

Contributions to the  
Use of Statistical Methods for  
Improving Continuous Production

Francesca Capaci

Quality Technology



LICENTIATE THESIS

November, 2017

# **Contributions to the Use of Statistical Methods for Improving Continuous Production**

Francesca Capaci

Main Supervisor: Erik Vanhatalo

Assistant supervisors: Bjarne Bergquist and Murat Kulahci

Luleå University of Technology

Department of Business Administration, Technology and Social Science

Division of Business Administration and Industrial Engineering

Quality Technology

Printed by Luleå University of Technology, Graphic Production 2017

ISSN 1402-1757

ISBN 978-91-7583-996-7 (print)

ISBN 978-91-7583-997-4 (pdf)

Luleå 2017

[www.ltu.se](http://www.ltu.se)

## ABSTRACT

---

Complexity of production processes, high computing capabilities, and massive data sets characterize today's manufacturing environments, such as those of continuous and batch production industries. Continuous production has spread gradually across different industries, covering a significant part of today's production. Common consumer goods such as food, drugs, and cosmetics, and industrial goods such as iron, chemicals, oil, and ore come from continuous processes. To stay competitive in today's market requires constant process improvements in terms of both effectiveness and efficiency. Statistical process control (SPC) and design of experiments (DoE) techniques can play an important role in this improvement strategy. SPC attempts to reduce process variation by eliminating assignable causes, while DoE is used to improve products and processes by systematic experimentation and analysis. However, special issues emerge when applying these methods in continuous process settings. Highly automated and computerized processes provide an exorbitant amount of serially dependent and cross-correlated data, which may be difficult to analyze simultaneously. Time series data, transition times, and closed-loop operation are examples of additional challenges that the analyst faces.

The overall objective of this thesis is to contribute to using of statistical methods, namely SPC and DoE methods, to improve continuous production. Specifically, this research serves two aims: [1] to explore, identify, and outline potential challenges when applying SPC and DoE in continuous processes, and [2] to propose simulation tools and new or adapted methods to overcome the identified challenges.

The results are summarized in three appended papers. Through a literature review, Paper A outlines SPC and DoE implementation challenges for managers, researchers, and practitioners. For example, problems due to process transitions, the multivariate nature of data, serial correlation, and the presence of engineering process control (EPC) are discussed. Paper B further explores one of the DoE challenges identified in Paper A. Specifically, Paper B describes issues and potential strategies when designing and analyzing experiments in processes operating under closed-loop control. Two simulated examples in the Tennessee Eastman (TE) process simulator show the benefits of using DoE techniques to improve and optimize such industrial processes. Finally, Paper C provides guidelines, using flow charts, on how to use the continuous process simulator, "The revised TE process simulator," run with a decentralized control strategy as a test bed for developing SPC and DoE methods in continuous processes. Simulated SPC and DoE examples are also discussed.

**Keywords:** Process industry, Continuous process, Statistical process control, Design of experiments, Process improvements, Simulation tool, Engineering process control.



# CONTENTS

ABSTRACT .....	I
CONTENTS.....	III
APPENDED PAPERS.....	V
THESIS STRUCTURE.....	VII
<b>PART I: THEORETICAL FOUNDATIONS</b>	
<b>1. INTRODUCTION.....</b>	<b>3</b>
1.1. SPC and DoE for quality control and improvement.....	3
1.2. Continuous processes.....	4
1.3. SPC and DoE in continuous processes.....	6
1.3.1. SPC challenges in continuous processes.....	6
1.3.2. DoE challenges in continuous processes.....	8
1.4. Research objective and scope.....	11
1.5. Introduction and authors' contributions to the appended papers .....	11
<b>PART II: EMPIRICAL WORK AND FINDINGS</b>	
<b>2. RESEARCH METHOD.....</b>	<b>15</b>
2.1. Introduction to the academic research.....	15
2.2. Summary and background of my research process.....	15
2.3. Methods used in the appended papers .....	19
2.4. Summary of methodological choices.....	22
<b>3. RESULTS.....</b>	<b>23</b>
3.1. SPC and DoE in continuous processes.....	23
3.2. Experimentation in closed-loop operations .....	27
3.3. The TE simulator for SPC and DoE methods development.....	29
3.4. Main contributions .....	30
<b>Part III: FUTURE RESEARCH</b>	
<b>4. FUTURE RESEARCH DIRECTIONS.....</b>	<b>35</b>
<b>Acknowledgments .....</b>	<b>37</b>
<b>References .....</b>	<b>38</b>
<b>Part IV: APPENDIX - APPENDED PAPERS (A-C)</b>	





## APPENDED PAPERS

---

*This licentiate thesis summarizes and discusses the following three full appended papers.*

- A<sup>1</sup>**    **Capaci, F., Vanhatalo, E., Bergquist, B., and Kulahci, M. (2017).** Managerial Implications for Improving Continuous Production Processes. *Conference Proceedings, 24th International Annual EurOMA Conference: Inspiring Operations Management*, July 1–5, 2017, Edinburgh (Scotland).
- B<sup>2</sup>**    **Capaci, F., Bergquist, B., Kulahci, M., and Vanhatalo, E. (2017).** Exploring the Use of Design of Experiments in Industrial Processes Operating Under Closed-Loop Control. *Quality and Reliability Engineering International*. Published online ahead of print. DOI: 10.1002/qre.2128.
- C**      **Capaci, F., Vanhatalo, E., Kulahci, M., and Bergquist, B. (2017).** The Revised Tennessee Eastman Process Simulator as Testbed for SPC and DoE Methods. *Submitted for publication*.

---

<sup>1</sup> Paper A was presented by Francesca Capaci on July 4, 2017, at the *24th International Annual EurOMA Conference: Inspiring Operations Management in Edinburgh, Scotland*.

<sup>2</sup> An early version of paper C was presented by Francesca Capaci on September 13, 2016, at the *16th International Annual Conference of the European Network for Business and Industrial Statistics (ENBIS-16)* in Sheffield, United Kingdom.



# THESIS STRUCTURE

---

The overall structure of this thesis is organized in four parts: theoretical foundations, empirical work and findings, future research, and the appended papers. Figure I illustrates the chapters included in each part, except for the appendix, which shows the order and type of the appended papers.

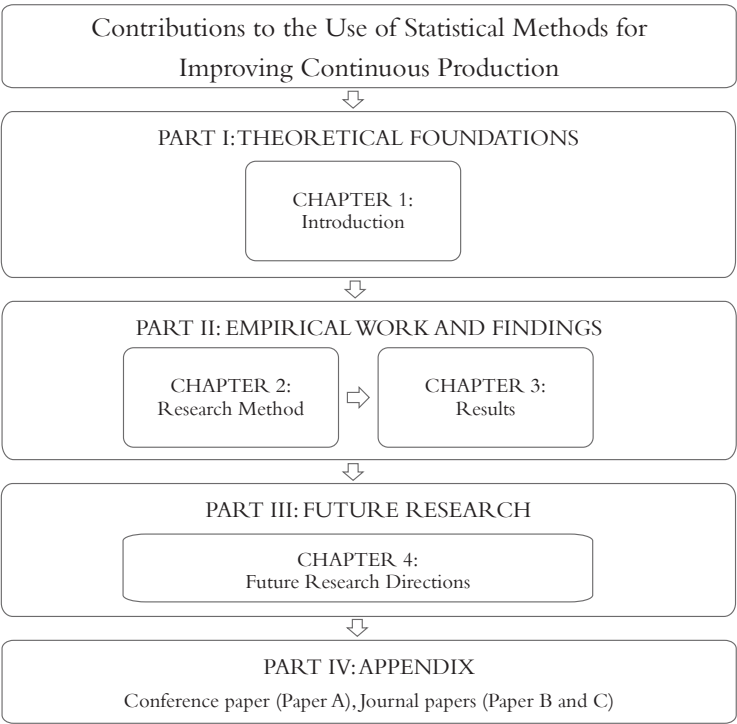


Figure I. Structure of the thesis, including its parts and chapters. The type and order of the appended papers are also shown.

Chapter 1 (*Introduction*) provides an introduction and the background to the research area. The research objective and scope are outlined. The chapter ends with a brief summary of the appended papers and the thesis structure. Chapter 2 (*Research Method*) summarizes the research process and the methodological choices made during the research. Chapter 3 (*Results*) outlines the main results, conclusions, recommendations, reflections, and contributions drawn from the results of this research. Chapter 4 describes the ideas and research questions that arose during the research process, and that I would like to investigate further in my doctoral studies.



## **PART I: THEORETICAL FOUNDATIONS**

*“He who loves practice without theory is  
like the sailor who boards ship without a rudder and  
compass and never knows where he may cast.”*

*Leonardo da Vinci*



# 1. INTRODUCTION

---

*Chapter 1 provides an introduction and background to the research area. The research objective and scope are outlined. The chapter ends with a brief summary of the appended papers and the thesis structure.*

## 1.1. SPC and DoE for quality control and improvement

Statistical process control (SPC) and Design of Experiments (DoE) are two important and well-established methodologies that include statistical and analytical tools to analyze quality problems and improve process performance. A manufacturing process uses a combination of resources (e.g., tools, operations, machines, information, and people) to transform a set of inputs (mainly raw materials) into a finished output (or product). Process inputs are controllable process variables, such as temperature, pressure, and feed rate, whereas the process output can be associated with one or more observable and measurable response variables. Response variables can be process performance indicators, such as cost or energy consumption and/or final product quality characteristics. Changing the (controllable) inputs usually induces a related change in the response variable(s). Other inputs, called noise factors, typically also affect the response variable(s), but they are impossible, difficult, or expensive to change or control (i.e., they are uncontrollable) (Montgomery, 2012b). Figure 1.1 illustrates a general model of a process, highlighting how SPC and DoE interact with process inputs and response variables for quality control and improvement.

Control charts are the main and most well-known tools of SPC techniques. Applied to the process response variable(s), control charts provide a means for online monitoring of the process performance. Alarms issued by control charts indicate the presence of so-called assignable causes, which can be investigated further. If their root cause can be uncovered, the assignable causes can be systematically eliminated, thus reducing unwanted process variability (Montgomery, 2012a).

A designed experiment allows for systematically changing the controllable inputs to study the effects on the process response variable(s) of interest. Factorial designs and fractional factorial designs are two major types of designed experiments, in which factors are varied together in such a way that all, or a subset of combinations of factor levels are tested. Therefore, most DoE methods are offline quality improvement tools, which aim to reveal potential causal relationships between the process inputs<sup>3</sup> and the response variable(s). Knowledge of the crucial process inputs is essential for characterizing a process

---

<sup>3</sup>“Controllable process inputs” are often called, for brevity, “process inputs.” Hereafter, “process inputs” refer to “controllable process inputs,” unless otherwise specified.

## THEORETICAL FOUNDATIONS

and optimizing its performance by steering it toward a target value and/or reducing the process variability (Montgomery, 2012b).

When the key process variables<sup>4</sup> have been identified, an online process control chart can be routinely employed for process surveillance to promptly adjust the process whenever unusual events drive the process toward out-of-control situations.

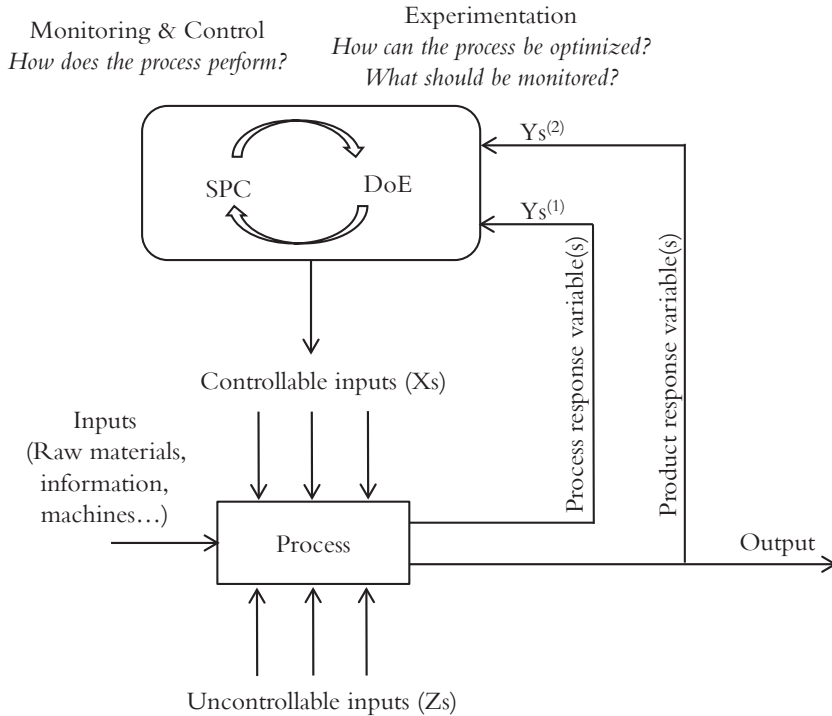


Figure 1.1. A general model of a process that highlights how SPC and DoE interact with process inputs and outputs. Adapted from Montgomery (2012b).

### 1.2. Continuous processes

Reid and Sanders (2012) classify production processes into two fundamental categories of operations: intermittent and repetitive operations. Depending on the product volume and degree of product customization, intermittent operations can be divided further into project processes and batch processes, while repetitive operations can be divided into line

---

<sup>4</sup>“Controllable process inputs” are sometimes also called “process variables,” “process factors,” “experimental factors,” or “experimental variables” in the DoE literature. Therefore, these terms are used interchangeably in this thesis.



## INTRODUCTION

processes and continuous processes (*ibid.*). Figure 1.2 presents a classification of production processes and their main characteristics.

Batch and continuous production represent the main process technologies in the process industry. The process industry is responsible for about 30% of the worldwide production (Lager, 2010) and involves industries such as pulp and paper, oil and gas, food and beverage, steel, and mining and material, among many others.

A common misconception is that the terms “process industry” and “continuous processes” are interchangeable, when in fact, they have different meanings (Abdulmalek et al., 2006). In line with this concept, Dennis and Meredith (2000) use the definition of the American Production and Inventory Control Society (APICS, 2008 p. 104) to define the process industry as:

*“production that adds value by mixing, separating, forming and/or performing chemical reactions by either batch or continuous mode”*

and a continuous process as:

*“a production approach with minimal interruptions in actual processing in any one production run or between production runs of similar products.”*

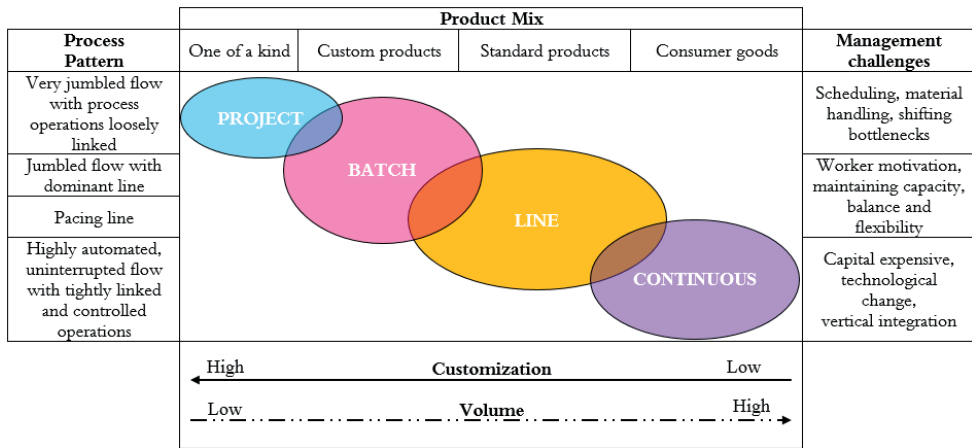


Figure 1.2. Classification and characteristics of production processes. Adapted from Hayes and Wheelwright (1979).

Three main features differentiate continuous processes from other types of manufacturing processes: the types of incoming materials, transformation processes, and outgoing materials (Lager, 2010). Incoming materials in continuous processes are usually raw materials, often stemming directly from natural resources and, therefore, with inherent

characteristics that can vary substantially (Fransoo and Rutten, 1994). Engineering process control (EPC) is often necessary to stabilize product the quality and process characteristics of continuous processes (Montgomery et al., 1994; Box and Luceño, 1997). The transformation process includes several operation units, such as tanks, reactors, mixing units that work in a continuous flow, and input-output relationships that might not be immediately clear (Hild et al., 2001; Vanhatalo, 2009). Finally, the outgoing materials (and often also the incoming materials) are non-discrete products, such as liquids, pulp, slurries, gases, and powders that evaporate, expand, contract, settle out, absorb moisture, or dry out (Dennis and Meredith, 2000). The nature of the handled materials makes these processes more sensitive to stoppages and interruptions because of the loss in production quality and long lead-times for startup (Duchesne et al., 2002; Abdulmalek et al., 2006; Lager, 2010).

### **1.3. SPC and DoE in continuous processes**

For decades, management improvement programs such as Robust Design, Total Quality Management, and Six Sigma have been promoting the use of statistical improvement methods such as SPC and DoE to improve processes and the quality of products (Bergquist and Albing, 2006; Bergman and Klefsjö, 2010). Although these methods are well established in the statistics and quality engineering literature, their application has been found to be relatively rare in industry (Deleryd et al., 1999; Bergquist and Albing, 2006; Tanco et al., 2010). The use of SPC and DoE in industrial applications within discrete manufacturing production environments faces barriers such as a lack of theoretical knowledge, change management, practical problems, and a lack of resources for internal training (*ibid.*). In addition to these barriers, the implementation of SPC and DoE methods in continuous processes is complicated by the need to promote and adapt such methods to continuous production environments, as well as the new and complex challenges they offer (Bergquist, 2015b; Vining et al., 2015).

#### **1.3.1. SPC challenges in continuous processes**

The literature on SPC and related fields, such as chemometrics and control engineering, identifies several challenges that may arise when using SPC in continuous processes. These challenges are summarized in this section.

##### *Multivariate nature of process data*

Researchers in different areas are increasingly focusing on the issue of managing big data, although SPC requires more application-oriented and methodological research to handle this modern challenge (Vining et al., 2015). The increasing availability of high-tech sensors and storage capacity make it possible to take measurements at multiple locations and with a high sampling frequency. Thus, the uninterrupted flow of continuous

## INTRODUCTION

processes can produce massive data sets in terms of both variables and observations, exhibiting varying degrees of auto- and cross-correlation (Saunders and Eccleston, 1992; Hild et al., 2001).

Historically, most SPC research and industrial applications have focused on univariate control charts, in which process and product variables are monitored individually. However, in data-rich environments, such as those of continuous processes, the univariate monitoring of each variable in separate control charts is often both inefficient and misleading (Kourti and MacGregor, 1995). Instead, the simultaneous monitoring of multiple process variables is needed, leading analysts to the field of multivariate SPC. Multivariate monitoring charts, based on latent variable techniques such as a principal component analysis (PCA) and partial least squares (PLS), have already been used successfully in industrial applications (e.g., see Kourti et al. 1996, and Ferrer 2014). The strength of latent variable methods is their dimensionality reduction properties. Using the cross-correlation of process variables, these methods reduce an original data set to a few linear combinations of process variables (so-called latent-variables) that can be considered as the main drivers of the process events (Frank and Friedman, 1993; MacGregor and Kourti, 1995). Commonly, a Hotelling  $T^2$  control chart simultaneously monitors the retained latent variables from the PCA/PLS model, while the squared prediction error (Q) chart monitors the model's residuals. However, control charts based on PCA can handle the cross-correlation, but cannot address the autocorrelation issue (Vanhatalo and Kulahci, 2016). An extension to PCA, called dynamic PCA, deals with autocorrelation by adding time-lagged variables (Ku et al., 1995).

### *Autocorrelated data*

In continuous production processes, the data sampling of automated data collection schemes is usually performed more frequently than the process dynamics and, consequently, the collected data are typically highly (and normally positively) autocorrelated (Hild et al., 2001; Vanhatalo and Bergquist, 2007). Autocorrelation in the data violates the basic assumption of time- independent observations, on which SPC methods rely, affecting both univariate and multivariate SPC techniques. Specifically, the positive autocorrelation affects the process variability estimation, which, in turn, leads to deflated control limits in the control charts and increased false alarm rates (Mastrangelo and Montgomery, 1995; Runger, 1996; Bisgaard and Kulahci, 2005).

The literature suggests two solutions to dealing with multivariate and autocorrelated data. The first is to use a standard multivariate control chart, adjusting the control limits to achieve the desired in-control alarm rate. However, this procedure requires an ad-hoc adjustment to each case, which is awkward and time consuming. The second solution requires “filtering out the autocorrelation” using a multivariate time series model, and then applying a multivariate control chart to the residuals from this

model (Harris and Ross, 1991). However, fitting a multivariate time series model for many variables is challenging owing to the large number of parameters that must be estimated. Moreover, fault detection and isolation become non-trivial problems.

The autocorrelation also affects the estimation of process capability indices, limiting the possibilities for assessing the process performance (Shore, 1997; Zhang, 1998; Sun et al., 2010; Lundkvist et al., 2012).

### *Closed-loop operations*

In continuous processes, EPC systems are often used to stabilize product quality and process characteristics. These process control systems are designed to maintain crucial process variables around their set-points by transferring the variability into upstream input process variables (i.e., manipulated variables) (MacGregor and Harris, 1990; Hild et al., 2001). When a process involves EPC, the fault detection of SPC charts applied to all output of the process could fail. However, the identification and elimination of potential assignable causes of variation may still be pursued by applying a monitoring control chart to both the manipulated variables and the controlled output. Thus, SPC and EPC can complement each other effectively. Indeed, the former attempts to control the process in the long-term by detecting and eliminating the occurrence of an assignable cause. The latter attempts to control the process in the short-term by transferring the variability to another variable.

The effectiveness of integrating SPC and EPC has already been shown in the literature (Montgomery et al., 1994; Box and Luceño, 1997). However, further research is needed to adjust the traditional SPC paradigm to monitor process output when EPC is in place (Box and Kramer, 1992).

### **1.3.2. DoE challenges in continuous processes**

This section summarizes the challenges that may emerge when applying DoE methods to continuous processes.

#### *Large-scale experimentation*

Continuous-process plants are usually spread out over a large area and operate around the clock. Thus, experimentation in full-scale continuous processes may involve the majority of the production staff, making coordination and information flow essential requirements. Experimental campaigns can carry on for a long time, jeopardizing the production plan and producing off-grade products. Therefore, time and costs are often significant constraints.

Continuous production process characteristics unavoidably affect the experimentation strategy. Therefore, planning, conducting, and analyzing experiments require proper adjustments in continuous process settings. An experimental campaign should always start with the careful planning of the activities preceding the experiments,

## INTRODUCTION

because they are critical to successfully solving the experimenters' problem (Coleman and Montgomery, 1993; Box et al., 2005). Vanhatalo and Bergquist (2007) provide a checklist for planning experiments in continuous-process settings, where limited numbers of experimental runs, easy-/hard-to-change factors, randomization restrictions, and design preferences are particularly relevant. Time restrictions and budget constraints force the analyst to consider experiments with few factors and runs, and replicating experiments may not always be possible (Bergquist, 2015a). Therefore, two-level (fractional) factorial designs are important, but analyzing unreplicated designs might not always be easy to accomplish owing to the impossibility of estimating the experimental random variation and/or the lack of degrees of freedom when calculating the model's unknowns (i.e., the factors' effects). When split-plot designs are needed, for example to reduce the transition times between runs, the analysis might be complicated further. Moreover, not resetting the factor levels leads to a correlation between adjacent runs and to designs called randomized-not-reset (RNR) designs (Webb et al., 2004). Further methodological research could be beneficial to improve the analysis methods used to understand the experimental results.

Owing to their sequential nature, the response surface methodology (RSM) and evolutionary operations (EVOP) are also appealing strategies in experiments involving continuous processes (Box and Wilson, 1951; Box, 1957). However, these techniques may need to be adjusted for closed-loop operations because, for example, it might not be immediately clear which variables can be optimized.

### *Closed-loop operations*

Conventional applications of DoE methods implicitly assume open-loop operations. In this process configuration, the potential effects of changes to process inputs can be observed directly in the process outputs (Montgomery, 2012b). Under closed-loop operations, process outputs that might be interesting responses are usually maintained around desired target values (i.e., the set-points). Hence, input-output causal relationships might not be immediately clear (Hild et al., 2001). The potential effects of changes to process inputs are displaced to other process variables (so-called manipulated variables) if the control loop works properly. Therefore, closed-loop operations require a different strategy for experimentation and analysis, which need further research to improve the understanding of the experimental results.

### *Process dynamics*

In a continuous process, production steps such as mixing, melting, reflux flows, or product state changes make the process dynamic. In a dynamic process, effects of changed process inputs on the process outputs develop gradually until the process stabilizes to a new steady state (Nembhard and Valverde-Ventura, 2003; Bisgaard and Khachatryan, 2011). The time needed for a response to reach a new steady state is called the transition

time (Vanhatalo et al., 2010), and its characterization is a crucial issue when experimenting in continuous processes.

To correctly estimate the effects of factor changes on the process response variables, it is important that the process reach a steady-state condition. Hence, transition times affect the run length of the experiments (Vanhatalo and Vännman, 2008). Knowing the transition times, the experimenter can avoid unnecessary long and costly run length or run length that are too short, yielding misleading estimates of the effects. However, to determine transition times in continuous processes is difficult, for several reasons. Changes to the process inputs often affect the process responses in several ways, and transition times may vary for different responses in terms of both length and behavior. For example, Vanhatalo et al. (2010) developed a method for estimating transition times in dynamic processes, combining PCA and transfer function-noise modeling. However, the proposed method is an offline method that needs to determine the transition times a priori during the planning phase of the experiment. Methodological research for an online estimation of transition times in continuous processes could help to solve experimentation challenges in these production environments.

#### *Autocorrelated and cross-correlated responses*

In continuous processes, the high sampling frequency induces a positive correlation in the process response variables (Hild et al., 2001). Ignoring the autocorrelation in the responses might lead to ineffective or erroneous analysis of the experimental results. For example, using the run averages of the response might be a poor alternative, because it likely leads to an incorrect estimation of the effects. Instead, a time series analysis seems to be a useful tool to analyze the experimental results, because the time series nature of the data and the autocorrelation can be taken into account. However, few attempts have been made to combine the benefits of DoE and time series analysis. As shown in Vanhatalo et al. (2010), the dynamic input–output relationships of a process can be modeled using transfer–function noise modeling and intervention analysis, improving the efficiency of the results (Bisgaard and Kulahci, 2011; Vanhatalo et al., 2013; Lundkvist and Vanhatalo, 2014).

In continuous production processes, process variables are often related to each other. These interrelationships make it difficult to identify variables that can be changed independently from one another and used as experimental factors. Moreover, a change in one experimental factor often affects several variables, because they are simply reflections of the same underlying event (Kourti and MacGregor, 1995; Kourti and MacGregor, 1996; Kourti, 2005). Hence, a multivariate analysis approach, using latent variable techniques such as PCA and PLS, should be preferred to a univariate approach. Interpreting the results of cross-correlated variables using a univariate approach can be considered analogous to a “one-at-a-time” experimentation approach in the presence of interaction effects (MacGregor, 1997; Vanhatalo and Vännman, 2008). However, latent

## INTRODUCTION

variable methods used in conjunction with DoE techniques need further research in order to overcome the issues highlighted in Section 1.3.1, which hold for DoE applications as well.

### 1.4. Research objective and scope

The overall objective of this thesis is to contribute to using SPC and DoE methods to improve continuous production. Specifically, this research serves two aims:

- I. to explore, identify, and outline potential challenges when applying SPC and DoE in continuous processes, and
- II. to propose simulation tools and new or adapted methods to overcome the identified challenges.

This thesis focuses on filling the academic gap between the methods available to researchers and practitioners in empirical sciences and the challenges offered by today's manufacturing environments, such as those of continuous industries. Rapid data collection from multiple and interconnected sources and massive data sets are common for such processes. In the last few decades, the concept of big data has attracted the attention of researchers in many fields, including machine learning, data mining, computer engineering, and cloud computing. However, research on SPC and DoE has been slower in providing answers to this new paradigm. When applying SPC and DoE methods, the importance of big data does not revolve around how much data are available, but on how to handle the data properly and the challenges they offer in addressing questions of interest. The SPC and DoE methods available for industrial practitioners are limited, insufficient, or non-existent for this new paradigm.

Therefore, the core of this research is built on a quality engineering perspective, with the aim of contributing to the development of statistical methodologies for quality and productivity improvements in continuous processes.

### 1.5. Introduction and authors' contributions to the appended papers

This section introduces the appended papers and highlights the relationships between them and the research aims. Contributions to the appended papers are also presented.

**Paper A: Managerial Implications for Improving Continuous Production Processes. Capaci, F., Vanhatalo, E., Bergquist, B., and Kulahci, M. (2017).**

Paper A outlines SPC and DoE implementation challenges described in the literature for managers, researchers, and practitioners interested in continuous production process improvements. Besides research gaps and state-of-the-art solutions, current challenges are also illustrated. This is the first appended paper since it serves to introduce the research topic and relates to aim I of the research.

