On Design of Experiments in Continuous Processes

ERIK VANHATALO

Doctoral Thesis number 18 in the subject
Quality Technology and Management
Division of Quality Technology, Environmental Management, and Social Informatics
Department of Business Administration and Social Sciences
Luleå University of Technology
Copyright © Erik Vanhatalo (2009)
ACKNOWLEDGEMENT

The research presented in this thesis was carried out at the Division of Quality Technology, Environmental Management and Social Informatics, Luleå University of Technology between November 2005 and November 2009. Indeed, this work would not have been possible without the important contributions from a large number of people and organizations.

I am indebted to my supervisors Dr. Bjarne Bergquist and Prof. Kerstin Vännman for their continuous guidance, support, and valuable cooperation during the work presented in this thesis. Thank you!

I thank LKAB for the financial support of this research, for the ongoing exchange of ideas and research results during the research project, and for providing me with a very interesting and challenging research case – the LKAB Experimental Blast Furnace (EBF). The financial support from the European Union, European Regional Development Fund, Produktion Botnia is gratefully acknowledged.

Many people at LKAB have made important contributions to the work presented in this thesis. First of all, I thank Gunilla Hyllander who has been the head of the ‘EBF methodology development project’ and worked tirelessly together with me to move the project forward. Gunilla has made important contributions to the research presented in this thesis. For valuable contributions, interesting discussions and support I thank: Mats Hallin, Anna Dahlstedt, Peter Sikström, Nicklas Eklund, Guangqing Zuo, Jonas Lövgren, Mikael Pettersson, Carina Brandell, Anna Brännmark, Sofia Nordquist, Per-Ola Eriksson, and Bo Lindblom. They have all made important contributions to the results presented in this thesis and made my visits to LKAB a pleasure.

I want to thank all my outstanding colleagues at the university for their friendship and support. Special thanks to Björn Kvarnström for all interesting collaborations, exchange of ideas, discussions, and ‘hockey talk’ during the years. For valuable feedback at, for example, pie seminars I thank Thomas Zobel, Peter Söderholm, Bengt Klefjö, Karin Schön, Åsa Wreder, Malin Albing, Anna-Karin Jonsson-Kvist, Klara Palmberg, Peder Lundqvist, Fredrik Backlund and Mari Runardotter.

Special thanks to Dr. Rickard Garvare for priceless feedback and comments on an early draft of this thesis.

Finally, but certainly not least, I thank my family and friends for their support and, at times, taking my mind off the ‘continuous process’ of research questions and questioning of my research. I dedicate this thesis to my mother Doris and my grandparents Elsa and Bertil, who always support and encourage me.

Erik Vanhatalo
Luleå, November 2009
ABSTRACT

Design of Experiments (DoE) includes powerful methods, such as factorial designs, to help maximize the information output from conducted experiments while minimizing the experimental work required for statistically significant results. The benefits of using DoE in industry are thoroughly described in the literature although the actual use of the methods in industry is far from being pervasive.

Continuous processes, frequently found in the process industry, highlight special issues that are typically not addressed in the DoE literature. The overall objective of this research is to increase the knowledge of DoE in continuous processes. More specifically, the aims of this research are [1] to identify, explore, and describe potential problems that can occur when planning, conducting, and analyzing experiments in continuous processes, and [2] to propose methods of analysis that help the experimenter in continuous processes tackle some of the identified problems.

This research has focused on developing analysis procedures adapted for experiments in continuous processes using a combination of existing DoE methods and methods from the related fields: multivariate statistical methods and time series analysis. The work uses real industrial data as well as simulations. The method is dominated by the study of the practical use of DoE methods and the developed analysis procedures using an industrial case – the LKAB Experimental Blast Furnace plant.

The results are presented in six appended papers. Paper A provides a tentative overview of special considerations that the experimenter needs to consider in the planning phase of an experiment in a continuous process. Examples of important experimental complications further discussed in the papers are: their multivariate nature, their dynamic characteristics, the need for randomization restrictions due to experimental costs, the need for process control during experimentation, and the time series nature of the responses. Paper B develops a method to analyze factorial experiments with randomization restrictions using principal components combined with analysis of variance. Paper C shows how the use of the multivariate projection method principal component analysis can reduce the monitoring problem for a process with many and correlated variables. Paper D focuses on the dynamic characteristic of continuous processes and presents a method to determine the transition time between experimental runs combining principal components and transfer function-noise models and/or intervention analysis. Paper E further addresses the time series aspects of responses from continuous processes and illustrates and compares different methods to analyze two-level factorials with time series responses to estimate location effects. In particular, Paper E shows how multiple interventions with autoregressive integrated moving average models for the noise can be used to effectively analyze experiments in continuous processes. Paper F develops a Bayesian procedure, adapted from Box and Meyer (1986), to calculate posterior probabilities of active effects for unreplicated two-level factorials, successively considering the sparsity, hierarchy, and heredity principles.

Keywords: Design of Experiments, Continuous process, Process industry, Multivariate statistical methods, Process monitoring and control, Time series analysis, Analysis of unreplicated factorials.
SWEDISH ABSTRACT

Försöksplanering omfattar kraftfulla metoder, exempelvis faktorförsök, för att maximera informationsutbytet vid experiment och samtidigt minimera de resurser som krävs för att nå statistiskt säkerställda resultat. Nyttan av att använda försöksplanering vid industriella experiment är väl beskriven i litteraturen men varken kännedomen om eller användningen av metoderna är lika utbredd i industrin.


Denna forskning fokuserar på att utveckla analysmetoder anpassade för experiment i kontinuerliga processer genom att kombinera befintliga metoder inom försöksplanering med metoder från de närbelägna områdena multivariat dataanalys och tidsserieanalys. Arbetet använder verklig industriell data samt simuleringar. Forskningsmetoden domineras av praktiskt användande och tester av metoder inom försöksplanering och de utvecklade analysmetoderna kring ett verkligt industriellt fall – LKAB:s experimentmasugn i Luleå.


Nyckelord: Försöksplanering, Kontinuerlig process, Processindustri, Multivariat dataanalys, Processövervakning och stryning, Tidsserieanalys, Analys av icke-upprepade faktorförsök.
CONTENTS

1. INTRODUCTION ........................................................................................................ 1
   1.1 Industrial experiments .......................................................................................... 1
   1.2 Design of Experiments ......................................................................................... 2
   1.3 Continuous processes ........................................................................................... 4
   1.4 Design of Experiments in continuous processes .................................................... 5
   1.5 Research objective and scope ............................................................................ 10
   1.6 The organization of the thesis ............................................................................ 10

2. RESEARCH METHOD .............................................................................................. 15
   2.1 An introduction ................................................................................................... 15
   2.2 A summary of the research process ..................................................................... 16
   2.3 Research purpose, scope, and strategy ................................................................. 21
   2.4 Data collection and analysis ................................................................................ 23
   2.5 Research quality .................................................................................................. 24

3. INTRODUCTION TO THE APPENDED PAPERS ............................................ 27
   3.1 Planning and performing experiments in continuous processes ................................ 27
      3.1.1 Experimental design and continuous processes ............................................. 29
      3.1.2 The need for process control and monitoring during experimentation .......... 33
   3.2 Analysis of experiments in continuous processes ............................................... 36
      3.2.1 Multivariate statistical analysis of experiments .............................................. 37
      3.2.2 Time series analysis and experiments in continuous processes ...................... 39
      3.2.3 Analysis of unreplicated two-level factorials ............................................... 42

4. CONCLUSIONS AND DISCUSSION .................................................................. 47
   4.1 Conclusions and recommendations .................................................................... 47
   4.2 Reflections on the research process .................................................................... 50
   4.3 Contribution ........................................................................................................ 52
   4.4 Implications and potential benefits for the experimental work at the EBF ............ 52

5. FUTURE RESEARCH ......................................................................................... 55

APPENDIX I – ABOUT THE BLAST FURNACE PROCESS .................................. 57
   A.1 The blast furnace process .................................................................................... 57
   A.2 The LKAB Experimental Blast Furnace (EBF) .................................................... 59

REFERENCES ........................................................................................................... 67

APPENDED PAPERS (A – F)
APPENDED PAPERS

This thesis includes the following six papers. The papers, which are appended in full, are summarized and discussed in the thesis.


1 Paper A was also presented by Erik Vanhatalo, as an invited paper, on October 9th, 2008 at the 52nd Annual Fall Technical Conference in Mesa, Arizona, USA.


3 Paper D was also presented by Erik Vanhatalo at the 9th Annual Conference of the European Network for Business and Industrial Statistics (ENBIS9) in Gotheburg, Sweden, on September 23rd, 2009.
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ARIMA</td>
<td>Autoregressive integrated moving average</td>
</tr>
<tr>
<td>ARMA</td>
<td>Autoregressive moving average</td>
</tr>
<tr>
<td>CUSUM</td>
<td>Cumulative sum</td>
</tr>
<tr>
<td>DoE</td>
<td>Design of experiments</td>
</tr>
<tr>
<td>EBF</td>
<td>The LKAB experimental blast furnace in Luleå, Sweden</td>
</tr>
<tr>
<td>EVOP</td>
<td>Evolutionary operation</td>
</tr>
<tr>
<td>EWMA</td>
<td>Exponentially weighted moving average</td>
</tr>
<tr>
<td>LKAB</td>
<td>The Swedish mining industry company Loussavaara Kiirunavaara AB</td>
</tr>
<tr>
<td>LTU</td>
<td>Luleå University of Technology</td>
</tr>
<tr>
<td>MANOVA</td>
<td>Multivariate analysis of variance</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PLS</td>
<td>Projection to latent structures by use of partial least squares</td>
</tr>
<tr>
<td>RSM</td>
<td>Response surface methodology</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

This chapter provides an introduction and background to the research area. The objective and scope of the research and the organization of the thesis are then presented.

1.1 Industrial experiments

An important way to gain knowledge about processes and products in industry is by experimenting. Experiments are also a fundamental part of research. However, conducting experiments in industry is normally expensive and knowledge about a process or product that can be gained in other ways is often a less costly alternative. Studying historical process operation or consulting process expertise may provide the needed information. Nonetheless, sometimes the only possible or best way to gain new knowledge about processes and products or to verify suspected process behavior is to perform experiments.

An experiment can be defined as “a test or a series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reasons for changes that may be observed in the output response,” see Montgomery (2009, p. 1). The process under experimentation can have one or many responses (Ys), see Figure 1.1. The purpose of an experiment is to measure the effects that the controllable factors4 (the Xs) have on the output response. In reality, though, there are often factors that are impossible, or too expensive, to control during an experiment (the Zs), so-called disturbance factors (or noise factors), that also affect the response.

Experiments are normally conducted on controlled systems (Cox and Reid, 2000). That is, the important features of the investigated materials, the nature of the studied manipulations of the system and the measurement procedures are all determined by the experimenter. By contrast, in an ‘observational study’ the investigator does not control all of these features even though the objective of the two types of studies may be identical (Cox and Reid, 2000). Hence, experiments make it possible to verify causality between experimental factors and process responses in a way that might be difficult through an observational study.

4 Variables such as experimental variables are often labeled “factors” in DoE literature. “Factors” and “variables” are used interchangeably in this thesis to label such entities that affect the system under experimentation.
1.2 Design of Experiments

The costs of an experiment always make it valuable to maximize the information output while at the same time minimize the resources required for producing this information. According to Wu and Hamada (2000, p.1) Design of Experiments (DoE) can be viewed as:

*a body of knowledge and techniques that enables an investigator to conduct better experiments, analyze data efficiently, and make the connections between the conclusions from the analysis and the original objectives of the investigation.*

Consequently, DoE is useful for an experimenter who wants to understand and improve a product or a process and wants to do it effectively and efficiently.

Two of the pioneers of the DoE field were Ronald A. Fisher and Frank Yates, who worked on problems in agriculture and biology at Rothamsted Experimental Station in the 1920s and 1930s, see Box (1980). Some of Fisher’s many contributions of statistical insight into the DoE field were to point out the importance of randomization of experimental treatments, introducing the use of the analysis of variance to judge the significance of effects, and not least factorial designs, see Fisher (1925; 1926). The essence of factorial designs is that several experimental factors are studied simultaneously instead of one at a time.

Much development has occurred within the DoE field since the 1930s. Steinberg and Hunter (1984) provide a review of the development up until the mid 1980s. Brief historical summaries over the development can also be found in, for

---

Wu and Hamada (2000) and many other authors, for example, Box et al. (2005), use the concept “Experimental design” to label the body of knowledge which is also often referred to as “Design of Experiments” (used throughout the thesis).
example, Montgomery (2009) and Wu and Hamada (2000). One of the most significant contributions was the introduction of fractional factorial designs, see Finney (1945). After World War II, the methods got a boost when they were developed to tackle problems in industrial processes (especially in the chemical industry). Key contributions were made by, for example, George E. P. Box leading to, for instance, development of response surface methods and sequential experimentation for process optimization, see Box and Wilson (1951). According to Steinberg and Hunter (1984), other important subjects within DoE that received attention during the 1970s were design optimality, computer-aided design, and mixture designs. Since the 1980s and the much debated work of G. Taguchi, discussed by, for example, Box et al. (1988), an increased focus has been on experimental designs for variation reduction in products and processes. According to Borror et al. (2000) and Montgomery (2009), the increased interest in quality improvement by Western industries together with Taguchi methods helped to expand the use of DoE. In particular, the methods became more widely used in discrete parts industries, such as automotive and electronics manufacturing. Today, DoE has grown far beyond the agricultural area and is now used in many areas of science and engineering.

DoE contains many statistical methods and therefore knowledge about statistics is a central part of understanding how the methods work. DoE along with statistical process control was early adopted by the quality movement and are often classified as important methodologies within quality management, see, for example, Hellsten and Klefsjö (2000), Deleryd et al. (1999), Xie and Goh (1999), and Powell (1995). Today, important forum for the development of DoE methods are, for example, the journals published by the American Society for Quality and other journals with the word quality in their names. DoE is also one of the important methodologies in the strategic quality improvement initiative Six Sigma, see Goh (2002). Since quality improvement is linked to reduction of variation in products and processes, not least due to Shewhart (1931), statistical thinking and quality improvement is closely connected (Snee, 1990).

DoE is well known by professionals in the fields of statistics and quality, but Goh (2001) argues that the use of DoE in industry is far from being pervasive. Studies of industrial use of DoE show mixed results. In Sweden, Gremyr et al. (2003) report that DoE is used by a little over 50 percent of the studied industries, while Bergquist and Albing (2006) present a much lower number in their study. Tanco et al. (2008) report that DoE was used by about 20 percent of the companies
in a survey of manufacturing industries in Spain. The differing results are probably
dependent on the different countries and samples, the types of respondents, and how
the definitions of the word ‘use’ are made in the studies. It is, however, clear that
there still exists substantial improvement potential in spreading the use of DoE in
many industries.

As mentioned, statistics is a central part of DoE, but when statistical methods
are applied it is important not to forget about non-statistical knowledge. Box et al.
(2005, p.13) claim that “statistical techniques are useless unless combined with
appropriate subject matter knowledge and experience.” Thus, both statistical skills
and process knowledge are needed to successfully design, conduct and analyze an
experiment. In this thesis, prior process knowledge is defined as knowledge of the
process or product that the experimenter or experimental team possess. Examples of
such knowledge can be theoretical knowledge of chemical reactions and physical
relations in a process or collected experience from running the process and from
previous experiments.

