. Liu. An on-line self-tuning algorithm of pi controller for the heating and cooling coil in buildings. In *Proceedings of the 11th Symposium on Improving Building Systems in Hot and Humid Climates*, pages 285–293, 1998.

[114] H.-P. Hong, S.-J. Park, S.-J. Han, K.-Y. Cho, Y.-C. Lim, J.-K. Park, and T.-G. Kim. A design of auto-tuning pid controller using fuzzy logic. In *Proceedings of the 1992 International Conference on Industrial Electronics, Control, Instrumentation, and Automation*, pages 971–976, 1992.

[115] A. Seif. On the adaptive pattern recognition control. In *Proceedings of CompEuro'92 Computer Systems and Software Engineering*, pages 706–709, 1992.

[116] T. Salsbury. A practical algorithm for diagnosing control loop problems. *Energy and buildings*, 29(3):217–227, 1999.

[117] T. I. Salsbury. A practical method for assessing the performance of control loops subject to random load changes. *Journal of Process Control*, 15(4):393–405, 2005.

[118] R. Howard. *Automated autocorrelation function analysis for detection, diagnosis and correction of underperforming controllers*. Ph.D. Thesis, University of Connecticut, Storrs, CT, 2009.

[119] R. Howard and D. Cooper. A novel pattern-based approach for diagnostic controller performance monitoring. *Control Engineering Practice*, 18(3):279–288, 2010.

[120] Rachelle Howard and Douglas J Cooper. Performance assessment of non-self-regulating controllers in a cogeneration power plant. *Applied Energy*, 86(10):2121–2129, 2009.

[121] E. H. Bristol. Adaptive control odyssey. In *Proceedings of the ISA Annual Conference*, pages 561–70, 1970.

[122] Edgar H Bristol. Pattern recognition: an alternative to parameter identification in adaptive control. *Automatica*, 13(2):197–202, 1977.

[123] F. E. Barwig, J. M. House, C. J. Klaassen, M. M. Ardehali, and T. F. Smith. The national building controls information program. In *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings*, 2002.

[124] M. M. Ardehali and T. F. Smith. Literature review to identify existing case studies of controls-related energy-inefficiency in buildings. In *the National Building Controls Information Program*, 2001.

[125] M. Veronesi and A. Visioli. A technique for abrupt load disturbance detection in process control systems. In *Proceedings 17th IFAC World Congress on Automatic Control*, pages 14900–14905, 2008.

[126] T. Hägglund and K. J. Åström. Supervision of adaptive control algorithms. *Automatica*, 36(8):1171–1180, 2000.

[127] T. Hägglund. Industrial implementation of on-line performance monitoring tools. *Control Engineering Practice*, 13(11):1383–1390, 2005.

[128] P. Kuehl and A. Horch. Detection of sluggish control loops-experiences and improvements. *Control engineering practice*, 13(8):1019–1025, 2005.

[129] M. Jelali. Control system performance monitoring, Essen, Germany, 2010. Habilitationsschrift.

[130] C. M. Bishop. *Pattern recognition and machine learning*, volume 1. Springer, New York, 2006.

[131] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of dna nucleotides. *Physical Review E*, 49(2):1685, 1994.

[132] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3(04):785–798, 1995.

[133] Z. Chen, K. Hu, P. Carpena, P. Bernaola-Galvan, H. E. Stanley, and P. C. Ivanov. Effect of nonlinear filters on detrended fluctuation analysis. *Physical Review E*, 71(1):011104, 2005.

[134] Z. Chen, P. C. Ivanov, K. Hu, and H. E. Stanley. Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E*, 65(4):041107, 2002.

[135] C. K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):82–87, 1995.

[136] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley. Effect of trends on detrended fluctuation analysis. *Physical Review E*, 64(1):011114, 2001.

[137] W. Palma. *Long-memory time series: theory and methods*, volume 662. John Wiley & Sons, 2007.

[138] D. E. Seborg, D. A. Mellichamp, T. F. Edgar, and F. J. Doyle III. *Process dynamics and control*. John Wiley & Sons, 2010.

[139] F. G. Shinskey. Process control: as taught vs as practiced. *Industrial & Engineering Chemistry Research*, 41(16):3745–3750, 2002.

[140] Y.-G. Wang and H.-H. Shao. Optimal tuning for pi controller. *Automatica*, 36(1):147–152, 2000.

[141] K. Åström, C. Hang, P. Persson, and W. Ho. Towards intelligent pid control. *Automatica*, 28(1):1–9, 1992.

[142] A Horch and A. J. Isaksson. A method for detection of stiction in control valves. In *Proceedings of the IFAC Workshop on On-line fault detection and supervision in the chemical process industry*, 1998.

[143] M. A. A. S. Choudhury, S. L. Shah, N. F. Thornhill, and D. Shook. Detection and Quantification of stiction in control valves. *Control Engineering Practice*, 14:1395–1412, 2006.

[144] M. Kano, H. Maruta, H. Kugemoto, and K. Shimizu. Practical model and detection algorithm for valve stiction. In *Preprints of the IFAC symposium on dynamics and control of process systems*, pages 5–7, 2004.

[145] H. Jiang, R. Patwardhan, and S.L. Shah. Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix. *Journal of Process Control*, 19(8):1347–1354, 2009.