*The idea for this paper came from Francesca Capaci when there was an opportunity to submit a contribution to the 24th International Annual EurOMA Conference. Francesca Capaci performed the four phases and eight review stages of the literature review process including searches for data collection, screening steps, and analysis of data. The co-authors commented throughout the emerging analysis steps. Francesca Capaci wrote the paper with contributions by all co-authors.*

**Paper B: Exploring the Use of Design of Experiments in Industrial Processes Operating Under Closed-Loop Control. Capaci, F., Bergquist, B., Kulahci, M., and Vanhatalo, E. (2017).**

Paper B can be described as an in-depth study of one of the challenges identified in Paper A and is related to aim II of the research. Paper B conceptually explores issues of experimental design and analysis in processes operating under closed-loop control and illustrates how DoE can help in improving and optimizing such processes. The Tennessee Eastman (TE) Challenge process simulator is used to illustrate two experimental scenarios.

*All the authors jointly developed the idea of exploring the use of DoE in systems operating under closed-loop control. Francesca Capaci then worked to understand the TE process simulator with the aim to find viable scenarios for conducting experiments. Francesca Capaci planned, simulated, and analyzed the experimental scenarios while all authors were involved in the discussions leading up to the results. Francesca Capaci wrote the paper with contributions by all co-authors.*

**Paper C: The Revised Tennessee Eastman Process Simulator as Testbed for SPC and DoE Methods. Capaci, F., Vanhatalo, E., Kulahci, M., and Bergquist, B. (2017).**

Paper C provides guidelines for how to use the revised TE process simulator, run with a decentralized control strategy, as a testbed for SPC and DoE methods in continuous processes. Flow charts give details on the necessary steps to get started in the Matlab Simulink® framework. The paper also explains how to create random variability in the simulator and two examples illustrate two potential applications in the SPC and DoE contexts. Paper C thus mainly relates to aim II of the research.

*The idea to use the revised TE process as testbed for SPC and DoE methods in continuous processes was jointly developed by all the authors. Francesca Capaci located the revised simulator, performed all work required to understand the details of the simulator, and was mainly responsible for the development of the idea on how to create random variability in the simulator. Francesca Capaci also developed the illustrated examples and was responsible for all simulations and analyses. All the authors were involved in the discussions leading up to the results. Francesca Capaci wrote the paper with contributions by all co-authors.*



## PART II: EMPIRICAL WORK AND FINDINGS

*“If we knew what it was we were doing,  
it would not be called research, would it?”*

Albert Einstein



## 2. RESEARCH METHOD

---

*This chapter summarizes the research process and the methodological choices made during the research.*

### 2.1. Introduction to the academic research

My first experience of statistical thinking as a well-recognized methodology for continuous improvement was in the bachelor's and master's degree courses in the industrial and management engineering program at Università degli Studi di Palermo (UniPa). During my time as a student, I got to learn about both technical and industrial applied statistical concepts, in addition to managerial, economic, and strategic aspects of business management. During my time at UniPa, quality technology and industrial applied statistics increasingly attracted my attention. Quality technology and Six Sigma, SPC, and DoE were courses I was intrigued by. In March 2014, I presented my master's thesis on a methodological study about metamodeling techniques in computer experiments. During the development of this work, I had the chance to learn other programming languages, such as R and Matlab, and to discover my research interest in industrial applied statistics. I then decided to look for a PhD position, and found an open position in an interesting topic at Luleå University of Technology (LTU). I applied, was admitted, and started my research in September 2014 after moving from Italy to Sweden.

My PhD position was part of a project aiming to develop industrial statistical methods for quality and productivity improvements in continuous production processes. The research project, funded by the Swedish Research Council and supervised by Bjarne Bergquist, Erik Vanhatalo, and Murat Kulahci, is still ongoing, and will formally end in December 2017. However, my research related to the project aims is planned to continue until September 2019.

### 2.2. Summary and background of my research process

Figure 2.1 illustrates a Gantt chart containing the main research activities thus far. The authors' contributions in developing the appended papers are also highlighted. In the chart, upside down triangles mark the beginning of a research study. Circles indicate conference presentations for the appended papers, and triangles and diamonds indicate papers that have been submitted (or accepted) for publication.

# EMPIRICAL WORK AND FINDINGS

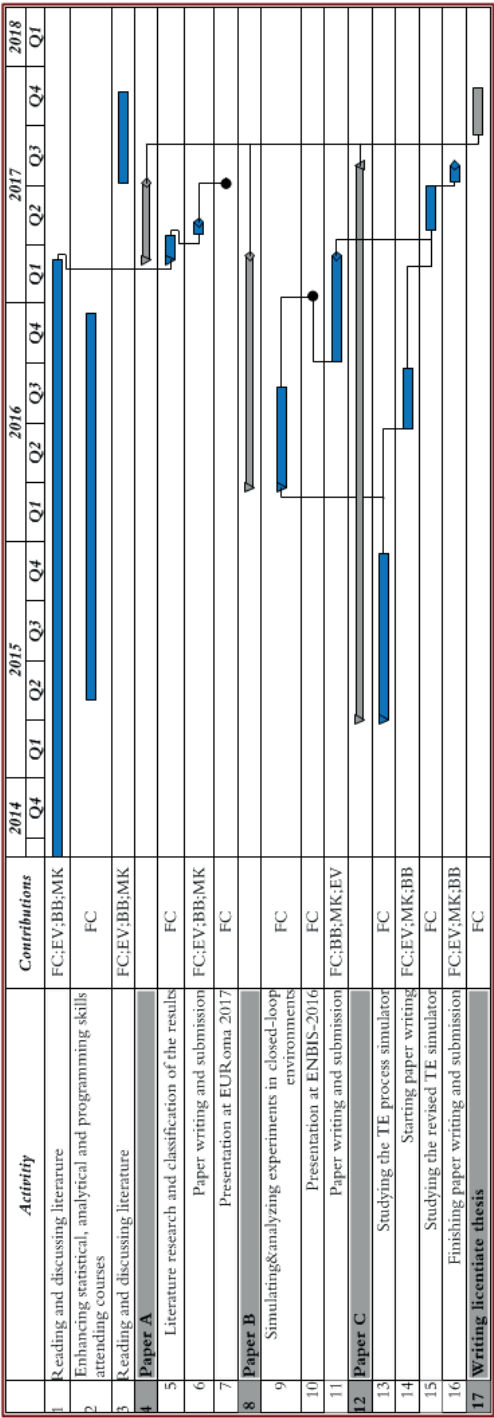


Figure 2.1. Gantt chart showing the main research activities from the beginning of my research education until the licentiate seminar. Authors' contributions in developing the appended papers are also highlighted (FC=Francesca Capaci; BB=Bjame Bergquist; EV= Erik Vanhatalo; MK=Murat Kulahci).

## RESEARCH METHOD

My research officially started at the beginning of September 2014 when my supervisors and I agreed on prioritizing activities aimed at enhancing my technical background and increasing my understanding of the research topic. In addition to attending courses, my focus was on reading and discussing literature connected to my research topic with my supervisors. This activity was organized as a weekly meeting, where my supervisors and I discussed a preselected article. During each meeting, we discussed the article's main message, as well as areas of special interest or aspects we did not understand. These discussions led to additional readings, selected from among known seminal works of my research field or from the articles' reference lists. The knowledge acquired during this literature study was later used to support the research study conducted in paper A.

In the first half of 2015, I realized there was a need for a simulator that could emulate a continuous production process. This simulator needed to offer a good balance between realistic simulations of a continuous process and the flexibility necessary for methodological research on SPC and DoE. The main reason a simulator was needed was that my research project does not involve industrial collaborators where SPC and DoE methods can be studied. Even with access to industrial processes, it would have been difficult to gain access to processes that would allow for full-scale methodological developments. When developing and testing SPC methods, data sets with specific characteristics, such as sample size, sampling time, and the occurrence of specific known faults, need to be available. Furthermore, DoE applications in full-scale industrial processes may unavoidably jeopardize the production plants, affecting their production goals. This would make it difficult to convince top management to start large and costly experimental campaigns. Thus, finding a realistic simulator became a priority.

Reading the literature, I discovered that many published articles in chemometrics, an important field of research connected to continuous processes, used the TE process as a testbed to illustrate new methods being developed (e.g., see Lee et al., 2004, Liu et al., 2015, and Rato et al., 2016). Downs and Vogel (1993) further supported this interest in an in-depth study of the TE process simulator. In fact, the authors originally proposed the TE process as a test problem providing a list of potential applications in a wide variety of topics such as plant control, optimization, education, non-linear control and, many others. Moreover, the TE process simulator can emulate many of the challenges frequently found in continuous processes, such as the multivariate nature of the data, process dynamics, and autocorrelated and cross-correlated responses. However, the TE process simulator has to be run with an implemented control strategy to overcome its unstable operation in an open-loop. The need to run the TE process simulator in a closed loop widened its usability to studying the challenges that arise when applying SPC and DoE in continuous processes operating under closed-loop control.

An internet search showed that there were many control strategies available to control and stabilize the TE process. Among these, the characteristics of the decentralized

## EMPIRICAL WORK AND FINDINGS

control strategy simulator proposed by Ricker (1996) was the most suitable for my research purposes, offering the following advantages:

- the simulator is implemented in the Matlab Simulink® interface and is fairly easy to use and free to access,
- the set-points of the controlled variables and the process inputs can be modified, as long as they are maintained within the restrictions of the decentralized control strategy, and
- the analyst can specify the characteristics of the simulated data (e.g., the length of an experiment, sampling frequency, types of process disturbances, etc.).

As my understanding of the TE process simulator grew, I started simulating and analyzing planned experiments in order to explore the use of DoE in industrial processes operating under closed-loop control. The results are summarized in Paper B. Analyzing the experimental data of the TE process highlighted an important limitation of the decentralized TE process simulator (Ricker, 1996), namely, its deterministic nature. The decentralized TE process variables are only affected by white Gaussian noise, mimicking typical measurement noise, so that repeated simulations with the same setup produce the same results, except for measurements errors. The value of a model containing only measurement noise is limited when running repeated simulations to assess the performance of an SPC method. Moreover, the impossibility of simulating the experimental error, essential for estimating the effects of experimental factors on the responses under study, renders important DoE concepts such as randomization and replication unusable.

The deterministic nature of the decentralized TE process simulator limited its usability for SPC and DoE applications. Hence, additional research was needed to understand how to overcome the limitations of the decentralized TE process simulator of Ricker (2005). This research led to finding out about a new release of the decentralized TE process simulator (Bathelt et al., 2015b), known as the revised TE process model (Bathelt et al., 2015a). Guidelines on how to use the revised TE process as a testbed for SPC and DoE methods are provided in Paper C. Among other possibilities, the revised simulator enables the disturbances introduced into the process to be scaled and the seed of each simulation to be changed. Scaling the random variation disturbances makes it possible to add variability to the results without overly distorting them. Moreover, the seed change of the random numbers forces the simulator to generate different results for repeated simulations with the same starting conditions. Combining these two features, I found a way to overcome the deterministic nature of the simulator, making the revised TE model suitable for testing SPC and DoE methods in continuous processes.

### 2.3. Methods used in the appended papers

The following subsections describe the methodological choices in the appended papers, as well as the relationships between the papers and the aims.

#### *Paper A*

Paper A relates to aim I of the research objective, and was motivated by the need to summarize the results of the literature searches and the results of a more systematic literature review. A systematic literature review is an essential step in any research process. Being familiar with the existing literature helps to, for example, determine what researchers already know about a research topic, summarize the research evidence from high quality studies, identify research gaps, and generate new ideas to fill these gaps (Tranfield et al., 2003; Briner and Denyer, 2012).

This article aimed to highlight the SPC and DoE implementation challenges described in the literature for managers, researchers, and practitioners interested in continuous production process improvement. The literature review was conducted in four phases, based on the eight review stages suggested by Briner and Denyer (2012), as shown in Table 2.1.

Table 2.1. Four phases and eight review stages of the literature review process (based on Briner and Denyer (2012)).

Phases	Review stages
To plan	1. Identify and clarify the addressed question(s)
	2. Determine the types of studies that will answer the question(s)
	3. Establish the audience
To conduct	4. Search the literature to locate relevant studies
	5. Sift through the studies and include or exclude following predefined criteria
To analyze	6. Extract the relevant information from the studies
	7. Classify the findings from the studies
To remember	8. Synthesize and disseminate the findings from the studies

Moreover, using Cooper's literature review taxonomy (Randolph, 2009), the review process's characteristics were outlined in the "to plan" phase, as follows:

- *focus*: to identify methods for SPC and DoE in continuous processes;
- *goal*: to classify central issues related to the identified methods;
- *perspective*: to present the review findings assuming a neutral position (i.e., reporting the results);
- *coverage*: to consider publications that are central or pivotal to achieving the goal;
- *organization*: to be structured around concepts (i.e., around the central issues identified by the findings of the literature review);
- *audience*: researchers and practitioners in the field, as well as top management.

## EMPIRICAL WORK AND FINDINGS

The other phases shown in Table 1 were conducted twice, in five steps, once for the SPC field and once for the DoE field. Specifically, the “to conduct” phase was realized in April 2017 using the Scopus database, limiting the search to publications in English in the past 30 years. Sequential searches were conducted using keywords and combined queries, such as (“statistical process control”) AND (“continuous process” OR “continuous production”) for the SPC literature searches, and (“design of experiments”) AND (“continuous process” OR “continuous production”) for the DoE literature searches. Starting from the search results (Step 1), the items were sequentially screened in two further steps (Steps 2 and 3), excluding all items not related to SPC and DoE applications in continuous processes and those that did not highlight potential challenges in applying SPC and DoE methods in continuous processes. Conference articles were excluded if a later journal article by the same authors and with the same title was found. In Step 4 (“to analyze” phase in Table 1), I classified the remaining publications in order to identify challenges or development needs for SPC and DoE in continuous production processes using a Microsoft Excel® worksheet and a color-coded system. In Step 5 (“to remember” phase in Table 1), I added publications known to be relevant, but that were missed by the searches. This final step provided the pivotal or central publications making up the representative sample on which the paper’s results were based.

### *Paper B*

Paper B relates to aim II of the research objective, aiming to overcome one of the challenges identified in the literature review, namely, how to run experiments in continuous processes operating under closed-loop control. The paper explores issues of experimental design and analysis in closed-loop environments, explaining how DoE can improve and optimize such processes for researchers and practitioners. Two experimental scenarios, using the decentralized TE process simulator (Ricker, 2005) as a testbed, exemplify the conceptual ideas outlined in the paper.

Design Expert® (version 9) was used to generate the experimental plans and to analyze the experimental results. The experiments were simulated using the Matlab Simulink® decentralized TE simulator (Ricker, 2005), together with Microsoft Excel® and Matlab scripts for extracting averages and saving results.

The first scenario explored the role of experimental factors acting as disturbances in closed-loop systems. A  $2^2$  randomized factorial design with three replicates was generated with the aim of estimating the location effects (main effects and the interaction) of two variables not involved in control loops on controlled variables and on associated manipulated variables.

The second scenario exemplified a screening design using the set-points of the controllers as experimental factors. A two-step sequential experiment was used to estimate the impact of the controllers’ set-points on the process operating cost. A  $2_{III}^{9-5}$  fully randomized fractional factorial design with four additional center points was



followed by a full-fold over in a new block to explain some aliased effects. The final design was of resolution IV.

In both experimental scenarios, the analysis of the numerical examples was based on calculating the averages of each experimental run in order to perform an analysis of variance (ANOVA). Vanhatalo et al. (2013) recommend removing apparent dynamic behavior at the beginning of each run to avoid a biased estimation of the effects. However, in the first experimental scenario, the initial observations were included to investigate whether the control loops were effective, because the control action may not succeed in immediately removing the impact on the controlled variable. On the other hand, in the second experimental scenario, a transition time of 24 hours was removed at the beginning of each run before calculating the run averages.

### *Paper C*

Paper C relates to both aim I and aim II of the research objective. The paper can be classified as a tutorial on the revised TE process simulator (Bathelt et al., 2015b), run using a decentralized control strategy, for researchers and practitioners who want to explore SPC and DoE in a continuous process operating in closed-loop.

The tutorial provides guidelines on the steps required to initialize the revised TE process simulator and to simulate data for SPC and DoE applications using flow charts. The flow charts were created using Bizagi modeler®, based on the business process modeling notation (BPMN) (e.g., see Chinosi and Trombetta (2012) and the BPMN archive (2011)). Furthermore, two simulated examples demonstrate the strategy for creating random variability in the simulator and potential SPC and DoE applications. The reader is referred to Paper C for further detail.

The first example demonstrates how closed-loop operations can affect the shift detection ability of control charts. Therefore, I used the Matlab Simulink® revised TE process model to simulate the Phase I and Phase II data, and the free software RStudio to analyze the collected data using a Hotelling  $T^2$  multivariate control chart.

The second example employs a response surface methodology approach, based on sequential experimentation, and using a subset of the controllers' set-points in the TE process to improve the overall performance indicator (i.e., the production cost). Here, Design Expert® (version 10) was used to generate the experimental designs and to analyze the experimental results. In addition, I used the revised Matlab Simulink® TE simulator together with Microsoft Excel® and Matlab Scripts® to simulate the experiments.

The sequential experimentation started with a  $2_V^{5-1}$  fully randomized fractional factorial design, with four additional center points in order to screen five controllers' set-points. Then, a central composite design was created by augmenting the resolution V fractional factorial design with 10 additional axial points, run in a new block, allowing for the estimation of a second-order model. The numerical optimization tool in Design

## EMPIRICAL WORK AND FINDINGS

Expert® (version 10) was used to search the design space to find the settings for the set-points that would produce the lowest predicted cost. Finally, three additional runs were simulated to confirm the results. During the sequential stages, the experimental results were analyzed using ANOVA tests by calculating the averages of each run after removing 24 hours of transition time, as suggested by Vanhatalo et al. (2013).

### 2.4. Summary of methodological choices

Table 2.2 provides an overview of the methodological choices made in the studies connected to the three appended papers and the papers' relationships with the research purposes.

Table 2.2. Overview of methodological choices in the appended papers.

	<b>Paper A</b>	<b>Paper B</b>	<b>Paper C</b>
<b>Aim of the research objective</b>	I	II	I, II
<b>Type of paper</b>	Literature review	Conceptual	Tutorial
<b>Target audience</b>	Managers, researchers, and practitioners	Researchers and practitioners	Researchers and practitioners
<b>Tool for data collection</b>	Scopus	TE process	Revised TE process
<b>Methods for data collection</b>	Searches using keywords and queries	DoE; Simulations	DoE; SPC; Simulations
<b>Data analysis</b>	Sequential screening and classification of publications identified during the searches	ANOVA	ANOVA; Multivariate Hotelling $T^2$ chart
<b>Illustration of the results</b>	Summary of classified SPC and DoE challenges in continuous processes	Two simulated examples	BPMN flow charts; Two simulated examples;
<b>Software used</b>	Microsoft Excel®	Design Expert® version 9; Matlab®; Matlab Simulink®; Microsoft Excel®;	Bizagi Modeler®; Design Expert® version 10; Matlab®; Matlab Simulink®; Microsoft Excel®; RStudio;

### 3. RESULTS

---

*This chapter summarizes the results presented in the three appended papers and links these results to the state-of-the-art in the research area. The chapter ends with a presentation of the main contributions of the research.*

#### 3.1. SPC and DoE in continuous processes

The technological advances characterizing today's production environments, such as those of continuous processes, requires adaptation and new development of SPC and DoE methods. High-tech operations, robots, the development of new and inexpensive sensors, and increased storage capacity provide an exorbitant amount of process data, requiring that researchers to adapt SPC and DoE methods to the challenges of the big data era (Vining et al., 2015).

Researchers must be aware of the needed research effort that this data rich environment brings to SPC and DoE so that practitioners can take full advantage of the methods by making proper adjustments. Moreover, these adjustments need managerial support, because they require resources and a company culture that understands the competitive advantage that SPC and DoE methods can offer (Hild et al., 1999; Bergquist, 2015b). In continuous production, any managerial attempt to improve products and processes in order to reduce waste, increase productivity, optimize resource consumption, or to produce in a sustainable way should consider SPC and DoE methods and support the adjustments necessary for the data rich environment. The author's view on the connection between these methods, challenges, and organizational decision-making is summarized in the thought map in Figure 3.1. That is, top management can find support in improving competitive advantage of companies by encouraging the use of SPC and DoE methods. Researchers and practitioners need to develop methods and provide answers to the challenges posed by continuous processes. These answers can support the decision-making process of top management, while development requires managerial support in joint efforts for continuous improvement.

The results of the literature review in Paper A are presented with this objective in mind. That is, the results help managers supporting the use of SPC and DoE methods, as well as making researchers and practitioners aware of SPC and DoE challenges in continuous processes, resulting in the need for methodological development.

## EMPIRICAL WORK AND FINDINGS

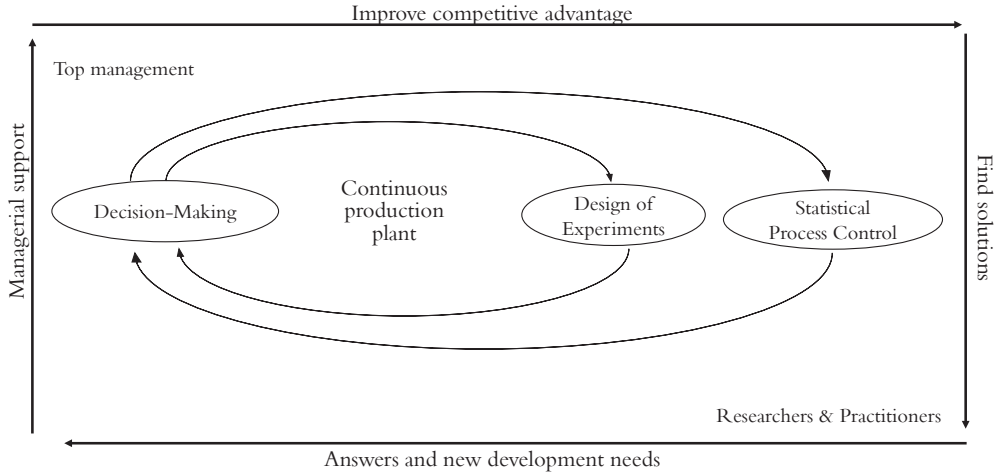


Figure 3.1. Connection between top managers and researchers, and practitioners to support SPC and DoE methods implementation and development.

### *SPC in continuous processes*

The existing literature in the SPC field recognizes the need for control charts that can handle multiple quality characteristics or multiple process variables simultaneously (Kourti and MacGregor, 1995). There are several options to consider here. The Hotelling  $T^2$  control chart is commonly used for multivariate data with 10 or fewer variables exhibiting moderate cross-correlation. For larger numbers of variables and observations in industrial processes, there are several multivariate SPC tools available (Shi and MacGregor, 2000; Qin, 2012; Ge et al., 2013). The choice of multivariate SPC methods should depend on assumed process characteristics: Gaussian/non-Gaussian, static/dynamic, and linear/non-linear. The dimensions and degree of autocorrelation of the data will also affect the choice. Ge et al. (2013) classify these methods into five categories: *Gaussian process monitoring methods* (e.g., latent structure variable methods, such as PCA/PLS), *non-Gaussian process monitoring methods* (e.g., independent component analysis), *non-linear process monitoring methods* (e.g., neural networks), *time varying and multimode process monitoring* (e.g., adaptive/recursive methods), and *dynamic process monitoring* (e.g., dynamic multivariate SPC methods). Among these methods, PCA-/PLS-based monitoring techniques are popular and important, having been used successfully in the process industry (e.g., see MacGregor and Kourti, 1995, Kourti et al., 1996, and Ferrer, 2014). Taking advantage of the many times high cross-correlation between process variables, SPC based on latent variable methods reduces the dimensions of the monitoring problem while retaining the majority of the content of the data (Frank and Friedman, 1993; Kourti and MacGregor, 1995). When data are both autocorrelated and cross-correlated, a recommended approach is to expand the data matrix by adding

## RESULTS

time-lagged versions of the original variables to transform the autocorrelation into cross-correlation. Performing PCA on this extended data matrix is called DPCA (Ku et al., 1995).

Typically, continuous processes provide multivariate autocorrelated and cross-correlated data that have been handled using PCA-/PLS-based monitoring techniques and their extensions. However, the literature review presented in Paper A highlighted some technical issues and development needs to improve the applicability of these methods. While the knowledge of the above-mentioned solutions can help managers to promote the adoption of these methods, researchers and practitioners should be aware of the following issues, because they need to be overcome. Relevant problems include the following:

- how to select the number of latent variables to retain and lags to add in DPCA (Himes et al., 1994; Ku et al., 1995; De Ketelaere et al., 2015; Vanhatalo et al., 2017),
- fault detection/isolation (Kourti and MacGregor, 1996; Dunia et al., 1996; Yoon and MacGregor, 2001), and
- how to handle outliers in data (Stanimirova et al., 2007; Serneels and Verdonck, 2008).

Moreover, Vanhatalo and Kulahci (2016) recently showed that control charts based on PCA can handle the cross-correlation, but that both PCA and its use in process monitoring are impacted by autocorrelation. Furthermore, autocorrelation affects the estimation of the covariance matrix, leading to an increased false alarm rate (Mastrangelo and Montgomery, 1995; Runger, 1996; Kulahci & Bisgaard, 2006). When assessing the process capability, the autocorrelation problem also extends to process capability analyses (Shore, 1997; Zhang, 1998; Sun et al., 2010; Lundkvist et al., 2012), but an extensive study of this issue is still lacking in the literature.

The results of paper A also show that another important SPC challenge comes from the need to monitor real continuous processes run under closed-loop operation. In a closed-loop operation, unwanted deviations in controlled process variables are mitigated by adjusting a manipulated variable (MacGregor and Harris, 1990; Hild et al., 2001). Therefore, closed-loop operations imply that the propagation of a disturbance through the process might not always be visible in the controlled response variable, but may instead be displaced to the manipulated variable. This behavior is well illustrated through the SPC example simulated in the revised TE process in Paper C. Here, the analysis of the results shows that the traditional approach of applying a control chart on (controlled) process output needs to be complemented with a control chart on the manipulated variables. The concurrent use of both control charts confirms the presence and effectiveness of the control system by analyzing the control chart for the controlled variables and identifying potential assignable causes by analyzing the control chart for the manipulated variables.

While the benefit of complementing SPC and EPC has already been explained in the literature (e.g., see Box and Kramer, 1992, Keats et al., 1996, and Box and Luceño, 1997), further research is needed to adjust the traditional SPC paradigm when EPC is in place.

### *DoE in continuous processes*

The results of the literature review conducted in paper A highlighted both the challenges and existing solutions when conducting experiments in continuous production processes. However, while the challenges affect all the experimental phases (i.e., planning, conducting, and analyzing the experiment), related solutions are not always available or, if available, need further development.

Following the recommendations of Coleman and Montgomery (1993) who highlight the critical importance of the planning phase, Vanhatalo and Bergquist (2007) provide a systematic approach to planning an industrial experiment in continuous processes. The authors provide a list of 12 important steps for the planning phase, where the need for both technical and organizational choices emerge due to the complexity of a large-scale experimentation. The choice of design preferences, need for restricted randomization, and factors levels are as critical as the need to assign responsibilities in coordinating the experiment or collecting relevant background information.

Vanhatalo and Bergquist (2007) also recommend identifying the presence of controlled variables in the planning phase, suggesting that closed-loop operations affect the entire experimentation strategy. In Paper A, experimentation under a closed-loop is classified as one of the important issues when conducting experiments in continuous processes, because conventional DoE methods implicitly assume open-loop operations (Montgomery, 2012b). Paper B explicitly focuses on exploring the use of DoE in closed-loop operations. Therefore, the reader is referred to Section 3.2, where experimentation under closed-loop operations is discussed in more detail.

Paper A also discusses issues that emerge when analyzing experiments conducted in continuous processes. Continuous processes are dynamic systems with inertia, meaning that the impact of an experimental factor change on the responses can take time to reach its full impact (Nembhard and Valverde-Ventura, 2003; Vanhatalo et al., 2010; Lundkvist and Vanhatalo, 2014). These transition times need to be considered in both the planning and the analysis phases. Moreover, responses from continuous processes are chronological sequences of observations, and their analysis relates to the field of time series (Bisgaard and Kulahci, 2011). He et al. (2015) review the methods that can be used to analyze dynamic process responses, and point out that there are several methods available to do so. These include, for example, functional analysis, time series analysis, and shape analysis. Nevertheless, it is fair to expect that these analysis methods will receive increasing attention from both academia and industry, owing to their considerable importance (He et al., 2015; Vining et al., 2015).

## RESULTS

Other challenges in the analysis phase are related to the need to use multivariate statistical analysis of the experiments. In continuous processes, the presence of many cross-correlated responses suggests that a univariate approach to analysis might be ineffective (MacGregor, 1997). PCA and PLS methods can be used to summarize the variation in the experimental response variables. Thus, the latent variables can be used as new responses to test the statistical significance of the effects of the experimental factors. El-Hagrasy et al. (2006), Vanhatalo and Vännman (2008), Baldinger (2012), and Souihi et al. (2013), among others, provide examples of multivariate analysis combined with DoE.

### 3.2. Experimentation in closed-loop operations

Continuous production processes often operate under closed-loop control for plant security and personal safety reasons (Box and MacGregor, 1974; 1976). As highlighted in Paper A, closed-loop operations affect all the experimental phases, i.e. planning, conducting, and analyzing the experiment. The continuous interference of the controllers indeed makes experimenting challenging as the control action can potentially eliminate the impact of the experimental factors on the process responses (Box and MacGregor, 1974; 1976). Paper B thus explores issues of experimental design and analysis in processes operating under closed-loop control and illustrates how DoE can help in improving and optimizing such processes.