1.3 Continuous processes

There are a number of ways to run production processes in industry. A rough
partitioning of different industrial production processes, used in this thesis, is into
non-continuous and continuous production. Non-continuous production can be
further divided into discrete and batch production. Discrete production (or parts
production) typically have distinct operations where each operation contributes with
specific outputs to achieve the overall process output (Hild et al., 2000). By contrast,
in continuous processes (frequently found in the process industries), the product
gradually and with minimal interruptions passes through a series of different
operations and exhibits characteristics such as liquids, powders, slurries and other
non-discrete states (Dennis and Meredith, 2000; Fransoo and Rutten, 1994). APICS
dictionary (2008, p. 25) describe a continuous process as:

*a production system in which the productive equipment is organized and sequenced
according to the steps involved to produce the product. This term denotes that
material flow is continuous during the production process. The routing of the jobs is
fixed and setups are seldom changed.*

APICS dictionary (2008, p. 104) in turn describe process industry as:

*the group of manufacturers that produce products by mixing, separating, forming,
and/or performing chemical reactions.*
Continuous production processes can be found in, for example, the pulp and paper industries, chemical industries, parts of the medical and food industries, as well as parts of the mining and steel industries. The blast furnace process, which is studied more closely in this work, is an example of a continuous process in the steel industry.

Continuous production is generally characterized by, for example: high-technological and complex production processes, capital-intensive production plants, low-technological products, low added value to products, high production speed, low equipment flexibility, large change-over times, large volumes of product, and divergent product flow (many products are produced from a few raw materials), see, for example, Rajaram and Robotis (2004), Dennis and Meredith (2000), Fransoo and Rutten (1994), and Kim and Lee (1993).

According to Fransoo and Rutten (1994), continuous processes are typically hard to control, leading to variable yield and reflux flows of material. The raw materials to the process industries often come from mining and agricultural industries and this means that the materials often are afflicted with natural variations.

1.4 Design of Experiments in continuous processes

There is a vast literature on DoE for use in industrial processes. However, in well-known and comprehensive textbooks about DoE, such as Box, Hunter and Hunter (2005), Montgomery (2009), and Wu and Hamada (2000), most illustrated applications of DoE are exemplified for non-continuous processes, that is, parts or batch production. An article search combining search strings such as: ‘design of experiments’, ‘experimental design’, ‘continuous process’, and ‘continuous production’ in databases and search engines like SCOPUS or Google Scholar gives very few hits. Such a result can depend on (at least) two reasons. Either there are no special issues that arise when planning, conducting, and analyzing experiments in continuous processes and the general recommendations in the existing literature apply, or the potential special issues in continuous processes have not been given much attention in the literature. This thesis argues in favor for the latter case and that planning, conducting, and analyzing experiments in continuous processes highlight special issues that are typically not addressed in DoE literature.

Prior work explicitly focusing on DoE in continuous processes is limited. Saunders and Ecclestone (1992) discuss the high degree of autocorrelation in many

---

6 http://www.scopus.com
7 http://www.scholar.google.com
continuous processes and provide some recommendations about the proper sampling interval of the responses. Saunders et al. (1995) and Cheng and Steinberg (1991) provide algorithms to maximize the number of level changes for experiments in processes with time trends (such as continuous processes). Hild et al. (2000) discuss how the characteristics of continuous processes can affect the use of statistical tools such as DoE within the Six Sigma framework.

Related topics within the DoE field that are relevant for continuous process and process industry problems are the methods of ‘Response Surface Methodology’ (RSM), introduced by Box and Wilson (1951), and ‘Evolutionary Operation’ (EVOP), introduced by Box (1957). RSM has been widely used in process industry (Myers et al., 2004), especially in the chemical industries, and typically applies smaller sequential experiments to attain optimal conditions for products and processes. The EVOP procedure is a method for continuous monitoring and improvement of full-scale plants. EVOP typically uses a small two-level factorial design, $2^k$, to introduce small changes in the levels of process variables and then runs a number of cycles of the design to investigate the significance of the effects (Myers and Montgomery, 2002). However, EVOP is suitable for processes with high-volume production over a reasonable extensive time period and the process also needs to stabilize rapidly after a process change (Hahn and Dershowitz, 1974).

Another important field of research connected to issues in continuous processes is ‘chemometrics’, where statistical methods (often multivariate) have been developed to analyze data in process industry settings (mainly the chemical industry), see Wold (1995) for a discussion of chemometrics. Multivariate statistical methods, such as Principal Component Analysis (PCA) and Projection to Latent Structures by use of Partial Least Squares (PLS) were further developed during the 1980s and 1990s to meet new challenges from the more richly instrumented processes in industry. Multivariate methods become more and more relevant today since the amount of data that is collected and stored in databases seems to be ever increasing. Han and Kamber (2001) claim that in such environments you operate in a ’data rich but information poor situation.’

1.4.1 Experimental challenges in continuous processes

When studying the existing DoE literature and literature from related areas a number of experimental challenges in continuous processes emerge that are summarized here.
INTRODUCTION

Dynamic systems

A prerequisite to correctly estimate the effect of the change in factor X on the response Y is that full impact of the factor changes on the process must have taken place. In continuous processes, the propagation of a disturbance (for instance when changing an experimental factor X) can take time which can lead to requirements of prolonged experimental runs compared to experiments in non-continuous production (Saunders and Eccleston, 1992). Black-Nembhard and Valverde-Ventura (2003) differentiate between ‘dynamic’ and ‘responsive’ systems. They explain that in a dynamic system, a period of delay will occur between the time that X is changed and the time that this change is realized in the output Y, while this change in Y is immediate in a responsive system. The time needed for the change in X to reach full impact in the process, in this thesis often referred to as the transition time, can also depend on which factor that is changed and how large the change is. The design of continuous processes often include, for example, tanks, reactors, chemical reactions, buffer systems, reflux flows, mixing, product state changes and so on, which typically make continuous processes dynamic systems.

Draper and Stoneman (1968) make the point that, in some situations, it can be desirable to keep the number of level changes at a minimum, since the time required for apparatus to return steady-state after changes can be considerable and depend on the number of factors that are changed. In fact, this is opposite to the recommendations of Saunders et al. (1995), Cheng and Steinberg (1991), and Meyer and Napier-Munn (1999) who focus on maximizing the number of level changes to keep bias from time trends in the process to affect the responses. John (1990) also discuss how factorial designs can be made robust against time trends. Meyer and Napier-Munn (1999) recommend that the sampling intervals should be as small as possible, but not smaller than the time it takes for the process to reach a new equilibrium state. Since recommended designs for time-dependent (autocorrelated) processes may have a large number of level changes they may be costly to implement (Martin et al., 1998). The experimenter may therefore need to balance optimal designs against practical and cost issues, for example, the time required for the process to return to steady-state after a change is made. Another related problem highlighted by Pan et al. (2004) is that many continuous processes are non-stationary and thus the steady-state assumption is often not reasonable for industrial data. Although the processes often are non-stationary (or not in statistical control), key DoE concepts like randomization, replication and blocking still make it possible to perform designed experiments in these processes, see Bisgaard et al. (2008).
Restrictions in factor levels and replication

According to Hild et al. (2000), it is seldom recommended to use bold levels of experimental factors in a continuous process as even small changes in factors can be exaggerated to unacceptable effects on the processing state as well as cause unwanted disruptions. This is in contrast to discrete processes where it is common to use fairly large intervals of experimental factor levels. However, the smaller the differences in factor levels are, the more difficult it is to detect the resulting effects of the factor changes. A large number of replicates might be needed in order for the effects to appear through the noise but since experiments in continuous processes normally are costly, the ability to replicate experimental runs is limited. The small changes in factor levels is an explanation of why the EVOP procedure is designed to run for many cycles before the effects can be found significant. As replication of experimental runs can be hard to realize, powerful analysis methods for unreplicated experiments may be of special importance for experiments in continuous processes.

Many cross- and autocorrelated responses

Hild et al. (2000) argue that not all process variables are independent from one another. The many inter-relationships limit which experimental factors and how many that can be manipulated in an experiment. Another obstacle for the experimenter is that data from continuous processes often are dominated by frequent on-line logging of process variables while the measurements of product characteristics usually are made less frequently and by off-line analysis (Hild et al., 2000). Therefore, the cause and effect relation between processing conditions and product characteristics can be hard to establish.

The frequent sampling of many continuous processes, combined with their dynamic behavior, often leads to positively autocorrelated responses, see Saunders et al. (1995), which can cause problems during analysis. For experiments run in continuous processes, responses of interest often need to be viewed in the form of time series. The autocorrelation indicates that time series analysis could be a useful tool. Time series analysis contains techniques where stochastic and dynamic models are developed to model the dependence between observations sampled at different times, see Box et al. (2008). By contrast, in discrete processes, responses are usually measured by single measurements on individual experimental units. It can be possible to remove (or reduce) the time-dependence of measurements by increasing the sampling interval between adjacent measurements, but this a costly strategy as the experiment is prolonged.
INTRODUCTION

In continuous processes the measured responses are typically not independent since a few underlying events often drive the process at any time. As it is often difficult to measure process events and reactions directly, many secondary responses such as temperatures, pressures and flows must be used as proxies for real process events. Hence, several of the measurements on process variables are merely different reflections of the same underlying event (Kourtì et al., 1996; Kourtì and MacGregor, 1995). The experimenter in a continuous process must therefore consider that responses are inter-connected, that is, a change in one variable often affects several other variables as well. Shifts in the processing state may be visible in multivariate representations but may, due to normal variation, not deviate significantly in univariate plots (Hild et al., 2000). Therefore, examining such responses one at a time makes interpretation difficult (Kourtì and MacGregor, 1995). Kourtì (2005), Duchesne and MacGregor (2000) and Wikström et al. (1998a, b) highlight this problem area and provide examples of how multivariate statistical techniques can be used for process analysis, and control and monitoring applications for continuous and batch processes. Examples of such techniques are the latent variable techniques PCA and PLS.

Process control

The experimenter in continuous processes must also be aware of that sometimes autonomous and automatic control systems are working to create process stability, which can counteract deliberate changes of X-factors. Hence, process responses may not be directly visible as changes in response variables (Y) but instead as changes in other X-factors, see Hild et al. (2000). The need for process control means that experiments in many process industries are performed under, so-called, ‘closed-loop’ conditions due to plant and personal safety reasons (Box and MacGregor, 1974).

Issues like the ones outlined above when performing experiments in continuous processes are typically not addressed in DoE literature. Clearly, these issues will affect the planning, execution and analysis of designed experiments in continuous processes. Hence, further research on DoE in continuous processes is a valuable contribution to improve the understanding of planning, conducting, and analyzing experiments in industries that operate continuous processes.
1.5 Research objective and scope

The overall objective of this research is to increase the knowledge of DoE in continuous processes. More specifically, the aims of this research are:

1. to identify, explore, and describe potential problems that can occur when planning, conducting, and analyzing designed experiments in continuous processes, and
2. to propose and develop methods for planning, conducting, and analyzing designed experiments that help the experimenter in continuous processes to tackle some of the above identified problems.

This research has an applied focus as the identified problems and the proposed methods have mainly been developed by studying the practical use of statistical methods in industry and by using real industrial data. That is, this research has a quality engineering perspective rather than an exclusively statistical standpoint. This research has focused on developing procedures adapted for experiments in continuous processes using a combination of existing DoE methods and methods from the related fields: multivariate statistical methods and time series analysis.

Much of the empirical evidence and data in support of this research have been collected by studying an experimental blast furnace plant (a continuous process), see Appendix I for further descriptions. There are many other types of continuous processes that could have been studied. This choice may limit the possibility to generalize the results somewhat due to the prerequisites for running a continuous blast furnace process. However, the issues outlined in Section 1.4 above fit the blast furnace processes well, and it is therefore argued that the experimental blast furnace process is a good representative for the population of continuous processes.

1.6 The organization of the thesis

This thesis includes a summary of the research method, an introduction to the appended papers, conclusions and discussion, and the six appended papers. The contents in the upcoming chapters are briefly summarized here and the relations between the papers are also outlined.

1.6.1 Research method

Chapter 2 provides a summary of the research method and process. This includes a description of the methodological choices made during the research and data collection activities.
1.6.2 Introduction to the appended papers

Chapter 3 serves as an introduction to the appended papers as well as a brief summary of the results. Chapter 3 also provides further theoretical foundations relevant for the appended papers and summarizes and discusses the results in the light of prior research. Chapter 4 gives conclusions and recommendation and discusses the contribution of the research. Chapter 5 presents recommendations for future research. Appendix I gives a brief introduction to the blast furnace process of ironmaking and background information about the LKAB Experimental Blast Furnace.

1.6.3 Paper A: Special considerations when planning experiments in a continuous process. Vanhatalo E. and Bergquist, B. (2007).

Paper A outlines a check-list for planning experiments in continuous processes based on prior recommendations for the experimental planning phase. A tentative list of special considerations that the experimenter needs to consider during the planning phase is developed. The paper primarily builds on the authors’ experiences of planning, conducting and analyzing experiments in collaboration with engineers at the experimental blast furnace. As a first paper it also presents an overview of special DoE issues that are important for continuous processes. Methods to address some of the important issues are developed in the following papers.

Data collection of the empirical material from the experimental blast furnace plant on which the paper builds, for example, interviews, observations, analysis of experimental data, and development of an experimental planning guide for LKAB, was mainly performed by Erik Vanhatalo. Bjarne Bergquist took active part in the analysis of the interviews and the empirical material and during discussions of the results with engineers at LKAB. The paper was mainly written by Erik Vanhatalo with contributions by Bjarne Bergquist.


Paper B discusses the planning and analysis of a specific experiment in the experimental blast furnace. It further illustrates some of the special issues that need to be considered for experiments in continuous processes outlined in Paper A and highlights, for example, their multivariate nature, dynamic characteristic, and the need for process control during experimentation. In particular the paper focuses on developing a method to analyze a factorial experiment with randomization
restrictions using a multivariate statistical method (PCA) combined with analysis of variance.

Both authors were involved in the planning of the experiment although Erik Vanhatalo handled all communication and discussions with the research engineer at charge of the experiment at LKAB (Gunilla Hyllander). The proposed analysis procedure was developed jointly by the authors and all calculations were performed by Erik Vanhatalo. The paper was mainly written by Erik Vanhatalo with contributions by Kerstin Vännman.

1.6.5 Paper C: Multivariate process monitoring of an experimental blast furnace. Vanhatalo, E. (2009).

Paper C is connected to one of the important issues discussed in papers A and B, namely the need for process control during experimentation. The control problem is hard to eliminate in continuous processes. However, to make well-informed control decisions (important for experimental results), the process state needs to be monitored. The paper focuses on the use of the multivariate projection method PCA to reduce the monitoring problem for richly instrumented industrial processes with many correlated variables. More specifically, a case study at the experimental blast furnace is outlined where principal components are used to monitor the thermal state of the process. The results show how the thermal state can be monitored by only studying a few principal components instead of many original variables. The paper also discusses the problem of multivariate monitoring of a process with frequently shifting operating modes and process drifts.


Paper D provides a deeper discussion of the dynamic characteristic of continuous processes outlined in papers A and B. After changes of experimental factors dynamic processes, such as continuous processes, undergo a transition time before full impact of the change has been reached. To minimize experimental time and reduce costs, knowledge about this transition time is important for the design and the analysis of the experiment. The paper proposes and illustrates a method to determine the transition time in a richly instrumented dynamic process combining principal component analysis and the time series analysis methods transfer function-noise modelling and intervention analysis.
INTRODUCTION

The conceptual ideas for the analysis method were sparked during a course in time series analysis and all authors helped develop the analysis method. The data analysis work in relation to the paper was made by Erik Vanhatalo with the assistance of Björn Kvarnström. Erik Vanhatalo mainly wrote the paper with contributions by the other authors.


Paper E focuses on the time series aspects of responses from continuous processes. The paper proposes and compares different methods to analyze time series responses and estimate location effects. Time series responses are simulated using dynamic propagations of the effects to mimic a situation that can occur in continuous processes. The results show how time series analysis and in particular multiple interventions with autoregressive integrated moving average models for the noise can be used to analyze two-level factorial experiments in a continuous process. Time series analysis of the responses is compared with, for example, ‘traditional’ analysis methods using averages as the single response in analysis of variance. The results indicate that by using intervention-noise models to estimate the significance of the effects, fewer spurious effects are found when the effects are small compared to the noise, and a larger number of the active effects are found when replication is limited. The results also show that using averages for each run as the single response is a straightforward and fairly robust analysis method, which is used to provide crude estimates of the effects needed to guide the analyst using the multiple intervention-noise models.