[146] M. A. A. S. Choudhury, M. Jain, and S. L. Shah. Stiction definition, modelling, detection and quantification. *Journal of Process control*, 18:232–243, 2008.

[147] R. Rengaswamy, T. Hägglund, and V. Venkatasubramanian. A qualitative shape analysis formalism for monitoring control loop performance. *Engineering Applications of Articial Intelligence*, 14(1):23–33, 2001.

[148] A. Singhal and T. I. Salsbury. A simple method for detecting valve stiction in oscillating control loops. *Journal of Process Control*, 15:371–382, 2005.

[149] M. A. A. S. Choudhury, S. L. Shah, and N. F. Thornhill. Diagnosis of poor control loop performance using higher order statistics. *Automatica*, 40(10):1719–1728, 2004.

[150] R. Srinivasan, R. Rengaswamy, S. Narasimhan, and R. M. Miller. Control loop performance assessment 2: Hammerstein model approach for stiction diagnosis. *Industrial and Engineering Chemistry Research*, 44:6719–6728, 2005.

[151] M. Jelali. Estimation of valve stiction in control loops using separable least-squares and global search algorithms. *Journal of Process Control*, 18(7):632–642, 2008.

[152] K.H. Lee, Z. Ren, and B. Huang. Novel closed-loop stiction detection and quantification method via system identification. In *Proceedings of the 4th International Symposium on Advanced control of industrial processes ADCONIP*, 2008.

[153] L.Z.X. Ivan and S. Lakshminarayanan. A new unified approach to valve stiction quantification and compensation. *Industrial and Engineering Chemistry Research*, 7(48):3474–3483, 2009.

[154] S. Karra and M. N. Karim. Comprehensive methodology for detection and diagnosis of oscillatory control loops. *Control Engineering Practice*, 17(8):939–956, 2009.

[155] H. Akaike. A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, 19(6):716–23, 1974.

[156] L. Ljung. *System Identification: Theory for the user*. Prentice Hall, Englewood Cliffs, NJ, 2 edition, 1999.

[157] A. Stenman, F. Gustafsson, and K. Forsman. A segmentation-based method for detection of stiction in control valves. *International Journal of Adaptive Control and Signal Processing*, 17(7):625–634, August 2003.

[158] D.B. Ender. Process control performance: Not as good as you think. *Control Engineering*, 40(10):180–190, 1993.

[159] Michel Ruel. Stiction: The hidden menace. *Control Magazine*, 13(11):1–11, 2000.

[160] B. Huang, M. Jelali, and A. Horch. Conclusions and future research challenges. In *Detection and diagnosis of stiction in control loops*, pages 359–366. Springer, 2010.

[161] Hailei Jiang, R.S. Patwardhan, and S.L. Shah. Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix. *Journal of Process Control*, 19(8):1347–1354, 2009.

[162] M. A. A. S.. Choudhury, N.F. Thornhill, and S.L. Shah. Modelling valve stiction. *Control Engineering Practice*, 13(5):641–658, 2005.

[163] Claudio Garcia. Comparison of friction models applied to a control valve. *Contr. Engineering Prac.*, 16(10):1231–1243, 2008.

[164] R. Wood and M. Berry. Terminal composition control of a binary distillation column. *Chemical Engineering Science*, 28(9):1707–1717, 1973.

[165] B.A. Ogunnaike and W.H. Ray. Multivariable controller design for linear systems having multiple time delays. *AIChE Journal*, 25(6):1043–1057, 1979.

[166] Q.P. He, J. Wang, and S.J. Qin. An alternative stiction-modelling approach and comparison of different stiction models. In *Detection and Diagnosis of Stiction in Control Loops*, pages 37–59. Springer, 2010.

[167] M. Forbes, M. Guay, and J. Forbes. Control design for first-order processes: shaping the probability density of the process state. *Journal of process control*, 14(4):399–410, 2004.

[168] B. Huang, S.L. Shah, and E. Kwok. Good, bad or optimal? performance assessment of multivariable processes. *Automatica*, 33(6):1175–1183, 1997.

[169] D. J. Chmielewski and A. M. Manthanwar. On the tuning of predictive controllers: Inverse optimality and the minimum variance covariance constrained control problem. *Industrial & Engineering Chemistry Research*, 43(24):7807–7814, 2004.

[170] C. Zhao, Y. Zhao, H. Su, and B. Huang. Economic performance assessment of advanced process control with lqg benchmarking. *Journal of process control*, 19(4):557–569, 2009.

[171] *MATLAB Statistics Toolbox User Guide (R2013a)*. The MathWorks Inc., Natick, Massachusetts, 2013.

[172] G. Gan, C. Ma, and J. Wu. *Data Clustering: Theory, Algorithms, and Applications*. Society for Industrial and Applied Mathematics, 2007.

[173] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.

[174] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[175] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 2002.

[176] J.E. Jackson. *A User's Guide to Principal Components*. Wiley, Hoboken, NJ, 1991.

[177] D. S. Coming and O. G. Staadt. Velocity-aligned discrete oriented polytopes for dynamic collision detection. *Visualization and Computer Graphics, IEEE Transactions on*, 14(1):1–12, 2008.