Ogata (2010, p. 7) defines an open-loop system as “a system where the output is neither measured nor fed back for comparison with a desired target.” To each desired target, there corresponds a fixed operating condition and the accuracy of the system depends on calibration. In an open-loop system, the experimenter can thus observe the potential impact of experimental factors changes in the process response(s). In this case, the purpose of DoE is essentially to reveal potential causal relationships between the experimental factors, i.e. process inputs, and the process response(s) (Figure 3.2). Instead, Ogata (2010, p. 6) defines a closed loop system as “a system that maintains a prescribed relationship between the output and the desired target by comparing them and using the difference as a means of control.” In a closed-loop system, given that the response is a controlled variable, an automatic controller compares the process response measurements to a target value, the so-called set-point value, and adjusts the measured deviation regulating a manipulated variable, namely, a process input (Figure 3.2). That is, in a closed-loop system, the causal relationships between manipulated variable and (controlled) process response are already established and known and are normally not the focus of the designed experiment. Moreover, the manipulated variables involved in control loops cannot be considered as potential experimental factors as they are not free to vary independently such as in an experimental design.

Results of Paper B shows that experimentation under closed-loop operation can still be valuable. When controllers are in place, at least two experimental scenarios exist.

## EMPIRICAL WORK AND FINDINGS

In the first scenario, the experimenter can consider any set of system inputs not involved in control loops as potential experimental factors. In this case, both the manipulated and the controlled variables are interesting process responses. The analysis of the manipulated variables can reveal if experimental factors affect important process phenomena controlled in the loops, whereas the analysis of the controlled variables can provide information about the effectiveness of the controllers.

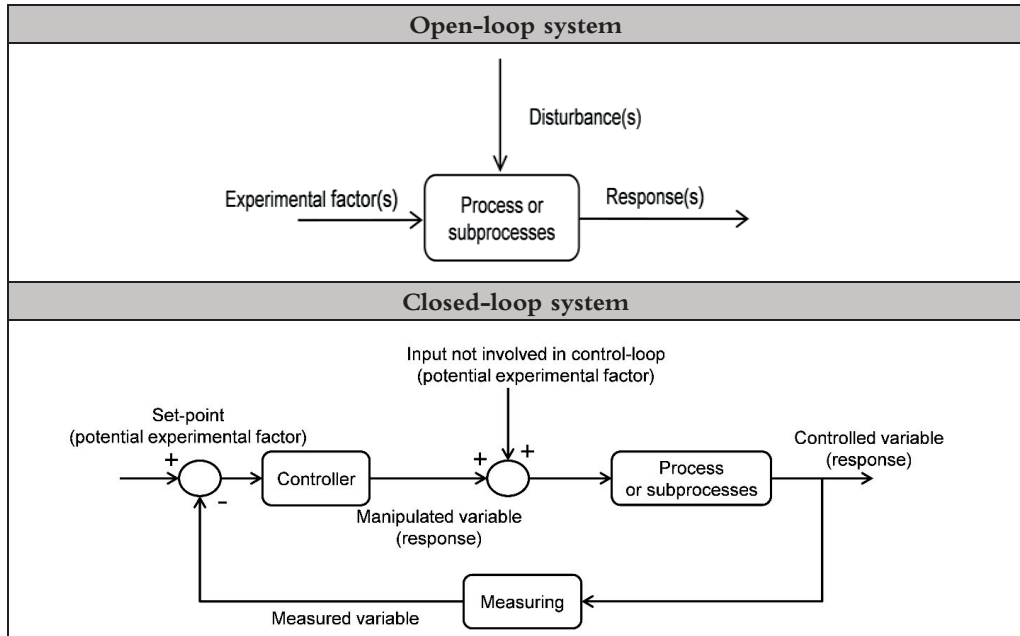


Figure 3.2. Schematic overview of an open-loop system and a closed-loop system (inspired by Ogata, 2010).

In the second scenario, the experimenter can change the set-point of the control loops (experimental factors) to study their effect on overall process performance indicators such as cost, quality and/or energy consumption (response). The suggested experimental scenarios can generate knowledge and contribute to process improvement in closed-loop systems, as they make it possible to study:

- the presence of the controllers if it suspected but unknown,
- the efficiency and cost effectiveness of the controllers,
- the impact of experimental factors on process phenomena, and
- how the set-points of the controlled loops affect process performance indicators.



## RESULTS

### 3.3. The TE simulator for SPC and DoE methods development

It is usually difficult to test new SPC or DoE methodological developments in real continuous process plants. Method development of SPC does not usually affect the production plan but the need to have datasets with specific characteristics, such as sample size, sampling time, and occurrence of a fault, could limit, or slow down the testing process. On the contrary, DoE method development may jeopardize the production plant and affect the production goals. Therefore, it may be inconvenient to invest time and money on lengthy experimental campaigns. Simulation tools can thus be instrumental in methodological research on both SPC and DoE.

Reis and Kennet (2017) map a wide variety of interactive resources and simulators that can be used to teach statistical methods such as Virtual Lab in Statistics, StatLab, PENSIM, RAYMOND, and many others. The authors classify the simulators based on three characteristics: (1) linear/non-linear elements in the simulation model, (2) time independent/dependent behavior, and (3) size of the simulator. In this classification scheme, the TE process is considered one of the most complex simulators as it is a large-scale, nonlinear, and dynamic simulator. Moreover, the TE process is open-loop unstable and needs to be run in closed-loop operations (Downs and Vogel, 1993).

The decentralized control strategy of the TE process (Ricker, 1996) is attractive from a SPC and DoE method development perspective because it can mimic the challenges frequently found in continuous processes (see Sections 1.3.1 and 1.3.2). Ricker (2005) devised a Matlab Simulink® model of the TE decentralized control strategy. Recently Bathelt et al. (2015a; 2015b) implemented a new version known as revised TE process simulator also using the Matlab Simulink® interface.

Table 3.1 provides a comparison of the main characteristics of the decentralized TE process simulator implementations by Ricker (2005) and Bathelt et al. (2015b).

Table 3.1. Comparison between the characteristics of the decentralized TE process by Ricker (2005) and by Bathelt et al. (2015b).

Characteristic	Decentralized TE process (Ricker, 2005)	Revised Decentralized TE process (Bathelt et al., 2015b)
Set simulation seed		X
Set simulation length and sampling frequency	X	X
Introduce process disturbances	X	X
Scale process disturbances		X
Monitor output of the disturbances		X
Random generator uses different state variables for process disturbances and measurements noise		X
Possibility to pause and resume the simulation using final process conditions	X	X
Repeatability of simulation results		X
Graphical User Interface	NA	NA

As shown in Table 3.1, the revised TE process simulator offers more flexibility to do simulations than the simulator originally developed by Ricker (2005). The new possibilities, described in Paper C, widen the usability of the revised TE process simulator making it more suitable for methodological tests of SPC and DoE methods. The results of the examples illustrated in Paper B and C, simulated using Ricker's simulator and the revised TE process respectively, highlight this concept as well. The possibility to change the seed of each simulation and to scale the disturbances make it possible introduce random variation needed for random simulations, which is essential for testing SPC and DoE methods. However, the revised TE process lacks a graphical user interface (GUI) and new users would probably find it challenging to understand the details of the revised TE process and to get it to run. Results of Paper C are illustrated with this awareness and aim to increase the possibility of interaction between a new user and the simulator. Therefore, flowcharts using the Business Process Modelling Notation (BPMN) provide a step-by-step description on how to get started with the simulator and simulate data for SPC and DoE applications.

### **3.4. Main contributions**

This research explicitly explores and describes challenges that are usually encountered when applying SPC and DoE methods in continuous production processes. Exploring and describing these challenges can be considered a contribution for supporting the use of SPC and DoE methods in continuous processes and for involving, on different levels, managers, practitioners, and researchers in their future methodological development.

An important contribution is a description of the benefits of using DoE methods in a closed-loop environment that is a different framework than the one usually found in textbooks or traditional DoE applications. Specifically, the results contribute to increase the understanding of DoE in these environments and to widen the applicability of DoE methods for closed-loop systems. The traditional open-loop experimental framework is adjusted to the closed-loop framework and implications are also discussed. In this adjustment, the two suggested closed-loop experimental strategies classify the potential experimental factors as either a set of system inputs not involved in control loops or the actual set-point of the control loops. In the former case, the manipulated variables and the controlled variables become the responses. In the latter case, typical responses include overall process performance indicators such as cost and/or quality.

A further contribution relates the use of SPC methods when engineering process control is place. Under closed-loop control, control actions may partly or completely displace the impact of a disturbance from the controlled variables to manipulated variables. The traditional approach of applying a control chart on the (controlled) process output then needs to be supplemented with a control chart on the manipulated variables.

Other important contribution of the research is to provide a detailed tutorial of a the TE process, a flexible simulation tool, that can be used to further develop SPC and

## RESULTS

DoE methods to overcome the challenges offered by continuous process environments. The TE process has been used for methodological work especially in the SPC field for a long time. However, the previous simulator's deterministic nature has likely hampered researchers in doing methodological work and making fair comparisons of methods. By following the ideas and recommendations presented in this research, the deterministic nature of the TE process can be overcome and its usability for method development substantially increased.



## **Part III: FUTURE RESEARCH**

*“Somewhere, something incredible is  
waiting to be known.”*

*Carl Sagan*



## 4. FUTURE RESEARCH DIRECTIONS

---

*This chapter presents ideas and new questions for future research that emerged during the research process. My plan is to pursue these future research directions as part of the PhD studies and beyond.*

I believe that SPC and DoE methods are far from being obsolete and that companies will not take full advantage of the big data transition without proper adjustments of these statistically based methodologies for learning and improvement. In the next years, it is reasonable to assume that the research community will increase the attention around the challenges offered by the modern manufacturing environments. Both methodological and application-oriented research is needed to effectively and efficiently solve problems coming from process industry.

The background and the development of this research allows me to undertake both methodological and applied future research directions. The fact that a more flexible simulator of the TE process is available allows more in-depth studies of the challenges described in the thesis such as multivariate nature of process data, process dynamics, closed-loop operations. Moreover, thus far most of the suggested multivariate SPC methods have been tested using pre-simulated training and testing data sets in the TE process. The characteristics of the revised TE simulator make it possible to revisit suggested multivariate SPC methods and to perform improved and more realistic comparative studies between existing methods or between existing and new methods.

At this stage of the research process, I personally would like to pursue the research related to the challenge of applying SPC and DoE methods in continuous processes under closed-loop operations. This direction is a natural continuation to build on the results already accomplished so far. My priority will be on development of DoE methods in closed-loop operation and therefore the next research steps planned are described below in three studies.

### *Study 1*

Paper B explores issues of experimental design and analysis in processes operating under closed-loop control. However, Paper B mainly focuses on *how* to conduct experiments in closed-loop systems and illustrate examples of *why* it may still be valuable to experiment in such process environments. I believe that the analysis of experimental results in these environments deserves further attention.

In Paper B, the experimental analysis was handled taking the time series nature of the data into account but used a simplified approach based on the run averages as scalar response values. However, analysis methods to model the dynamic relations between several experimental factors and the time series response(s) under closed-loop operations would be interesting to explore. From this perspective, the first scenario described in Paper B is of special interest (see also Section 3.2). By a more in-depth study of the control engineering theory, I would like to develop a two-step analysis method for

## FUTURE RESEARCH

analyzing two-level factorial designs with time series responses. In the first step, the closed-loop system will be transformed to the relative open-loop system by filtering back the effect of the control action on the controlled output. In the second step, the effect of the experimental factors changes on the manipulated variables, on the controlled output, and on the “back-filtered” output will be analyzed and compared. Analysis methods based on transfer function-noise models for a multiple input setting and the standard ANOVA approach will be compared.

The two-steps analysis method will first be performed on process data simulated by using a small-scale simulator in the Matlab Simulink® interface and then, if possible, using the revised TE process. Because this study requires knowledge of engineering control theory, system identification, and DoE, collaborations with PhD students or researchers who have a control engineering background is a potential way forward.

### *Study 2*

Further interesting research relates on how to adapt and apply sequential experimentation techniques, such as RSM and EVOP, in processes run under closed-loop control. Adjustments of the experimentation strategy are also needed in this framework. For example, the analyst needs to decide on which response variable(s) to optimize.

This research study will build on the results of Paper B and “Study 1”. The two experimental scenarios described in Paper B (see also Section 3.2) will be used to conduct experiments in closed-loop systems. RSM and EVOP methodologies will be then adjusted for the two experimental scenarios.

In the first scenario, input variables not involved in control loops will be the experimental factors. In this case, using the results of “Study 1”, the closed-loop will be transformed to the relative open-loop system. An optimization procedure based on the RSM or EVOP approach will be implemented for the “back-filtered” output. The best setting of the experimental factors will be the one that provides the response value closest to the target value (i.e. the set-point of the control loop). It is fair to assume that this strategy will minimize the needed control action and therefore the overall cost for controlling and stabilizing the process. Data will be analyzed using the most promising analysis method identified in “Study 1”.

In the second scenario, the set-points of the control loops will be the experimental factors. An optimization procedure based on the RSM or EVOP approach will be implemented for a process performance indicator such as process cost or product quality. Data will also be analyzed using the most promising analysis method identified in “Study 1”. The best setting of the experimental factors will be the one that optimize the process performance indicator.

For both scenarios, data will be simulated using a using a small-scale simulator in the Matlab Simulink® interface and, if possible, using the revised TE process.



### *Study 3*

The closed-loop operations environment also offers interesting research opportunities in the SPC field. In Paper C, the SPC example shows that the concurrent use of control charts based on controlled process outputs and manipulated variables is important to judge if a process operates in statistical process control. However, this result may be difficult to generalize because process variables may behave differently depending on the implemented control strategy and the type of fault. I think it would be interesting to understand how SPC charts behave for different faults and for different implemented control strategies. Such a “scenario analysis” could provide useful guidelines for practitioners on how to use SPC charts when different automated control strategies are in place and different kind of faults occur.

Process data for the different scenarios will be simulated using a small-scale simulator in the Matlab Simulink® interface and then, if possible, using the revised TE process.

## **Acknowledgments**

The author and the supervision team gratefully acknowledge the financial support from the Swedish Research Council under grant number 340-2013-5108.

## References

- Abdulmalek, F.A., Rajgopal, J. and Needy, K.L., (2006). A Classification Scheme for the Process Industry to Guide the Implementation of Lean. *Engineering Management Journal*, **18**(2): 15-25.
- APICS dictionary (2008). 12th ed. Edited by: Blackstone, J. H. Alexandria, VA, APICS-  
- The Association for Operations Management.
- Baldinger, A., Clerdent, L., Rantanen, J., Yang, M. and Grohgan, H., (2012). Quality by Design Approach in the Optimization of the Spray-Drying Process. *Pharmaceutical development and technology*, **17**(4): 389-397.
- Bathelt, A., Ricker, N.L. and Jelali, M., (2015a). Revision of the Tennessee Eastman Process Model. *IFAC-PapersOnLine*, **48**(8): 309-314.
- Bathelt, A., Ricker, N.L. and Jelali, M., (2015b). Tennessee Eastman Challenge Archive. Available: <http://depts.washington.edu/control/LARRY/TE/download.html>.
- Bergman, B. and Klefsjö, B., (2010). *Quality from Customer Needs to Customer Satisfaction*. Lund, Studentlitteratur AB.
- Bergquist, B., (2015a). Analysis of an Unreplicated  $2^2$  Factorial Experiment Performed in a Continuous Process. *Total Quality Management & Business Excellence*, **26**(9-10): 1083-1094.
- Bergquist, B., (2015b). Some Ideas on Why Factorial Designs are Seldom used for Full-Scale Experiments in Continuous Production Processes. *Total Quality Management & Business Excellence*, **26**(11-12): 1242-1254.
- Bergquist, B. and Albing, M., (2006). Statistical Methods – Does Anyone Really Use Them? *Total Quality Management*, **17**(8): 961-972.
- Bisgaard, S. and Khachatryan, D., (2011). Quasi-experiments on Process Dynamics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **60**(4): 497-517.
- Bisgaard, S. and Kulahci, M., (2011). *Time Series Analysis and Forecasting by Example*. Hoboken, New Jersey, John Wiley & Sons.
- Bisgaard, S. and Kulahci, M., (2005). Quality Quandaries: The Effect of Autocorrelation on Statistical Process Control Procedures. *Quality Engineering*, **17**(3): 481-489.
- Box, G.E.P., (1957). Evolutionary Operation: A Method for Increasing Industrial Productivity. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **6**(2): 81-101.
- Box, G.E.P. and Luceño, A., (1997). *Statistical Control by Monitoring and Feedback Adjustment*. New York, Wiley.

## REFERENCES

- Box, G.E.P. and MacGregor, J.F., (1976). Parameter Estimation with Closed-Loop Operating Data. *Technometrics*, **18**(4): 371-380.
- Box, G.E.P. and MacGregor, J.F., (1974). The Analysis of Closed-Loop Dynamic-Stochastic Systems. *Technometrics*, **16**(3): 391-398.
- Box, G.E.P. and Wilson, K.B., (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society. Series B (Methodological)*, **13**(1): 1-45.
- Box, G.E.P. and Kramer, T., (1992). Statistical Process Monitoring and Feedback Adjustment: A Discussion. *Technometrics*, **34**(3): 251-267.
- Box, G.E., Hunter, J.S. and Hunter, W.G., (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*. New York, John Wiley & Sons.
- Briner, R.B. and Denyer, D., (2012). Systematic Review and Evidence Synthesis as a Practice and Scholarship Tool. *Handbook of evidence-based management: Companies, classrooms and research*, 112-129.
- Chinosi, M. and Trombetta, A., (2012). BPMN: An Introduction to the Standard. *Computer Standards & Interfaces*, **34**(1): 124-134.
- Coleman, D.E. and Montgomery, D.C., (1993). A Systematic Approach to Planning for a Designed Industrial Experiment. *Technometrics*, **35**(1): 1-12.
- De Ketelaere, B., Hubert, M. and Schimitt, E., (2015). Overview of PCA-Based Statistical Process-Monitoring Methods for Time Dependent, High Dimensional Data. *Journal of Quality Technology*, **47**(4): 318-335.
- Deleryd, M., Garvare, R. and Bengt Klefsjö, (1999). Experiences of Implementing Statistical Methods in Small Enterprises. *The TQM Magazine*, **11**(5): 341-351.
- Dennis, D. and Meredith, J., (2000). An Empirical Analysis of Process Industry Transformation Systems. *Management Science*, **46**(8): 1058-1099.
- Downs, J.J. and Vogel, E.F., (1993). A Plant Wide Industrial Process Control Problem. *Computers & Chemical Engineering*, **17**(3): 245-255.
- Duchesne, C., Kourti, T. and MacGregor, J.F., (2002). Multivariate SPC for Startups and Grade Transitions. *AIChE Journal*, **48**(12): 2890-2901.
- Dunia, R., Qin, S.J., Edgar, T.F. and McAvoy, T.J., (1996). Identification of Faulty Sensors using Principal Component Analysis. *AIChE Journal*, **42**(10): 2797-2812.
- El-Hagrasy, A.S., D'Amico, F. and Drennen, J.K., (2006). A Process Analytical Technology Approach to Near-infrared Process Control of Pharmaceutical Powder Blending. Part I: D-optimal Design for Characterization of Powder Mixing and Preliminary Spectral Data Evaluation. *Journal of Pharmaceutical Sciences*, **95**(2): 392-406.

## REFERENCES

- Ferrer, A., (2014). Latent Structures-Based Multivariate Statistical Process Control: A Paradigm Shift. *Quality Engineering*, **26**(1): 72-91.
- Frank, I.E. and Friedman, J.H., (1993). A Statistical View of some Chemometrics Regression Tools. *Technometrics*, **35**(2): 109-135.
- Fransoo, J.C. and Rutten, W.G., (1994). A Typology of Production Control Situations in Process Industries. *International Journal of Operations & Production Management*, **14**(12): 47-57.
- Ge, Z., Song, Z. and Gao, F., (2013). Review of Recent Research on Data-Based Process Monitoring. *Industrial & Engineering Chemistry Research*, **52**(10): 3543-3562.
- Harris, T.J. and Ross, W.H., (1991). Statistical Process Control Procedures for Correlated Observations. *The Canadian Journal of Chemical Engineering*, **69**(1): 48-57.
- Hayes, R.H. and Wheelwright, S.C., (1979). Link Manufacturing Process and Product Life Cycles. *Harvard business review*, **57**(1): 133-140.
- He, Z., Zhou, P., Zhang, M. and Goh, T.N., (2015). A Review of Analysis of Dynamic Response in Design of Experiments. *Quality and Reliability Engineering International*, **31**(4): 535-542.
- Hild, C., Sanders, D. and Cooper, T., (2001). Six Sigma<sup>★</sup> on Continuous Processes: How and Why it Differs. *Quality Engineering*, **13**(1): 1-9.
- Hild, C., Sanders, D. and Ross, B., (1999). The Thought Map. *Quality Engineering*, **12**(1): 21-27.
- Himes, D.M., Storer, R.H. and Georgakis, C., (1994). Determination of the Number of Principal Components for Disturbance Detection and Isolation. *Proceedings of the American Control Conference*, **2**: 1279-1283.
- Keats, J.B., Montgomery, D.C., Runger, G.C. and Messina, W., (1996). Feedback Control and Statistical Process Monitoring. *International Journal of Reliability, Quality and Safety Engineering*, **03**(03): 231-241.
- Kourti, T., Lee, J. and Macgregor, J.F., (1996). Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control. *Computers and Chemical Engineering*, **20**: S745-S750.
- Kourti, T. and MacGregor, J.F., (1996). Multivariate SPC Methods for Process and Product Monitoring. *Journal of Quality Technology*, **28**(4): 409-428.
- Kourti, T., (2005). Application of Latent Variable Methods to Process Control and Multivariate Statistical Process Control in Industry. *International Journal of Adaptive Control and Signal Processing*, **19**(4): 213-246.

## REFERENCES

- Kourti, T. and MacGregor, J.F., (1995). Process Analysis, Monitoring and Diagnosis, using Multivariate Projection Methods. *Chemometrics and Intelligent Laboratory Systems*, **28**:3–21.
- Ku, W., Storer, R.H. and Georgakis, C., (1995). Disturbance Detection and Isolation by Dynamic Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **30**(1): 179–196.
- Kulahci, M. and Bisgaard, S., (2006). Challenges in Multivariate Control Charts with Autocorrelated Data. *ISSAT International Conference on Reliability and Quality in Design*, 215–219.
- Lager, T., (2010). Series On Technology Management-Vol. 17. *Managing Process Innovation: From Idea Generation to Implementation*. London, Imperial College Press.
- Lee, G., Han, C. and Yoon, E.S., (2004). Multiple-Fault Diagnosis of the Tennessee Eastman Process Based on System Decomposition and Dynamic PLS. *Industrial & Engineering Chemistry Research*, **43**(25): 8037–8048.
- Liu, K., Fei, Z., Yue, B., Liang, J. and Lin, H., (2015). Adaptive Sparse Principal Component Analysis for Enhanced Process Monitoring and Fault Isolation. *Chemometrics and Intelligent Laboratory Systems*, **146**:426–436.
- Lundkvist, P. and Vanhatalo, E., (2014). Identifying Process Dynamics through a Two-Level Factorial Experiment. *Quality Engineering*, **26**(2): 154–167.
- Lundkvist, P., Vännman, K. and Kulahci, M., (2012). A Comparison of Decision Methods for  $C_{pk}$  when Data are Autocorrelated. *Quality Engineering*, **24**(4): 460–472.
- MacGregor, J.F., (1997). Using on-Line Process Data to Improve Quality: Challenges for Statisticians. *International Statistical Review / Revue Internationale de Statistique*, **65**(3): 309–323.
- MacGregor, J.F. and Harris, T.J., (1990). Discussion. *Technometrics*, **32**(1): 23–26.
- MacGregor, J.F. and Kourti, T., (1995). Statistical Process Control of Multivariate Processes. *Control Engineering Practice*, **3**(3): 403–414.
- Mastrangelo, C.M. and Montgomery, D.C., (1995). SPC with Correlated Observations for the Chemical and Process Industries. *Quality and Reliability Engineering International*, **11**(2): 79–89.
- Montgomery, D., (2012a). *Statistical Process Control: A Modern Introduction*. Hoboken, New Jersey, Wiley.
- Montgomery, D.C., (2012b). *Design and Analysis of Experiments*. New York, John Wiley & Sons.

## REFERENCES

- Montgomery, D.C., Keats, J.B., Runger, G.C. and Messina, W.S., (1994). Integrating Statistical Process Control and Engineering Process Control. *Journal of Quality Technology*, **26**(2): 79-87.
- Nembhard, H.B. and Valverde-Ventura, R., (2003). Integrating Experimental Design and Statistical Control for Quality Improvement. *Journal of Quality Technology*, **35**(4): 406-423.
- Object Management Group Inc., (2011). Documents Associated with Business Process Model and Notation™ (BPMN™) Version 2.0. Available: <http://www.omg.org/spec/BPMN/2.0/>.
- Ogata, K., (2010). *Modern Control Engineering*. Boston, Pearson.
- Qin, S.J., (2012). Survey on Data-Driven Industrial Process Monitoring and Diagnosis. *Annual Reviews in Control*, **36**(2): 220-234.
- Randolph, J.J., (2009). A Guide to Writing the Dissertation Literature Review. *Practical Assessment, Research & Evaluation*, **14**(13): 1-13.
- Rato, T., Reis, M., Schmitt, E., Hubert, M. and De Ketelaere, B., (2016). A Systematic Comparison of PCA-based Statistical Process Monitoring Methods for High-dimensional, Time-dependent Processes. *AIChE Journal*, **62**(5): 1478-1493.
- Reid, R.D. and Sanders, N.R., (2012). *Operations Management. An Integrated Approach*. Hoboken, New Jersey, Wiley.
- Reis, M. and Kenett, R.S., (2017). A Structured Overview on the use of Computational Simulators for Teaching Statistical Methods. *Quality Engineering*, **29**(4): 730-744.
- Ricker, L.N., (2005). Tennessee Eastman Challenge Archive. Available: <http://depts.washington.edu/control/LARRY/TE/download.html>.
- Ricker, L.N., (1996). Decentralized Control of the Tennessee Eastman Challenge Process. *Journal of Process Control*, **6**(4): 205-221.
- Runger, G.C., (1996). Multivariate Statistical Process Control for Autocorrelated Processes. *International Journal of Production Research*, **34**(6): 1715-1724.
- Saunders, I.W. and Eccleston, J.A., (1992). Experimental Design for Continuous Processes. *Australian Journal of Statistics*, **34**(1): 77-89.
- Serneels, S. and Verdonck, T., (2008). Principal Component Analysis for Data Containing Outliers and Missing Elements. *Computational Statistics & Data Analysis*, **52**(3): 1712-1727.
- Shi, R. and MacGregor, J.F., (2000). Modeling of Dynamic Systems using Latent Variable and Subspace Methods. *Journal of Chemometrics*, **14**(5-6): 423-439.
- Shore, H., (1997). Process Capability Analysis when Data are Autocorrelated. *Quality Engineering*, **9**(4): 615-626.

## REFERENCES

- Souhihi, N., Dumarey, M., Wikström, H., Tajarobi, P., Fransson, M., Svensson, O., Josefson, M. and Trygg, J., (2013). A Quality by Design Approach to Investigate the Effect of Mannitol and Dicalcium Phosphate Qualities on Roll Compaction. *International Journal of Pharmaceutics*, **447**(1): 47-61.
- Stanimirova, I., Daszykowski, M. and Walczak, B., (2007). Dealing with Missing Values and Outliers in Principal Component Analysis. *Talanta*, **72**(1): 172-178.
- Sun, J., Wang, S. and Fu, Z., (2010). Process Capability Analysis and Estimation Scheme for Autocorrelated Data. *Journal of Systems Science and Systems Engineering*, **19**(1): 105-127.
- Tanco, M., Viles, E., Jesus Alvarez, M. and Ilzarbe, L., (2010). Why is Not Design of Experiments Widely used by Engineers in Europe? *Journal of Applied Statistics*, **37**(12): 1961-1977.
- Tranfield, D., Denyer, D. and Smart, P., (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, **14**(3): 207-222.
- Vanhatalo, E., Bergquist, B. and Vännman, K., (2013). Towards Improved Analysis Methods for Two-Level Factorial Experiment with Time Series Responses. *Quality and Reliability Engineering International*, **29**(5): 725-741.
- Vanhatalo, E., Kvarnström, B., Bergquist, B. and Vännman, K., (2010). A Method to Determine Transition Time for Experiments in Dynamic Processes. *Quality Engineering*, **23**(1): 30-45.
- Vanhatalo, E., (2009). On Design of Experiments in Continuous Processes. Luleå Tekniska Universitet.
- Vanhatalo, E. and Bergquist, B., (2007). Special Considerations when Planning Experiments in a Continuous Process. *Quality Engineering*, **19**(3): 155-169.
- Vanhatalo, E. and Kulahci, M., (2016). Impact of Autocorrelation on Principal Components and their use in Statistical Process Control. *Quality and Reliability Engineering International*, **32**(4): 1483-1503.
- Vanhatalo, E., Kulahci, M. and Bergquist, B., (2017). On the Structure of Dynamic Principal Component Analysis used in Statistical Process Monitoring. *Chemometrics and Intelligent Laboratory Systems*, **167**:1-11.
- Vanhatalo, E. and Vännman, K., (2008). Using Factorial Design and Multivariate Analysis when Experimenting in a Continuous Process. *Quality and Reliability Engineering International*, **24**(8): 983-995.
- Vining, G., Kulahci, M. and Pedersen, S., (2015). Recent Advances and Future Directions for Quality Engineering. *Quality and Reliability Engineering International*, **32**(3): 11-21.