This paper was initiated by Erik Vanhatalo who developed the simulation model in Matlab® for the dynamic effects and time series responses, performed all the analyses, and did the main part of the writing of the paper. Bjarne Bergquist and Kerstin Vännman were both involved in the discussions leading up to the simulation of the time series, the proposed analysis methods, the setup of the study, and were also involved in the writing process.


Paper F focuses on the analysis of unreplicated factorials. The paper is not limited to continuous processes only. However, in continuous processes, issues like process stability concerns, cost of experimentation, and the relatively small differences in factor levels can lead to small unreplicated designs and in these cases powerful
analysis methods are of special importance. The viability of the sparsity, hierarchy, and heredity principles are first studied by analyzing experiments found in the literature. The results are then used for prior probability assessment in a Bayesian procedure, adapted from Box and Meyer (1986), to calculate posterior probabilities of active effects for unreplicated two-level factorials. A three-step approach is outlined using the results concerning the sparsity, hierarchy, and heredity principles. Individual prior probabilities for each effect being active are specified in three steps, successively considering sparsity, hierarchy, and heredity and posterior probabilities are calculated for each step.

The paper was originally initiated by Bjarne Bergquist who located and performed the analysis of the experiments found in the literature. Erik Vanhatalo worked to adapt the Box and Meyer (1986) approach to use the proposed three-step procedure and developed a calculation application in Matlab® with the help of Magnus Lundberg Nordenvaad. Magnus Lundberg Nordenvaad did the main development of the Markov chain Monte Carlo integration procedure needed for the Bayesian analysis method. Analysis of the examples in the paper and the writing of the paper were performed by Bjarne Bergquist and Erik Vanhatalo jointly.
2. RESEARCH METHOD

This chapter provides a summary of the research method and process including descriptions of the methodological choices and data collection activities made during the research.

2.1 An introduction

I first came in contact with DoE in 2002 during courses at Luleå University of Technology (LTU) within the master’s programme of Industrial and Management Engineering. I immediately found quality technology and applied statistics a very interesting area and I was happy when I got the chance to focus on this area in my PhD studies at LTU. Prior to me becoming a Ph.D. student I held a teaching position at the university for about 18 months where I taught, for example, introductory courses in DoE, statistical process control, and multivariate statistical methods for engineering students.

The research presented in this thesis began in late 2005 when I started as a Ph.D. student. The arrangement of my research project meant that I together with my supervisors Bjarne Bergquist and Kerstin Vännman became involved in a collaboration project with the Swedish mining industry company Loussavaara-Kirunavaara AB (LKAB).

LKAB had been running an Experimental Blast Furnace plant (hereafter the EBF) in Luleå since 1997, mainly for product development experiments and customer experiments. When initially discussing the forming of the collaboration project, the research engineers at the EBF (EBF engineers) had expressed an interest to improve their experimental work and they were also interested in testing factorial designs at the EBF. The typical experimental designs used at the EBF at the start of the project, were different forms of one-factor-at-a-time experiments.

The collaboration project was viewed as an excellent opportunity to conduct research on the use of DoE in continuous processes. Given that the EBF is specifically designed for experimental purposes it would present opportunities to study and take part in the planning and analysis of several experiments as well as to learn from the experimental experiences of the EBF engineers. This led to the start of the ‘Experimental Blast Furnace methodology development project’ (hereafter the EBF project) in November 2005. Two years into the project (in late 2007) it was decided to prolong the collaboration for another two years until November 2009. Hence, the research presented in this thesis has been conducted within the

8 More information about the LKAB company can be found at http://www.lkab.com
frame of this collaboration project and mainly with the EBF as the studied case. Further descriptions of the blast furnace process and the EBF are given in Appendix I and in the appended papers.

2.2 A summary of the research process

This section gives a summary and background to the main activities in the research process, and how the work presented in the appended papers came about.

The EBF project was formally started in November 2005. Gunilla Hyllander has been the head of the steering group at LKAB throughout the collaboration project, which also included research engineers from LKAB as well as me, Bjarne Bergquist and Kerstin Vännman. This steering group has been the forum for discussing and deciding on research activities, discussing emerging results, and an ongoing exchange of ideas between the researchers at the university and EBF engineers.

In late 2005 I began by searching the literature through database searches to find prior research of relevance. Initially little was found that explicitly focused on DoE in continuous processes. The search and study of literature has been an ongoing process during the entire research process where specific questions during the research have intensified the literature search concerning specific subject matter areas, for example multivariate statistical methods and time series analysis. As my understanding of the topic grew, I was able to locate more literature related to DoE in continuous processes. However, there seems to be little out there to find.

In January 2006 interviews with EBF engineers were performed to find a base and starting point on which to build further research activities. The purpose of the interviews were to create an understanding of how EBF engineers, at the starting point of the research project, were planning, conducting, and analyzing experiments carried out in the EBF and to gather their experiences and thoughts about using DoE in the EBF. Therefore, five engineers (a substantial part of all people involved in the planning of the experiments) were interviewed. The interviews were semistructured, following the seven-step interview process described by Kvale (1997, p. 85), since a structure provided by a list of questions (61 in total) was desirable. Merriam (1998, p. 72-75) further discusses semistructured interviews. Most questions were open-ended to make it possible for the respondents to

---

9 On behalf of LKAB, the project steering group has, with some minor variations over time, consisted of Gunilla Hyllander, Mats Hallin, Anna Dahlstedt, Peter Sikström, Nicklas Eklund, Guangqing Zuo, Carina Brandell, Jonas Lövgren, Mikael Pettersson, Anna Brännmark, and Per-Ola Eriksson.
elaborate and explain specific matters concerning the EBF process. Descriptions of
the seven steps of the interview process along with the questions can be found in

With the understanding created from the interviews the next step was to plan,
conduct and analyze a pilot test of a factorial experiment in the EBF. The purpose
of this experiment was to investigate the potential of using factorial experiments as
experimental designs in the EBF. The design, analysis and results of this experiment
are not elaborated at any length in the appended papers. Briefly, the experiment was
a $2^3$ factorial design with center points, testing the two process variables: ‘blast
volume’ and ‘moisture content of the blast air.’ The experiment required seven days
of operation in the EBF, and used 24 hours for each run. One of the conclusions
from the interviews and from the experience of running a factorial experiment in
the EBF was that a more structured way of planning experiments in the EBF was
needed. Therefore a new experimental planning guide was developed during the
spring and summer of 2006. This guide was developed in collaboration with the
EBF engineers and by incorporating recommendations found in the literature.
Further refinement of the planning guide came from using it to plan upcoming
experiments in the EBF. The essence of the planning guide is described by the
thirteen-step checklist given in Paper A.

The results from the pilot test of factorial designs in the EBF showed promising
potential but also raised new questions. An important concern was the minimum
length required for a factorial run in the EBF. There were also questions about the
proper analysis procedure for the experiments due to the multivariate nature of the
responses. Nonetheless, a new and somewhat more complex factorial experiment
was planned and conducted in October 2006. The second experiment tested two
experimental factors, one at three levels and one at two levels, see Paper B for more
details. This time experiences from the previous activities in the collaboration
project were used in the planning phase. Furthermore, after this second experiment,
the experience of working with the new planning guide was evaluated and flaws in
the guide were corrected to produce a template for future use at LKAB.

The period from November 2006 to May 2007 was spent reflecting over the
experiences from the two experiments in the EBF plant. The large number of
responses from the experiments in the EBF together with the recommendations of
multivariate analysis tools for similar situations in process industry found in the
literature led to a focus on multivariate statistical methods for the analysis. The
multivariate analyses of process data from the EBF experiments were made in close
collaboration with EBF engineers to be able to select important process responses and appropriate periods of operation in the EBF. The work up until this point led to the writing of two of the appended papers. Paper A comments on the learnings from the first two experiments in the EBF, the interviews, our frequent visits at the EBF plant, and the development of the planning guide. Paper B focuses on the design and analysis of the second (3×2) factorial experiment in the EBF and proposes an analysis method combining PCA [see e.g. Jackson (2003)] and analysis of variance [see e.g. Griffith et al. (1989)] to analyze the experiment that had randomization restrictions. The analysis method outlined in Paper B has also been used to analyze other experiments at the EBF.

When the decision to prolong the collaboration project with LKAB was made in late 2007 we had time to reflect and decide on some prioritized areas for the continued research. Processes in process industry often need to be controlled due to personal, and plant safety reasons and to control the quality of the product. Process control may be needed also during experiments in these processes. This was also the case at the EBF, and the need for process control during the experiments was one of the problems that had drawn our attention already at the start of the research process. It had become clear that process control of the thermal state of the EBF was unavoidable during the experiments and we had already worked to incorporate recommended control strategies in the experimental planning guide. Discussions in the steering group for the EBF project also concerned having some common decision criteria for when to perform control actions in the EBF process thus making control actions less subjective. Furthermore, it was clear that there were many process variables that needed to be monitored to form a basis for a control decision, which made monitoring complicated. Again, the problem was a multivariate one. Therefore, with start in the spring of 2008, work was initiated to investigate if multivariate process monitoring could aid the situation to achieve an overview of, for example, the thermal state in the process. Paper C written in late 2008 illustrates the method chosen for monitoring the thermal state of the EBF using principal components and outlines some further challenges to be able to use the method in the EBF. The monitoring method outlined in Paper C was tested on-line in an experimental campaign during the fall of 2008 and some further tests were made in the fall of 2009.

Around September 2008 we also started to take interest in time series analysis and the techniques available in that field. During the whole research project we had been handling responses from the EBF that were in form of time series. Up until
this point, the dynamics of the EBF process had been handled by simply excluding observations of the responses during the transition time between the experimental runs. The transition times had been estimated using the EBF engineers experience adding some margin to be on the ‘safe side.’ Time series analysis provided us with formal methods to approach this problem. The initial work concerned the use of the time series techniques transfer function-noise modeling and intervention analysis [see e.g. Box et al. (2008)] combined with principal components to more formally assess the transition times for experiments in the EBF and at the same time handle the multivariate nature of responses. This work was made in the spring of 2009 and the results are presented in Paper D.

During the summer and fall of 2009 we continued to work on the time series aspects of the responses from continuous processes. During the work with transfer function-noise models to estimate the transitions times the idea to use these models to analyze the entire time series from an experiment was initiated. In particular, it was interesting to compare a time series analysis approach with other ways of analyzing experiments with time series responses, such as using averages of the response in each run. Paper E describes this work where we chose to focus on two-level factorial experiments with time series responses. As experimental data from two-level factorials in the EBF were limited, a simulation program was built in Matlab® to be able to simulate responses from a continuous process using the EBF as inspiration for the dynamics of the effects and responses in the simulations. The results of this work are given in Paper E.

A question that had been discussed throughout the whole research project was effective analysis procedures for unreplicated factorials. The question is by no means limited to continuous processes but grew stronger in our minds since it soon became evident that large experiments with many replicates typically become too costly in industrial settings, not least for continuous processes. Hence, powerful analysis procedures for unreplicated experiments are of special importance for large-scale industrial experiments. Already in 2006 we started to think about how the analysis of unreplicated experiments could benefit from incorporating prior knowledge in form of the governing principles of sparsity, hierarchy, and heredity. The first idea was to specify a method using the normal probability plot. However, in the spring of 2009 the work was directed towards a Bayesian method that would allow us to incorporate the principles into the prior probabilities and create a more formal procedure. This work is presented in Paper F. Paper E also discusses the analysis of
unreplicated factorial designs, but focuses on experiments with time series responses. The whole research process is also summarized in Figure 2.1.

Figure 2.1 An overview of the main research activities during the research process.
RESEARCH METHOD

2.3 Research purpose, scope, and strategy

This section discusses and positions the research in relation to methodological aspects and comments on data collection and analysis activities.

Zikmund (2000), although with a focus on business research method, argues that research can be classified on the basis of its purpose and describe three categories: to explore, to describe, or to explain the phenomena under study. Marshall and Rossman (2006) describe how different research purposes are connected to these categories, although they focus on qualitative research methods. Inspired by the classification in Marshall and Rossman (2006), Table 2.1 shows how the purpose of the research presented in this thesis can be classified according to the three categories.

<table>
<thead>
<tr>
<th>Exploratory</th>
<th>Descriptive</th>
<th>Explanatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>To identify and explore</td>
<td>To document and describe</td>
<td>To explain why the identified</td>
</tr>
<tr>
<td>potential problems that can</td>
<td>problems that can occur</td>
<td>problems make it more difficult to</td>
</tr>
<tr>
<td>occur when planning,</td>
<td>when planning, conducting, and</td>
<td>perform experiments and how the</td>
</tr>
<tr>
<td>conducting, and analyzing</td>
<td>analyzing experiments in</td>
<td>proposed analysis methods can help</td>
</tr>
<tr>
<td>experiments in continuous</td>
<td>continuous processes.</td>
<td>the situation.</td>
</tr>
<tr>
<td>processes.</td>
<td>To describe the proposed methods</td>
<td></td>
</tr>
<tr>
<td>To explore how different</td>
<td>of analysis and how they can be</td>
<td></td>
</tr>
<tr>
<td>methods of analysis can</td>
<td>used to tackle the identified</td>
<td></td>
</tr>
<tr>
<td>help the experimenter in</td>
<td>problems.</td>
<td></td>
</tr>
<tr>
<td>continuous processes to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tackle some of the above</td>
<td>To generate hypotheses and</td>
<td></td>
</tr>
<tr>
<td>identified problems.</td>
<td>important directions for further</td>
<td></td>
</tr>
<tr>
<td>To generate hypotheses and</td>
<td>research.</td>
<td></td>
</tr>
<tr>
<td>important directions for</td>
<td></td>
<td></td>
</tr>
<tr>
<td>further research.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The research approach characterizes how the study will be performed. A distinction is often made between induction and deduction. An inductive approach tries to construct a general rule from a specific case, while deduction departs from a general rule to try to explain a specific case (Molander, 1988), see also Box et al. (2005, Chapter 1) for a discussion of induction and deduction. Using an abductive approach the analysis of empirical data can be preceded or combined with the study of literature (Alvesson and Sköldberg, 1994). This research started with the author having some pre-understanding of the DoE area from courses and teaching. Literature and empirical data mainly from the EBF have then been iteratively approached and successively been reinterpreted in the light of the other. Hence, the
research has been run much in accordance with the abductive approach, see Figure 2.2.

![Deduction, Induction, Abduction Diagram](image)

**Figure 2.2** Deduction, induction and abduction according to Alvesson and Sköldberg (1994, p. 45) and the approach used in this research. The Figure is inspired by Söderholm (2005).

The research strategy can be viewed as the framework for the collection and analysis of data, see Bryman (2001). This research has relied heavily on the EBF and the EBF engineers as the source to find potential problems when planning, conducting, and analyzing experiments in a continuous process. The EBF has also been the source of industrial data and constitute the case around which the proposed analysis methods have been developed and tested (see especially papers B, C, and D).

It is possible to view the research strategy as dominated by a single case study of an industrial case (the EBF). Yin (2003) is an excellent reference for case study research but has an explicit focus on social science research. This research studies the use of statistical methods in an industrial context but the work mainly relates to the development and testing of analysis methods inspired by the industrial context; a strategy that certainly does not fit the typical description of a case study. Nonetheless, I believe that certain aspects of the recommendations for research design and methods in Yin (2003) are worth considering and I refer to them when appropriate. For example, using a single industrial case was judged to be a good strategy since the EBF provided the opportunity to closely follow the experimental work in a continuous process specifically designed for experimental purposes. The EBF can therefore be argued to be a unique case, and hence be a reason for choosing a single case, see Yin (2003).