## REFERENCES

- Webb, D.F., Lucas, J.M. and Borkowski, J.J., (2004). Factorial Experiments when Factor Levels are Not Necessarily Reset. *Journal of Quality Technology*, **36**(1): 1-11.
- Yoon, S. and MacGregor, J.F., (2001). Fault Diagnosis with Multivariate Statistical Models Part I: Using Steady State Fault Signatures. *Journal of Process Control*, **11**(4): 387-400.
- Zhang, N.F., (1998). Estimating Process Capability Indexes for Autocorrelated Data. *Journal of Applied Statistics*, **25**(4): 559-574.



## **PAPER A**

### **Managerial Implications for Improving Continuous Production Processes**

**Capaci, F., Vanhatalo, E., Bergquist, B.,  
and Kulahci, M. (2017)**

#### **Published as:**

Capaci, F., Vanhatalo, E., Bergquist, B., and Kulahci, M. (2017). Managerial Implications for Improving Continuous Production Processes. *Conference Proceedings, 24th International Annual EurOMA Conference: Inspiring Operations Management*, July 1-5, 2017, Edinburgh (Scotland).



# Managerial implications for improving continuous production processes

*Francesca Capaci ([francesca.capaci@ltu.se](mailto:francesca.capaci@ltu.se))  
Luleå University of Technology, Luleå, Sweden*

*Erik Vanhatalo  
Luleå University of Technology, Luleå, Sweden*

*Bjarne Bergquist  
Luleå University of Technology, Luleå, Sweden*

*Murat Kulahci  
Luleå University of Technology, Luleå, Sweden and,  
Technical University of Denmark, Kongens Lyngby, Denmark*

## Abstract

Data analytics remains essential for process improvement and optimization. Statistical process control and design of experiments are among the most powerful process and product improvement methods available. However, continuous process environments challenge the application of these methods. In this article, we highlight SPC and DoE implementation challenges described in the literature for managers, researchers and practitioners interested in continuous production process improvement. The results may help managers support the implementation of these methods and make researchers and practitioners aware of methodological challenges in continuous process environments.

**Keywords:** Productivity, Statistical tools, Continuous processes

## Introduction

Continuous production processes (CPPs), often found in, e.g., pulp and paper, chemical, steel, or other process industries, constitute a significant part of goods production. In a CPP, the product is gradually and often with minimal interruption refined through different process steps (Dennis and Meredith, 2000). Raw materials in these processes often stem directly from natural resources and characteristics of inputs such as ores or wood will therefore vary substantially. CPPs are often large-scale and tend to include interconnected process steps and complex flows. Continuous production environments are typically inflexible producing only one or a few products, require large investments, and occupy a large area. Wear and varying raw material characteristics are examples of frequent disturbances, making engineering process control (EPC) necessary to stabilize product quality and process characteristics (Montgomery et al., 1994; Box and Luceño, 1997). Although EPC keeps quality characteristics on target, CPPs require continuous improvements to remain competitive (Hild et al., 2001).

The main possibilities to learn and improve any process come from the analysis of observational and experimental process data. While first principles support correlations among observational data, process analyst usually needs experiments to discover causal relationships in industrial processes (Montgomery, 2012).

In this article, we focus on statistical process control (SPC) and design of experiments (DoE) since they constitute two fundamental process improvement methodologies. The purpose of SPC is to monitor the process and reduce process variation through identification and elimination of assignable causes of variation. In the SPC field univariate and multivariate control charts constitute the most important improvement tools. Alarms issued by control charts indicate the presence of potential assignable causes (i.e., unusual events). Root-cause analysis is the next step to uncover reasons for these events and if possible, to eliminate their causes. SPC is a long-term improvement methodology, while EPC is a short-term control strategy that transfers variability from the controlled variable to manipulated variables (MacGregor and Harris, 1990). The purpose of DoE is to plan, conduct and analyse experiments to improve products and processes in a systematic and statistically sound manner.

Since their introduction in the early twentieth century, management controlled improvement programs such as Robust Design, Total Quality Management, and Six Sigma have been promoting these methodologies. Their apparent omission from the currently popular lean program descriptions, as well as methods within popular data analytics and machine learning, indicate that textbook implementation of these methods may be ill-suited for today's production environment. It is becoming increasingly apparent that standard SPC and DoE methods need to be adapted to challenges such as rapid data collection from multiple and interconnected sources and massive datasets (Vining et al., 2015), which are common for CPPs. We argue that DoE and SPC are far from obsolete and that companies will not take full advantage of the big data transition without such proper statistically based methodologies for learning and improvements. However, practitioners must be aware of the challenges that this data rich environment brings to SPC and DoE.

McAfee et al. (2012) highlight leadership and decision-making as important management challenges in the big data era. If managers of CPPs understand SPC and DOE challenges, they can support pairing their data with effective improvement methods. Hild et al. (1999) suggest using thought maps to promote improvement methods and critical thinking. While managers need to be aware of techniques such as DoE and SPC to reduce resources, to meet customer requirements and, perhaps most important, they should also promote their use (Lendrem et al., 2001; Bergquist and Albing, 2006; Tanco et al., 2010).

The purpose of this article is to highlight challenges and development needs described in the literature for SPC and DoE in CPPs. We also provide some examples of state-of-the-art solutions to current challenges.

## **Method**

Literature searches were conducted in April 2017 using the Scopus database, limited to publications in English in the last 30 years (1987->). Table 2 and 3 show sequential search steps and keywords used. We examined reference lists of selected publications in Search 4 to minimize the risk of missing relevant publications, following recommendation by Randolph (2009).

*Table 2 – Search terms and number of publications in each step in the SPC search.*

Search #	Search terms and queries	Step 1	Step 2	Step 3	Step 4	Step 5
Search 1	("statistical process control") AND ("continuous process" OR "continuous production")	136	32	14	Classification	23 (7)
Search 2	("statistical process monitoring") AND ("continuous process" OR "continuous production")	16	2	0		
Search 3	("statistical process monitoring") AND ("process industry")	9	4	3		
Search 4	References of selected publications in Search 1, 2 and 3	436	64	35		

The initial sample from Step 1 is the number of publications found using the keywords in Scopus. Duplicates were deleted in each search. In Step 2, the initial sample was reduced by screening titles, author keywords, and sources. Conference articles were excluded if a later journal article of the same authors and with the same title was found. Many publications were rejected after abstracts were read in Step 3. We then classified challenges or development needs for DoE and SPC in CPPs in Step 4. Publications were further analysed in Step 5 to identify the central or pivotal publications on which our results are mainly based. Additional relevant publications known by the authors (indicated in brackets at Step 5 in Tables 2 and 3) were also added and analysed.

*Table 3 – Search terms and number of publications in each step in the DoE search.*

Search #	Search terms and queries	Step 1	Step 2	Step 3	Step 4	Step 5
Search 1	("design of experiments") AND ("continuous process" OR "continuous production")	49	27	8	Classification	20 (11)
Search 2	("experimental design") AND ("continuous process" OR "continuous production")	50	25	15		
Search 3	("experimental design") AND ("process industry")	12	7	2		
Search 4	References of selected publications in Search 1, 2 and 3	877	66	40		

### **SPC challenges in continuous production processes**

The literature review revealed many technical solutions to challenges arising when using SPC in continuous processes. The aim of this section is to provide an overview of challenges and potential strategies that managers can promote. Technical details are therefore not be completely covered in this article.

#### ***Process transitions and data acquisition***

Operating conditions frequently change due to grade changes, restarts or process adjustments and process inertia leads to transition phases. Data storage should be designed as to preserve the history of transitions phases and interrelation of process variables during transitions (Kourti, 2003). Process transitions may involve loss of production time and increased costs due to produced sub-grade products. The monitoring phase in SPC should begin after the transition is complete (Duchesne et al., 2002). Moreover, properly stored historical data is crucial to gain process knowledge.

#### ***Multivariate nature of process data***

Important reactions such as phase changes from ore to metal are difficult to measure accurately. Instead, engineers try to measure a multitude of secondary variables such as temperatures and pressures as proxies to the real, hidden process events. Technological development continuously reduces sensor costs and increases data storage capacity.

Today measuring, e.g., a reactor temperature at multiple locations is easily achieved. With many underlying phenomena, the analyst soon has hundreds of cross-correlated variables that need simultaneous monitoring. A univariate approach with each variable in separate control charts is inefficient and often misleading.

Fortunately, there are many multivariate SPC tools available (see, e.g., Shi and MacGregor, 2000; Qin, 2012, and Ge et al., 2013). These methods can be classified in five categories: *Gaussian process monitoring methods* (e.g. latent structure variable methods), *non-Gaussian process monitoring methods* (e.g. independent component analysis), *non-linear process monitoring methods* (e.g. neural networks), *time varying and multimode process monitoring* (e.g. adaptive/recursive methods), and *dynamic process monitoring* (e.g. dynamic multivariate SPC methods). The choice of multivariate SPC method depends on assumed process characteristics: Gaussian/non-Gaussian, static/dynamic, and linear/non-linear. Data characteristics such as if data are two or multidimensional or if data can be assumed to be time independent also affect the choice. An important multivariate process monitoring technique is to use a few linear combinations of the process variables (the so-called latent variables). Multivariate monitoring based on latent variables such as Principal Component Analysis (PCA) and Partial Least Square (PLS) are popular and important especially due to their dimensionality reduction properties (Frank and Friedman, 1993; MacGregor and Kourti, 1995). Kourti et al. (1996) provide a review of examples with industrial applications of latent variable monitoring techniques in process plants such as a chemical smelter, a polymerization process, a pulp digester, and others. Ferrer (2014) illustrates how latent variable methods for process understanding, monitoring and improvement can be used effectively in a petrochemical CPP. Latent variable techniques use the process variables' cross-correlation. Process monitoring uses a few linear combinations of the process variables (the so-called latent variables). Commonly, a Hotelling  $T^2$  control chart simultaneously monitors the retained latent variables from the PCA/PLS model whereas the squared prediction error ( $Q$ ) chart monitors the model's residuals. When the charts signal an out-of-control observation, these composite statistics are often decomposed into the original variables for fault identification (Himes et al., 1994; Ku et al., 1995; Kourti and MacGregor, 1996; Yoon and MacGregor, 2001; De Ketelaere et al., 2015)

### ***Serial correlation (autocorrelation)***

Process variables in CPPs are often highly (and positively) autocorrelated due to high sampling rates and process dynamics. This challenge is increasing due to sensor development and availability of almost unlimited data storages. Serial correlation usually means that the current observation is similar to the previous one. Since autocorrelation affects the estimation of the process' variability, autocorrelation can lead to increased false alarm rates in both univariate and multivariate control charts or incorrectly estimated process capability indices (Tracy et al., 1992; Runger, 1996; Mastrangelo et al., 1996; Bisgaard and Kulahci, 2005; Jarrett and Pan, 2007).

Two ways to handle SPC of multivariate, autocorrelated data have been suggested. The first employs a standard univariate or multivariate control chart but with adjusted control limits to achieve the desired in-control alarm rate. The second requires 'filtering out the autocorrelation' through a univariate or multivariate time series model and applying a control chart to the residuals from this model. However, fitting a multivariate time series model with many variables is difficult.

Latent variables based SPC is recommended for cases with multiple and highly cross-correlated process variables. Vanhatalo and Kulahci (2015) show that autocorrelated process variables still affect the monitoring performance of PCA based control charts

since the principal components also are autocorrelated. Control charts based on PCA/PLS are well equipped to deal with cross-correlated, independent, and stationary data but will be affected by autocorrelation. De Ketelaere et al. (2015) review extensions of PCA/PLS based monitoring methods available for more complex process and data characteristics, see Figure 1. Specifically, dynamic PCA/PLS have been promoted for handling the autocorrelation by adding time-lagged variables (Ku et al., 1995) to transform autocorrelation into the cross-correlation that is suitable for PCA/PLS.

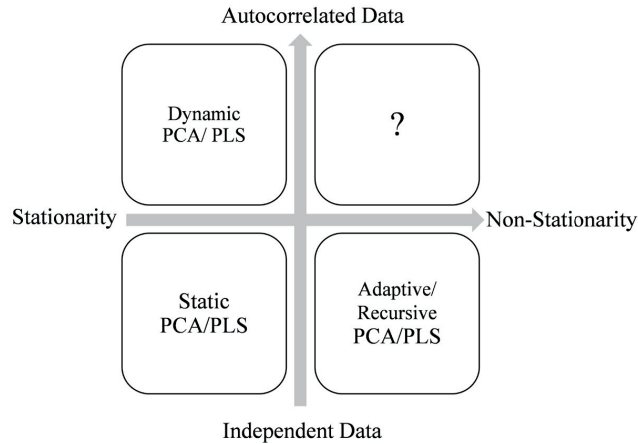


Figure 1 – Process and data challenges and available PCA/PLS methods.

Process capability analyses are important and popular for assessing process performance, frequently used in six sigma companies and promoted by various management and industrial systems standards. However, positive autocorrelation would lead to an overestimation of process capability indices (Shore, 1997; Zhang, 1998; Sun et al., 2010; Lundkvist et al., 2012).

The literature seems to lack a comprehensive solution to assessing process capability from processes with autocorrelated and multivariate data. Pan and Huang (2015) develop two multivariate process capability indices for autocorrelated data and compare their performance via a simulation study and, Mignoti and Oliveira (2011) propose an adjustment of multivariate capability indices to handle autocorrelation.

### ***Presence of engineering process control***

Fault detection using SPC control charts could fail when EPC is applied. Integrating SPC and EPC requires applying control charts to manipulated and not to controlled process variables. Box and Kramer (1992) provide a comprehensive discussion on the interface between EPC and SPC and Montgomery et al. (1994) demonstrate the effectiveness of integrating SPC and EPC in process surveillance. Contributions related to this challenge for most CPPs can also be found in Box and Luceño (1997), Janakiram and Keats (1998), Capilla et al. (1999), Tsung (2000) and in Huang and Lin (2002).

### **DoE challenges in continuous production processes**

The literature seems unanimous on the benefits of using DoE but also on the need of managerial support for increased use of DoE in industry (Tanco et al., 2009; Bergquist, 2015b). In this section, we describe specific challenges when applying DoE in CPPs but also suggest remedies.

### ***Large scale and costly experimentation***

Operations in CPP plants typically occur around the clock with few operators in charge. Full-scale experiments may thus involve the majority of the production staff, making managerial support, coordination, and information flow essential. Moreover, the often lengthy experimental campaigns can jeopardize the production plan. Previously unexplored factor settings may lead to production of low-grade products. Time and costs are therefore unavoidable constraints. Nevertheless, the need for improvements often make experimentation necessary. Relevant examples include Wormbs et al (2004) who describe experimentation to evaluate production methods of milk using a three factors, two-levels full factorial design in a dairy company and, Gonnissen et al. (2008) who show how a continuously produced powder mixture can be optimized using DoE.

We have found two best practices that managers can promote: (i) support and allocate resources to the planning phase of the experiment and (ii) create awareness of experimental strategies suitable for large scale experimentation.

Montgomery (2012) and Box et al. (2005) highlight the planning activities preceding the actual experiments. However, recognizing that the planning phase is seldom a taught skill, Coleman and Montgomery (1993) provide a systematic approach to plan an industrial experiment. Later, Vanhatalo and Bergquist (2007) adapt this approach to CPPs. Beside a well-chosen design, the planning phase should include, e.g., a clear problem statement, background such as expert knowledge or previous experiments, and someone responsible for coordination and information flow. Of special importance for CPPs is a list of experimental restrictions such as the number of possible experimental runs, easy/hard-to-change factors, randomization restrictions and design preferences.

Due to restrictions, cost, and time constraints, experiments in CPPs typically involve few factors, runs and replicates (Vanhatalo and Bergquist, 2007). Two-level (fractional) factorial designs are especially important to reduce the number of runs and factor level changes (Bergquist, 2015a). Box-Behnken designs also require few runs and are particularly suitable when extreme regions of the experimental space need to be avoided (Stazi et al., 2005; Kamath et al., 2011; Iyyaswami et al., 2013). Needs for restricted randomization, for instance to minimize transition times, may require split-plot designs (Sanders and Coleman, 1999; Bjerke et al., 2008; Vanhatalo and Vännman, 2008).

Response surface methodology (Box and Wilson, 1951; Myers et al., 2004) and evolutionary operation (Box, 1957) are two useful sequential experimental strategies when the goal is process optimization. Kvist and Thyregod (2005) demonstrate evolutionary operation for optimizing an industrial enzyme fermentation process.

### ***Closed loop process operation***

Applying EPC means running CPPs under closed-loop control, which complicates experimental design and analysis. Conventional DoE methods make the implicit assumption of open-loop operation in which effects of changes of experimental factors on responses may be studied directly. In closed-loop, many potentially interesting variables are kept around a certain values (set-points) to achieve desired product quality and/or for plant safety reasons. Potential effects of experimental factors on controlled variables are masked when manipulated variables are adjusted to counteract their deviations from set-points (Figure 2).

Capaci et al. (2017) suggest two closed-loop experimental strategies that classify the potential experimental factors as either a set of system inputs not involved in control loops or the actual control loop set-points, see Figure 2. In the former case, the manipulated



variables become the responses. The experimenter can also use controlled variables as responses to study controller effectiveness. In the latter case, typical responses include overall process performance indicators such as cost and/or quality.

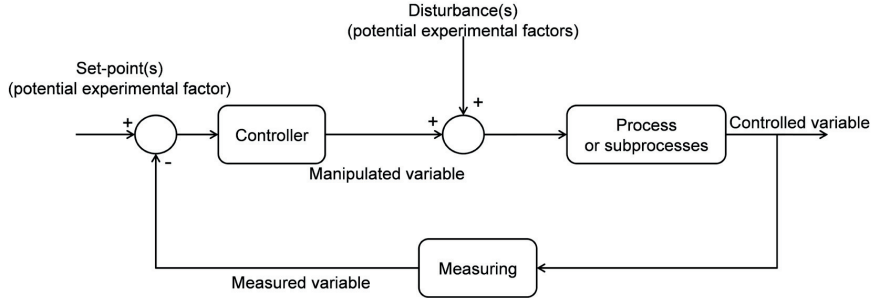


Figure 2. Schematic overview of process operating under closed-loop control

### ***Process transitions and time series responses***

High sampling frequencies in CPPs produce time series responses. Moreover, process dynamics often cause effects of experimental factors to develop gradually and then stabilize (Nembhard and Valverde-Ventura, 2003; Bisgaard and Khachatryan, 2011). These process transitions need consideration. Vanhatalo et al. (2013) develop possible analysis methods for experiments with time series responses. If the analyst can estimate the transition time (see for example Vanhatalo et al., 2010), the analyst can (i) use averages of the response in each run after eliminating transition time or (ii) use transfer function-noise modelling. However, transition times may prolong experimentation since it may be unclear when the process reaches steady state. Lundkvist and Vanhatalo (2014) apply a version of the second method to model time series of factors and responses of a full-scale blast furnace experiment. He et al. (2015) provide a recent review of additional available methods to analyse dynamic process responses in DoE.

### ***Multivariate responses***

Cross-correlations among responses often make multivariate analysis methods effective. Applications of multivariate projection methods such as PCA and PLS have been used to reduce the dimensionality and restrict the loss of information compared to univariate response analysis. A multivariate analysis approach also controls the Type I error rate. Vanhatalo and Vännman (2008) use principal components as new responses for a blast furnace experiment. El-Hagrasy et al. (2006), Baldinger (2012) and Souihi et al., (2013) provide additional multivariate analysis examples in DoE.

### ***Conclusions and discussions***

In this article, we focus our attention on discussing challenges of employing SPC and DoE for improving CPPs. Existing challenges do not mean that these methods cannot be used or should be discouraged. Similar or other challenges will be encountered also in other data analytics methods as in machine learning or neural networks. Managers of CPPs environments need to be aware that data-rich environments produce challenges for most employed methods. This is true also in applying SPC and DoE. We are aware that many of the mentioned challenges are not unique for CPPs and lie outside of the general managerial knowledge domain. A managerial implication is thus to guide analysts to a proper choice of tools by posing questions of how to address these challenges. We recommend that managers should solicit the competence of a statistically trained data

analyst until process engineers gain such competence. This is especially true during SPC method selection, or when designing and analysing experiments.

Our literature review has revealed challenges in using SPC and DoE in CPPs, but also many remedies to overcome those challenges. Applications of SPC in CPPs are often multivariate, need to deal with autocorrelation and process transitions, as well as to work alongside EPC procedures. DoE may need to deal with the large-scale, closed-loop operation and multivariate time series responses. An important message is also that SPC and DoE methods can be applied readily using proper adjustments presented in the literature. We also recommend managers to make sufficient resources available to engineers and analysts to adapt methods and to acquire software that can support application. Software are continuously developing to meet some of the challenges we highlight in this article. Examples of commercial software that can aid the application of SPC in CPPs are Prosensus® ([www.prosensus.com](http://www.prosensus.com)), Simca® ([www.umetrics.com](http://www.umetrics.com)), and Unscrambler X® ([www.camo.com](http://www.camo.com)). Available DoE software include JMP® ([www.jmp.com](http://www.jmp.com)), Design Expert® ([www.statease.com](http://www.statease.com)), and Modde® ([www.umetrics.com](http://www.umetrics.com)). For the more experienced analyst free software such as the R statistics software are interesting alternatives.

### Acknowledgments

The authors gratefully acknowledge the financial support from the Swedish Research Council under grant number 340-2013-5108.

### References

- Baldinger, A., Clerdent, L., Rantanen, J., Yang, M. and Grohgan, H., (2012), "Quality by design approach in the optimization of the spray-drying process", *Pharm. Dev. Technol.*, (17): 389-397.
- Bergquist, B., (2015a), "Analysis of an unreplicated 2<sup>2</sup> factorial experiment performed in a continuous process", *Total Qual. Manage. & Bus. Excel.*, (26): 1083-1094.
- Bergquist, B., (2015b), "Some ideas on why factorial designs are seldom used for full-scale experiments in continuous production processes", *Total Qual. Manage. & Bus. Excel.*, (26): 1242-1254.
- Bergquist, B. and Albing, M., (2006), "Statistical methods—does anyone really use them?", *Total Qual. Manage.*, (17): 961-972.
- Bisgaard, S. and Khachatryan, D., (2011), "Quasi-experiments on process dynamics", *J. of the Royal Stat. Soc.: Series C (Applied Statistics)*, (60): 497-517.
- Bisgaard, S. and Kulahci, M., (2005), "Quality quandaries: the effect of autocorrelation on statistical process control procedures", *Qual. Engng.*, (17): 481-489.
- Bjerke, F., Langsrud, Ø and Aastveit, A.H., (2008), "Restricted randomization and multiple responses in industrial experiments", *Qual. Reliab. Engng. Int.*, (24): 167-181.
- Box, G.E., Hunter, J.S. and Hunter, W.G., (2005), *Statistics for experimenters: design, innovation, and discovery*, Wiley: New York.
- Box, G.E.P., (1957), "Evolutionary Operation: a method for increasing industrial productivity", *J. of the Royal Stat. Soc.: Series C (Applied Statistics)*, (6): 81-101.
- Box, G.E.P. and Luceño, A., (1997), *Statistical Control By Monitoring and Feedback Adjustment*, Wiley: New York.
- Box, G.E.P. and Wilson, K.B., (1951), "On the experimental attainment of optimum conditions", *J. of the Royal Stat. Soc.: Series B (Methodological)*, (13): 1-45.
- Box, G.E.P. and Kramer, T., (1992), "Statistical process monitoring and feedback adjustment: a discussion", *Technometrics*, (34): 251-267.
- Capaci, F., Bergquist, B., Kulahci, M. and Vanhatalo, E., (2017), "Exploring the use of design of experiments in industrial processes operating under closed - loop control", *Qual. Reliab. Engng. Int.*, (In Press).
- Capilla, C., Ferrer, A., Romero, R. and Hualda, A., (1999), "Integration of statistical and engineering process control in a continuous polymerization process", *Technometrics*, (41): 14-28.
- Coleman, D.E. and Montgomery, D.C., (1993), "A systematic approach to planning for a designed industrial experiment", *Technometrics*, (35): 1-12.

- De Ketelaere, B., Hubert, M. and Schmitt, E., (2015), "Overview of PCA-based statistical process-monitoring methods for time dependent, high dimensional data", *J. of Qual. Tech.*, (4): 318-334.
- Dennis, D. and Meredith, J., (2000), "An empirical analysis of process industry transformation systems", *Management Science*, (46): 1058-1099.
- Duchesne, C., Kourti, T. and MacGregor, J.F., (2002), "Multivariate SPC for startups and grade transitions", *AIChE J.*, (48): 2890-2901.
- El-Hagrasy, A.S., D'Amico, F. and Drennen, J.K., (2006), "A Process Analytical Technology approach to near-infrared process control of pharmaceutical powder blending. Part I: D-optimal design for characterization of powder mixing and preliminary spectral data evaluation", *J. Pharm. Sci.*, (95): 392-406.
- Ferrer, A., (2014), "Latent structures-based multivariate statistical process control: A paradigm shift", *Qual. Engng.*, (26): 72-91.
- Frank, I.E. and Friedman, J.H., (1993), "A statistical view of some chemometrics regression tools", *Technometrics*, (35): 109-135.
- Ge, Z., Song, Z. and Gao, F., (2013), "Review of recent research on data-based process monitoring", *Ind. Engng. Chem. Res.*, (52): 3543-3562.
- Gonnissen, Y., Goncalves, S., De Geest, B., Remon, J. P. and Vervaet, C., (2008), "Process design applied to optimise a directly compressible powder produced via a continuous manufacturing process", *European J. of Pharm. and Biopharm.*, (68): 760-770.
- He, Z., Zhou, P., Zhang, M. and Goh, T.N., (2015), "A review of analysis of dynamic response in design of experiments", *Qual. Reliab. Engng. Int.*, (31): 535-542.
- Hild, C., Sanders, D. and Ross, B., (1999), "The thought map", *Qual. Engng.*, (12): 21-27.
- Hild, C., Sanders, D. and Cooper, T., (2001), "Six Sigma\* on continuous processes: how and why it differs", *Qual. Engng.*, (13): 1-9.
- Himes, D.M., Storer, R.H. and Georgakis, C., (1994), "Determination of the number of principal components for disturbance detection and isolation", *Proc. of the Am. Control Conf.*: 1279-1283.
- Huang, C. and Lin, Y., (2002), "Decision rule of assignable causes removal under an SPC-EPC integration system", *Int. J. Syst. Sci.*, (33): 855-867.
- Iyyaswami, R., Halladi, V.K., Yarramreddy, S.R. and Bharathaiyengar, S.M., (2013), "Microwave-assisted batch and continuous transesterification of karanja oil: process variables optimization and effectiveness of irradiation", *Biomass Conv. and Bioref.*, (3): 305-317.
- Janakiram, M. and Keats, J.B., (1998), "Combining SPC and EPC in a hybrid industry", *J. of Qual. Tech.*, (30): 189-200.
- Jarrett, J. E. and Pan, X., (2007), "The quality control chart for monitoring multivariate autocorrelated processes", *Comput. Stat. Data Anal.*, (51): 3862-3870.
- Kamath, H.V., Regupathi, I. and Saidutta, M., (2011), "Optimization of two step karanja biodiesel synthesis under microwave irradiation", *Fuel Process Tech.*, (92): 100-105.
- Kourti, T., (2003), "Abnormal situation detection, three-way data and projection methods; robust data archiving and modeling for industrial applications", *Annual Reviews in Control*, (27): 131-139.
- Kourti, T., Lee, J. and Macgregor, J.F., (1996), "Experiences with industrial applications of projection methods for multivariate statistical process control", *Comput. Chem. Engng.*, (20): S745-S750.
- Kourti, T. and MacGregor, J.F., (1996), "Multivariate SPC methods for process and product monitoring", *J. Qual. Tech.*, (28): 409-428.
- Ku, W., Storer, R.H. and Georgakis, C., (1995), "Disturbance detection and isolation by dynamic principal component analysis", *Chem. and Intel. Lab. Syst.*, (30): 179-196.
- Kvist, T. and Thyregod, P., (2005), "Using Evolutionary Operation to improve yield in biotechnological processes", *Qual. Reliab. Engng. Int.*, (21): 457-463.
- Lendrem, D., Owen, M. and Godbert, S., (2001), "DOE (design of experiments) in development chemistry: potential obstacles", *Organic Process Res. & Devel.*, (5): 324-327.
- Lundkvist, P. and Vanhatalo, E., (2014), "Identifying process dynamics through a two-level factorial experiment", *Qual. Engng.*, (26): 154-167.
- Lundkvist, P., Vännman, K. and Kulachi, M., (2012), "A comparison of decision methods for  $C_{pk}$  when data are autocorrelated", *Qual. Engng.*, (24): 460-472.
- MacGregor, J.F. and Harris, T.J., (1990), "Discussion", *Technometrics*, (32): 23-26.
- MacGregor, J.F. and Kourti, T., (1995), "Statistical process control of multivariate processes", *Control Engng. Pract.*, (3): 403-414.
- Mastrangelo, C. M., Runger, G. C. and Montgomery, D. C., (1996), "Statistical process monitoring with principal components", *Qual. Reliab. Eng. Int.*, (12): 203-210.
- McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D. and Barton, D., (2012), "Big data: the management revolution", *Harvard Bus. Rev.*, (90): 61-67.

- Mingoti, S.A. and Oliveira, Fernando Luiz Pereira de, (2011), "On capability indices for multivariate autocorrelated processes", *Brazilian J. of Op. & Prod. Manage.*, (8): 133-152.
- Montgomery, D.C., (2012), *Design and analysis of experiments*, Wiley: New York.
- Montgomery, D.C., Keats, J.B., Runger, G.C. and Messina, W.S., (1994), "Integrating statistical process control and engineering process control", *J. of Qual. Tech.*, (26): 79-87.
- Myers, R.H., Montgomery, D.C., Vining, G.G., Borror, C.M. and Kowalski, S.M., (2004), "Response Surface Methodology: a retrospective and literature survey", *J. of Qual. Tech.*, (36): 53-77.
- Nembhard, H.B. and Valverde-Ventura, R., (2003), "Integrating experimental design and statistical control for quality improvement", *J. of Qual. Tech.*, (35): 406-423.
- Pan, J. and Huang, W.K., (2015), "Developing new multivariate process capability indices for autocorrelated data", *Qual. Reliab. Engng. Int.*, (31): 431-444.
- Qin, S.J., (2012), "Survey on data-driven industrial process monitoring and diagnosis", *Annual Rev. in Control*, (36): 220-234.
- Randolph, J.J., (2009), "A guide to writing the dissertation literature review", *Pract. Assess. Res. & Eval.*, (14): 1-13.
- Runger, G., (1996), "Multivariate statistical process control for autocorrelated processes", *Int. J. Prod. Res.*, (34): 1715-1724.
- Sanders, D. and Coleman, J., (1999), "Considerations associated with restrictions on randomization in industrial experimentation", *Qual. Engng.*, (12): 57-64.
- Shi, R. and MacGregor, J.F., (2000), "Modeling of dynamic systems using latent variable and subspace methods", *J. of Chem.*, (14): 423-439.
- Shore, H., (1997), "Process capability analysis when data are autocorrelated", *Qual. Engng.*, (9): 615-626.
- Souhi, N., Dumarey, M., Wikström, H., Tajarobi, P., Fransson, M., Svensson, O., Josefson, M. and Trygg, J., (2013), "A quality by design approach to investigate the effect of mannitol and dicalcium phosphate qualities on roll compaction", *Int. J. Pharm.*, (447): 47-61.
- Stazi, F., Palmisano, G., Turconi, M. and Santagostino, M., (2005), "Statistical experimental design-driven discovery of room-temperature conditions for palladium-catalyzed cyanation of aryl bromides", *Tetrahedron Lett.*, (46): 1815-1818.
- Sun, J., Wang, S. and Fu, Z., (2010), "Process capability analysis and estimation scheme for autocorrelated data", *J. of Syst. Sc. and Syst. Engng.*, (19): 105-127.
- Tanco, M., Viles, E., Ilzarbe, L. and Jesus Alvarez, M., (2009), "Barriers faced by engineers when applying design of experiments", *The TQM J.*, (21): 565-575.
- Tanco, M., Viles, E., Jesus Alvarez, M. and Ilzarbe, L., (2010), "Why is not design of experiments widely used by engineers in Europe?", *J. of Appl. Stat.*, (37): 1961-1977.
- Tracy, N.D., Young, J.C. and Mason, R.L., (1992), "Multivariate control charts for individual observations", *J. of Qual. Tech.*, (24): 88-95.
- Tsung, F., (2000), "Statistical monitoring and diagnosis of automatic controlled processes using dynamic PCA", *Int. J. Prod. Res.*, (38): 625-637.
- Vanhatalo, E., Bergquist, B. and Vännman, K., (2013), "Towards improved analysis methods for two-level factorial experiment with time series responses", *Qual. and Rel. Engng. Int.*, (29): 725-741.
- Vanhatalo, E., Kvarnström, B., Bergquist, B. and Vännman, K., (2010), "A method to determine transition time for experiments in dynamic processes", *Qual. Engng.*, (23): 30-45.
- Vanhatalo, E. and Vännman, K., (2008), "Using factorial design and multivariate analysis when experimenting in a continuous process", *Qual. Reliab. Engng. Int.*, (24): 983-995.
- Vanhatalo, E. and Bergquist, B., (2007), "Special considerations when planning experiments in a continuous process", *Qual. Engng.*, (19): 155-169.
- Vanhatalo, E. and Kulahci, M., (2015), "Impact of autocorrelation on principal components and their use in statistical process control", *Qual. Reliab. Engng. Int.*, (32): 1483-1503.
- Vining, G., Kulahci, M. and Pedersen, S., (2015), "Recent advances and future directions for quality engineering", *Qual. Reliab. Engng. Int.*, (32): 863-875.
- Worms, G., Larsson, A., Alm, J., Tunklint-Aspelin, C., Strinning, O., Danielsson, E. and Larsson, H., (2004), "The use of Design of Experiment and sensory analysis as tools for the evaluation of production methods for milk", *Chem. and Intel. Lab. Syst.*, (73): 67-71.
- Yoon, S. and MacGregor, J.F., (2001), "Fault diagnosis with multivariate statistical models part I: using steady state fault signatures", *J. of Process Control*, (11): 387-400.
- Zhang, N.F., (1998), "Estimating process capability indexes for autocorrelated data", *J. of Appl. Stat.*, (25): 559-574.

## **PAPER B**

### Exploring the Use of Design of Experiments in Industrial Processes Operating Under Closed-Loop Control

**Capaci, F., Bergquist, B., Kulahci, M.,  
and Vanhatalo, E. (2017).**

#### **Published as:**

Capaci, F., Bergquist, B., Kulahci, M., and Vanhatalo, E. (2017). Exploring the Use of Design of Experiments in Industrial Processes Operating Under Closed-Loop Control. *Quality and Reliability Engineering International*. Published online ahead of print. DOI: 10.1002/qre.2128.



# Exploring the Use of Design of Experiments in Industrial Processes Operating Under Closed-Loop Control

Francesca Capaci,<sup>a,\*†</sup> Bjarne Bergquist,<sup>a</sup> Murat Kulahci<sup>a,b</sup> and Erik Vanhatalo<sup>a</sup>

Industrial manufacturing processes often operate under closed-loop control, where automation aims to keep important process variables at their set-points. In process industries such as pulp, paper, chemical and steel plants, it is often hard to find production processes operating in open loop. Instead, closed-loop control systems will actively attempt to minimize the impact of process disturbances. However, we argue that an implicit assumption in most experimental investigations is that the studied system is open loop, allowing the experimental factors to freely affect the important system responses. This scenario is typically not found in process industries. The purpose of this article is therefore to explore issues of experimental design and analysis in processes operating under closed-loop control and to illustrate how Design of Experiments can help in improving and optimizing such processes. The Tennessee Eastman challenge process simulator is used as a test-bed to highlight two experimental scenarios. The first scenario explores the impact of experimental factors that may be considered as disturbances in the closed-loop system. The second scenario exemplifies a screening design using the set-points of controllers as experimental factors. We provide examples of how to analyze the two scenarios. © 2017 The Authors Quality and Reliability Engineering International Published by John Wiley & Sons Ltd

**Keywords:** Design of Experiments; engineering control; feedback adjustment; simulation; Tennessee Eastman process

## 1. Introduction

Industrial processes often involve automated control systems to reduce variation of quality characteristics or variables affecting plant safety. Sometimes, the control relies on human intervention, such as subjective evaluation of the process state followed by an operator's control action. Processes operating under such control regimes are operating under some form of closed-loop control. Experimenting in these processes will be challenging due to controllers' continuous interference, see Box and MacGregor.<sup>1,2</sup> Because the control action will potentially eliminate the impact of experimental factor changes, experimentation in closed-loop systems may be seen as futile. However, we argue that well designed and properly analyzed experiments run under such conditions can yield valuable information.

This article relates to system identification, which aims at building mathematical models of dynamic systems based on observed data from the system, see Ljung.<sup>3</sup> Experimental design in that sense typically concerns the selection of a proper input signal disturbance to discover the causal relationships between the disturbance and the responses or manipulated variables. This way, system identification allows for the estimation of model parameters to optimize a feedback controller, see, e.g. Jansson.<sup>4</sup> Typically, experimental design research in the system identification field studies 'optimal' input signals to model the system.

In this article, we are primarily concerned with factor screening, factor characterization or process improvement and optimization rather than modeling process dynamics through factors that are already known to affect the response. Similar to system identification experiments, allowable factor ranges are usually restricted, the experiments could be run in full-scale production and the number of experimental runs are limited. However, compared to system identification, the experiments we consider are run for longer periods of time and, most importantly, they have a more overarching purpose of improving or optimizing a process rather than to guarantee stability of a control loop.

<sup>a</sup>Luleå University of Technology, Luleå, Sweden

<sup>b</sup>Technical University of Denmark, Kongens Lyngby, Denmark

\*Correspondence to: Francesca Capaci, Luleå University of Technology, Luleå, Sweden.

†E-mail: francesca.capaci@ltu.se

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Closed-loop environments add complexity to experimental design and analysis because the control strategy affects the choice of experimental factors. For example, some input variables are manipulated within control loops and therefore may not be suitable as experimental factors. Moreover, even though closed-loop operation is common, we argue that Design of Experiments (DoE) literature typically rests on the implicit assumption that the studied system is operating in open loop, hence allowing the experimental factors to freely affect the response(s). However, as pointed out by, e.g. Vanhatalo and Bergquist<sup>2</sup> and Hild *et al.*,<sup>6</sup> process control systems are designed to maintain the important process variables at their set-points with low variability. Hence, control loops may counteract deliberate changes of experimental factors and thereby displace the effect from typical responses to manipulated variables. An analysis implication is that these manipulated variables instead may have to be used as responses to understand the experimental factors' impact on the system.

The purpose of this article is therefore to explore experimental design and analysis issues in processes operating under closed-loop control and to illustrate how DoE can add value in improving or optimizing such processes. We will pursue this through the help of a process simulator. Process simulators in general have limitations in mimicking the behavior of a real process, but they also offer the flexibility required for methodological developments without jeopardizing plant safety or product quality.

A well-known simulator in the engineering control community is the Tennessee Eastman (TE) challenge chemical process simulator first described by Downs and Vogel.<sup>7</sup> The TE simulator has been primarily used in the development of different process control strategies and for the development of statistical process monitoring methods mainly in chemometrics literature, see for example Kruger *et al.*<sup>8</sup> In this article, we run the TE process with a decentralized control strategy to simulate and illustrate experiments in a closed-loop system.

The remainder of this article is organized as follows: Section 2 establishes important concepts and provides a general comparison of open loop and closed-loop systems from a DoE perspective. Section 3 provides a general description of the TE process simulator and the chosen control strategy. Section 3 also outlines the two experimental scenarios we illustrate in closed-loop operation of the process. The experimental scenarios are elaborated and analyzed in Sections 4 and 5, respectively. Finally, conclusions and discussion are provided in Section 6.

## 2. Experiments run in open vs. closed-loop systems

Experiments imply that one or many input variables (experimental factors) are allowed to vary to affect the output (response(s)) with the aim of revealing potential causal relationships (effects) between factors and responses, and providing estimates of these effects. The response could be also affected by random disturbances, see Figure 1.

In a process operating under closed-loop control, unwanted variable deviations are mitigated by adjusting a manipulated variable, see Figure 2.

From an experimental perspective, the manipulated variables involved in control loops are not potential experimental factors. In fact, because manipulated variables are involved in control loops, the control engineers have an idea, e.g., from a past experiment, how the manipulated variables affect the response. In relation to Figure 2, the experimental factors in a closed-loop setting should be viewed as disturbances to the system operating under closed-loop control. The potential effects of a disturbance on the controlled variable(s) are therefore typically masked and displaced to one or several manipulated variables if the control system is working

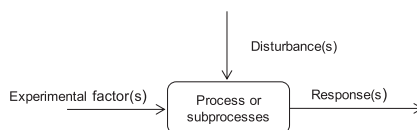


Figure 1. Experimental paradigm for open-loop operation. Figure inspired by Montgomery.<sup>9</sup>

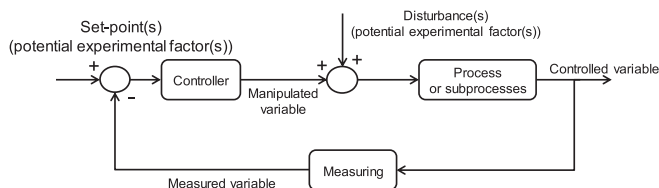


Figure 2. Schematic overview of a process under closed-loop control.



properly. This constitutes the first message we would like to convey in this article. That is, if the control action is ignored, the experimental factor changes will likely not affect the response (the controlled variable) significantly. An erroneous conclusion from the lack of detectable reaction would then be, depending on the effectiveness of the control action, that the factor is unimportant. However, if the presence of the controller is suspected or known, controlled variables may be used as responses primarily to test the presence and the effectiveness of the controllers. Manipulated variables may thus be considered as responses to study the impact of the experimental factors on the system and its dynamics due to the displacement of the potential effects from controlled to manipulated variables.

We classify experimental factors for processes operating under closed-loop control as (i) either a set of system inputs not involved in any control loop (should be viewed as disturbances in Figure 2) or (ii) the actual set-point values in the control loops. In the former scenario, both the manipulated and controlled variables can be used as experimental responses, while in the latter case more natural responses may be overall process performance indicators such as cost and/or product quality.

### 3. The Tennessee Eastman process simulator

Downs and Vogel<sup>7</sup> introduced the TE chemical process simulator for studying and developing engineering control design. The process is open loop unstable meaning that it will deviate and stop after a certain time period without any active control. With an appropriate control strategy, however, the process will remain stable. Several different control strategies for the TE process have been proposed; see for example McAvoy,<sup>10</sup> Lyman and Georgakis,<sup>11</sup> and Ricker.<sup>12</sup> The TE process has also been used as a test-bed for methodological development of multivariate statistical process monitoring.<sup>8,13–16</sup>

In the remainder of this section, we will describe some of the details of the TE process to facilitate the understanding of the experimental scenarios we use.

#### 3.1. Process description

The TE process is a chemical process for which the components, kinetics, processing and operating conditions have been modified for proprietary reasons, see Downs and Vogel.<sup>7</sup> Following four irreversible and exothermic reactions, the process produces two liquid products from four gaseous reactants. With an additional byproduct and an inert product, eight components are present in the process. The process has five major unit operations: a reactor, a product condenser, a vapor–liquid separator, a recycle compressor and a product stripper as shown in a simplified process overview in Figure 3. A more detailed process map is given in the original reference.<sup>7</sup>

The physical inputs to the process consist of four gaseous streams, out of which three are fed to a reactor. After the reaction, the product mixture flows into a condenser, in which most of the gas is condensed. Some non-condensable components remain as vapors and the two phases are separated in the vapor–liquid separator. Vapor is partially recycled and purged together with the inert product and the byproduct. The product stripper separates remaining reactants from the products. The reactants are recycled, and the products exit the process from the stripper.

The TE process simulator has 12 manipulated variables (XMV) and 41 measured variables (XMEAS). Out of 41 measured variables, 22 are measured directly while the remaining 19 variables can be calculated by the composition of the directly measured streams. In addition to XMV and XMEAS, operating costs, production and product quality data are also recorded.

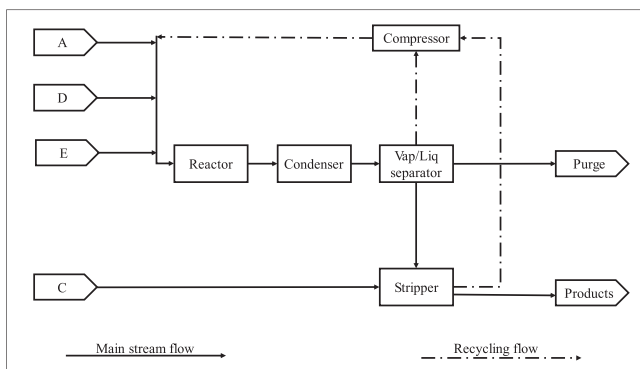


Figure 3. A schematic overview of the TE process.

The TE process has six different operating modes based on the production ratio of the two products and the production rate. Mode 1 is the most commonly used base case in research articles, which we also employ in this article. Five operating constraints need to be fulfilled to avoid process shutdown. There is also a possibility to activate 20 pre-set process disturbances (IDVs) during process operation. Downs and Vogel<sup>7</sup> provide more information on manipulated and measured variables, operating constraints, disturbances and the different operating modes.

### 3.2. Implemented process control strategy

A control strategy is a prerequisite for experimentation in the TE process because it is open loop unstable. Ricker<sup>12</sup> developed a decentralized control strategy for the TE process for improved performance, especially for the maximization of the production rate. The decentralized approach partitions the TE plant into 19 sub-units to each of which a controller is added. Tables I and II list the control loops, controlled variables, their set-points and manipulated variables. Note that we also provide  $XMV(i)$  and  $XMEAS(j)$ ; the  $i^{\text{th}}$  manipulated variable and the  $j^{\text{th}}$  measured variable given in Tables III and IV of the original article by Downs and Vogel<sup>7</sup> for ease of comparison. The manipulated variables listed with different codes, such as  $F_p$ ,  $r_7$  etc. come from the decentralized control strategy settings given in Ricker.<sup>12</sup>

We use a Matlab/Simulink decentralized control simulator (available at: [http://depts.washington.edu/control/LARRY/TE/download.html#MATLAB\\_5x](http://depts.washington.edu/control/LARRY/TE/download.html#MATLAB_5x)). In this configuration, all constraints are satisfied and the process can operate without undesired shutdowns. Moreover, the set-point values for some controlled variables and the values of inputs (X MVs) not involved in control loops may be varied, thereby allowing for experimentation.

The override loops 18 and 19 are exceptions to the control procedure described in Section 2. These control loops are only active when abnormal conditions occur that require an operating strategy change. Severe disturbances such as an introduction of the feed loss of A (IDV 6) activate the override loops. The production index  $F_p$  and the compressor recycle valve  $XMV(5)$  are not manipulated when the process operates without disturbances. All variables that can be manipulated except for the stripper steam valve  $XMV(9)$  and the agitator speed  $XMV(12)$  are involved in control loops in the decentralized control strategy. Consequently,  $XMV(9)$  and  $XMV(12)$  may be varied during experimentation and should then be viewed as disturbances in Figure 2.

### 3.3. Chosen experimental scenarios in the TE process

Two experimental scenarios in the TE process will illustrate experimentation in a process under closed-loop control. The first scenario will demonstrate an experiment when the system is disturbed by experimental factors. Input variables not involved in control loops can act as such disturbances and therefore be defined as experimental factors. The second scenario will demonstrate the use of the set-points of the control loops as experimental factors.

**3.3.1. Scenario 1.** The aim of this scenario is to demonstrate and visualize how experimental factor variation effects are distributed among the controlled and manipulated variables and how these effects can be analyzed. Recall that the stripper steam valve  $XMV(9)$

**Table I.** Control loops for the decentralized control strategy (Ricker<sup>12</sup>)

Loop	Controlled variable		Manipulated variable	
	Name	Code	Name	Code
1	A feed rate (stream 1)	XMEAS(1)	A feed flow	XMV(3)
2	D feed rate (stream 2)	XMEAS(2)	D feed flow	XMV(1)
3	E feed rate (stream 3)	XMEAS(3)	E feed flow	XMV(2)
4	C feed rate (stream 4)	XMEAS(4)	A and C feed flow	XMV(4)
5	Purge rate (stream 9)	XMEAS(10)	Purge valve	XMV(6)
6	Separator liquid rate (stream 10)	XMEAS(14)	Separator pot liquid flow	XMV(7)
7	Stripper liquid rate (stream 11)	XMEAS(17)	Stripper liquid product flow	XMV(8)
8	Production rate (stream 11)	XMEAS(41)	Production index	$F_p$
9	Stripper liquid level	XMEAS(15)	Ratio in loop 7	$r_7$
10	Separator liquid level	XMEAS(12)	Ratio in loop 6	$r_6$
11	Reactor liquid level	XMEAS(8)	Set-point of loop 17	s.p. 17
12	Reactor pressure	XMEAS(7)	Ratio in loop 5	$r_5$
13	Mol % G (stream 11)	XMEAS(40)	Adjustment to the molar feed rate of E	$E_{\text{adj}}$
14	Amount of A in reactor feed, $y_A$ (stream 6)	XMEAS(6)	Ratio in loop 1	$r_1$
15	Amount of A + C in reactor feed, $y_{AC}$ (stream 6)	XMEAS(6)	Sum of ratio in loop 1 and 4	$r_1 + r_4$
16	Reactor temperature	XMEAS(9)	Reactor cooling water flow	XMV(10)
17	Separator temperature	XMEAS(11)	Condenser cooling water flow	XMV(11)
18	Maximum reactor pressure	XMEAS(7)	Production index	$F_p$
19	Reactor level override	XMEAS(8)	Compressor recycle valve	XMV(5)

**Table II.** Set-point values in the control loops for the decentralized control strategy (Ricker<sup>12</sup>)

Loop	Controlled variable	Set-point	
		Base case values	Units
1	A feed rate (stream 1)	0.2505	kscmh
2	D feed rate (stream 2)	3664.0	kg h <sup>-1</sup>
3	E feed rate (stream 3)	4509.3	kg h <sup>-1</sup>
4	C feed rate (stream 4)	9.3477	kscmh
5	Purge rate (stream 9)	0.3371	kscmh
6	Separator liquid rate (stream 10)	25.160	m <sup>3</sup> h <sup>-1</sup>
7	Stripper liquid rate (stream 11)	22.949	m <sup>3</sup> h <sup>-1</sup>
8	Production rate (stream 11)	100	%
9	Stripper liquid level	50	%
10	Separator liquid level	50	%
11	Reactor liquid level	75	%
12	Reactor pressure	2705	kPa
13	Mol % G (stream 11)	53.724	mol%
14	Amount of A in reactor feed, $y_A$ (stream 6)	54.95	%
15	Amount of A + C in reactor feed, $y_{AC}$ (stream 6)	58.57	%
16	Reactor temperature	120.40	°C
17	Separator temperature	80.109	°C
18	Maximum reactor pressure	2950	kPa
19	Reactor level override	95	%

**Table III.** Potential experimental factors in scenario 1. Input variables not involved in control loops

Variable name	Code	Base case value (%)	Low limit (%)	High limit (%)
Compressor recycle valve	XMV(5)	22.210	0	100
Stripper steam valve	XMV(9)	47.446	0	100
Agitator speed	XMV(12)	50.000	0	100

**Table IV.** Potential experimental factors of the TE process: set-point values of the control loops

Loop	Controlled variable	Base set-point
7	Stripper liquid rate (production)	22.949 m <sup>3</sup> h <sup>-1</sup>
9	Stripper liquid level	50%
10	Separator liquid level	50%
11	Reactor liquid level	75%
12	Reactor pressure	2705 kPa
13	Mole % G	53.724 mol%
14	Amount of A in reactor feed ( $y_A$ )	54.95%
15	Amount of A + C in reactor feed ( $y_{AC}$ )	58.57%
16	Reactor temperature	120.40 °C

and the agitator speed XMV(12) are the only two manipulated variables not involved in control loops. Moreover, if the process is run without introducing any of the pre-set disturbances (IDVs), the compressor recycle valve XMV(5) is not manipulated and can be considered as another possible experimental factor. Because the TE simulator is designed the way it is, these factors not involved in control loops can be seen as potential experimental factors (disturbances), and an experiment can be designed to evaluate their impact on the system. We would like to note that in a real process the experimental factors need not only come from a list of numeric input variables not involved in control loops but can rather be drawn from a variety of potential disturbances to the system, such as different raw materials, methods of operation etc. Our choice here is convenient because XMV(5, 9, and 12) can be changed rather easily in the simulation model.

Three experimental factors are thus available in this scenario. Response variables will be the controlled variables as well as the manipulated variables in the control loops (see Section 2). Table III presents base case values of XMV(5, 9 and 12) and their allowed ranges in operating Mode 1 of the TE process.

3.3.2. *Scenario 2.* The aim of this scenario in the TE process is to explore the set-points of the controllers to reveal their potential impact on the process operating cost. That is, to see causal relationships between the process' operating conditions and an important process performance indicator. By changing the set-points, the second experimental scenario indirectly uses the levels of the controlled variables as experimental factors. However, some of the set-points are actually controlled in a cascaded procedure based on directives generated by other controllers. Thus, only a subset of the controlled variables may be considered potential experimental factors. Table IV lists the controlled variables that may be used as potential experimental factors and their set-point values for operating Mode 1.

## 4. Scenario 1: design and analysis

This section and Section 5 through examples illustrate the two experimental scenarios explained above. We would like to clarify that the aim of these examples is not to show the 'best' experimental designs or analysis procedures but rather to illustrate issues related to experimentation in closed-loop operation.

### 4.1. A two-level factorial design

Scenario 1 involves a  $2^2$  randomized factorial design with three replicates with the aim of estimating location effects (main effects and interaction) of the stripper steam valve XMV(9) and of the agitator speed XMV(12) on controlled variables and associated manipulated variables. Control loops 9, 10, 11, 12 and 16 (see Table I) include constraints implemented for securing plant safety and adequate control actions to avoid shutdown.

The run-order of the experiments and the averages of the controlled and manipulated variables are given in Table V. The TE process was run for 36 h under normal operating conditions, i.e., the base case values for operating Mode 1, before starting the first experimental run. This 'warm-up phase' allows for the process to reach a steady-state condition before the manipulated variables are changed. Thereafter, every run lasted 50 h, and all 12 runs were run in sequence during continuous operation of the process. We did not apply any of the possible pre-set disturbances (IDVs) during experimentation. Including the warm-up phase, the entire experiment contained 636 h of simulated operation (real simulation time is only 115 s on a computer using an Intel® Core™ i5-4310 U processor running at 2.0 GHz with 16 GB of RAM.) The controlled and manipulated variables were sampled every 12 min.

Due to the process' continuous nature, the experimental factors and responses need to be viewed as time series. For example, Figure 4 illustrates the impact of the experimental factors on the controlled and manipulated variables in Loop 16 which controls the reactor temperature, XMEAS(9), by adjusting the reactor cooling water flow, XMV(10).

As seen in Figure 4, the experiment has a substantial impact on the manipulated variable – reactor cooling water flow, XMV(10). However, even though the levels of the experimental factors are changing, the controlled reactor temperature XMEAS(9) exhibits a random variation around its set-point value, indicating that the impact on this controlled variable is small or non-existent. A similar behavior has been observed also for loops 9, 10 and 11.

### 4.2. Statistical analysis

In the first scenario, the manipulated variables of loops 9, 10, 11, 12 and 16 are considered as the main response variables. A simple but reasonable way to analyze the experiments with time series responses is to ignore the time series aspect of the responses and to calculate the average value for each run in Table V, see Vanhatalo *et al.*<sup>17</sup>. Vanhatalo *et al.*<sup>18</sup> recommend removing apparent dynamic behavior at the beginning of each run. However, the initial observations are here included to investigate if the control loops are effective because the control action may not succeed to remove the impact on the controlled variable instantly. The run averages can be used to perform analysis of variance (ANOVA). Table VI presents a summary of the ANOVA based on the averages in Table V. The analysis was performed using the software Design-Expert® version 9.

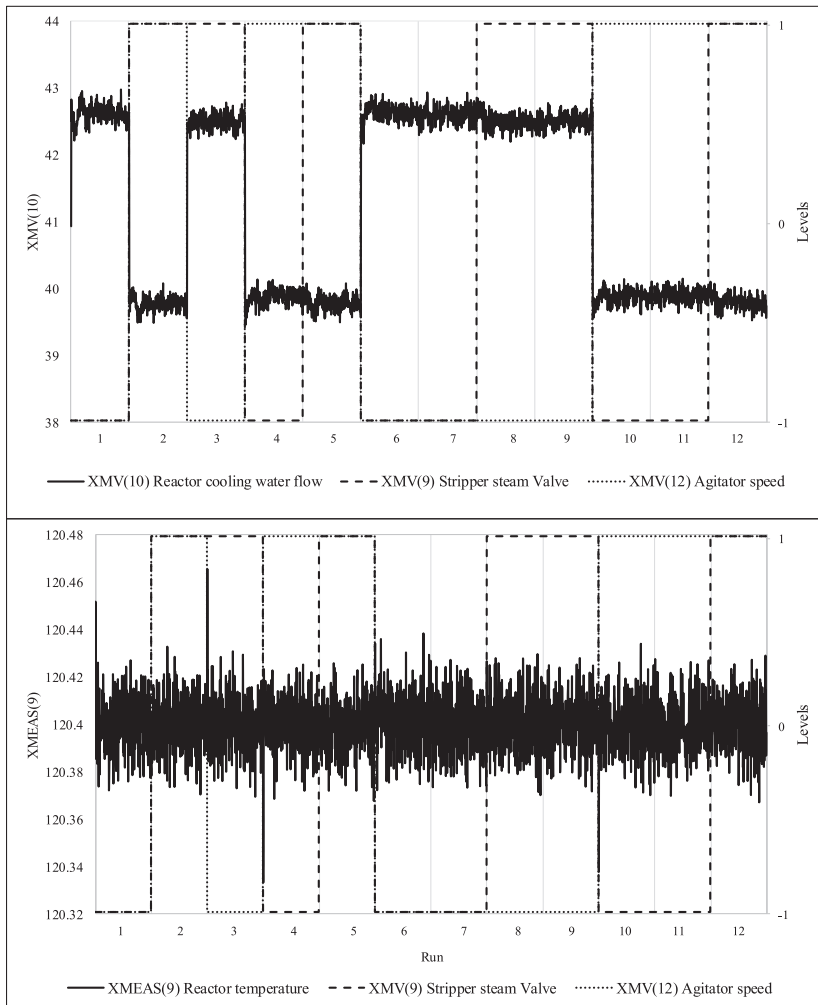
Based on the high  $p$ -values for the controlled variables in Loops 11, 12 and 16, the results do not indicate that the experimental factors affect their related controlled variables. However, as revealed by the low  $p$ -values for the manipulated variables in Loops 12 and 16 in Table VI, the experimental factors affect process phenomena controlled by these loops. Furthermore, Loops 9 and 10 fail to remove the full impact of the experimental factor variation on the controlled variables as indicated by the low  $p$ -values on the controlled variables. There is no evidence that the experimental factors are affecting process phenomena controlled by Loop 12. Furthermore, the low  $p$ -value of the main effect of the stripper steam valve XMV(9) on the stripper liquid level in Loop 9, XMEAS(15), is explained by the inclusion of the transition time. The run averages are affected because the control action of Loop 9 is delayed.