This research can also be considered to include elements of action research, as the author (and his supervisors) participated in, for example, the planning and the analysis of experiments at the EBF plant. Whether action research is a research strategy by itself or not is not obvious in the literature. However, I view action research as a method to perform a case study, see also Gummesson (2000). Coughlan and Coghlan (2002, p. 236) argue that “action research generates emergent theory,
in which the theory develops from a synthesis of that which emerges from the data and that which emerges from the use in practice of the body of theory which informed the intervention and research intention.” The action research process is undertaken in a spirit of collaboration and co-inquiry and aims to stimulate change in organizations, develop self-help competencies, and to add to scientific knowledge (Shani and Pasmore, 1985). I believe these descriptions of action research fits well to the collaborative nature within the EBF project.

2.4 Data collection and analysis

Within the research project, sources used during data collection include interviews, documentation, and direct and participant observations, which are all common sources of evidence in case studies, see Yin (2003). Since the EBF project was run in close collaboration between the author, his supervisors, and the EBF engineers, data were collected by, for example, observations, participation and discussions about the various choices and activities within the project. These discussions were recorded and documented in protocols from monthly project meetings and project reports. The meetings were used to discuss practical issues, to discuss and decide future research activities, and to present and discuss emerging results. These types of qualitative data were important to identify potential problems with using DoE in the EBF process. Furthermore, the experiments in the EBF, doing data analysis work using process data from the EBF, and simulations have been highly important sources of evidence. Process data from the EBF were gathered from process databases with the help of EBF engineers. The research has thus used a combination of quantitative and qualitative data, where the quantitative data have been used to develop analysis methods in the industrial context.

Analysis of the data generated through the research project has been made on two main levels. Firstly, on a more detailed level, many research activities demanded separate analyses; for example, the interviews, the different experiments and the analysis of process data and simulations in the appended papers. These analyses are described in the appended papers. It is worth pointing out that much of the work deals with the development of analysis methods through the analysis of process data from the EBF. Figure 2.3 provides a summary of the main data collection and analysis activities used in connection to each paper.

---

10 The protocols can be viewed, after consideration of possible secrecy issues, by contacting the author of this thesis.
Secondly, on a summarized level, the analysis of the empirical data and results from the separate research activities was made by comparing empirical evidence against the theory by reflection and drawing of rational conclusions. The analysis can hence be described as an iterative process where the empirical results in the research project either strengthened or overthrew our prevailing understanding about DoE in continuous processes. Discussions between the author and his supervisors as well as with EBF engineers at both informal and formal project meetings have been an important part of this analysis process.

### 2.5 Research quality

Important criteria for assessing the quality of research are validity, reliability and replication (Bryman, 2001). Somewhat simplified, these three concepts are explained in Figure 2.4.
Another important criterion for assessing the quality of research is the domain to which the results can be generalized, which is referred to as external validity by Yin (2003).

Yin (2003, p. 34) describes tactics useful within a case study research to secure research quality. I believe that some of these tactics are relevant to strengthen the quality of this research. A strengthening of the validity of the conclusions from the research based on the EBF case, is provided by the multiple sources of evidence that were used during data collection, such as experiments, interviews, data analysis of process data, and observations. The many different sources of evidence gathered using the EBF case makes data triangulation possible, see Yin (2003).

To further strengthen the validity, separate reports describing ongoing work and emerging results have been produced within the EBF project and continuous discussions were held with key informants at the EBF plant. Examples of such separate reports are protocols from project meetings, monthly reports, and internal feedback reports. The frequent discussions helped us to focus on problems that were considered important to the engineers doing experiments in a continuous process.

A general weakness of research that relies heavily on the study of a single industrial case, here in form of the EBF, is that the results may be hard to generalize, or in other words, the results may have poor external validity (Yin, 2003). Instead of statistical generalization (made possible by studying many different industrial contexts), this research has to rely on analytical generalization which tries to generalize the results to broader theory. The results are analytically compared to previous theory about DoE, continuous processes, and the methods of analysis. Analytical generalization is hence used to increase the external validity of the results.
In addition, the ongoing study of literature has been used to guide decisions concerning data collection activities during the research project.

A challenge in case study research is to produce reliable results, or more specifically, results that can be repeated (Yin, 2003). In the EBF project, most decisions and data collection activities were recorded in protocols from meetings and stored in a database. Documentation of activities and decisions within the EBF project has thus been used to increase the reliability. However, it is difficult to replicate the study because of the elements of action research, the collaborative nature of data collection within the study, and the uniqueness of the EBF setting. The specific data analyses, using, for example, process data from the EBF, described in the appended papers are however possible to replicate for an outsider, which strengthens the reliability of the results.
3. INTRODUCTION TO THE APPENDED PAPERS

The purpose of this chapter is to serve as an introduction to the appended papers. The chapter is organized to introduce the core elements of the thesis, summarize the results and to provide theoretical foundations for the appended papers.

3.1 Planning and performing experiments in continuous processes

An experiment can be separated into three phases: pre-experimental planning, performing the experiment, and post-experimental analysis. The experimenter needs to think about the whole experimental process when planning the experiment. That is, how is the experiment going to be conducted and analyzed? The analysis will also depend on the type of experimental design chosen. This section will however focus on the issues concerned with planning and performing experiments in continuous processes. Analysis of the experiments will be discussed more thoroughly later on.

Indeed, the planning of the experiment will determine the possible knowledge that can be created through its realization. Therefore I believe that the planning phase is the most important of the three phases. In fact, Hahn (1984) argues that formal statistical analysis of data may not be needed to draw conclusions if the experiment is well-designed and well-executed, but that statistical analyses can be used to fine-tune conclusions and produce quantitative estimates.

Guidelines for planning designed experiments can be found in the literature but they are often of a general nature giving thoughtful tips for any experimenter planning an experiment. No prior work with a focus on planning for experiments in continuous processes has been found. More general guidelines can often be found in one of the first chapters of typical DoE textbooks, see, for example, Phadke (1989), Schmidt and Launsby (1994), Dean and Voss (1999), Wu and Hamada (2000), and Montgomery (2009). There are some variations but the guidelines are usually something like those given in Montgomery (2009), see Table 3.1. Guidelines usually include recommended steps throughout all three phases of an experiment.

The most detailed checklist that was found is given by Coleman and Montgomery (1993) and can be viewed as an expansion of Table 3.1. They present a master guide sheet with twelve steps (see also Paper A) to structure the planning process of an experiment, including recommendations for each step with illustrations from a CNC machining process of jet engine impellers. Hahn (1977) also provides similar and important recommendations for the planning phase of an experiment.
Barton (1997) discusses experimental planning but focus on graphical tools to aid the planning process.

Table 3.1 Seven steps of designing an experiment. Source: Montgomery (2009, p. 14), who argues that steps 2 and 3 are often done simultaneously or in reverse order.

<table>
<thead>
<tr>
<th>Step</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Recognition of and statement of the problem</td>
</tr>
<tr>
<td>2</td>
<td>Selection of the response variable(s)</td>
</tr>
<tr>
<td>3</td>
<td>Choice of factors, levels, and ranges</td>
</tr>
<tr>
<td>4</td>
<td>Choice of experimental design</td>
</tr>
<tr>
<td>5</td>
<td>Performing the experiment</td>
</tr>
<tr>
<td>6</td>
<td>Statistical analysis of the data</td>
</tr>
<tr>
<td>7</td>
<td>Conclusions and recommendations</td>
</tr>
</tbody>
</table>

Paper A of this thesis builds on the recommendations given foremost by Coleman and Montgomery (1993), but sets out to identify those special issues that the experimenter needs to consider in a continuous process. The results in Paper A are summarized in a thirteen-step checklist and the special issues are discussed using the EBF process as the studied case. The planning steps in the presented checklist in Paper A are very similar to those given by Coleman and Montgomery (1993) with one additional step, namely to plan for the process control strategy. Besides this, the main contribution of Paper A is to discuss those special considerations that are needed in connection to each step. Paper A discusses a number of special considerations that are not normally highlighted in the more general recommendations for the planning phase. The most important issues are summarized here.

The need for process control during ongoing experiments is identified as an important experimental complication. It is vital that control strategies are developed during the planning phase, especially if the control includes human deliberations (as is the case at the EBF plant).

Moreover the dynamic characteristic of the EBF process (and many other continuous processes) leads to costly transition times between changes of experimental treatments, see Figure 3.1. In conjunction with the use of factorial
designs the accumulated transition times can become too costly. Hence, randomization restrictions are often needed for experiments in continuous processes.

**Figure 3.1** An illustration of the need for a transition time between experimental runs in an experiment in a continuous process.

... Many responses (often highly autocorrelated) are needed to capture the effect of the experimental treatments. A multivariate response situation needs to be considered already during the planning phase as it undoubtedly affects the analysis of the experiment. The multivariate situation also makes it more complicated to follow the recommendations in literature of detailed planning of, for example, anticipated effects and to foresee and document suspected interactions.

Furthermore, experiments in continuous processes typically mean large-scale and long-term experimentation (around the clock). Coordination, information, and control issues become even more important (and complicated) in such cases. The scale, complexity, and the many involved people in the experiments make it hard to perform pilot tests of, for example, factor levels and the experimenter must also plan for breakdown of important process equipment. The continuous nature of the process makes every incident more severe since it can come to affect a long period of process operation and results in large costs.

### 3.1.1 Experimental design and continuous processes

Industrial experimentation is expensive not least in full-scale continuous processes. Therefore, factorial designs, especially two-level factorials, are often interesting designs that produce information at a relatively low cost. Montgomery (2009) considers two-level factorial designs to be the cornerstone of industrial experimentation. Two-level factorials also form the basis for fractional factorial designs which are valuable for screening experiments (Box *et al.*, 2005). Fractional
factorials are arranged so that less likely interactions are aliased (varied in the same pattern) with factors or interactions considered more likely to be active. The resolution of a fractional factorial design provides important information about the alias structure of the design. In a resolution III (three) design, for example, main effects are aliased with two-factor interactions. See, for example, Myers and Montgomery (2002) for more on design resolution. Czistrom (1999) describes the advantages of factorial designs (compared to one-factor-at-a-time experiments) for testing two or more experimental factors:

- they require less resources for the amount of information that is obtained,
- the effect estimate for each factor (or interaction) is more precise given the same number of observations,
- the interaction effects between two or more factors are systematically estimated, and
- the experiment produces information in a larger region of the factor space.

Figure 3.2 gives an example of short notation (used in Paper F) and explanations for a two-level 1/8 fractional factorial design testing six factors in eight runs with resolution III.

![Figure 3.2 An example of short notation of a two-level factorial design.](image)

The appended papers to this thesis have an explicit (Paper B, E, and F) as well as an implicit (Paper D) focus on the use and analysis of factorial designs in continuous processes.

Randomization is one of the core principles of well-designed experiments and should be used whenever possible, see, for example, Young (1996) and Bjerke (2002). That is, both the allocation of experimental material and the order of the runs in the experiment should be determined randomly. Randomization is used to avoid bias or systematic error to affect the conclusion of the experiment (Cox and Reid, 2000). Randomization should be considered of special importance when experimenting in processes that are non-stationary in nature, see Bisgaard et al. (2008).
Although randomization is desirable, industrial experimentation often produces situations where complete randomization of the run order might not be feasible because of cost or time concerns. When the randomization of the run order is restricted, the experiment is said to have a split-plot structure. Split-plot designs have been discussed at length in the literature and are considered important for industrial experimentation, see, for example, Kowalski et al. (2007), Naes et al. (2007), Federer and King (2007), Tyssedal and Kulahci (2005), Sanders and Coleman (2003), Bisgaard (2000), Box and Jones (1992), and Wooding (1973).

As an illustration, suppose that a $2^3$ factorial experiment will be performed with factors A, B, and C. Factor A is, for some reason, difficult (costly) to change while factors B and C are easy (inexpensive) to change. In a split-plot structure, we can choose to fix the level of the hard-to-change factor (A) and then run all or a fraction of all combinations of the other factors before changing the level of A. In the current example the hard-to-change factor (A) is labelled the ‘whole-plot factor’ and the easy to change factors (B and C) ‘sub-plot factors.’

In a split-plot experiment, there are normally two levels of randomization. The whole-plot treatments are randomly assigned to the whole-plots and then the sub-plot treatments are randomly assigned to sub-plots using a separate randomization pattern for each whole-plot. The implication on the experimental analysis of split-plot experiments is that different types of errors (variance components) must be considered during analysis. Whole-plot factors are associated with a larger error (whole-plot error) than the sub-plot factors and the interactions between whole-plot and sub-plot factors (sub-plot error), see Bisgaard and de Pinho (2004). Consequently, the whole-plot factors are estimated with a lower precision while the opposite is true for the sub-plot factors and the interactions between whole- and sub-plot factors, see Goos et al. (2006) and Box and Jones (1992).

Bisgaard (2000) and Bingham and Sitter (1999; 2001) discuss the extension of split-plot experiments to fractional factorial experiments and Vining and Kowalski (2008), Vining et al. (2005) and Goos et al. (2006) deal with response surface designs run in a split-plot structure.

Indeed, split-plot designs are of special interest in continuous processes. Paper A outlines the need to consider split-plot designs when planning experiments in continuous processes, since propagations of, for instance, effects of experimental factors can take time, which can lead to requirements of prolonged experimental

\[\text{Numenclature is due to the early agricultural heritage of the DoE field where 'plots' of land received different experimental treatments.}\]
runs. The randomization order in continuous processes may therefore need to be restricted to avoid many and costly transition periods between experimental runs, producing a split-plot design. Experimental factors afflicted with longer transition times can be suitable whole-plot factors while experimental factors with shorter transition times are candidates for sub-plot factors.

Paper B describes an experimental design somewhere between a split-plot design and a completely randomized design performed in the EBF and illustrates the need to restrict the randomization order when performing a factorial design in a continuous process. The analysis of the experiment in Paper B also illustrates the difference in the results between a split-plot analysis and analysis according to a completely randomized design.

Paper B further exemplifies how the split-plot structure of the experiment can be used to an advantage to increase the precision in estimating effects of special importance. Whole-plot factors can be those factors where experimenter can accept estimates with lower precision in order to achieve higher precision for other main factors and interactions (sub-plots), see also Box and Jones (1992). Paper B also shows how the restriction and direct control of the run order may be used instead of a trial run or pilot experiment to divert uncertainty about the choice of factor levels.

Furthermore, the concept of an ‘adaptive design’ is used in Paper B as a response to practical considerations of running the experiments in the EBF. That is, for the particular experiment the planned time for each run was 24 hours of process operation but was held open for extension if a disturbance occurred, for example, process equipment failures. These types of disturbances were not uncommon in the EBF process. The use of the adaptive design is considered a sound strategy in a continuous process. It often may make less sense to interrupt the specific run and move on to the next run before collecting enough data from ‘normal’ operation at the specific setting than to wait for the disturbance to pass and continue to collect data from stable operation in the specific run. Due to the continuous nature of the process, many times it may take less time to let the disturbance pass and remove the disturbed period from the data than to move on to the next run with the associated transition time.

Finally, an experimenter that considers running an experiment in a continuous process plant (especially for full-scale operation) should consider the EVOP procedure. Even though Box (1957) viewed EVOP as a way of running the plant for continuous improvement rather than a one-shot experiment, using an approach inspired by EVOP (repeated small factor changes using a small factorial design over a
long time period) may be the only viable experimental design for full-scale production plants.

3.1.2 The need for process control and monitoring during experimentation

As highlighted in Papers A and B and specifically discussed in Paper C, process control during the experiment may be unavoidable. In the EBF case the thermal state of the process needs to be monitored and controlled for personal and plant safety reasons. In many plants autonomous and automatic control systems are constantly working to create process stability even during the experiments. Process control may also be non-automated (manual) and performed by operators. An adaption of Figure 1.1 may therefore be appropriate to give a more realistic representation of an experiment in many industrial processes, see Figure 3.3.

![Diagram of Process (or system)](image)

Figure 3.3 An adaptation of Figure 1.1 to show a general model of an industrial process under experimentation. Here \( X \)s label experimental factors varied according to a pre-determined experimental design or held-constant factors. The \( C \)s label control variables that are varied to maintain process control during the experiment. The \( Z \)s are the uncontrollable (noise) factors.