### 4.3. Concluding remarks for scenario 1

When experimenting in a closed-loop system, the analyst should expect that the impact of the experimental factors could be partly or completely displaced from the controlled variables to manipulated variables. This is true despite using inputs not involved in control loops as experimental factors, if the experimental factors affect the phenomena controlled in the loops. However, as illustrated, the analysis may reveal potential ineffectiveness of the controllers to completely or instantly remove disturbances acting on controlled variables. We therefore recommend viewing the responses as two important and closely related groups: [1] controlled variables and [2] manipulated variables when analyzing an experiment in a closed-loop system as illustrated above.

Table V. Run order and averages of the observations for the manipulated and controlled variables in the 2<sup>2</sup> factorial experiment

Run	Experimental factors		Manipulated variables (XMV <sub>s</sub> )					Controlled variables (XMEAS <sub>s</sub> )				
	XMV(9) %	XMV(12) %	Loop 9	Loop 10	Loop 11	Loop 12	Loop 16	Loop 9	Loop 10	Loop 11	Loop 12	Loop 16
			$r_7$	$r_6$	s.p. 17	$r_5$	XMV(10)	j = 15	j = 12	j = 8	j = 7	j = 9
1	40	40	0.2273	0.2489	80.3247	0.0034	42.5970	49.97	49.97	74.98	2705.03	120.40
2	60	60	0.2281	0.2500	80.2536	0.0034	39.7900	50.34	50.09	74.94	2705.04	120.40
3	60	40	0.2281	0.2500	80.2754	0.0034	42.4825	50.06	49.97	75.01	2705.03	120.40
4	40	60	0.2272	0.2489	80.3755	0.0034	39.8809	49.62	50.00	75.02	2704.99	120.40
5	60	60	0.2281	0.2501	80.2909	0.0033	39.7862	50.37	50.19	74.99	2704.95	120.40
6	40	40	0.2272	0.2488	80.3795	0.0033	42.6090	49.62	50.02	74.98	2705.02	120.40
7	40	40	0.2272	0.2489	80.3306	0.0033	42.5969	49.92	50.04	74.99	2705.00	120.40
8	60	40	0.2281	0.2501	80.2541	0.0033	42.4961	50.16	50.04	74.98	2704.99	120.40
9	60	40	0.2281	0.2501	80.2469	0.0033	42.5057	49.98	49.97	75.02	2705.01	120.40
10	40	60	0.2272	0.2487	80.3858	0.0033	39.8842	49.81	49.93	74.99	2705.05	120.40
11	40	60	0.2272	0.2489	80.3315	0.0033	39.8970	49.97	49.83	74.96	2704.95	120.40
12	60	60	0.2281	0.2500	80.2335	0.0033	39.8065	50.36	50.00	75.06	2705.01	120.40



**Figure 4.** Overview of experimental factors' impact on variables related to control loop 16. The manipulated variable, XMV(10), is given in the top chart and controlled variable, XMEAS(9), in the bottom chart. The levels, in coded units, of the experimental factors XMV(9) and XMV(12) are superimposed on the plots. The duration of each experiment is 50 h.

## 5. Scenario 2: design and analysis

The second scenario illustrates a different way of running experiments in closed-loop controlled processes. Now, we consider the set-points of the control loops as experimental factors. Our major concern is no longer to reveal cause and effect relationships between inputs and important measured variables in the process. These should have been identified already in the engineering control design phase. Instead, we are exploring the set-points of the controllers to see causal relationships between the process operating conditions and process performance indicators with the aim of optimizing the process.

### 5.1. A screening experiment

In this case, we focus on the process operating cost as an important response. We have nine possible set-points to change (see Table IV), and we wish to test their impact on the process operating cost using a two-step sequential experiment. The starting point

Table VI. The <i>p</i> -values of the estimated effects based on ANOVA. Cells with bold text indicate the significant effects based on a significance level of 5%														
Controlled variables XMEAS( <i>j</i> )														
Main effects and interaction	Manipulated variables													
	Loop 9	Loop 10	Loop 11	Loop 12	Loop 16	Loop 9	Loop 10	Loop 11	Loop 12	Loop 16				
	<i>r</i> <sub>7</sub>	<i>r</i> <sub>6</sub>	sp17	<i>r</i> <sub>5</sub>	XMV(10)	<i>j</i> = 15	<i>j</i> = 12	<i>j</i> = 8	<i>j</i> = 7	<i>j</i> = 9				
XMV(9)	<0.0001	<0.0001	<0.0002	0.7011	<0.0001	0.001	0.0937	0.4997	0.9031	0.9588				
XMV(12)	0.9958	0.3744	0.5351	0.4029	<0.0001	0.1414	0.8736	0.8567	0.4668	0.1363				
XMV(9)*XMV(12)	0.2754	0.9865	0.5564	0.9997	0.2759	0.0692	0.0412	0.7048	0.8390	0.6849				

is a  $2^{9-5}_{III}$  fully randomized fractional factorial design with four additional center points. This resolution III design is then followed by a full fold-over in a new block to entangle some aliased effects. The final design, i.e., the original plus the fold-over, is of resolution IV.

Some factor setting combinations will invoke a process shutdown and some shutdown limits are also given in the Downs and Vogel<sup>7</sup> paper. The base case value of each factor (rounded to the nearest integer) was chosen as either the low or high factor level in the design. The other level of each variable was defined by trial and error by either adding to or subtracting from the base case value while trying to keep the process from shutting down. Table VII provides the low and high levels of each experimental factor (set-point) used in the experiment.

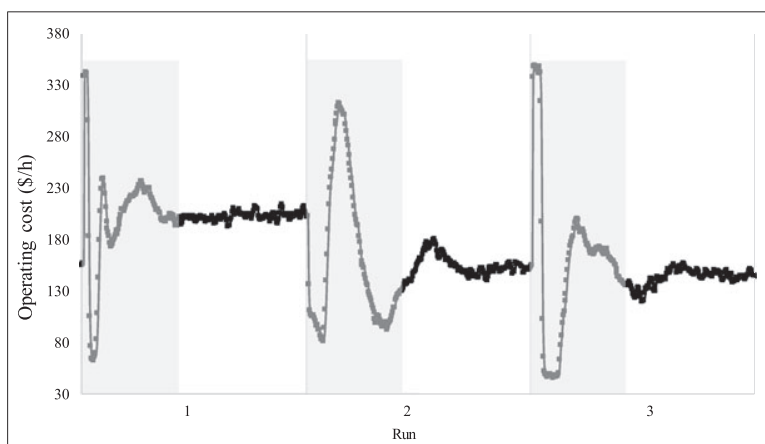
Furthermore, we chose to keep X MVs (5, 9 and 12) fixed at their base case values given in Table III during the experiment because they are not involved in the loops but do affect the process behavior.

A 'warm-up phase' of 36 h was once again used before the start of the first run of the experiment. During this phase, the experimental factors (set-points) were fixed to their base case values for operating Mode 1. The 40 runs of the experiment are given in Table VIII. Each experimental run lasted 50 h. Including the warm-up phase, the entire experiment contained 2036 h of operation (simulation time was 147 s for all runs). From the TE simulator, the process operating cost (\$/h) can be extracted, and we have the operating cost for every 12 min. Figure 5 illustrates the impact of the experimental factors on the process operating cost during the first three experiments in run order.

Loop	Controlled variable	Base set-point	Low level	High level
7	Stripper liquid rate (production)	22.949 m <sup>3</sup> h <sup>-1</sup>	21 m <sup>3</sup> h <sup>-1</sup>	23 m <sup>3</sup> h <sup>-1</sup>
9	Stripper liquid level	50%	50%	60%
10	Separator liquid level	50%	35%	50%
11	Reactor liquid level	75%	70%	75%
12	Reactor pressure	2705 kPa	2600 kPa	2705 kPa
13	Mole % G	53.724 mol%	54 mol%	62 mol%
14	Amount of A in reactor feed ( $y_A$ )	54.95%	55%	65%
15	Amount of A + C in reactor feed ( $y_{AC}$ )	58.57%	50%	59%
16	Reactor temperature	120.40 °C	120 °C	127 °C

Block 1: $2^{9-5}_{III}$ experimental design				Block 2: Full fold-over			
Run order	Standard order	Operating cost (\$/h)	Operating cost (\$/h) (after removing transition time)	Run order	Standard order	Operating cost (\$/h)	Operating cost (\$/h) (after removing transition time)
1	14	201.11	201.68	21	38	139.46	130.84
2	2	156.51	154.51	22	26	130.55	131.72
3	9	148.60	143.56	23	34	152.75	146.08
4	4	127.37	140.00	24	27	156.25	157.61
5	6	185.37	172.01	25	35	182.89	170.58
6	20	124.19	129.89	26	22	125.28	126.76
7	1	139.87	141.24	27	30	175.37	157.19
8	17	133.27	131.09	28	39	120.70	131.02
9	11	123.56	129.74	29	33	151.78	150.66
10	12	255.76	215.15	30	24	166.46	155.20
11	8	175.52	187.61	31	29	129.43	142.91
12	16	164.44	160.05	32	31	186.93	167.84
13	18	127.84	130.15	33	28	166.30	167.94
14	15	147.23	142.59	34	36	164.98	165.41
15	19	130.64	132.81	35	37	128.14	132.72
16	5	104.70	109.27	36	21	145.67	140.70
17	13	181.27	161.61	37	32	104.34	115.46
18	3	128.85	127.87	38	23	174.01	166.69
19	7	182.26	177.45	39	25	213.02	198.88
20	10	117.49	127.62	40	40	127.23	135.06





**Figure 5.** The operating cost during the first three runs of the experiment. Note the dynamic behavior of the response during the first part of each run. The shaded areas highlight the removed observations before calculating the run averages. The duration of each experiment is 50 h.

## 5.2. Statistical analysis

The aim of the experiment is to find set-points which reduce the long-term operating cost. In contrast to scenario 1, it makes sense to remove transition time from the runs and then use the remaining observations to calculate run averages. To keep the observations during the transition time in the calculation of run averages will lead to an underestimation of the location effect of the factors and interactions, see Vanhatalo et al.<sup>17</sup> The process operating cost exhibits some transition time before reaching the steady state as illustrated in Figure 5. A visual inspection of the operating cost reveals that 24 h can be considered as a reasonable transition time (grey shaded area in Figure 5), and thus the observations obtained during the first 24 h of all runs were removed before calculating the run averages, see Table VIII.

Table IX presents an ANOVA table of the 40-run experimental design in Table VIII based on a significance level of 5%. We have also repeated the analysis including the transition time. The results of that analysis are not reported in this article, but with the transition time included, the same main effects turn out to be active, but the significant interaction effects differ. As seen in Table IX, seven main effects and eight two-factor interaction alias strings are active (interactions of order three or higher are ignored). It is perhaps not surprising that most factors affect the operating cost because control loops aim to control important process phenomena which tend to affect the production cost. Moreover, the interconnectedness of the different control loops is demonstrated by the many significant interactions.

Note that the curvature test is significant and that the model exhibits significant lack of fit, suggesting that a higher order model is appropriate. The fitted model in Table IX is thus ill-suited for optimization and prediction but provides a starting point for future response surface experimentation. The many significant two-factor interaction alias strings would need further investigation to decide which among the aliases are actually active. However, as we mentioned earlier, the main purpose of this article is not necessarily to provide an optimization procedure on a simulated process but rather to draw attention to possibilities and pitfalls in experimentation under closed-loop operation. Hence, for demonstration purposes, we simply assume that the first interactions of the interaction strings in Table IX are the important ones, ignoring the interactions in brackets. We proceed to use the estimated model to provide suggested factor settings for the lowest operating cost within the experimental region. In this case, the lowest cost will be at a corner point on the multidimensional hyperplane. The settings of the factors and the predicted operating cost at this point (104.5 \$/h) are provided in Table X. The significant curvature, the lack of fit tests and the  $R^2$  for prediction indicate that the predictive ability of the model is poor. A confirmation run in the TE process simulator using the suggested factors settings gives the long-term average operating cost 109.1 \$/h. The 4.6 \$/h discrepancy between the predicted cost and the confirmation run is likely due to the models' poor predictive ability. Nevertheless, this rough analysis provides a significant improvement of the process operating cost. A simulation of the process keeping the factors settings at the base set-points values given in Table VII gives a long-term average operating cost of 170.2 \$/h. Hence, running the process at the suggested factors settings leads to a substantial cost reduction of 61.1 \$/h. Further reduction of the operating cost is likely possible but outside the scope of this article.

## 5.3. Concluding remarks for scenario 2

The second scenario illustrates how designed experiments can be used to improve process performance indicators using the set-points of variables controlled in closed-loop. This scenario also exemplifies the importance of considering, and here removing, the transition time during analysis. We want to point out that the set-points of the controllers in this example and in real life in general affect important process operating conditions. The experimenter should therefore expect that improper choices of factor levels of the

**Table IX.** ANOVA and estimated effects based on the averages of the response after removing the transition time. The model includes only terms significant at 5% level. Aliased two-factor interaction aliases that based on the heredity principle are less likely given in italic text within brackets. The control loop numbers are indicated by (#) in the factor names

Source	Sum of squares	df	Mean square	F value	Prob > F	Estimated standardized effects
Block	15.11	1	15.11			
Model	18 431.18	15	1228.75	83.97	<0.0001	
A: #7—Production	3946.30	1	3946.30	269.68	<0.0001	11.11
D: #11—Reactor level	321.10	1	321.10	21.94	0.0001	3.17
E: #12—Reactor pressure	3131.68	1	3131.68	214.01	<0.0001	−9.89
F: #13—Mole %G	4085.75	1	4085.75	279.21	<0.0001	−11.30
G: #14— $y_A$	443.48	1	443.48	30.31	<0.0001	−3.72
H: #15— $y_{AC}$	2444.72	1	2444.72	167.07	<0.0001	8.74
J: #16—Reactor temp	126.25	1	126.25	8.63	0.0076	−1.99
AD ( <i>BH CG FG</i> )	124.36	1	124.36	8.50	0.0080	1.97
AF ( <i>BG CH DE</i> )	207.73	1	207.73	14.20	0.0011	−2.55
AG ( <i>BF CD EH</i> )	78.98	1	78.98	5.40	0.0298	−1.57
AH ( <i>BD CF EG</i> )	151.98	1	151.98	10.39	0.0039	−2.18
AJ	532.42	1	532.42	36.38	<0.0001	−4.08
FJ	282.93	1	282.93	19.34	0.0002	2.97
GJ	619.92	1	619.92	42.36	<0.0001	4.40
HJ	1933.59	1	1933.59	132.14	<0.0001	7.77
Curvature	3415.43	1	3415.43	233.40	<0.0001	
Residual	321.93	22	14.63			
Lack of Fit	305.16	16	19.07	6.82	0.0129	
Pure Error	16.77	6	2.80			
Cor Total	22 183.66	39				
				$R^2$		83.1%
				Adjusted $R^2$		72.1%
				$R^2$ prediction		67.2%

**Table X.** Suggested settings of the set-points of the control loops to provide the lowest operating cost of the estimated model within the experimental region

Loop	Set-point	Suggested setting
7	Stripper liquid rate (production)	21 m <sup>3</sup> h <sup>−1</sup>
9	Stripper liquid level	Not in model, use base case
10	Separator liquid level	Not in model, use base case
11	Reactor liquid level	70%
12	Reactor pressure	2705 kPa
13	Mole % G	62 mol%
14	Amount of A in reactor feed ( $y_A$ )	65%
15	Amount of A + C in reactor feed ( $y_{AC}$ )	50%
16	Reactor temperature	120 °C
Resulting predicted process operating cost:		104.5 \$/h

set-points may lead to unexpected process behavior or even shutdown. Special care should be taken in choosing the levels because the window of operability may be irregular or unknown.

## 6. Conclusion and discussion

This article explores important issues in designing and analyzing experiments in the presence of engineering process control. The closed-loop operation increases process complexity and influences the strategy of experimentation. Two experimental scenarios

based on the TE process simulator are used to answer the questions *why* and *how* to conduct and analyze experiments in closed-loop systems.

Even though we have prior experience with experiments lasting several weeks in continuous processes, the 2038 h of experimentation we use in our examples may admittedly be considered unrealistically long in practice. This is, however, beside the point because the examples we provide are for demonstration purposes, and we did not necessarily focus on shortening the duration of the experiments.

The first experimental scenario illustrates how the experimental factors not directly involved in control loops impact the closed-loop system and how the controllers affect the analysis. The controllers adjust manipulated variables to limit or eliminate the experimental factor effects on the controlled variable(s). We note that this will only occur if the experimental factors affect phenomena/variables governed by the closed-loop system. The effect on the controlled variables is partly or fully transferred to the manipulated variables depending on the effectiveness of the controllers. Hence, both the controlled and manipulated variables should be used as responses. Analyzing the effects of experimental factors on controlled variables may give important information about the effectiveness of the engineering process control. The effects on the manipulated variables instead reveal whether the experimental factors affect important process behavior.

In the second scenario, the experimental factors are the set-points of the controlled variables. The set-points are target values for the controlled variables and are typically closely tied to important process operating conditions. A level change of the set-points can therefore be considered equivalent to shifting the location of the process. Overall process performance indicators such as operating cost or product quality may then be suitable responses.

Using two scenarios we have illustrated that DoE can generate knowledge and aid process improvement in closed-loop systems. More specifically, DoE can be used to study:

- if the engineering process control is efficient and cost effective;
- if experimental factors affect important process phenomena; and
- how controlled variable set-points affect important process performance indicators.

We believe simulation software like the TE process offer great opportunities for methodology development in experimentation in closed-loop systems. In this article, we simply provide some basic ideas and approaches, but more research is needed for further development of experimentation and analysis methods for better process understanding and optimization in closed-loop systems.

## Acknowledgments

The authors gratefully acknowledge the financial support from the Swedish Research Council under grant number 340-2013-5108.

## References

1. Box GEP, MacGregor JF. The analysis of closed-loop dynamic-stochastic systems. *Technometrics* 1974; **16**(3):391–398.
2. Box GEP, MacGregor JF. Parameter estimation with closed-loop operating data. *Technometrics* 1976; **18**(4):371–380.
3. Ljung L. System Identification: Theory for the User (2nd. edn). Prentice-Hall: Upper Saddle River, NJ, 1999.
4. Jansson H. Experiment design with applications in identification for control. Department of Signals, Sensors, and Systems, Royal Institute of Technology (KTH), Stockholm, Sweden, 2004. Doctoral Thesis.
5. Vanhatalo E, Bergquist B. Special considerations when planning experiments in a continuous process. *Quality Engineering* 2007; **19**(3):155–169. doi:10.1080/08982110701474100.
6. Hild C, Sanders D, Cooper T. Six Sigma\* on continuous processes: how and why it differs. *Quality Engineering* 2001; **13**(1):1–9. doi:10.1080/08982110108918618.
7. Downs JJ, Vogel EF. A plant wide industrial process control problem. *Computers & Chemical Engineering* 1993; **17**(3):245–255.
8. Kruger U, Zhou Y, Irwin GW. Improved principal component monitoring of large-scale processes. *Journal of Process Control* 2004; **14**(8):879–888. doi:10.1016/j.jprocont.2004.02.002.
9. Montgomery DC. Design and Analysis of Experiments (8th. edn). Wiley: New York, 2012.
10. McAvoy TJ, Ye N. Base control for the Tennessee Eastman problem. *Computers & Chemical Engineering* 1994; **18**(5):383–413. doi:10.1016/0098-1354(94)88019-0.
11. Lyman PR, Georgakis C. Plant-wide control of the Tennessee Eastman problem. *Computers & Chemical Engineering* 1995; **19**(3):321–331. doi:10.1016/0098-1354(94)00057-U.
12. Ricker LN. Decentralized control of the Tennessee Eastman challenge process. *Journal of Process Control* 1996; **6**(4):205–221. doi:10.1016/0959-1524(96)00031-5.
13. Ku W, Storer RH, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 1995; **30**(1):179–196. doi:10.1016/0169-7439(95)00076-3.
14. Lee G, Han C, Yoon ES. Multiple-fault diagnosis of the Tennessee Eastman process based on system decomposition and dynamic PLS. *Industrial & Engineering Chemistry Research* 2004; **43**(25):8037–8048. doi:10.1021/ie049624u.
15. Hsu C, Chen M, Chen L. A novel process monitoring approach with dynamic independent component analysis. *Control Engineering Practice* 2010; **18**:242–253. doi:10.1016/j.conengprac.2009.11.002.
16. Liu K, Fei Z, Yue B, Liang J, Lin H. Adaptive sparse principal component analysis for enhanced process monitoring and fault isolation. *Chemometrics and Intelligent Laboratory Systems* 2015; **146**:426–436. doi:10.1016/j.chemolab.2015.06.014.
17. Vanhatalo E, Bergquist B, Vännman K. Towards improved analysis methods for two-level factorial experiment with time series responses. *Quality and Reliability Engineering International* 2013; **29**(5):725–741. doi:10.1002/qre.1424.

18. Vanhatalo E, Kvarnström B, Bergquist B, Vännman K. A method to determine transition time for experiments in dynamic processes. *Quality Engineering* 2010; **23**(1):30–45. doi:10.1080/08982112.2010.495099.

#### *Authors' biographies*

**Francesca Capaci** is a doctoral candidate in Quality Technology at the Department of Business Administration and Social Sciences at Luleå University of Technology (LTU) in Sweden. She received her Master's degree in Industrial Engineering and Management at Università Degli Studi di Palermo (Italy) in 2014. Her research area is focused around experimental strategies and statistical process control for continuous processes. Since 2014, she has also been teaching Design of Experiment and Statistical Process Control at LTU. She is a member of the European Network for Business and Industrial Statistics (ENBIS).

**Bjarne Bergquist** is Professor and Chair of Quality Technology and Management at Luleå University of Technology, Sweden. His current research interests are focused on the use of statistically based methods and tools for improvements, particularly in the continuous process industries, as well as on statistically based railway maintenance. He has a background in Materials Science but has lately mainly published the field of operations management, quality technology and quality improvements.

**Murat Kulahci** is a professor in the Department of Business Administration, Technology and Social Sciences at Luleå University of Technology in Sweden and an associate professor in the Department of Applied Mathematics and Computer Science, at the Technical University of Denmark. His research focuses on design of physical and computer experiments, statistical process control, time series analysis and forecasting, and financial engineering. He has presented his work in international conferences, and published over 70 articles in archival journals. He is the co-author of two books on time series analysis and forecasting.

**Erik Vanhatalo** is an assistant professor and senior lecturer of quality technology and management at Luleå University of Technology (LTU). He holds an MSc degree in industrial and management engineering from LTU and a PhD in quality technology and management from LTU. His current research is focused on the use of statistical process control, experimental design, time series analysis and multivariate statistical methods, especially in process industry. He is a member of the European Network for Business and Industrial Statistics (ENBIS).

## **PAPER C**

The Revised Tennessee Eastman Process Simulator as  
Testbed for SPC and DoE Methods

**Capaci, F., Vanhatalo, E., Kulahci, M.,  
and Bergquist, B. (2017).**

*Submitted for publication*



# The Revised Tennessee Eastman Process Simulator as Testbed for SPC and DoE Methods

Francesca Capaci<sup>1</sup>, Erik Vanhatalo<sup>1</sup>, Murat Kulahci<sup>1,2</sup>, and Bjarne Bergquist<sup>1</sup>

<sup>1</sup>*Luleå University of Technology, Luleå, Sweden*

<sup>2</sup>*Technical University of Denmark, Kongens Lyngby, Denmark*

## Abstract

Engineering process control and high-dimensional, time-dependent data present great methodological challenges when applying statistical process control (SPC) and design of experiments (DoE) in continuous industrial processes. Process simulators with an ability to mimic these challenges are instrumental in research and education. This article focuses on the revised Tennessee Eastman process simulator providing guidelines to use it as a testbed for SPC and DoE methods. We provide flowcharts that will help new users get started in the Simulink/Matlab framework, and illustrate how to run stochastic simulations for SPC and DoE applications using the Tennessee Eastman process.

**Keywords:** Simulation; Tutorial; Statistical process control; Design of Experiments; Engineering process control, Closed-loop.

## 1. Introduction

Continuous production during which the product is gradually refined through different process steps and with minimal interruptions (Dennis and Meredith 2000) is common across different industries. Today these processes manufacture both consumption goods such as food, drugs, and cosmetics, and industrial goods such as steel, chemicals, oil, and ore. Full-scale continuous production plants present analytical challenges since they are characterized by, for example, high-technological and complex production machinery, low flexibility, engineering process control (closed-loop operations) and high production speed. Automated data collection schemes producing multi-dimensional and high-frequency data generate additional analytical challenges. However, these processes still need to be improved continuously to remain competitive. Statistical process control (SPC) and design of experiments (DoE) techniques are essential in these improvement efforts.

The literature on the use of SPC and DoE in process industrial applications is extensive. However, a majority of these examples fail to capture essential challenges that analysts face when applying these methods in modern continuous processes. Recent SPC literature highlights the need to adapt SPC practices to the new manufacturing environments with massive datasets, multistep production processes, or greater computing capabilities (Ge et al. 2013, Ferrer 2014, Vining et al. 2015). Similarly, features of continuous processes unavoidably affect experiments and how experimental design strategies should be adapted, see, e.g., Vanhatalo and Bergquist (2007) and Capaci et al. (2017).

Methodological work to upgrade current SPC and DoE methods to address the continuous production challenges is needed, but it is often overly complicated to do methodological development using real processes. Tests of SPC or DoE methods in full-scale plants tend to require considerable resources and may jeopardize the production goals. Simulators may offer a reasonable trade-off between the required flexibility to perform tests and the limitations in mimicking the behavior of a real process.

Reis and Kenett (2016) map a wide range of simulators that can be used to aid the teaching of statistical methods to reduce the gap between theory and practice. They classify existing simulators based on various levels of complexity and guide educators to choose a proper simulator depending on the needed sophistication. Reis and Kenett (2016) classify



the Tennessee Eastman (TE) process simulator (Downs and Vogel 1993) as one of the more complex simulators (medium-/large-scale nonlinear dynamic simulator) suggesting its use for advanced applications in graduate or high-level statistical courses. Downs and Vogel (1993) originally proposed the TE process as a test problem providing a list of potential applications in a wide variety of topics such as plant control, optimization, education, non-linear control and, many others. However, older implementations of the TE process that we have come across have a fundamental drawback in that the simulations are deterministic, apart from measurement error that is added. An almost deterministic simulator is of limited value in methodological development since random replications as in Monte Carlo simulations are not possible. However, the revised TE process by Bathelt et al. (2015a) does provide flexibility enough to create random errors in simulations. Especially after this latest revision, we believe that the TE process simulator can help bridge the gap between theory and practice as well as provide a valuable tool for teaching. However, as argued by Reis and Kenett (2016), the TE process simulator together with other advanced simulators lack an interactive graphical user interface (GUI), which means that the user still needs to be able to have some programming skills.

In this article, we therefore aim to provide guidelines for how to use the TE process simulator as a testbed for SPC and DoE methods. We use the revised TE process presented in Bathelt et al. (2015a) run with a decentralized control strategy (Ricker 1996). Flowcharts based on the Business Process Modelling Notation (BPMN) illustrate the required steps to implement the simulations (Chinosi and Trombetta 2012). Finally, we provide examples of SPC and DoE applications using the TE process.

The next section of this article provides a general description of the revised TE process simulator and the chosen control strategy. The following two sections describe how to run simulations for SPC and DoE applications, respectively. We then present two simulated SPC and DOE examples in the TE process (Sections 5 and 6). Conclusions and discussion are provided in the last section.

## **2. The Tennessee Eastman Process Simulator**

The TE process simulator emulates a continuous chemical process originally developed for studies of engineering control design. The control engineering community has developed

different control strategies for the TE process. We argue that, among these, the decentralized control strategy proposed by Ricker (1996) is attractive from a method development perspective since it mimics most of the challenges that continuous processes offer sufficiently well.

Over the years, the TE process has been popular within the chemometrics community, and simulated TE process data have been used extensively for methodological development of multivariate statistical process control (MSPC) methods. For instance, the TE process simulator has been used for work on integrating dynamic principal component analysis (DPCA) into process monitoring, see Ku et al. (1995), Rato and Reis (2013), and Vanhatalo et al. (2017). Other TE process simulator examples for multivariate monitoring include Kruger et al. (2004), Lee et al., (2004), Hsu et al., (2010), and Liu et al., (2015). Instead, examples of DoE applications using the TE process are limited. Capaci et al. (2017) illustrate the use of two-level factorial designs using the TE process run under closed-loop control. However, the previous simulator's deterministic nature of the TE process has likely hampered researchers in doing methodological work.

We intend to illustrate how the new revised simulation model of the decentralized TE process implemented by Bathelt et al. (2015b) can be manipulated to allow stochastic simulations and replications. The simulator has the following additional advantages:

- the simulator is implemented in the Simulink/Matlab® interface and can be obtained for free,
- the set-points of the controlled variables and the process input can be modified as long as they are maintained within the restrictions of the decentralized control strategy,
- the analyst can specify the characteristics of the simulated data as, for example, length of experimentation, sampling frequency, type and magnitude of process disturbances, and
- the simulation speed is fast. For example, to simulate the SPC example in this article with 252 hours of operation takes less than a minute (56.26 seconds) on a computer using an Intel® Core™ i5-4310U processor running at 2.0 GHz with 16 GB of RAM.

## 2.1. Process Description

The TE process plant involves five major units: a reactor, a condenser, a vapor-liquid separator, a product stripper and a recycle compressor (Downs and Vogel 1993). The plant produces two liquid products from four gaseous reactants through four irreversible and exothermic reactions. It also produces an inert product and a byproduct purged as vapors from the system through the vapor-liquid separator (Figure 1).

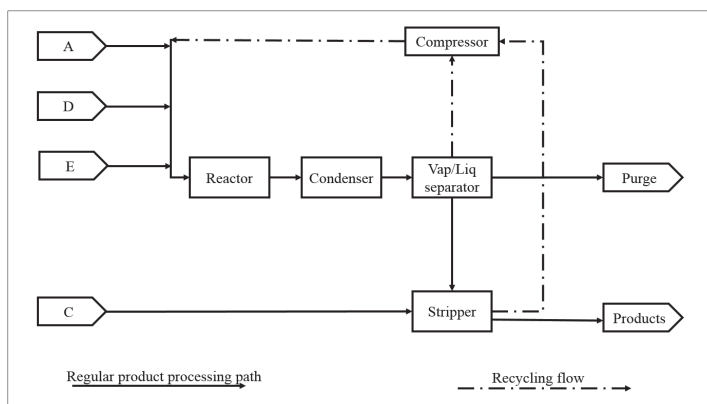


Figure 1. A simplified overview of the TE process flow.

Reactants A, D and E flow into a reactor where the reaction takes place. The output from the reactor is fed to a condenser. Some non-condensable vapors join the liquid products, but the following vapor-liquid separator again splits the substances into separate flows. Vapor is partially recycled and partially purged together with the inert product and the byproduct. The stripper separates the remaining A, D and E reactants from the liquid and another reactant C is added to the product. The final products then exit the process, and the remaining reactants are recycled.

The TE process has 12 manipulated variables (X MVs) and 41 measured variables (X MEAs). Tables in Downs and Vogel (1993) provide detailed information about all the process variables and the cost function that provides the process operating cost in \$/h. The combination of three G/H mass ratios and four production rates of the final products define six different operating modes of the process. The user can also choose to activate 20 preset process disturbances (IDVs).

The TE process is open-loop unstable and shuts down rapidly without engineering controller systems. A control strategy is therefore necessary for process stability. To avoid shutdowns and for securing plant safety, the control strategy should abide by five operating constraints related to the reactor pressure, level and temperature, the product separator level, and the stripper base level. Even with controllers working correctly, the TE process is sensitive and may shut down depending on the controller tuning and the set-points of the controlled variables.

## 2.2. Decentralized Control Strategy

The decentralized control strategy partitions the plant into sub-units and designs a controller for each one, with the intent of maximizing the production rate. Ricker (1996) identified nineteen feedback control loops to stabilize the process. Table 1 provides the control loops and the related controlled and manipulated variables. The original article by Ricker (1996) provides detailed information about the design phases of the decentralized control strategy.

Table 1. Controlled and manipulated variables in the 19 loops of the decentralized control strategy. The manipulated variables with codes such as  $F_p$  and  $r_7$  come from the decentralized control strategy settings (Ricker 1996). XMV(i) and XMEAS(j) are numbered according to the original article by Downs and Vogel (1993).

Loop	Controlled variable	Code	Manipulated variable	Code
1	A feed rate (stream 1)	XMEAS(1)	A feed flow	XMV(3)
2	D feed rate (stream 2)	XMEAS(2)	D feed flow	XMV(1)
3	E feed rate (stream 3)	XMEAS(3)	E feed flow	XMV(2)
4	C feed rate (stream 4)	XMEAS(4)	A and C feed flow	XMV(4)
5	Purge rate (stream 9)	XMEAS(10)	Purge valve	XMV(6)
6	Separator liquid rate (stream 10)	XMEAS(14)	Separator pot liquid flow	XMV(7)
7	Stripper liquid rate (stream 11)	XMEAS(17)	Stripper liquid product flow	XMV(8)
8	Production rate (stream 11)	XMEAS(41)	Production index	$F_p$
9	Stripper liquid level	XMEAS(15)	Ratio in loop 7	$r_7$
10	Separator liquid level	XMEAS(12)	Ratio in loop 6	$r_6$
11	Reactor liquid level	XMEAS(8)	Setpoint of Loop 17	s.p. 17
12	Reactor pressure	XMEAS(7)	Ratio in loop 5	$r_5$
13	Mol % G (stream 11)	XMEAS(40)	Adjustment of the molar feed rate of E	$E_{adj}$
14	Amount of A in reactor feed, $y_A$ (stream 6)	XMEAS(6)	Ratio in loop 1	$r_1$
15	Amount of A+C in reactor feed, $y_{AC}$ (stream 6)	XMEAS(6)	Sum of loops 1 and 4 ratio	$r_1 + r_4$
16	Reactor temperature	XMEAS(9)	Reactor cooling water flow	XMV(10)
17	Separator temperature	XMEAS(11)	Condenser cooling water flow	XMV(11)
18	Maximum reactor pressure	XMEAS(7)	Production index	$F_p$
19	Reactor level override	XMEAS(8)	Compressor recycle valve	XMV(5)

### 2.3. The Revised TE Simulation Model

Ricker (2005) devised the decentralized TE control strategy as a Simulink/Matlab<sup>®</sup> code. Bathelt et al. (2015b) recently developed a revised version of the simulation model. The revision is an update of Ricker's (2005) code that widens its usability by allowing for customization of the simulation by modifying a list of parameters in the process model function. Below we describe how to initialize the revised TE simulator and how to use the model function parameters to achieve intended simulator characteristics.

#### *Initialization of the revised TE model*

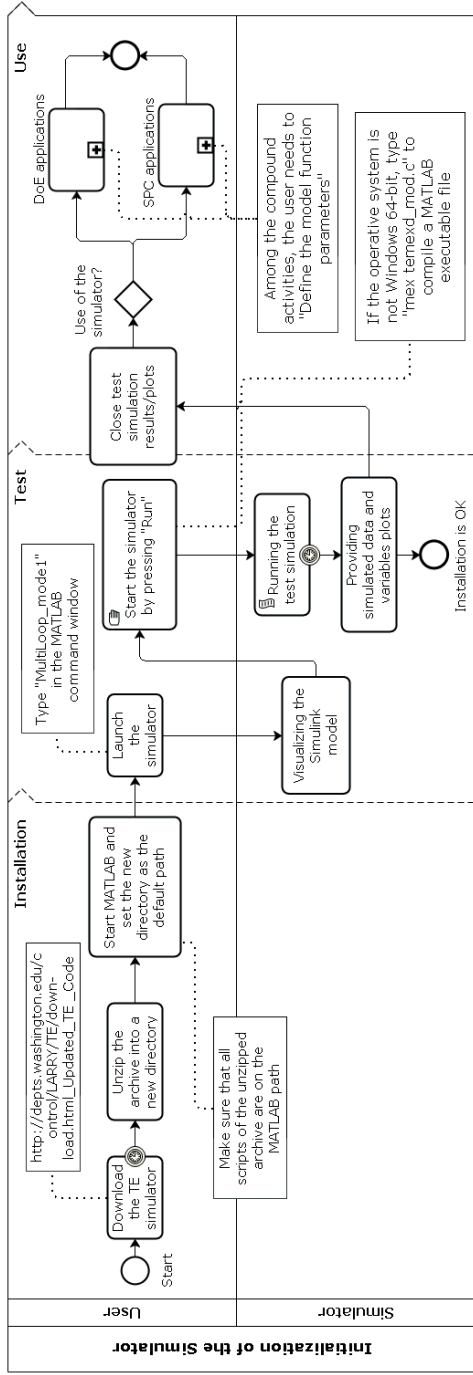
The files of the revised model are available as a Simulink/Matlab<sup>®</sup> code at the Tennessee Eastman Archive (Updated TE code by Bathelt et al. (2015b)). Figure 2 illustrates the workflow to initialize the simulator through a simulation test, using the BPMN standard (Chinosi and Trombetta 2012). During the simulation, four online plots display the reactor pressure, process operating cost, production flow, and product quality trend. When the simulation ends, the simulator provides datasets of X MVs and X MEAs as well as the related plots. Installation is then completed.

The simulator can be run in both operating Mode 1 and 3. Operating Mode 1, which we use in this article, seems to be the most commonly used in the literature. The model "MultiLoop\_model1" runs the process at Mode 1 when the set-points of the input variables not involved in control loops and of the controlled variables are set up according to the base case values given in Tables 2 and 6.

In Figure 2, "DoE applications" and "SPC applications" consist of different compound activities, expanded later, that the user must follow depending on which method is being applied. The definition of the model function parameters is one of these activities and can be done following the instructions below.

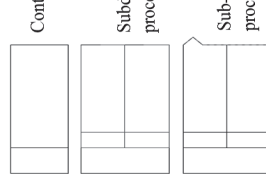
Table 2. Base case set-points of the input variables not involved in control loops for operating Mode 1 (Capaci et al. 2017).

Variable name	Code	Base case value (%)	Low limit (%)	High limit (%)
Compressor recycle valve	XMV(5)	22.210	0	100
Stripper Steam Valve	XMV(9)	47.446	0	100
Agitator Speed	XMV(12)	50.000	0	100



# LEGEND

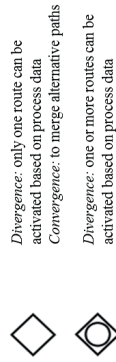
## Swimlanes



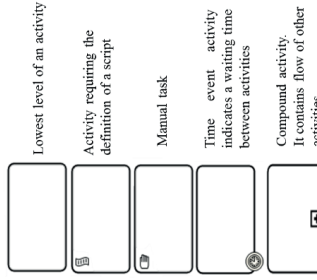
## Events (circles)



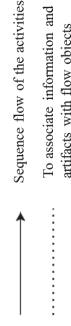
## Gateway (diamonds)



## Activities (Rounded rectangles)



## Connecting Objects



## Artifacts

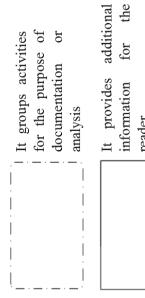


Figure 2. Main tasks required to initialize the simulator for operating Mode 1. Note that some symbols in the legend might be unused in this flowchart. Legend inspired by [http://resources.bizagi.com/docs/BPMN Quick Reference Guide\\_ENG.pdf](http://resources.bizagi.com/docs/BPMN Quick Reference Guide_ENG.pdf)

### *Using the model function parameters to customize the simulation*

The model function “temexd\_mod” contains the “TE code” and it is located in the “TE Plant” block of the Simulink model. A double-click on “temexd\_mod” opens a dialog window. In the field “S-function parameters,” the user can define three model function parameters separated by commas. Square brackets are used for undefined parameters. The simulation can be customized to fit different needs by changing these parameters. Table 3 provides more details of the model function parameters.

Table 3. Description and settings of the parameter list for the process model function “temexd\_mod” (Bathelt et al. 2015a). An example of settings for parameter 3 is given.

Parameter list of “temexd_mod”	Description		Setting	
1	An array of the initial values of the 50 states of the model. The user can specify a vector of 50 states of the model to run the simulator in a specific operating mode		Empty: default values of process operating Mode 1 are used (Downs and Vogel 1993).	
2	Initial value (seed) of the random generator		Every integer number greater than 0 is valid.	
3	Model structure flag		Example	$(11100010)_2 = (226)_{10}$
	Bit	Description		
	0	Additional measurements points	0	Integer value equivalent to the binary number activating/deactivating the bit of the model structure flag
	1	Monitoring outputs of the disturbances	1	
	2	Monitoring outputs of the reaction and process	0	
	3	Monitoring outputs of the component's concentration	0	
	4	Deactivation of measurement noise	0	
	5	Random generator uses different state variables for process disturbances and measurements noise	1	
	6	Solver-independent calculations of the process disturbances	1	
	7	Disturbances are scaled by the value of the activation flags	1	
	15	Reset model structure to original structure of Ricker (2005)	0	

Parameter 1 relates to the initial values of the model states. Since we wish to run the process in Mode 1, we assume hereafter that this parameter is set as empty unless otherwise specified. The default “xInitial” array is therefore used when we launch the simulator. Parameters 2 and 3 enable the customizations introduced in the revised TE code.

The possibility to change the seed of each simulation (parameter 2) creates the opportunity to avoid deterministic simulations, but only when the user activates process disturbances (IDVs) of the type random variation in the model, see Table 4. Parameter 3 allows for activating/deactivating the model flags listed in Table 3. Each model flag

corresponds to a bit that can be switched using the binary notation. The value of parameter 3 corresponds to the decimal integer of the binary number obtained after setting the value of each bit. For example, the binary number  $(11100010)_2$  is equivalent to the parameter value of  $(226)_{10}$ , which produces the exemplified model structure given in Table 3. Note that for the right conversion from a binary to a decimal number, the binary number must be written starting from the highest to the lowest bit position (from 15 to 0).

As a rule of thumb, model flags 5 and 6 should be active during the simulation while the user can set the other model flags to adjust the model to the simulation needs. Further details of the model flag structure are given in Bathelt et al. (2015a).

## **2.4. Creating random simulations in the revised TE process simulator**

The TE process is complex (Reis and Kenett 2016) and in that sense mimics a real chemical process. While the high degree of complexity makes it useful as a testbed for methodological development, the same complexity imposes some limitations. As already stated, without customization, the TE simulator provides output that does not differ much from a deterministic simulation where all measurement error is set to zero.

Figure 3 shows a schematic overview of the revised TE simulation model highlighting potential sources of random variation. Note that when random disturbances of type “random variation” are turned off, the TE process variables are only affected by white Gaussian noise mimicking typical measurement noise. Thus, repeated simulations with the same setup will produce, except for measurements errors, identical results, which limit the model’s value when running repeated simulations. These are for instance used when assessing the performance of an SPC method or when replicates of experimental runs are needed.

To overcome this limitation, we suggest running the simulator with added measurement noise and one or more of the random disturbances (IDVs) listed in Table 4 activated. Indeed, the possibility to scale random variation disturbances allows the user to add variability without overly distorting the results. Moreover, the simulator can generate different results for repeated simulations with the same starting conditions by changing the seed of the random numbers, making the revised TE model more suitable for methodological tests of SPC and DoE methods.



Please note that the choice of the scale factor(s) to adjust the random variation depends on the random disturbance(s) introduced in the simulation model and the aim of the simulation study. The random disturbances vary in both magnitude and dynamics, and hence have a different impact on the process. We, therefore, leave the choice of disturbances and the scale factor(s) to the user but explain the ideas behind our choices in our examples.

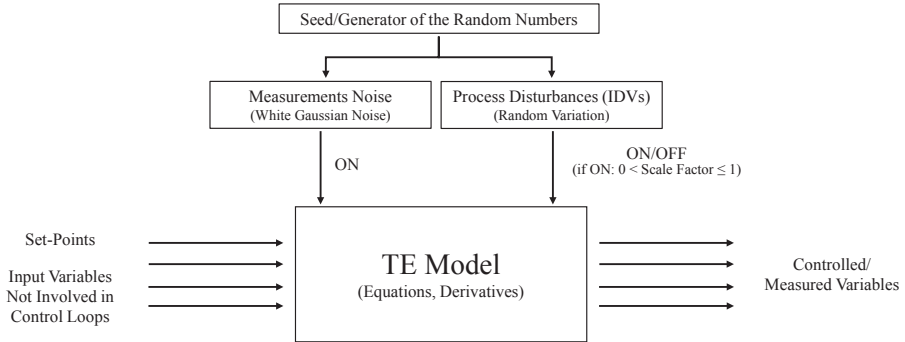


Figure 3. Schematic overview of the revised TE simulation model with a focus on potential sources of random variation.

### 3. The TE Process Simulator in the SPC Context

SPC applications require historical in-control data (Phase I dataset) and an online collection of data to perform Phase II analysis. Samples from Phase I and Phase II are typically collected in one shot in the TE process simulator. Using the BPMN standard, the upper half of Figure 4 presents the tasks required to simulate Phase I and Phase II data. Table 4 lists possible process disturbances that can be used as faults in Phase II. Note that the revised TE model adds eight “random variation” disturbances to the simulator, IDV(21)–IDV(28). A characteristic of the revised simulator valuable for SPC applications is that the analyst can now scale all process disturbances by setting their disturbance activation parameter values between 0 and 1.

The TE process simulator can emulate three important SPC challenges that frequently occur in continuous processes:

*Multivariate data:* The 53 variables available in the TE process (12 X MVs and 41 X MEAs), some of which are highly cross-correlated, allow for studies of multivariate monitoring methods. The TE process has been used extensively within the chemometrics literature to test monitoring applications and fault detection/isolation methods based on

latent structures techniques such as principal component analysis (PCA) and partial least square (PLS). The simulator will not produce missing data, but the analyst may remove data manually if needed.

*Autocorrelated data:* The user can choose the variables' sampling rate in the TE process, but for most choices, the resulting data will be serially correlated (autocorrelated). Autocorrelation will require adjustment of the control limits of control charts since the theoretical limits typically will be under/overestimated. This faulty estimation will affect in-control and out-of-control alarm rates (Bisgaard and Kulahci 2005, Kulahci and Bisgaard 2006), and this also extends to process capability analysis affecting both univariate and multivariate techniques.

*Closed-Loop Operation:* Engineering process control is constantly working to adjust process outputs through manipulated variables in the closed-loop operation of the TE process simulator. The closed-loop operation provides an interesting SPC challenge. Control charts applied to controlled outputs could fail to detect a fault and might erroneously indicate an in-control situation. The traditional SPC paradigm to monitor the process output when engineering process control is in place requires proper adjustments, and the TE process simulator provides a good testbed for this challenge.

#### **4. The TE Process Simulator in DoE Context**

The lower part of Figure 4 provides a guide on how to simulate data in TE process for testing DoE methods for continuous processes operating under closed-loop control. Note that one of the early tasks is to activate one or more process disturbances of type “random variation,” see Table 4, to overcome the deterministic nature of the simulator. Two experimental scenarios can, for example, be simulated using the TE process simulator (Capaci et al. 2017). In the first scenario, the experimental factors can include the three manipulated variables not involved in control loops, X<sub>MV</sub>(5), X<sub>MV</sub>(9), and X<sub>MV</sub>(12), see also Table 2, while the responses can include both manipulated and controlled variables. In the second scenario, the experimenter can use the set-point values of the control loops as experimental factors and the operating cost function as a response. However, a cascaded procedure based on directives generated by the decentralized control strategy will make some set-points dependent. Therefore, the experimenter only has the

subset of the nine set-points given in Table 5 available as experimental factors in the second scenario.

Table 4. The 28 process disturbances available (Downs and Vogel 1993, Bathelt et al. 2015a).

Variable Number	Process Variable	Type
IDV(1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV(2)	B composition, A/C ratio constant (stream 4)	Step
IDV(3)	D feed temperature (stream 2)	Step
IDV(4)	Water inlet temperature for reactor cooling	Step
IDV(5)	Water inlet temperature for condenser cooling	Step
IDV(6)	A feed loss (stream 1)	Step
IDV(7)	C header pressure loss- reduced availability (stream 4)	Step
IDV(8)	A,B,C proportion in stream 4	Random variation
IDV(9)	D feed temperature (stream 2)	Random variation
IDV(10)	A and C feed temperature(stream 4)	Random variation
IDV(11)	Water inlet temperature for reactor cooling	Random variation
IDV(12)	Ater inlet temperatur for condenser cooling	Random variation
IDV(13)	Variation coefficients of reaction kinetics	Random variation
IDV(14)	Reactor cooling water valve	Sticking
IDV(15)	Condenser cooling water valve	Sticking
IDV(16)	Variation coefficient of the steam supply of the heat exchanger of the stripper	Random variation
IDV(17)	Variation coefficient of heat transfer in reactor	Random variation
IDV(18)	Variation coefficient of heat transfer in condenser	Random variation
IDV(19)	Unknown	Unknown
IDV(20)	Unknown	Random variation
IDV(21)	A feed temperature (stream 1)	Random variation
IDV(22)	E feed temperature (stream 3)	Random variation
IDV(23)	A feed flow (stream 1)	Random variation
IDV(24)	D feed flow (stream 2)	Random variation
IDV(25)	E feed flow (stream 3)	Random variation
IDV(26)	A and C feed flow (stream 4)	Random variation
IDV(27)	Reactor cooling water flow	Random variation
IDV(28)	Condenser cooling water flow	Random variation

Table 5. Base-level set-points of available experimental factors in the TE process for operating Mode 1.

Loop	Controlled variable	Base set-point
7	Stripper liquid rate (production)	22.949 m <sup>3</sup> h <sup>-1</sup>
9	Stripper liquid level	50 %
10	Separator liquid level	50 %
11	Reactor liquid level	75 %
12	Reactor pressure	2705 kPa
13	Mole % G	53.724 mol%
14	Amount of A in reactor feed (y <sub>A</sub> )	54.95 %
15	Amount of A+C in reactor feed (y <sub>AC</sub> )	58.57 %
16	Reactor temperature	120.40 °C

The TE process simulator allows the user to pause, analyze the experiment, and make new choices based on the results. Thus, sequential experimentation, a cornerstone in experimental studies, is possible to simulate. The experimenter can repeat the experimental runs and expand the experiment with an augmented design since the seeds for the random disturbances can be changed. Hence, TE process simulator can emulate potential experimentation strategies such as response surface methodology (Box and Wilson 1951)

and evolutionary operation (Box 1957). Even though cost and time concerns are not important when experiments are run in a simulator, and the number of experimental factor levels and replicates are practically limitless compared to a real-life experiment, there are only a few potential experimental factors available. The simulator may aid studies on the robustness and the analysis of an experiment where the number of experimental runs is limited, such as unreplicated designs with a minimum number of runs.

Below we highlight three challenges for the analyst when applying DoE in the TE process. These challenges are also commonly found in full-scale experimentation in continuous processes:

*The closed-loop environment:* The TE process experimenter must select experimental factors and analyze process responses while considering the presence of feedback control systems (Capaci et al. 2017). The decentralized control of the TE process will mask relationships between process input and output, and feedback control loops will limit the possibility to vary all the process inputs freely. Furthermore, the experimenter must restrict potential experimental factor changes within constrained operating regions to avoid any process shuts downs.

*Transition times between runs:* The time required for different responses to reach a new steady state in the TE process will differ depending on the factors and the magnitude of the change. The characterization of transition times is crucial to minimize their effect on the experimental results as well as to allocate the time needed for the treatments to take full effect (Vanhatalo et al. 2010). Long transition times between steady-state conditions add to the costs of randomizing the runs in a real experiment. The literature suggests using split-plot designs to restrict factor changes in this situation. Moreover, it is common to avoid resetting the levels of factors between consecutive runs where the factors are to be held at the same level for time and cost concerns. However, maintaining the factor level settings between adjacent runs and disregarding resetting lead to a correlation between neighboring runs and to designs called randomized-not-reset (RNR) designs (Webb et al. 2004). These can also be studied in the TE process.

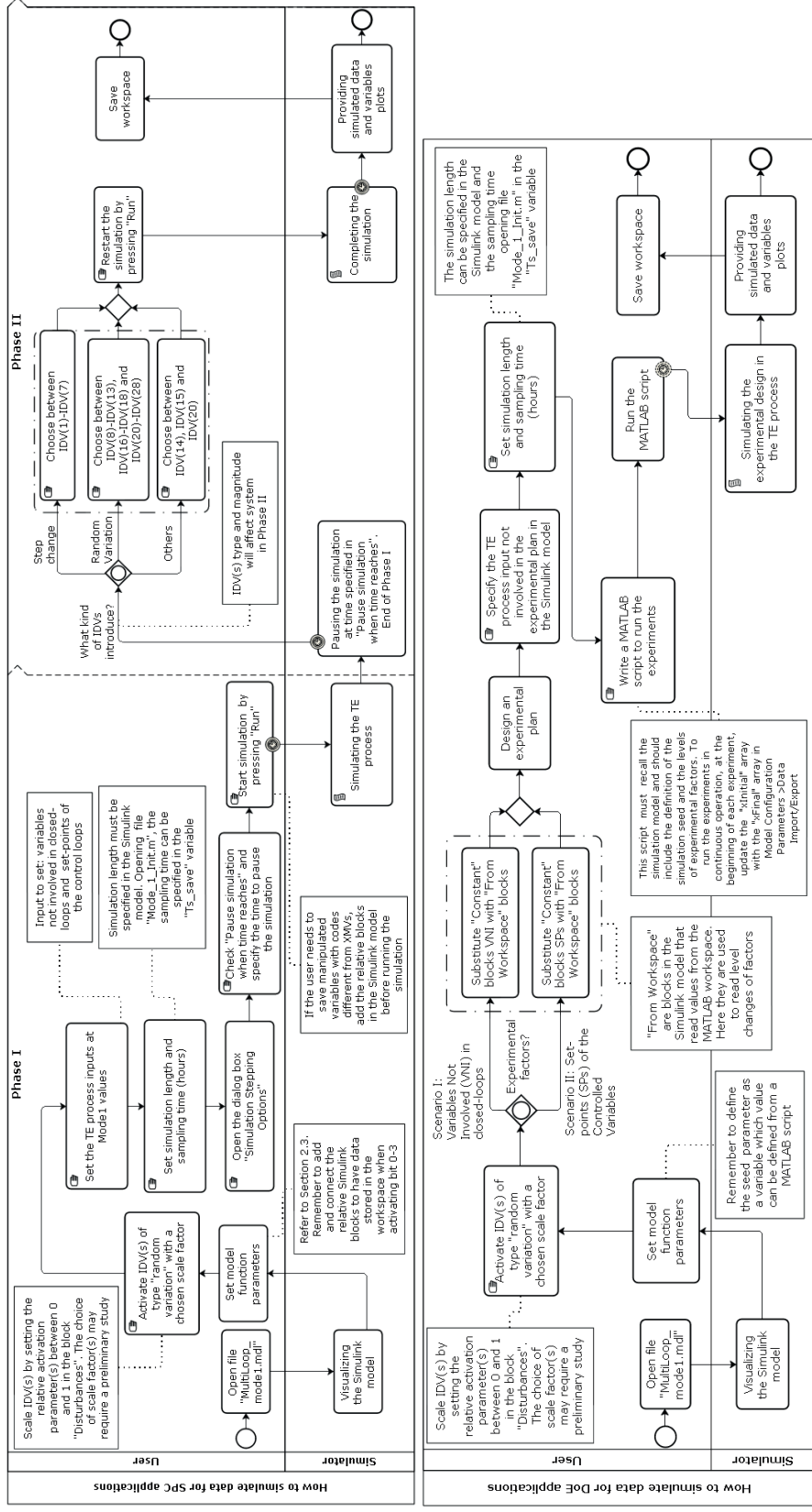


Figure 4. Overview of tasks required to simulate data for SPC (upper) and DoE (bottom) applications. Symbols are explained in Figure 2.

*Time series data for factors and responses:* the continuous nature, the dynamic behavior, and the transition times of the TE process lead to the fact that experimental factors and responses become time series. The analysis of the experiments from the TE process allows for considering the time series nature of factors and responses. The response time series need to be summarized in averages or standard deviations to fit in a standard analysis such as the Analysis of Variance (ANOVA). Transfer function-noise modeling may be used to model the dynamic relations between experimental factors and the response(s) (Lundkvist and Vanhatalo 2014).

## **5. Example 1: The TE process simulator and SPC**

Note that the aim of the examples provided here is not to describe the most complex scenarios available nor is it to suggest the “best solution” to the illustrated challenges. The examples are provided to show how the TE process can act as a testbed for developing and testing methodological ideas. In the first example, we illustrate how closed-loop operation can affect the shift detection ability of control charts. In particular, this example demonstrates how control charts applied to the (controlled) output could fail to detect a fault and might therefore erroneously indicate an in-control situation.