As pointed out by Hild et al. (2000) control actions can lead to that process responses are not directly visible as changes in typical responses but instead as changes in control variables. In relation to Figure 3.3 this means that the response due to an experimental treatment may be displaced from typical responses \( (Y) \) to control variables \( (C) \). Analyzing data from a process subjected to feedback control (automated or manual) is often referred to analysis under closed-loop operation, see, for example, Box and MacGregor (1974). An implication is that sometimes these control variables must be used as responses. A prerequisite to do so is, however, an unbiased control. When people are involved in the control actions, as is the case at the EBF, a subjective dimension is added to control decisions which further
complicates the matter. Particularly, it becomes hard to secure that the same control actions are made by different people given a certain process situation.

Another complicating issue is the need to analyze many responses jointly to determine the current process state and judging the need for control actions. This is a consequence of the multivariate nature of continuous processes. Furthermore, the control of a continuous process is complicated due to its dynamic characteristic. The experimenter should anticipate some time-lag for the process control actions to reach full effect, just as for responses to experimental factor changes. In addition, there may be a time-delay in the measurement of responses used as information about the process state. Paper C deals with the situation at the EBF plant where process control actions are made by operators based on information about the process state given by certain responses from the process. A starting point for unbiased control actions is process monitoring that signals when something out of the ordinary is occurring in the process. The need for process monitoring relates this research, Paper C in particular, to the area of statistical process control (or statistical quality control). It is a well-developed area and an introduction can be found in, for example, Montgomery (2005). Distinctive tools that form the basis for statistical process control are control charts such as the Shewhart (Shewhart, 1931), cumulative sum, CUSUM, (Page, 1954), and the exponentially weighted moving average, EWMA (Roberts, 1959). Briefly the control charts are used to monitor processes and designed to signal when a process shift has occurred or when the variability in the process is unusually large or small.

Multivariate monitoring and control have received increasing attention not least due to the development of computers and software the last decades. The focus on multivariate monitoring and control is especially apparent within the area of chemometrics and many examples of multivariate process monitoring and control come from process industries and continuous processes. Wise and Gallagher (1996) provide just one of many descriptions of how process industries often are richly instrumented with sensors routinely collecting measurements on many process variables, such as temperatures, pressures and physical properties. Hence, process industries often need to monitor a multitude of variables and they face a multivariate monitoring situation.

---

The typical terminology is to say that a process that operates with only chance causes of variation present is in statistical control but when assignable causes are present the process is said to be out of control, see Montgomery (2005).
INTRODUCTION TO THE APPENDED PAPERS

Overviews of the development of monitoring and control using multivariate statistical methods the last decades are given by, for example, Bersimis et al. (2007), Kourtı (2005), and Qin (2003). Multivariate extensions of the Shewhart, CUSUM, and EWMA charts are examples of methods. Multivariate projection methods such as PCA and PLS are often described as important methods used in multivariate monitoring applications.

Paper C illustrates the working process and development of principal component models to monitor the thermal state of the EBF process. Further discussion of the use of principal components for process monitoring can be found in, for example, Bisgaard and Kulahci (2006a), Kourtı (2005), Mastrangelo et al. (1996), Kresta (1991), and Wise et al. (1990). PCA was first introduced by Karl Pearson (1901) and today the aim of PCA is often data reduction and interpretation, but can also be aimed to detect outliers among other things (Wold et al., 1987). PCA is used to explain the variance-covariance structure of a number of variables by constructing linear combinations of these original variables to form principal components. Further descriptions of PCA can be found in, for example, Johnson and Wichern (2002) and in Paper C.

For process monitoring purposes Paper C outlines the use of the Hotelling’s $T^2$ chart, two-dimensional score plots using a Hotelling $T^2$ control ellipse, and a measure of the residuals from the PCA model called DModX (distance to the model). Contribution plots of the original variables’ contribution to the measures above are used to diagnose deviating observations. Using this approach the monitoring problem was reduced to following only a few principal components instead of many process variables from the EBF. However, a problem that is stressed in Paper C is the difficulty to choose a good reference data set to define normal process operating conditions during the building of the reference PCA model. The choice of the reference data set is made more difficult due to the frequently shifting operating modes of the EBF, caused by different experimental setups and the dismantling of the blast furnace between the experimental campaigns. Possible ways to adapt the PCA models to process drifts and shifting operating modes are proposed in the paper in form of online adaptation of model averages and manual calibration. The adaptation problem has been discussed in the literature, and similar and more sophisticated methods to adapt multivariate models to process drifts and shifts have been suggested by, for example, Lane et al. (2003), Li et al. (2000), and Wold (1994).
In relation to the experimental work in the EBF plant, the multivariate approach to monitoring described in Paper C provides a number of potential benefits:

- a quick overview of the thermal state in the EBF process, and a summary of the information in many original process variables which is important to make correct and timely control decisions,
- a way to standardize how the thermal state is assessed, which is one step towards an unbiased decision of when to perform control actions even though human deliberations may still be needed to determine what the appropriate action is, and
- formal decision criteria to determine when the process is operating normally, which could be used to decide on when a new experiment is to be started or to select data observations to be included in the analysis of the experiment.

3.2 Analysis of experiments in continuous processes

The statistical analysis of the experiments tests the hypotheses of treatment effects on the response of interest. For experiments with replications of experimental runs, analysis of variance (ANOVA) is typically used for analysis, see Griffith et al. (1989). The effects of two-level factorials can also be compared with their standard errors, see Box et al. (2005). The ANOVA partitions the total sum of squares (calculated from all observations in the experiment) into sums of squares for the main effects, interaction effects, and the sum of squares due to error, see Montgomery (2009). The mean squares, obtained by dividing the sum of squares by its degrees of freedom, of the main effects and interactions are compared with the mean square for errors through ratios. Under the assumption that the error terms are normally and independently distributed with constant variance, the ratios of mean squares follow the F-distribution. When designs are replicated, F-tests can be used to draw conclusions about the significance of effects. More details on ANOVA in connection with analysis of factorial designs can be found in, for example, Montgomery (2009, pp. 166-173).

Multivariate analysis of variance (MANOVA) can be described as the multivariate extension of ANOVA. MANOVA is used for making comparisons among mean vectors arranged according to treatment levels (Johnson and Wichern, 2002). The comparison can therefore be made for several responses simultaneously instead of making separate ANOVAs for each response. If there are many responses to analyze, it is recommended to start with a MANOVA to see if there are any
significant differences within the group of responses. By doing this the overall significance level is known and can be kept at the desired level. Individual ANOVAs can then be used to test the main and interaction effects on individual responses. Similar to ANOVA, MANOVA partitions the total variation into components attributable to the main and interaction effects and to the error. Unreplicated experiments have no internal estimate of the experimental error and therefore require other types of analysis. Analysis of unreplicated factorial designs will be discussed more thoroughly later in connection to Paper F.

3.2.1 Multivariate statistical analysis of experiments

Multivariate statistical analysis refers to all statistical methods that simultaneously analyze multiple measurements on an object (Hair et al., 1998). Typically, an analysis of more than two variables simultaneously can be considered multivariate. Today, it is common for industries to have to deal with large amounts of measurement data, such as temperatures, pressures and physical properties, which often are logged on-line (Yang, 2004). This is a common situation for continuous processes which is discussed above and in the appended papers.

In a situation with many crosscorrelated variables to analyze, a one-variable-at-a-time approach to analysis is often ineffective, inefficient, and can contribute to the drawing of wrong conclusions. MacGregor (1997) argues that interpreting results from a univariate approach to analysis under the presence of correlation among responses is analogous to the inferior ‘one-factor-at-a-time approach’ to experiments under the presence of interactions; Daniel (1959) also makes this point.

The field of multivariate analysis incorporates many techniques which are not all elaborated here. A good reference is Johnson and Wichern (2002), who explain many of the commonly used multivariate techniques. The research in this thesis has made frequent use of latent variable techniques, and PCA in particular. The two latent variable techniques PCA and PLS reduces the dimensionality of the data by projecting the information in the data into low-dimensional spaces defined by a small number of latent variables (Kourti et al., 1996). PCA is discussed and described in more depth in papers B, C, and D and is therefore not elaborated here. Jackson (1980; 1981; 2003) also provide a good introduction to PCA. Eriksson et al. (2006) describe PLS as a regression extension of PCA with the aim to connect two data matrices (X and Y) to each other. While PCA can be described as a maximum variance least square projection of one of the matrices X or Y, PLS is a maximum covariance model of the relationship between X and Y (Eriksson et al., 2006). The
mathematics behind the PLS technique can be found in, for example, Eriksson et al. (2006) or Höskuldsson (1988).

Paper B illustrates the important connection between designed experiments (using factorial designs) and multivariate statistical analysis and shows how the multivariate characteristic of continuous processes affects the analysis of the experiments. Paper B proposes an analysis method for experiments in the EBF process that combines PCA, MANOVA, and ANOVA. PCA is used to derive latent, uncorrelated variables (principal components) that summarize the ‘strongest signals’ in the response data. The principal components are then used as new responses to test for statistical significance of main and interaction effects. The many responses from each run in the EBF process are time series with observations available each second or each hour. Consequently, the principal components also become time series from each run. Since the original responses are highly autocorrelated, so are the principal components. Therefore the approach used in Paper B is to calculate averages of the values of the principal components from each run before applying MANOVA and ANOVA to the averages. The assumption of independent and normally distributed observations for MANOVA and ANOVA is reasonably achieved by calculating averages for each run. The proposed analysis method is summarized in Figure 3.4. The analysis in Paper B also compares the result between the assumption of a completely randomized design and a split-plot design since the design used in Paper B lies somewhere between the two extremes.

Figure 3.4 A summary of the analysis method outlined in Paper B.

Two examples of similar analysis approaches as the one proposed in Paper B have been found in the DoE literature. Ellekjer et al. (1997) use PCA and normal probability plots of the estimated effects based on principal components to analyze split-plot experiments. Bjerke et al. (2008) use PCA as a first step when analyzing an experiment with restricted randomization in the food industry.
3.2.2 Time series analysis and experiments in continuous processes

As mentioned above, responses from continuous processes are often available as time series, and thus the analysis of experiments in continuous processes relates to the field of time series analysis. A time series can be defined as “a time-oriented or chronological sequence of observations on a variable of interest” (Montgomery et al., 2008, p. 2). Often adjacent observations in the time series are dependent. Box et al. (2008) describe time series analysis as concerned with the techniques to develop stochastic and dynamic models to analyze this dependence. Time series analysis covers many advanced methods and textbooks with broad coverage are, for example, Box et al. (2008), Montgomery et al. (2008), and Wei (2006), where Box et al. delivers the most advanced and comprehensive discussion.

This research applies two specific areas within time series analysis: [1] autoregressive integrated moving average (ARIMA) models, and [2] transfer function-noise models and intervention analysis. ARIMA models can be described as autoregressive moving average (ARMA) models extended to be able to describe a nonstationary time series, \( Z_t \). Often, \( Z_t \) can be made stationary by differencing. That is, the differenced time series \( (1-B)^d Z_t \) follows the stationary ARMA(\( p,q \)) model. The general ARIMA(\( p,d,q \)) model can be represented by:

\[
\Phi(B)(1-B)^d Z_t = \theta_o + \Theta(B)a_t
\]

where \( B \) is the backshift operator on \( t \), \( \Phi(B) = (1-\phi_1 B - \ldots - \phi_p B^p) \) is the autoregressive (AR) operator, \( p \) is the order of the AR operator, \( \Theta(B) = (1-\theta_1 B - \ldots - \theta_q B^q) \) is the moving average (MA) operator, \( q \) is the order of the MA operator, \( d \) is the order of differencing applied, and \( \epsilon_t \) is Gaussian white noise. The parameter \( \theta_o \) is related to the mean of the process when \( d = 0 \), but is called the deterministic time trend when \( d \geq 1 \). A thorough discussion of ARIMA models is given by, for example, Box et al. (2008).

A discrete transfer function model uses pairs of observations \( (X_t, Y_t) \), available at equispaced intervals in time, from time series of an input \( X_t \) and an output \( Y_t \) and creates a dynamic model of the relation between the input and the output, see Figure 3.5. Sometimes both \( X \) and \( Y \) are essentially continuous but observed only at discrete times. Even under controlled situations, \( Y \) is affected by other influences than \( X \). The combined effect on \( Y \) of these additional influences is referred to as the noise, see Box et al. (2008). A model that can describe real data should therefore consist of a transfer function to describe the deterministic dynamic relation between \( X \) and \( Y \) as well as a stochastic noise model. The noise series, \( N_t \), left after the
transfer function has been determined is typically represented by an ARIMA model thus creating a transfer function-noise model.

![Diagram of Dynamic System](image)

**Figure 3.5** The dynamic relation between an input to, and output from a system can be represented by a transfer function. The figure is adapted from Box et al. (2008, p. 440).

Intervention analysis (Box and Tiao, 1976) can be viewed as transfer function-noise modeling using qualitative step or pulse variables as inputs (instead of continuous X variables) to indicate the presence or absence of an event of some kind. Transfer function-noise models and intervention analysis are further discussed in Papers D and E, see also Box et al. (2008), Montgomery et al. (2008), Wei (2006), and Jenkins (1979).

Paper D returns to the observation discussed above and in connection to papers A and B, namely that continuous processes are dynamic systems with inertia. Process dynamics must be considered already during experimental planning since it affects the required length of the experimental runs in the process. Bisgaard and Kulahci (2007) briefly discuss the problem of studying what they call ‘regime changes’ in industrial processes and point to the use of transfer functions and intervention analysis to study the transition periods. Paper D outlines a method to analyze process dynamics and formally estimate the transition time required between runs in a dynamic process with many responses. The proposed method combines the time series analysis techniques transfer function-noise models and intervention analysis with PCA. Similar to the approach in Paper B, PCA is performed in order to summarize the variation in the responses. The principal components are then used in conjunction with transfer function-noise models (quantitative X) or intervention analysis (qualitative X) to model the propagation of a change of experimental factors in the EBF process. The steps used for model building proposed in Paper D (except the use of principal components) builds on the recommendations in Montgomery et al. (2008), and Bisgaard and Kulahci (2006b; 2006c).

Paper D once again highlights the multivariate nature of continuous processes and connects the use of multivariate statistical methods to time series analysis.
techniques. The knowledge of the transition times is an important input to the choice of experimental design. In the presence of different transition times for the experimental factors, the transition time may be an input to the choice of, for example, whole- and sub-plot factors in a split-plot design.

Paper D focuses on estimating the transition times between two runs when changing only one experimental factor. In Paper E, the idea of using transfer function-noise models to analyze time series of responses is extended to the analysis of a complete two-level factorial design with time series responses. As discussed above, one of the key features of experiments in continuous processes is that the responses often must be viewed as time series.

The analysis of experiments with time series responses seems to be scarcely discussed in the DoE literature. The only article found that explicitly focuses on the analysis of experiment with time series responses is Hau et al. (1996) who used regression analysis with the response series in each run as the dependent variable and time as the independent variable. The two estimated regression parameters for each run (the overall mean and trend) were then analyzed as ordinary single-response experiments.

Paper E focuses on the analysis of two-level factorials and essentially suggests, outlines, and compares three analysis methods for the analysis of location effects in two-level factorials with time series responses:

1. Calculating the average of the response in each run of the experimental design and use the averages as the single response in, for example, an ANOVA.
2. Fitting an appropriate ARIMA model to each run and, in the case that the time series can be described by an ARMA model, use the estimated mean for each run as the new single response.
3. Letting binary indicator variables represent the inputs to the dynamic system related to main and interaction effects and fit an intervention-noise model to the time series from the entire experiment. The effects are estimated through the estimated parameters of the significant transfer functions.

Adjusted versions of methods 1 and 2 above are also tested by removing observations during the estimated transition time in each run.