The example focuses on control loops 9-12 and 16 (Table 1). These loops regulate the process operating constraints needed to secure plant safety and to avoid unwanted shutdowns. Five process inputs ( $r_5$ ,  $s.p.17$ ,  $XMV(10)$ ,  $r_6$ , and  $r_7$ ), i.e., the manipulated variables, control the related TE process outputs ( $XMEAS$  7-9, 12 and 15). We here refer to control loops 9-12 and 16, and their related variables as critical control loops, critical controlled variables (C- $XMEAS$ ) and critical manipulated variables (C- $XMVs$ ) respectively.

### **5.1. Selecting and scaling disturbances**

After a preliminary study of the available process disturbances of type random variation (Table 4) to introduce in the TE process, we further analyzed the behavior of  $IDV(8)$  and  $IDV(13)$ .  $IDV(8)$  varies the proportion of the chemical agents (A, B, C) in stream 4 of the process, mimicking a reasonably realistic situation, whereas  $IDV(13)$  deviates the coefficients of reaction kinetics propagating its impact through the whole process. We performed four sets of 20 simulations each with a scale factor of the disturbances equal to

0.25, 0.5, 0.75 and 1 to understand the impact of IDV(8) and IDV(13) on process behavior. Each simulation, run with a randomly selected seed, lasted 200 hours in the TE process and the output of the random disturbances was collected with a sample time of 12 minutes. We maintained inputs not involved in control loops and set-points of the control loops at the base case values of operating Mode 1 (Tables 2 and 5).

Based on the averages and standard deviations presented in Table 6, and to achieve random variation in the TE process, we ran Phase I and Phase II data collection with both IDV(8) and IDV(13) active, with randomly selected scale factors between 0 and 0.25 and, 0 and 0.5 respectively.

We then performed a preliminary simulation with the same simulator settings for the random disturbances to select the magnitude of the step size (fault) for Phase II. Table 6 shows the magnitude of the step size for the scale factor equal to 0.25, 0.5, 0.75 and, 1. We, therefore, introduced a step change in the water inlet temperature of the reactor coolant in Phase II, i.e., IDV(4), with a randomly selected scale factor between 0.25 and 0.5.

Table 6. Step size of IDV(4) for different scale factor values. Averages and standard deviations of IDV(8) and IDV(13) based on 20 simulations.

Variable number	Process variable	Scale factor	Average	Step size	Standard deviation			
IDV(4)	Cooling water inlet temperature of reactor	$0.25n$ $n=0,1,\dots,4$	35	$+1.25n$	-			
IDV(8)	Proportion of A in stream 4	$0.25n$	48.51		$n=1$	$n=2$	$n=3$	$n=4$
	Proportion of B in stream 4	$0.25n$	0.50		0.372	0.748	1.126	1.50
	Proportion of C in stream 4	$0.25n$	50.99		0.038	0.074	0.110	0.15
IDV(13)	Variation coefficient of reaction kinetics $A + C + D \rightarrow G$	$0.25n$	1		0.374	0.751	1.130	1.51
	Variation coefficient of reaction kinetics $A + C + D \rightarrow G$		1		0.03n			

## 5.2. Data Collection

The TE process was first run for 144 hours at normal operating conditions (Phase I), i.e., base case values for operating Mode 1. A step change in the cooling water inlet temperature of the reactor (IDV4) was then introduced in the process for 108 hours (Phase II). The randomly selected scale factors of disturbances IDV(4), IDV(8) and IDV(13) in this simulation were 0.32, 0.1 and 0.25 respectively. Values on C-XMEAS and C-XMV were collected in sequence during continuous operation of the process with a sampling time of 12 minutes.

### 5.3. Multivariate process monitoring

For illustration purposes, consider a standard Hotelling  $T^2$  multivariate control chart for individual observations for the five critical controlled variables of the TE process (C-XMEAS). The Phase I sample was produced by excluding the start-up phase of the process. The critical controlled variables exhibit a dynamic behavior for about 36 hours or 180 samples at the start of the simulation. After this “warm-up phase,” the TE process was deemed to have reached the steady state.

Samples of C-XMEAS collected during steady-state operation provide a more stable estimation of the sample covariance matrix,  $\mathbf{S}$ , and thus of the  $T^2$  values. We discarded the first 180 observations and used datasets of 540 samples both in Phase I and Phase II to build the Hotelling  $T^2$  chart, see Figure 5. The standard sample covariance matrix was used to form the  $T^2$  chart. The theoretical Phase I and Phase II upper control limits were based on the  $\beta$  and  $F$  distributions, and on the assumption that observations are time-independent (Montgomery 2012). This assumption is unrealistic because of the observed positive autocorrelation in the critical controlled variables (and as a result in the  $T^2$  values), and consequently, the upper control limits could be adjusted. The point we want to make here is, however, still relevant using the theoretical control limits.

There are a few  $T^2$  observations above the control limit in the Phase II sample based on C-XMEAS (top panel in Figure 5), but an analyst might as well conclude that there is little evidence to deem the process out-of-control. Moreover, a visual inspection of the C-XMEAS univariate plots in Figure 6 seems to support this conclusion, as the critical and controlled variables appear to be insensitive to the step change in the cooling water inlet temperature for the reactor (IDV4). However, this conclusion is incorrect. Since the TE process is run in closed-loop operations, the analyst should know that the engineering process control seeks to displace most of the variability induced by the step change (fault) to some manipulated variable(s). In fact, the correct conclusion in this scenario is that the process is still working at the desired targets thanks to the feedback control loops. If the C-XMV's are studied, the analyst would probably deem the process to be out-of-control. That the process is disturbed becomes evident by studying the Hotelling  $T^2$  chart based only on the five C-XMV's (bottom panel in Figure 5). While one may consider the process in statistical process control during Phase I, it is out-of-control in Phase II. Moreover, a visual



inspection of the univariate C-XMV plots in Figure 6 suggests that an increase in the flow of the reactor cooling water, XMV(10) compensates for the effect of the introduced fault in the inlet temperature. Such a control action could, of course increase waste of water and/or energy while trying to maintain product properties on target.

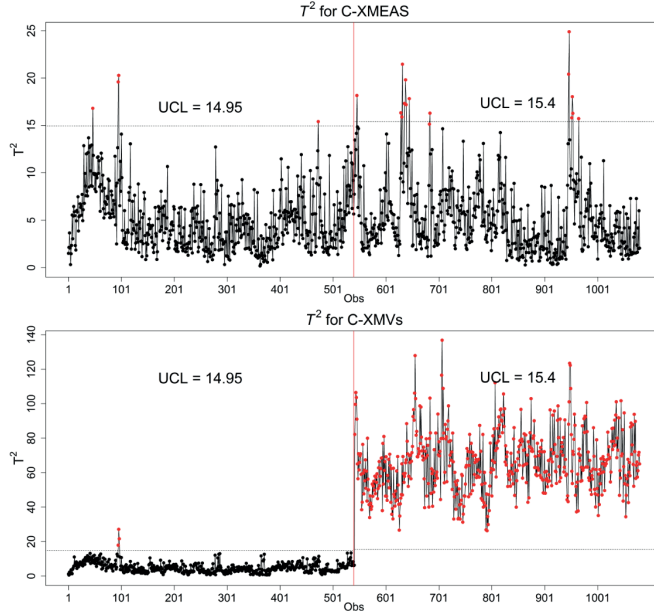


Figure 5. Hotelling  $T^2$  chart based on individual observations for the C-XMEAS (top) C-XMV (bottom). The vertical red line divides phase I and phase II data.

#### 5.4. Closing remarks

The example above shows a possible application of how to use the TE process as a testbed for SPC methods. As the TE process is run under closed-loop control, control actions may partly or completely displace the impact of a disturbance from the controlled variables to manipulated variables. The traditional approach of applying a control chart on the (controlled) process output then needs to be supplemented with a control chart on the manipulated variables. The concurrent use of both of these control charts allows for [1] confirming the presence and effectiveness of the controller by analyzing the control chart for the controlled variables and [2] identifying potential assignable causes by analyzing the control chart for the manipulated variables.

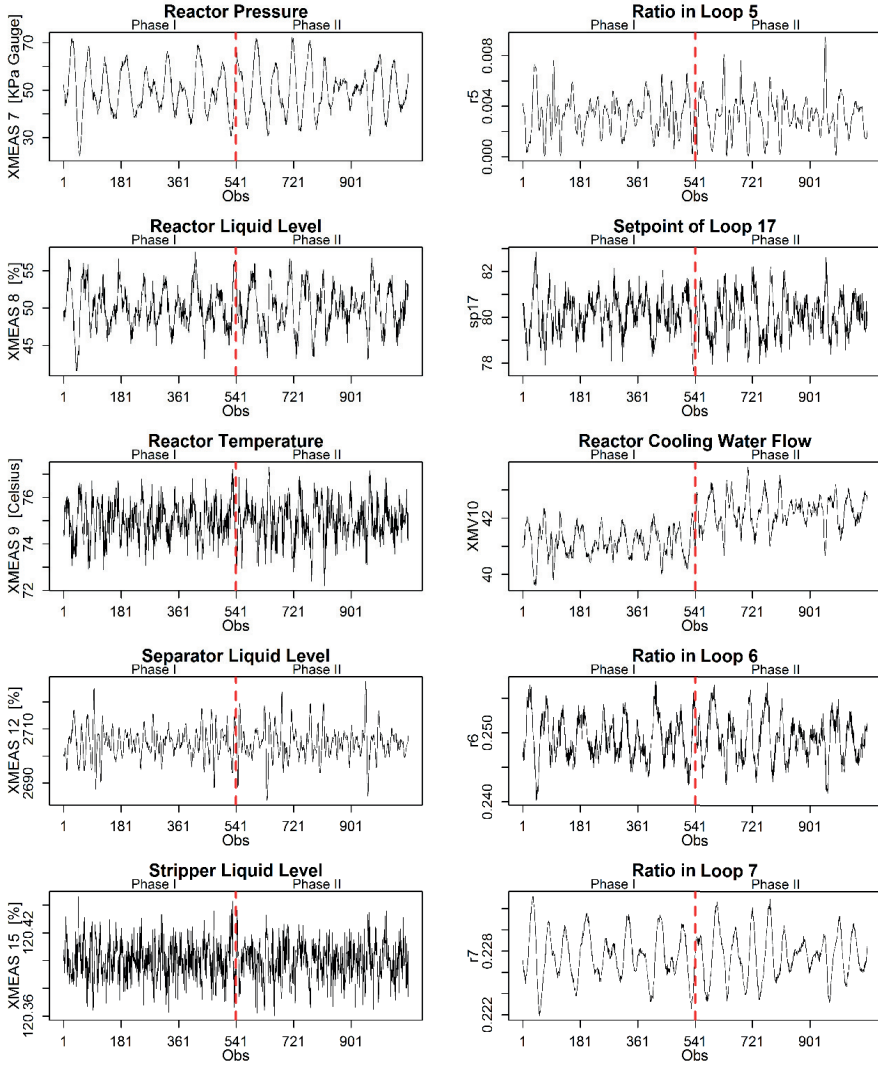


Figure 6. Univariate time series plots for C-XMEAS (left column) and C-XMVs (right column) during both Phase I and II.

## 6. Example 2: DoE in TE process simulator

This example illustrates a response surface methodology approach based on sequential experimentation using a subset of the set-points of the control loops in the TE process. The example starts with a two-level fractional factorial design, which is augmented to a central

composite design, followed by confirmation runs in the simulator. The example describes how to use the TE process for experimentation. Hence, we conduct a simplified analysis of the experimental results applying ANOVA on the average values of the factors and the response of each experimental run, as suggested by Vanhatalo et al. (2013).

Closed-loop process performance may improve by exploring the relationships between the set-points of the controlled variables and an overall performance indicator such as production cost. Consider an experiment where we first want to identify reactor set-points that affect the operating cost and then try to minimize this cost. Our experimental factors are in this case the five set-points of the controlled variables in loops 11, 12, 14, 15 and 16. The response is the process operating cost. Table 7 presents the set-points of the starting condition, the average operating cost (long-term value) given these set-points and the chosen levels of the set-points in the two-level experimental design. Note that the choices of experimental factor levels were found using trial and error by changing the base case values and in the meantime trying to keep the process from shutting down. The input variables that were not involved in control loops were set at operating Mode 1 values (Table 2) in all simulations.

Table 7. Long-term average operating cost at the starting set-point settings. Low and high level of the set-points used as experimental factors.

Loop	Controlled variable	Starting set-point setting	Low level	High level
7	Stripper liquid rate (production)	22.949 m <sup>3</sup> h <sup>-1</sup>	-	-
9	Stripper liquid level	50 %	-	-
10	Separator liquid level	50 %	-	-
11	Reactor liquid level	75 %	70 %	75 %
12	Reactor pressure	2705 kPa	2600 kPa	2705 Kpa
13	Mole % G (product quality)	62 mol%	-	-
14	Amount of A in reactor feed ( $y_A$ )	54.95 %	55 %	65 %
15	Amount of A+C in reactor feed ( $y_{AC}$ )	58.57 %	50 %	59 %
16	Reactor temperature	120.40 °C	120 °C	127 °C
Long-term average operating cost		147.60 \$/h		

## 6.1. Selecting and scaling disturbances

Real processes are often disturbed by unknown sources. The random process variation in the simulator needs to be comparable to disturbances affecting a real process. We used the random disturbances IDV(8) and IDV(13) to add random disturbances to the process. The impact of IDV(8) on the operating costs of the process was studied using 10 simulations with the starting values of the set-points given in Table 7. The scale factor of IDV(8) was then increased by steps of 0.1 in each run. Each simulation, run with a random seed, lasted

200 hours in the simulator and the operating cost was sampled every 12 minutes. We repeated the procedure for IDV(13), increasing the scale factor by steps of 0.1 in each run. Visual inspection of the resulting cost time series led us to the conclusion that scale factors between 0.1 and 0.4 produce reasonable random variability.

The scale factors of IDV(8) and IDV(13) were determined to 0.31 and 0.1 respectively throughout the simulations after drawing random numbers from a uniform distribution between 0.1 and 0.4. From another set of 20 simulations with these selected scale factors, the average (long-term) operating costs were 147.60 \$/h with a standard deviation of 36.75 \$/h. Visual inspection shows that the process operating cost exhibits a transition time of approximately 24 hours before reaching the steady state. We, therefore, removed observations of the cost function during the first 24 hours before calculating the average and standard deviation of the process operating cost.

## **6.2. Experimental design and analysis**

Analyses reported in this section were all made using Design Expert<sup>®</sup> version 10.

### *Phase I: Screening*

We chose a  $2_{IV}^{5-1}$  fully randomized fractional factorial design with four additional center runs to screen the five factors (set-points) in Table 7. The experiment started by a “warm-up phase” where the TE process was run for 36 hours (180 samples) using the starting set-points settings in Table 7. After these 36 hours, the TE process was deemed to have reached steady state. When steady state was reached, all runs were conducted in sequence according to their run order during continuous process operation. The simulation runs were 50 hours (250 samples) and the simulation seed was randomly changed before each run. The operating cost was sampled every 12 minutes.

We calculated response averages for each run to analyze the response time series of the cost. We then removed the observations of the transition time before calculating the run averages to avoid a biased estimation of the main effects and their interactions (Vanhatalo et al. 2013). The transition time during some runs were determined to be approximately 24 hours through visual inspection. Some settings thus had an effect on process stability, which meant that the run averages were based on the run’s last 26 hours (130 samples).

Table 8 shows the run order during the experiment and the averages of the process operating cost for each run.

Table 9 presents an ANOVA table of active effects (at 5% significance level) based on the first 20 experimental runs of Table 8. Four main effects and two two-factor interactions have statistically significant effects on the operating cost. We also included the main effect of factor E in the model due to effect heredity. However, the significant curvature suggests that a higher order model may be needed.

Table 8. Run order, standard order of the runs and average operating cost after removing the transition time at the beginning of each run. The “c” in standard order marks the center points.

Block 1: $2_V^{5-1}$						Block 2: Augmented plan		
Run order	Standard order	Operating Cost (\$/h)	Run order	Standard order	Operating Cost (\$/h)	Run order	Standard order	Operating Cost (\$/h)
1	9	163.66	11	17c	128.83	21	27	147.82
2	14	162.79	12	19c	126.96	22	22	135.91
3	12	155.84	13	3	144.00	23	21	126.96
4	10	175.84	14	13	175.24	24	28	189.13
5	7	127.38	15	2	180.51	25	29	168.99
6	18c	131.29	16	1	140.62	26	26	136.39
7	20c	123.26	17	16	159.84	27	24	117.51
8	6	145.13	18	8	136.36	28	30	163.25
9	5	151.53	19	11	158.39	29	23	147.95
10	4	129.90	20	15	136.09	30	25	140.13

Table 9. ANOVA and estimated effects based on the first 20 runs in Table 8. Third order and higher interactions are ignored.

Source	Sum of Squares	df	Mean Square	F Value	Prob > F	Estimated Standardized Effects (\$/h)
Model	3941.46	7	563.07	29.65	< 0.0001	
A: Reactor Liquid Level	151.89	1	151.89	8.00	0.0164	3.08
B: Reactor Pressure	1360.21	1	1360.21	71.64	< 0.0001	-9.22
C: Amount of A in the reactor feed ( $y_A$ )	184.90	1	184.90	9.74	0.0097	-3.40
D: Amount of A+C in the reactor feed ( $y_{AC}$ )	1093.39	1	1093.39	57.58	< 0.0001	8.27
E: Reactor Temperature	21.22	1	21.22	1.12	0.3131	-1.15
CE	225.63	1	225.63	11.88	0.0055	3.76
DE	904.21	1	904.21	47.62	< 0.0001	7.52
Curvature	2017.86	1	2017.86	106.27	< 0.0001	3.08
Residual	208.87	11	18.99			
Lack of Fit	174.46	8	21.81	1.90	0.3234	
Pure Error	34.41	3	11.47			
Cor Total	6168.18	19				
$R^2$						63.90%
Adjusted $R^2$						42.84%
$R^2$ prediction						34.71%

### Phase 2 – Second-order model

Augmenting the resolution V fractional factorial design with 10 additional axial points run in a new block produced a central composite design, allowing for estimation of a second order model. We simulated the second block of experimental runs in sequence as a continuation of the first 20 runs and used the same procedure to calculate run averages as in the first block. We did not impose any block effect in the simulations. The analysis of the 30-run augmented design gives the second order model shown in the ANOVA table (Table 10). The residual analysis indicated that the 15<sup>th</sup> run (standard order #2) could be an outlier. However, as we did not find a reasonable explanation for this run's behavior, we chose to include it in the model despite the slight decrease in the  $R^2$ ,  $R^2$  adjusted, and  $R^2$  predicted statistics. Table 10 thus presents the ANOVA table of the augmented design in Table 8 (5% significance level). The non-significant lack of fit and the high values of the  $R^2$  statistics indicate that the model has a good fit and predictive ability.

Table 10. ANOVA and estimated effects for the augmented design using observations in both blocks. The model includes only those terms significant on a 5% significance level. Third order and higher interactions are ignored.

Source	Sum of Squares	df	Mean Square	F Value	Prob > F	Estimated Standardized Effects
Block	0.49	1	0.49			
Model	9896.57	10	989.66	51.10	< 0.0001	
A: Reactor Liquid Level	188.15	1	188.15	9.72	0.0060	2.80
B: Reactor Pressure	1809.73	1	1809.73	93.45	< 0.0001	-8.68
C: Amount of A in the reactor feed ( $y_A$ )	159.56	1	159.56	8.24	0.0102	-2.58
D: Amount of A+C in the reactor feed ( $y_{AC}$ )	1924.18	1	1924.18	99.36	< 0.0001	8.95
E: Reactor Temperature	37.30	1	37.30	1.93	0.1821	-1.25
CE	225.63	1	225.63	11.65	0.0031	3.76
DE	904.21	1	904.21	46.69	< 0.0001	7.52
C <sup>2</sup>	160.68	1	160.68	8.30	0.0100	2.40
D <sup>2</sup>	2772.06	1	2772.06	143.14	< 0.0001	9.95
E <sup>2</sup>	2453.31	1	2453.31	126.68	< 0.0001	9.36
Residual	348.58	18	19.37			
Lack of Fit	314.17	15	20.94			
Pure Error	34.41	3	11.47			
Cor Total	10245.64	29				
R <sup>2</sup>						96.60%
Adjusted R <sup>2</sup>						94.71%
R <sup>2</sup> prediction						89.49%

We then minimized the operating cost within the experimental design region spanned by the low and high levels of the factors in Table 8 based on the model in Table 10. The numerical optimization tool in the Design Expert® was used to search the design space, and Table 11 presents the settings of the reactor set-points that result in the lowest predicted cost.

Table 11. The suggested setting of the reactor set-points to obtain lowest operating cost.

Loop	Controlled variable	Suggested set-points setting
7	Stripper liquid rate (production)	Not in model (refer to Table 8)
9	Stripper liquid level	Not in model (refer to Table 8)
10	Separator liquid level	Not in model (refer to Table 8)
11	Reactor liquid level	70.37 %
12	Reactor pressure	2701.30 kPa
13	Mole % G (product quality)	Not in model (refer to Table 8)
14	Amount of A in reactor feed ( $y_A$ )	63.67 %
15	Amount of A+C in reactor feed ( $y_{A+C}$ )	52.25 %
16	Reactor temperature	124.25 °C
Estimated process operating cost		117.07 \$/h

### *Phase 3: Confirmation runs*

Three additional confirmation runs were simulated in the TE process using the suggested set-points (Table 11). The average cost of these runs was 117.07 \$/h. An average operating cost of 117.16 \$/h represents a reduction of 30.44 \$/h compared to the operating cost when starting set-point values are used, a reduction that most production engineers would deem considerable.

### **6.3. Closing remarks**

The sequential experimentation example illustrates exploring DoE methodologies in processes where engineering process control is present using the TE process simulator as a testbed. The example shows how a continuous process operating in closed-loop can be improved by shifting the set-points of the controllers. Experimental plans can help to explore the relationship between set-points and overall process performance indicators such as process cost or product quality. Note that the change in operating conditions invoked by the recommended change of the set-points may require re-tuning of the controllers in the system. We have not done that. That is, we assume that the control configuration and settings can still maintain the stability of the system in the new operating condition based on the new set-points. In our approach, we use DoE as a systematic solution to reduce the cost of the TE process based on an existing control system without

redesigning it. As such, it resembles ideas in the so-called retrofit self-optimizing control approach from the engineering control domain described by Ye et al. (2016).

## **7. Conclusions and Discussion**

The TE process simulator is one of the more complex simulators available that offers possibilities to simulate a nonlinear, dynamic process and operates in closed-loop useful for both methodological research and teaching. In this article, we provide guidelines for using the revised TE process simulator, run with a decentralized control strategy, as a testbed for new SPC and DoE methods. In our experience, understanding the details of the TE process simulator and getting it to run may be challenging for new users. The main contribution of this article is the flowcharts coupled with recommended settings of the TE process that will help a new user of the simulator to get started. Another contribution is the suggested approach of how to induce random variation in the simulator. The possibility of introducing random variability in the simulator improves the usability of the TE process simulator in the SPC and DoE contexts. This way, independent simulations can now be produced for SPC applications, and independent replicates can be run in an experimental application.

In the two examples provided, we illustrate some of the challenges that an analyst normally face when applying SPC and DoE in continuous processes operating under closed-loop. We would like to reiterate that the illustrated examples are only examples of applications for which the TE process simulator can be used. We believe that the revised TE process simulator offers ample opportunities for studying other and more complicated scenarios that will mimic real-life applications.

## **Acknowledgements**

The authors gratefully acknowledge the financial support from the Swedish Research Council under grant number 340-2013-5108.

## **References**

Bathelt, A., Ricker, N.L. and Jelali, M. (2015a). Revision of the Tennessee Eastman Process Model. *IFAC-PapersOnLine*, 48(8):309-314.



- Bathelt, A., Ricker, N.L. and Jelali, M. (2015b). *Tennessee Eastman Challenge Archive*. January, Available from: [http://depts.washington.edu/control/LARRY/TE/download.html#Updated\\_TE\\_Code](http://depts.washington.edu/control/LARRY/TE/download.html#Updated_TE_Code).
- Bisgaard, S. and Kulahci, M. (2005). Quality Quandaries: The Effect of Autocorrelation on Statistical Process Control Procedures. *Quality Engineering*, 17(3):481-489.
- Box, G.E.P. (1957). Evolutionary Operation: A Method for Increasing Industrial Productivity. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 6(2):81-101.
- Box, G.E.P. and Wilson, K.B. (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(1):1-45.
- Capaci, F., Bergquist, B., Kulahci, M. and Vanhatalo, E. (2017). Exploring the use of Design of Experiments in Industrial Processes Operating Under Closed-Loop Control. *Quality and Reliability Engineering International*, In Press.
- Chinosi, M. and Trombetta, A. (2012). BPMN: An Introduction to the Standard. *Computer Standards & Interfaces*, 34(1):124-134.
- Dennis, D. and Meredith, J. (2000). An Empirical Analysis of Process Industry Transformation Systems. *Management Science*, 46 (8):1058-1099.
- Downs, J.J. and Vogel, E.F. (1993). A Plant Wide Industrial Process Control Problem. *Computers & Chemical Engineering*, 17(3):245-255.
- Ferrer, A. (2014). Latent Structures-Based Multivariate Statistical Process Control: A Paradigm Shift. *Quality Engineering*, 26(1):72-91.
- Ge, Z., Song, Z. and Gao, F. (2013). Review of Recent Research on Data-Based Process Monitoring. *Industrial & Engineering Chemistry Research*, 52(10):3543-3562.
- Hsu, C., Chen, M. and Chen, L. (2010). A Novel Process Monitoring Approach with Dynamic Independent Component Analysis. *Control Engineering Practice*, 18(3):242-253.
- Kruger, U., Zhou, Y. and Irwin, G.W. (2004). Improved Principal Component Monitoring of Large-Scale Processes. *Journal of Process Control*, 14(8):879-888.
- Ku, W., Storer, R.H. and Georgakis, C. (1995). Disturbance Detection and Isolation by Dynamic Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 30(1):179-196.
- Kulahci, M. and S. Bisgaard. (2006). *Challenges in Multivariate Control Charts with Autocorrelated Data*. ISSAT International Conference on Reliability and Quality in Design.
- Lee, G., Han, C. and Yoon, E.S. (2004). Multiple-Fault Diagnosis of the Tennessee Eastman Process Based on System Decomposition and Dynamic PLS. *Industrial & Engineering Chemistry Research*, 43(25):8037-8048.
- Liu, K., Fei, Z., Yue, B., Liang, J. and Lin, H. (2015). Adaptive Sparse Principal Component Analysis for Enhanced Process Monitoring and Fault Isolation. *Chemometrics and Intelligent Laboratory Systems*, 146:426-436.
- Lundkvist, P. and Vanhatalo, E. (2014). Identifying Process Dynamics through a Two-Level Factorial Experiment. *Quality Engineering*, 26(2):154-167.
- Montgomery, D. (2012). *Statistical process control: A modern introduction*. 7th ed., Hoboken (NJ), Wiley.
- Rato, T.J. and Reis, M.S. (2013). Fault Detection in the Tennessee Eastman Benchmark Process using Dynamic Principal Components Analysis Based on Decorrelated Residuals (DPCA-DR). *Chemometrics and Intelligent Laboratory Systems*, 125:101-108.
- Reis, M. and Kenett, R.S. (2017). A Structured Overview on the use of Computational Simulators for Teaching Statistical Methods. *Quality Engineering*, 29(4):730-744.
- Ricker, L.N. (1996). Decentralized Control of the Tennessee Eastman Challenge Process. *Journal of Process Control*, 6(4):205-221.
- Ricker, L.N. (2005). *Tennessee Eastman Challenge Archive*. Available from: [http://depts.washington.edu/control/LARRY/TE/download.html#Decentralized\\_control](http://depts.washington.edu/control/LARRY/TE/download.html#Decentralized_control)
- Vanhatalo, E., Bergquist, B. and Vännman, K. (2013). Towards Improved Analysis Methods for Two-Level Factorial Experiment with Time Series Responses. *Quality and Reliability Engineering International*, 29(5):725-741.
- Vanhatalo, E., Kvamström, B., Bergquist, B. and Vännman, K. (2010). A Method to Determine Transition Time for Experiments in Dynamic Processes. *Quality Engineering*, 23(1):30-45.
- Vanhatalo, E. and Bergquist, B. (2007). Special Considerations when Planning Experiments in a Continuous Process. *Quality Engineering*, 19(3):155-169.

- Vanhatalo, E., Kulahci, M. and Bergquist, B. (2017). On the Structure of Dynamic Principal Component Analysis used in Statistical Process Monitoring. *Chemometrics and Intelligent Laboratory Systems*, 167:1-11.
- Vining, G., Kulahci, M. and Pedersen, S. (2015). Recent Advances and Future Directions for Quality Engineering. *Quality and Reliability Engineering International*, 32(3):11-21.
- Webb, D.F., Lucas, J.M. and Borkowski, J.J. (2004). Factorial Experiments when Factor Levels are Not Necessarily Reset. *Journal of Quality Technology*, 36(1):1-11.
- Ye, L., Y. Cao, X. Yuan and Z. Song. (2016). *Retrofit Self-Optimizing Control of Tennessee Eastman Process*. 2016 IEEE International Conference on Industrial Technology (ICIT).