The analysis methods are compared by simulating two-level factorial experiments using a simplified model of a dynamic continuous process. In the simulations, the process is assumed to follow an ARMA(1,1) model during normal
operation. Dynamic effects are then allowed to affect the resulting time series during the simulated experiment, for example:

\[
\gamma_t = \delta + \tau_t^{(A)} + \tau_t^{(B)} + \ldots + \tau_t^{(AB)} + \ldots + \tau_t^{(ABC)} + \phi_t \gamma_{t-1} + \epsilon_t - \Theta_{t-1}
\]

(3.2)

where, \( \tau_t^{(A)}, \tau_t^{(B)}, \ldots, \tau_t^{(AB)}, \ldots, \tau_t^{(ABC)} \) are the contribution to the mean of the time series at time \( t \) due to possible effects of the main factors and interactions. The dynamic change pattern of the effects, \( \tau_t^{(A)}, \tau_t^{(B)}, \ldots, \tau_t^{(AB)}, \ldots, \tau_t^{(ABC)} \), can be gradual and the effects are parameterized so that the change pattern of the effects follow a step response of a first-order dynamic system, see Box et al. (2008) and Vu and Esfandiari (1997).

Paper E provides the results from comparisons of the analysis methods for the cases with only one active effect of different size in a replicated 2^3 experiment, three active effects of different sizes in a replicated 2^3 experiment, and with three active effects in an unreplicated 2^3 experiment. In each case, ten simulated experiments were analyzed. The results indicate that the effect estimates for the studied analysis methods are fairly similar irrespectively of the method used. Using averages of the response in each run adjusted for the transition time proves to be a fairly robust and straightforward analysis method, while intervention-noise models are more comprehensive, gives fewer spurious effects when the effects are small compared to the noise, and a larger number of the active effects are found when replication is limited. Furthermore, intervention-noise models provide the possibility to model effect dynamics.

### 3.2.3 Analysis of unreplicated two-level factorials

Even if two-level factorials and fractional factorials are used, the experimental design can become too large and costly if it is replicated. Therefore, unreplicated two-level factorials are often used to generate information at a low cost. However, unreplicated designs provide no independent estimate of the experimental error.

As described in Paper F, the analysis of unreplicated two-level factorials is traditionally made by studying a normal probability plot or half-normal probability plot to determine which of the effects that seem to divert from the reference distribution of inert\(^1\) effects, see, for example, Daniel (1959). Other more formal methods of analysis have been proposed in the literature. For example, it is possible to select a number of effects or contrasts, prior to the experiment, that are unlikely to be active.

\(^1\)Frequently used nomenclature calls effects that are significantly larger than most of the other effects 'active' effects, while those effects that seem to be measuring only random noise are called 'inert.'
to be active and use them to estimate the experimental error, see Finney (1945). Another alternative is to sort the contrasts based on their absolute sizes and use some fraction of the smallest effects to estimate the distribution of inert effects, see, for example, Lenth (1989). Much work has been focused on developing objective methods of analysis for unreplicated experiments, for example, Daniel (1959), Zahn (1975), Box and Meyer (1986), Voss (1988), Benski (1989), Lenth (1989), Berk and Picard (1991), Le and Zamar (1992), Box and Meyer (1993), Dong (1993), and Venter and Steel (1996). Hamada and Balakrishnan et al. (1998) provide a comprehensive review and comparison of these and other methods. The research area of the analysis of unreplicated factorials remains an active one, and further contributions are given by, for example, Sandvik-Wiklund and Bergman (1999), Chen and Kunert (2004), and Costa and Pereira (2007). The results in foremost Hamada and Balakrishnan et al. (1998) and Chen and Kunert (2004) show that there is no clear ‘winner’ among the methods. Some methods are good when there are only a few active effects but perform worse when there are many active effects.

Most analysis methods for unreplicated factorials rest on the implicit hypotheses called the ‘effects sparsity principle’, which states that, in general, only a few of the effects in a factorial experiment will be active, see Box and Meyer (1986). Two other important hypotheses for the analysis are the ‘effects hierarchy principle’, which states that lower order effects are more likely to be important than higher order effects (Wu and Hamada, 2000), and the ‘effects heredity principle’, which implies that an interaction is more likely to be active if at least one of its parent factors are active. These hypotheses are often referred to but have seldom been validated in the literature. Paper F sets out to investigate the viability of the three hypotheses important for analysis of unreplicated experiments by studying experiments found in the literature. All three principles are found to be viable. The results presented in Paper F largely agrees with those presented by Li et al. (2006).

In Paper F it is argued that during analysis of unreplicated experiments, prior knowledge (such as knowledge about the three principles) could make a contribution during analysis and increase power. Paper F focuses on incorporating the prior knowledge about the three principles in the Bayesian\textsuperscript{14} approach to analyze unreplicated two-level factorials presented by Box and Meyer (1986), which in its original form only considers effects sparsity. In the adapted Box and Meyer method

\textsuperscript{14} In Bayesian inference, the unknown parameters are regarded as stochastic variables with a prior distribution. From the observations the posterior distributions of the parameters are calculated using Bayes’ rule. In ‘classical’ (or frequentist) inference, the unknown parameters are regarded as deterministic.
proposed in Paper F, we let $T = (T_1, \ldots, T_n)$ be a vector of $n$ estimated effects. In accordance with Box and Meyer we assume that that an “active” effect is distributed $N(0, k^2 \sigma^2)$, with the implicit assumption that $k > 1$, while an “inert” effect is distributed $N(0, \sigma^2)$. That is, $\sigma$ is the standard deviation of an inert effect and $k$ is the “inflation factor” for the standard deviation of an active effect. To allow for individual prior probabilities for each effect we let $\alpha_i = (\alpha_1, \ldots, \alpha_n)$ be a vector of prior probabilities that the effects are active (Box and Meyer use a scalar $\alpha = 0.2$).

Under the assumption that the effects $T_i, i = 1, 2, \ldots, n$, are independent and identically distributed from the Gaussian mixture $(1 - \alpha_i)N(0, \sigma^2) + \alpha_iN(0, k^2 \sigma^2)$, the posterior probability that effect $i$ is active, given $T_i$ and $\sigma$, is:

$$P(i \text{ active} | T_i, \sigma) = \frac{\alpha_i \exp\left(\frac{-T_i^2}{2k^2\sigma^2}\right)}{\frac{\alpha_i}{k} \exp\left(\frac{-T_i^2}{2k^2\sigma^2}\right) + (1 - \alpha_i)\exp\left(\frac{-T_i^2}{2\sigma^2}\right)}$$

(3.3)

The conditioning on $\sigma$ in (3.3) needs to be removed and Paper F shows how this can be achieved by integrating (3.3) over the posterior distribution of $\sigma$, $p(\sigma | T)$. Box and Meyer (1986) propose numerical integration to calculate the posterior probabilities. Stephenson et al. (1989) show how the posterior probabilities can be calculated analytically for up to 15 effects as well as by numerical integration. The solution in Paper F involves the application of Bayes’ rule, see Gelman et al. (2004):

$$p(\sigma | T) = \frac{p(\sigma)p(T | \sigma)}{p(T)} \propto p(T | \sigma)p(\sigma)$$

(3.4)

and numerical integration using a Markov chain Monte Carlo approach. In Paper F, the Metropolis algorithm is used to perform the numerical integration required to calculate the posterior probabilities. The Metropolis algorithm is a special case of the Metropolis-Hastings algorithm discussed by, for example, Chib and Greenberg (1995) and Gelman et al. (2004).

Paper F outlines a three-step method that successively considers the sparsity, hierarchy, and heredity principles and calculates the posterior probabilities that effects are active. The principles are incorporated in the adapted Box and Meyer (1986) method by adjusting the prior probabilities, $\alpha_i$ (see Eq. 3.3), for the effects. The method in Paper F extends the Box and Meyer (1986) approach by also considering effects hierarchy and heredity. These principles have been incorporated
in other Bayesian algorithms for variables selection for more general regression models, see Chipman (1996), and Chipman et al. (1997).

In addition, the Bayesian method in Paper F also allows for the consideration of process knowledge when specifying the prior probabilities for the effects. However, process knowledge and statistical analysis skills do not always reside in the same person and therefore Box and Liu (1999) stress the collaboration between statisticians and experimenters during design and analysis. Knowledge about both statistics and the process itself is important to successfully design, conduct and analyze an experiment.
CONCLUSIONS AND DISCUSSION

4. CONCLUSIONS AND DISCUSSION

This chapter presents the conclusions from the research and provides recommendations for the experimenter in a continuous process. The research method and the contribution of the research are also discussed. Lastly some specific implications for the experimental work at the EBF are discussed.

4.1 Conclusions and recommendations

To conduct experiments in a continuous process (even if the process is in pilot scale) means large-scale experimentation around the clock for an extended time period. There are usually many people involved and the experimental environment is complex. As the experimenter moves away from the ‘laboratory experiment’, or from making an experiment limited to a specific machine or process section, and plans an experiment that involves a larger plant, activities like coordination of people, information and communication as well as logistics planning and handling become important for experimental success. The appended papers, Paper A in particular, provide an overview of special considerations needed when planning, performing and analyzing experiments in continuous processes. The papers also propose methods of analysis to deal with some of the complications that emerge when experimenting in a continuous process and analyzing the results. This section outlines the conclusions and provides recommendations to experimenters in continuous processes.

A decisive step of the experimental planning process is to choose a suitable design for the experiment. In this step one of the main complications in continuous processes presents itself, namely their dynamic characteristics. The dynamic nature of continuous processes results in transition periods between the runs of the experiments for the effects to reach full impact in the process. Hence, process dynamics affects the minimum time required for each run in the experiment. To estimate this transition time between runs becomes important during the planning phase but also during the analysis. The transition time can be estimated using prior knowledge of the process or by visually studying the behavior of time series of responses from the process, see Paper A. A more formal method to model process dynamics and estimate the transition time using the time series techniques transfer function-noise models and intervention analysis is presented in Paper D.

Completely randomized experiments result in many changes of experimental factors, which help to avoid possible time trends and disturbances from distorting the results too much. However, process dynamics in continuous processes in
combination with many transition periods produced by randomized experiments may result in too costly experiments. The cost is magnified since conducting experiments in a continuous process usually means experimenting in a full-scale process or in an environment closely resembling full scale production. Experiments with restricted randomization, such as split-plot designs, may therefore be considered to be of special importance for experiments in continuous processes, see also papers A and B. By using split-plot designs, the number of required factor level changes can be reduced at the price of lower precision when judging the significance of whole-plot effects. An experimental factor that requires a long transition time for process stabilization when changed can therefore be a natural choice for a whole-plot factor. However, since many continuous processes are non-stationary by nature and randomization is important to reduce the possible bias from an unstable process, the experimenter needs to weigh cost concerns versus the concerns of the validity of the results. Split-plot experiments require a somewhat more complex analysis, which implies that experimenters in continuous processes might benefit from being familiar with the analysis of split-plot designs.

As pointed out in Papers A, B, and C, process control during the experiment may be unavoidable since, for example, the thermal state of the process needs to be controlled for personal and plant safety reasons. The need for process control during ongoing experiments is not unique for continuous processes but the continuous nature of the process makes the control actions more critical since they may affect a long period of process operation and thus the experiment. Continuous processes often operate in closed-loop during the experiments. This is a complicating issue during analysis of data from the experiment. Process control practice and strategies should, therefore, be considered already during the planning phase of the experiment. It is concluded that developing a process control strategy during the planning phase and then following it, when conducting the experiment, is an important success factor when performing experiments in continuous processes. Process control actions and automatic control loops may displace the variation in responses due to the experimental factors to variation in control variables. Control variables can therefore be important responses for experiments in continuous processes.

Controlling a continuous process is a complicated matter. Due to the dynamic characteristic of a continuous process, the experimenter should anticipate some time-lag also for process control actions to reach full effect in the process. In addition, there is usually a time-delay in some of the responses used to decide
control actions. Furthermore, there are often many variables that have to be considered jointly to make the control decision. If the control includes human deliberations at specific processing states, as in the EBF case, a subjective dimension is added to control decisions which further complicates the matter. Process monitoring is needed to make correct and timely control decisions. Paper C discusses the development of multivariate monitoring of the EBF process using principal components to reduce the number of variables to monitor, improve monitoring results, decrease subjectivity in control actions, and in the end improve experimental validity. However, an exhaustive investigation of the importance of the effects of process control on the results and suitable strategies to tackle the need for control actions during experiments has not been conducted during this research.

An experimenter that performs lengthy experiments in a complex process, closely resembling a full scale production plant, should be prepared for the possibility of disturbances (sometimes of a critical nature) during the experiment. In continuous processes, disturbances in individual runs can, due to the dynamic nature of the process, come to affect a long period of operation. Good reliability of process equipment and proper maintenance are hence important issues, but also to, beforehand, develop strategies to tackle such disturbances if they do occur. The adaptive design strategy, exemplified in Paper B, is an example of such a strategy. In brief, this means that the experimenter can choose to prolong individual runs during the experiment if disturbances occur. Alternatively, leaving unplanned time at the end of the experiment to be able to compensate for disturbances can be a good idea.

The experimenter in continuous processes often cannot measure the actual phenomena taking place in the process due to a variety of practical issues. Instead, many secondary responses such as flows, pressures, and temperatures are measured and the experimenter needs to use prior process knowledge to interpret what is occurring inside the process. The many responses are typically crosscorrelated and frequent logging of many responses together with the dynamic characteristic of continuous processes causes a high degree of autocorrelation (often positive). Therefore, as shown in Papers A, B, C, and D, multivariate statistical methods make an important contribution during the analysis of experiments in continuous processes. This research has especially shown how latent variable techniques (in particular PCA) can be used to extract the strongest signals in response data when there are many responses to analyze. The latent variables can then be used as new responses in the following analyses of the experiments or in process analysis before the experiment. The multivariate nature of response data can also become
problematic during the planning phase, see Paper A. An abundance of responses and possible interactions can make it impractical to maintain the detail (for example predicting effects and interactions) in the experimental planning process. However, it is still critical that prior process knowledge is used in the planning process.

In continuous processes, restrictions on the ranges in which factors may be varied are frequent. With a weak signal sent into the system, the corresponding effect on the process output or performance can be difficult to detect, especially if the noise of the process is further amplified by process control activities. In addition, as split-plot designs often need to be used in continuous processes, the experimenter should also expect a lower precision when measuring whole-plot effects. Furthermore, experiments in continuous processes are expensive and making many replications of experimental runs is therefore not always realistic. Paper F, although not limited to continuous processes, shows how prior process knowledge, or knowledge about the sparsity, heredity, and hierarchy principles, can be used to increase the power of the analysis of unreplicated factorial designs. Underlining the importance of prior process knowledge in continuous processes is the fact that the experimenter will need to consult process knowledge to interpret the meaning of, for example, principal components, if multivariate statistical methods are to be used during analysis.

This research also demonstrates some advantages of viewing responses from experiments in continuous processes as time series. Again, this is due to the dynamic characteristic of the processes and that experimenters in continuous processes normally are interested in the performance of the process during experimental runs with an experimental setup which is fixed for some period of time. As shown in Papers D and E time series analysis thus becomes a useful tool to analyze experiments in continuous processes, to model process dynamics, and to establish transition times. In particular, Paper E shows that using an intervention-noise model to analyze an experiment with time series responses constitutes a more comprehensive method that seems to result in fewer spurious effects and higher power for unreplicated experiments than using a more simplified analysis based on averages and ANOVAs.

4.2 Reflections on the research process

The research presented in this thesis has made extensive use of the EBF process either as the studied case or as inspiration for developments of analysis methods. Hence, the empirical evidence for planning, conducting, and analyzing experiments
in continuous processes is exclusively based on a blast furnace process (in pilot scale). This choice has both advantages and drawbacks.

By focusing on one specific case (the EBF), an in-depth understanding of the special considerations that are needed to plan, conduct and analyze experiments in the process has been attained. The pilot-scale, and the fact that the EBF is specifically designed for experimental purposes, also provided unique opportunities to study and be a part of many experiments during the research project. Although the EBF is in pilot scale, it is by no means a small plant. Running the EBF requires similar deliberations, personnel, and machinery as running a full-scale furnace, but the volumes handled are of course much smaller. If the choice had been made, instead, to study several different continuous processes (during the same amount of time) it would, according to this author, have been at the cost of the depth of understanding of the phenomena that the experimenter needs to consider in the specific processes. The collaboration between the author and the EBF engineers has been valuable to create an understanding of the problems that are encountered when trying to apply DoE methods and related analysis tools in a complex continuous process setting. It is my belief that this understanding could not have been acquired using, for example, interviews, questionnaires or simulations only. However, I am aware that the strength of the approach at the same time is its weakness, since some of the results presented in this thesis cannot without reflection be transferred to other continuous processes. Instead I have to rely on analytic generalization. However, I believe that the proposed analysis methods in papers B, D, E, and F are general and not limited to the studied industrial case.

It is my conviction, from studying the literature, that the special considerations that this research reports regarding planning, conducting, and analyzing experiments based on the EBF case apply for many continuous processes. Important experimental complications for continuous processes have been found and verified by studying the EBF. Among these are, for example, the problems of running large and fully randomized experiments, their dynamic characteristics, the multivariate nature of data, the need for process control during the experiments, and the time series aspects of the responses.

However, it is not unlikely that I would have found some additional circumstances and complications of importance if I had been studying, for example, a paper mill, a pelletizing plant, or a chemical process. Especially, I believe that the even larger scale of a full-scale production plant with its specific complications, such as having to sell the product produced during the experiment, further complicates
ON DESIGN OF EXPERIMENTS IN CONTINUOUS PROCESSES

experiments in continuous processes. The EVOP procedure proposed by Box (1957) is one way to perform experiments in full-scale production plants without sacrificing the possibility to market and sell the products produced during experiments.

4.3 Contribution

Approaching the end of this thesis it is indeed time to ask the question: Is there a contribution in all this? Below I put forward what I believe to be the main contributions of the research presented in this thesis.

This research explicitly explores and describes special considerations and problems that can be encountered when planning, conducting and analyzing experiments in a dynamic continuous process. These types of considerations and problems have been scarcely described in DoE literature. Using the EBF case to discuss experimental challenges and demonstrate many of the proposed analysis methods in the appended papers hopefully adds to create a better understanding of the practical use of DoE methods in industry. The identified special considerations and problems can be seen as a theoretical contribution to the DoE field regarding the use in practice of DoE methods, such as the use and analysis of factorial designs.

This research also identifies the need to, and illustrates the benefits achieved by, combining methods from four rather distinct fields: DoE, multivariate statistical methods, time series analysis, and statistical process control and monitoring to deal with some of the identified problems. In particular, this research shows how multivariate approaches to analysis and monitoring, using time series analysis to determine transition times and analyze experiments, and a Bayesian approach to analysis of unreplicated experiments can be used to tackle some of the problems in continuous processes. Although this research does not deal with theoretical development of the specific applied analysis methods per se, I believe that the development of adapted analysis procedures for experiments in continuous processes is an important contribution to the DoE field and provide powerful aids for the experimenter in continuous processes.

4.4 Implications and potential benefits for the experimental work at the EBF

In this section I provide some implications of the research results for the experimental work at the EBF and outline possible benefits that can be achieved by adopting the work and recommendations presented in this thesis.
A systematic approach to planning the experiments at the EBF is important to handle the complex environment. An experimental planning guide, incorporating the special considerations discussed in Paper A, is currently used at the EBF and the reactions from EBF engineers have been positive.

The research has also shown the benefits of using factorial designs in those cases where more than one experimental factor is of interest. Conducting experiments in the EBF is costly and thus the cost and time savings produced by varying several experimental factors at the same time (compared to one-factor-at-a-time experiments) are important benefits. The dynamic characteristic of the EBF process does, however, mean that the frequent factor changes resulting from factorial designs can lead to many (costly) transition periods in the furnace. A method to formally estimate these transition times in the blast furnace is presented in Paper D. With knowledge of the transition times, the engineer can make an informed decision on the appropriate experimental design in the EBF. I believe that the costs and the dynamic characteristic of the EBF motivate that the EBF engineers should be familiar with the benefits and implications of split-plot designs when planning experiments with more than one experimental factor.

The research presented in this thesis has specifically highlighted the benefits of a multivariate approach to the analysis of process data from the EBF. The simultaneous analysis of the many and crosscorrelated responses from the process can lead to more valid results and conclusions of the experiments at the EBF. Furthermore, I believe that the multivariate approach to analysis can save time during the analysis phase. I do not, however, want to indicate that univariate analyses of certain important responses are unnecessary, only that they, preferably, should be preceded by a multivariate approach to get a more complete picture.

Performing experiments in a process that needs to be controlled is complicated. Disturbances such as cool periods of operation in the furnace reduce the time the EBF is ‘producing’ valid experimental data. The need for monitoring and control of the thermal state of the EBF process during the experiments can, at times, make the interpretation of experimental results difficult. The control decisions are made by operators and research engineers based on the interpretation of signals provided by many single responses from the process. Paper C shows how these signals of the thermal state can be viewed in just a few control charts of principal components. I believe that a successful implementation of this technique can reduce the number of plots that need to be studied and hopefully result in faster and better judgments of the thermal state and in more timely control decisions. However, I believe that
future work to develop strategies and methods of analysis for response data collected from a process under control (closed-loop) would be valuable.

Many process responses from the EBF are time series that the analysts need to handle in some way. Established analysis methods at the EBF often use, for example, averages and standard deviations of the responses to analyze the experiments. Papers D and E of this thesis show the benefits achieved by using time series analysis techniques to analyze process dynamics and analyze experimental results. I believe that time series techniques can be used to extract even more information from the experiments, and reduce the possible negative effect that the noise in the process can have on the analysis results. Since the experimental time in the EBF is costly, I argue that more powerful analysis methods, although more complicated, are warranted to make the most out of the information output from the conducted experiments. Added importance comes from the fact that the experiments often have a limited number of replications. Indeed, there still remains issues concerning, for example, how to handle process disturbances and missing values in the response time series from the EBF, that need to be addressed. The Bayesian analysis method presented in Paper F can also be used to analyze experiments at the EBF, where the costs and time concerns do not allow for many replications of experimental runs. The Bayesian approach allows for incorporation of prior knowledge, which the analyst may have regarding activity of effects, to increase the power of the analysis.
5. FUTURE RESEARCH

Research should also be viewed as a continuous process. Ideas and new questions for future research have come up during the research process and this chapter presents those implications for future research that I find the most interesting.

A natural continuation of this research could be to test the external validity of the results presented here which relies heavily on the study of a single industrial case (the EBF). To closely study how experiments are performed in other continuous process industry settings, for example, the pulp and paper industry or the chemical industry, is probably a good idea to verify the results, discover new potential complications, and learn more about experiments in continuous processes. To study experiments in full-scale continuous processes would also be important to uncover additional possible complications. I suspect that other statistical and quality engineering methods like statistical process control and capability analysis also are affected by the special characteristics and problems found in continuous processes. A future study including these methods may therefore be valuable.

One of the recommendations in this thesis is to consider split-plot designs when planning experiments in continuous processes. I would find it interesting to study how frequently split-plot type designs are used in industry, how often they should have been used, and how often these designs are in fact analyzed correctly. Based on personal communication, I suspect that in many cases actual split-plot experiments are analyzed as if they were fully randomized,

A complication that has been encountered during the research presented in this thesis is the need to control the process during experimentation. It has been found that process control decisions can cause ambiguity regarding the experimental results. I suspect that experiments performed under closed-loop conditions are common in process industry settings. Paper C discusses a multivariate method for process monitoring, but this method does not eliminate the possible bias in the experimental results due to control actions. For automated control loops (without human deliberations) the control variable(s) can be used as a response, but how should a manual control variable affected by subjective operator decisions be treated? I believe that further research on the analysis of closed-loop experiments and how the need for process control affects experimental procedures, analysis and results is highly motivated for the future.

The time series nature of the responses from continuous processes is specifically treated in Papers D and E, where time series analysis methods and multivariate
statistical methods are combined to estimate transition times and analyze experiments. I think that the combination of DoE and time series analysis for dynamic processes deserves future attention to develop more effective and efficient methods of analysis and to make them more easily available for the engineer. A future challenge would be to try to use the information conveyed during transition periods between runs to shorten the required length of the experiment and thereby cut costs.

Recently, it has come to my attention that the study of the dynamic relation between an input to and output from a dynamic system, related to Papers D and E of this thesis, is also discussed within the field of ‘system identification’ within control and systems engineering. Söderström and Stoica (1989) describe system identification as “the field of modeling dynamic systems from experimental data.” The purpose of the experiments is often to gain knowledge to design better controllers for the process. Papers D and E especially focus on modeling the dynamic response based on step changes of the input variables to the system, which is motivated as DoE often use factorials with two-levels of each factor. In system identification literature additional types of input ‘signals’ to the system commonly used are: impulses, sinusoids, ramps etc. To study the system identification literature more closely therefore seems motivated to capture possible ideas and work important for DoE in continuous processes.
APPENDIX I – ABOUT THE BLAST FURNACE PROCESS

The appendix provides a brief introduction to the blast furnace process of ironmaking. Although this is not a thesis about the blast furnace process or its metallurgical aspects, this introduction provides background information of value for the appended papers. An introduction of the EBF plant is also made here. Further descriptions are found in the appended papers.

A.1 The blast furnace process

Today, there are two main processes available for the production of steel products, namely the Blast Furnace and the Electrical Arc steelmaking processes. The former still remains the primary source for the world’s steel production. In the Blast Furnace process, coke and coal or oil are used as reductant sources (or fuel) and sinter and/or pellets as iron-bearers. (Geerdes et al., 2004)

In brief, reduction of iron oxide passes through three steps, Fe$_2$O$_3$ (Hematite) $\rightarrow$ Fe$_3$O$_4$ (Magnetite) $\rightarrow$ FeO (Wustite) $\rightarrow$ Fe. Carbon monoxide (CO) or hydrogen in the hot ascending gas in the furnace partially reduces the iron oxides in the upper part of the furnace. The remaining oxygen content of the iron-bearing materials is removed by ‘direct reduction’ in the high temperature zone (Biswa, 1981). A diagram of a blast furnace is presented in Figure A.1.

![Diagram of a blast furnace](image)

**Figure A.1** A diagram of a blast furnace with important terms indicated. Source: Zuo (2000, p. 1), with permission from the author.

---

The Electrical Arc Steelmaking process uses electric energy to melt scrap metal.
Biswas (1981) characterizes the blast furnace as a high temperature counter current chemical reactor for the reduction and smelting of iron ore into hot metal. The blast furnace produces hot metal in liquid form which is then transported to the steel plant for further refinement. This typically involves removing elements such as sulfur, silicon, carbon, manganese and phosphorus from the hot metal; see Geerdes et al. (2004).

As indicated in Figure A.1, the burden composed of iron-bearing materials, together with coke and fluxes is charged from the top of the furnace. Pre-heated air (often enriched with oxygen) and supplementary fuel (coal or oil) are injected through the tuyeres (cooled copper canonical pipes) in the bottom of the furnace. The combustion of coke and injected fuel with oxygen occurs in front of the tuyeres and this creates the reducing gas and heat needed for the reactions and melting of the burden. (Zuo, 2000)

Alternating layers of coke and iron bearers are charged in the top of the furnace. The pre-heated air (up to more than 1200° C), called blast, is blown into the furnace through a number of tuyeres, and the number of tuyeres varies depending on the size of the furnace. In the gasification in front of the tuyeres, the oxygen in the blast is transformed into carbon monoxide and the resulting gas has a flame temperature between 2100 and 2300° C. When the gas reaches the top of the furnace, the temperature has dropped to 100-150° C. (Geerdes et al., 2004)

The counter current principle (hot gas ascends while the burden descends) of the blast furnace process can be considered the driving force of the process. According to Geerdes et al. (2004), the hot gas ascending through the furnace shaft performs a number of important tasks:

- heats the coke in the lower part of the furnace,
- melts the iron-bearing material in the furnace,
- heats up the material in the upper part of the furnace, and
- removes oxygen from the iron oxides in the burden.

An obvious aim for a modern blast furnace is to produce as much raw iron as possible at the lowest possible cost. Thus, minimizing coke consumption becomes an important task since coke is the most expensive component in the burden. This motivates substituting coke with less expensive auxiliary fuel in form of coal powder or oil and injecting them through the tuyeres.

In general, the efficiency of the blast furnace process is considered to be the reductant rate per metric ton hot metal. This is monitored by continuously
measuring the chemical composition of the top gas in the furnace. The percentage of the CO gas that has been transformed into CO₂ is expressed as the gas utilization; see Geerdes et al. (2004) and Biswas (1981).

\[
\text{CO gas utilization} = \frac{\% \text{ CO}_2 \text{ in top gas}}{(\% \text{ CO} + \% \text{ CO}_2 \text{ in top gas})}
\]  

(A.1)

Zuo (2000) argues that the blast furnace process is quasi-stable. The process is controlled not only by measures taken by operators but also by constantly complex variations in the blast furnace, for example, changes of the composition and distribution of the burden and in the gas as well as the position of the cohesive zone in the furnace. Hence, Zuo (2000, p. 2) further argues that the modeling and control of the process in practice is difficult and arduous because of the following four characteristics:

1. The blast furnace is a continuous process and any irregular fluctuations in the process will disturb the steady-state condition reached, lasting from several hours to days with a negative effect on production.
2. Time-lag of furnace responses to adjustments of operational parameters.
3. Black box. The difficulties (often due to the high-temperature and dusty environment) of online measurement for showing transport phenomena and reactions occurring inside the furnace make parts of the blast furnace operation a black box with multiple input and output variables.
4. Dynamics and non-linearity. Changing one parameter in the blast furnace often causes a chain of changes. The quantitative relationships between process variables are time-dependent and probably non-linear.

A.2 The LKAB Experimental Blast Furnace (EBF)

In 1997, Luossavaara-Kiirunavaara AB (LKAB), a Swedish producer of iron ore products, pellets in particular, inaugurated a pilot scale blast furnace. The EBF was specifically designed for experimental purposes in connection to product development and has many possibilities for measurement during operation. Today, the EBF constitutes a vital tool in the research and development work at LKAB. The EBF represents the, previously missing, link between tests in lab-scale and full-scale tests of, for example, iron ore pellets in commercial blast furnaces. LKAB’s customers also have the possibility to use the EBF for tests and evaluation of processing conditions and raw materials before taking the next step towards full-scale tests. The EBF is also an important tool in external research projects with the
aim to study and develop the blast furnace process, for example, in projects with the aim to lower the coke consumption and the CO$_2$ emissions from the blast furnace processes. Figure A.2 shows a picture of the EBF plant.

![The LKAB Experimental Blast Furnace (EBF). Source: LKAB with permission.](image)

Although the experimental cost per run and risks associated with performing experiments are great even in this pilot scale, they are substantially lower than they would have been in full-scale operation. Volume-wise the EBF is much smaller than a commercial blast furnace, but running the EBF requires similar deliberations, personnel, and machinery as running a full-scale furnace. The EBF has much of the same measurement possibilities as commercial blast furnaces, but, in addition, the EBF has burden probes for extraction of burden materials for analysis, such as, semi-
reduced iron ore pellets from the process, see Figure A.3. The EBF is typically run for two experimental campaigns per year. The length of the campaigns varies, but often lies around two months. Each campaign may, in turn, consist of several specific experiments with different aims. After each campaign, the EBF is quenched by nitrogen and, after cooling, excavated and material samples from different layers in the furnace can be analyzed.

The customers of iron ore pellets generally want their blast furnaces to run efficiently and effectively with few disturbances and that the resulting iron is of a high grade. The experiments performed in the EBF include response variables related to the quality of the produced iron, that is, the chemical composition of the pig iron and the slag, which is determined by off-line analyses. Since LKAB is a producer of iron ore pellets to be used in blast furnaces, there is a natural interest to evaluate the performance of the EBF while a specific product (e.g. pellet type) is being charged into the furnace. Therefore, responses related to energy efficiency and stability of the process itself, such as, top gas composition, burden descent rate, CO gas utilization, pressure drops, temperatures in the top of the furnace and in the shaft, and cooling effects are highly important during the analysis of experiments in the EBF. Figure A.4 presents a schematic outline of the EBF process together with examples of measurement possibilities, and Table A.1 gives some facts about the
EBF. Hallin et al. (2002) provide further details about the EBF plant. The experimental work at the EBF is further discussed in Paper A.

![Schematic outline of the EBF process](image)

**Figure A.4** Schematic outline of the EBF process inspired by Zuo (2000). A few examples of possible responses are underlined.

**Table A.1** Examples of specifications of the EBF. Source: Hallin et al. (2002, p. 311).

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working volume</td>
<td>8.2 m³</td>
</tr>
<tr>
<td>Hearth diameter</td>
<td>1.2 m</td>
</tr>
<tr>
<td>Hearth height</td>
<td>5.9 m</td>
</tr>
<tr>
<td>Number of tuyeres</td>
<td>3 (diameter 54 mm)</td>
</tr>
<tr>
<td>Top pressure</td>
<td>up to 1.5 bar</td>
</tr>
<tr>
<td>Injection through tuyeres</td>
<td>coal, oil, slag formers</td>
</tr>
<tr>
<td>Blast volume</td>
<td>up to 2000 Nm³/h</td>
</tr>
<tr>
<td>Blast heating</td>
<td>pebble heaters</td>
</tr>
<tr>
<td>Maximum blast temperature</td>
<td>1300 °C</td>
</tr>
<tr>
<td>Furnace crew, excluding sampling and research staff</td>
<td>5/shift</td>
</tr>
<tr>
<td>Tapping volume</td>
<td>~ 1.5 t/tap</td>
</tr>
<tr>
<td>Tap time</td>
<td>5–15 min</td>
</tr>
<tr>
<td>Fuel rate</td>
<td>~ 500 kg/t hot metal</td>
</tr>
</tbody>
</table>
A.2.1 Examples of factors during an experiment in the EBF

Reconsider Figure 3.3 in Chapter 3, which provides a general model for a continuous process under experimentation. To provide some further background information, this section exemplifies some typical examples of important experimental factors, $X$, control factors, $C$, held-constant factors, disturbance factors, $Z$, and responses, $Y$, when performing experiments in the EBF process.

**Experimental factors, $X$**

Experimental factors are often different raw materials tested under constant or varied processing conditions. Table A.2 provide some examples of these and other experimental factors in the EBF.

**Control factors, $C$ (and held-constant factors)**

The thermal state of the blast furnace must be controlled during the experiments. Furthermore, disturbances of a critical nature may also bring about a manipulation of control variables during the experiment. The process is typically controlled by the coke rate, adding or subtracting coke in the burden mixture, but can also be controlled by other variables, see examples given in Table A.3. The choice is often made to control the process by manipulating, for example, the coke rate and/or the amount of auxiliary fuel injected through the tuyeres. Depending on the choice of control variable(s), the remaining variables are normally kept at a constant level during the experiment.

**Disturbance factors, $Z$**

The number of possible disturbance factors that can affect an experiment of the magnitude and complexity as those performed in the EBF are large. Table A.4 presents some examples of disturbances that can and have affected experiments run in the EBF.

**Responses, $Y$**

As discussed above, responses from the EBF are primarily related to two general classes: process responses and output responses. There are also responses related other more ‘qualitative’ signals of furnace stability and to lab tests on the extracted material through the burden probes. Table A.5 provides some typical examples.
### Table A.2 Examples of experimental factors in the EBF.

<table>
<thead>
<tr>
<th>Examples of experimental factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron ore pellet types and mixtures of pellets and sinter</td>
</tr>
<tr>
<td>Blast volume (or other regulators of the production rate)</td>
</tr>
<tr>
<td>Moisture content of blast air</td>
</tr>
<tr>
<td>Type of auxiliary fuel injected (e.g., carbon powder type)</td>
</tr>
<tr>
<td>Charging technique of iron bearers and coke</td>
</tr>
</tbody>
</table>

### Table A.3 Typical control variables (and held-constant factors).

<table>
<thead>
<tr>
<th>Examples of control variables and held-constant factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coke rate</td>
</tr>
<tr>
<td>Coal powder (or oil) injection through tuyeres</td>
</tr>
<tr>
<td>Oxygen content of blast air</td>
</tr>
<tr>
<td>Blast air temperature</td>
</tr>
<tr>
<td>Moisture content of blast air</td>
</tr>
<tr>
<td>Furnace top pressure</td>
</tr>
<tr>
<td>Charging method of iron bearers and coke</td>
</tr>
</tbody>
</table>

### Table A.4 Examples of disturbance factors that can affect an experiment in the EBF.

<table>
<thead>
<tr>
<th>Examples of disturbance factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moisture content of coke</td>
</tr>
<tr>
<td>Furnace irregularities such as:</td>
</tr>
<tr>
<td>- Channeling (uneven gas distribution and flow through the burden)</td>
</tr>
<tr>
<td>- Scaffolds and scabs (build-up of materials at the shaft wall)</td>
</tr>
<tr>
<td>- Hanging (obstructed downward flow of the burden)</td>
</tr>
<tr>
<td>- Slips (sudden rapid downward movements of the burden)</td>
</tr>
<tr>
<td>Breakdown of vital machinery that disturb the material flow</td>
</tr>
<tr>
<td>Variations in the incoming raw materials</td>
</tr>
<tr>
<td>Human factors, for example, control decisions</td>
</tr>
</tbody>
</table>
Table A.5 Examples of typical responses used to analyze experiments in the EBF.

<table>
<thead>
<tr>
<th>Examples of typical responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process responses</strong></td>
</tr>
<tr>
<td>• CO gas efficiency [%]</td>
</tr>
<tr>
<td>• Direct reduction rate [%]</td>
</tr>
<tr>
<td>• Burden permeability index [no unit]</td>
</tr>
<tr>
<td>• Differential pressures measured in the shaft [mbar]</td>
</tr>
<tr>
<td>• Production rate [t/hour]</td>
</tr>
<tr>
<td>• Burden descent rate [cm/min]</td>
</tr>
<tr>
<td>• Radial temperature index in furnace top [° C]</td>
</tr>
<tr>
<td>• Temperature at the shaft wall [° C]</td>
</tr>
<tr>
<td>• Silicon in pig iron [weight %]</td>
</tr>
<tr>
<td>• Sulphur in pig iron [weight %]</td>
</tr>
<tr>
<td>• Carbon in pig iron [weight %]</td>
</tr>
<tr>
<td>• Hot metal temperature [° C]</td>
</tr>
<tr>
<td>• Iron content in slag [weight %]</td>
</tr>
<tr>
<td>• Magnesium oxide in slag [weight %]</td>
</tr>
<tr>
<td>• Slag basicity, CaO/SiO2 [no unit]</td>
</tr>
<tr>
<td>• Disintegration strength tests through tumbling</td>
</tr>
<tr>
<td>• Chemical analyses</td>
</tr>
<tr>
<td>• Sieve fractions</td>
</tr>
<tr>
<td>Other signals of furnace stability, such as furnace irregularities</td>
</tr>
<tr>
<td>• Channeling, scaffolds, scabs, hanging, and slips</td>
</tr>
</tbody>
</table>
REFERENCES


REFERENCES


REFERENCES


REFERENCES


ON DESIGN OF EXPERIMENTS IN CONTINUOUS PROCESSES

Operation and Maintenance Engineering; Division of Quality and Environmental Management, Luleå University of Technology, Luleå. Doctoral Thesis.


REFERENCES


ON DESIGN OF EXPERIMENTS IN CONTINUOUS PROCESSES


PAPER A

Special Considerations when Planning Experiments in a Continuous Process

Vanhatalo, E. and Bergquist, B. (2007)

Published as:

Available at:
Due to the publisher's restrictions of the use of the post-print version of the article, Paper A is not published electronically. The paper can be accessed through the link below.

http://dx.doi.org/10.1080/0898210701474100
PAPER B

Using Factorial Design and Multivariate Analysis When Experimenting in a Continuous Process


Published as:

Available at:
Due to the publisher’s restrictions of the use of the post-print version of the article, Paper B is not published electronically. The paper can be accessed through the link below.

http://dx.doi.org/10.1002/qre.935
PAPER C

Multivariate Process Monitoring of an Experimental Blast Furnace

Vanhatalo, E. (2009)

Published as:

Available at:
Due to the publisher’s restrictions of the use of the post-print version of the article, Paper C is not published electronically. The paper can be accessed through the link below.

http://dx.doi.org/10.1002/qre.1070
PAPER D

A Method to Determine Transition Time for Experiments in Dynamic Processes


Submitted for publication
A Method to Determine Transition Time for Experiments in Dynamic Processes

Erik Vanhatalo¹, Björn Kvarnström¹, Bjarne Bergquist¹, and Kerstin Vännman²
¹ Division of Quality Technology, Environmental Management, and Social Informatics
² Department of Mathematics
1,2 Luleå University of Technology, SE-97187, Luleå, Sweden
Correspondence to Erik Vanhatalo: E-mail: erik.vanhatalo@ltu.se, Phone: +46 920 49 17 20

Abstract: Process dynamics is an important consideration already during the planning phase of designed experiments in dynamic processes. After changes of experimental factors dynamic processes undergo a transition time before reaching a new steady-state. To minimize experimental time and reduce costs, knowledge about this transition time is important for the design and the analysis of the experiment. In this article, we propose a method to analyze process dynamics and estimate the transition time by combining principal component analysis and transfer function-noise modeling or intervention analysis. We illustrate the method by estimating transition times for a planned experiment in an experimental blast furnace.

Keywords: Time series analysis, Transition time, Principal component analysis, Transfer function-noise model, Intervention analysis, Design of Experiments, Blast furnace.

1. Introduction

Dynamic processes are frequently found in continuous process industries, where process steps such as: mixing, melting, chemical reactions, silos, and reflux flows contribute to their dynamic characteristics. Planning, conducting, and analyzing experiments in dynamic processes highlight special issues that the experimenter needs to consider. Examples of such issues are: process dynamics (inertia), a multitude of responses, large-scale and costly experimentation, and many involved people, see Hild et al. (2000) and Vanhatalo and Bergquist (2007). In this article we focus on the dynamic characteristic of continuous processes and argue that process dynamics must be considered already during the experimental planning phase.

In a dynamic process, a delay, here called the transition time, will occur between the change of an experimental factor until the response is affected, whereas in a responsive system, this change is almost immediate, see, for example, Saunders and Eccleston (1992) and Black Nembhard and Valverde-Ventura (2003). Consequently, time series of the process responses need to be studied after each experimental treatment is applied to allow for the possible effect of the experimental treatment to manifest itself. By contrast, responses can often be measured on individual experimental units directly after the experimental treatment has been applied in many parts production processes.

When planning an experiment in a dynamic process it is important to have some knowledge of the transition time caused by process dynamics. The transition time affects the required length of each experimental run in the process and long transition times may call for restrictions in the randomization order of the experimental runs, see, for example, Vanhatalo...
and Vännman (2008). Knowledge about the transition time will help the experimenter to avoid experimental runs that are either too short for a new steady-state to be reached, and thus wrongly estimating the treatment effects, or unnecessarily long, which increases costs. By knowing the required transition time, the experimenter can choose a better design that produces the needed information at a lower cost, without jeopardizing experimental validity.

Determining the transition time in a dynamic continuous process can be difficult due to a number of reasons. Continuous processes are often heavily instrumented to measure different aspects of the process and product being processed, and multiple process responses are often needed to capture the effect of experimental treatments. The transition time may also vary for different responses and treatment changes may affect some responses but not all. Process variables are typically correlated and often react to the same underlying event, see Kourti and MacGregor (1995). In such cases latent variable techniques such as principal component analysis (PCA) can be used to achieve data reduction and aid interpretation, see, for example, Wold et al. (1987) and Jackson (2003).

The sample rates of the process measurements in continuous processes in industry are often frequent enough to estimate process dynamics, usually with a higher sampling frequency than the frequency of the process oscillations. Slow process drifts and oscillations combined with high sampling frequencies lead to positively autocorrelated responses. The autocorrelation indicates that time series analysis could be a useful analysis tool. Time series analysis contains techniques where stochastic and dynamic models are developed to model the dependence between observations sampled at different times. The time series analysis techniques transfer function-noise modeling and intervention analysis have been proposed to model the dynamic relation between an input time series \(X_t\) and an output series \(Y_t\) and between a known intervention at time \(t\) and an output series \(Y_t\) respectively, see, for example, Box et al. (1994), Wei (2006), and Montgomery et al. (2008).

In this article, we propose a method to determine the transition time for experimental runs in a dynamic process. The proposed method combines multivariate statistical methods and time series analysis techniques to analyze process dynamics and estimate the transition times in dynamic processes. Section 2 introduces the method where PCA summarizes the systematic variation in a multivariate response space. Transfer function-noise modeling or intervention analysis is then used to model process dynamics and determine the transition time between an input time series event and output time series response using principal component scores. The approach is illustrated in Section 3 using data from a continuously operating experimental blast furnace.

2. Proposed method

Continuous processes running in, for example, process industries usually have servers filled with old process data, and any important process that has run for some time should have undergone trials or experiments to improve it. Often, past interventions that resemble the changes planned for an upcoming experiment can be found. If similar past interventions does
not exist, a trial run (if possible) designed to induce the transition in the process may be justified to estimate the transition times.

Normally, process analysis of continuous processes is a multivariate task, and determining the transition time can be rather difficult. For this problem we use PCA to create a few, independent, linear combinations of the original variables that together summarize the main variability in the response space. We then use time series analysis on the principal component scores to investigate the transition time. The following sections introduce PCA and the times series analysis techniques transfer function–noise models and intervention analysis.

2.1 Principal component analysis (PCA)

PCA can reduce the dimensionality of the response space by extracting a few new, latent, uncorrelated variables called principal components (linear combinations of the original variables) that together explain the main variability in the data, see, for example, Johnson and Wichern (2002) and Jackson (2003). The use of the PCA technique to summarize the response space is outlined below.

Let \( \mathbf{y} = [y_1, \ldots, y_m] \) represent a random response vector describing an \( m \)-dimensional response variable with covariance matrix \( \Sigma \). Let \( \Sigma \) have the eigenvalue–eigenvector pairs \((\lambda_1, \mathbf{p}_1), (\lambda_2, \mathbf{p}_2), \ldots, (\lambda_m, \mathbf{p}_m)\). The \( m \) principal components (PCs) are formed as linear combinations of the original responses:

\[
P_{C_1} = \mathbf{p}_1' \mathbf{y} = p_{11}y_1 + p_{12}y_2 + \ldots + p_{1m}y_m \\
P_{C_2} = \mathbf{p}_2' \mathbf{y} = p_{21}y_1 + p_{22}y_2 + \ldots + p_{2m}y_m \\
\vdots \\
P_{C_m} = \mathbf{p}_m' \mathbf{y} = p_{m1}y_1 + p_{m2}y_2 + \ldots + p_{mm}y_m
\]

The PCs are orthogonal to one another, ordered according to their variances. The first PC has the largest variance, the second PC the second-largest variance and so on, where the eigenvalues, \( \lambda_a, a = 1, 2, \ldots, m \), are the variances of the PCs. The eigenvectors, \( \mathbf{p}_a, a = 1, 2, \ldots, m \), have unit length, \( \mathbf{p}_a' \mathbf{p}_a = 1 \), and are called the PC loading vectors. PCA is scale-dependent and the responses are usually scaled to unit variance, using standardized variables, before the PCA. The PCs of the standardized variables are obtained by calculating the eigenvalues and eigenvectors of the correlation matrix of \( \mathbf{y} \) instead of the covariance matrix, see, for example, Johnson and Wichern (2002). In the applications studied here we scale to unit variance before the PCA and, hence, use the correlation matrix of \( \mathbf{y} \) to derive the eigenvectors and eigenvalues.

The correlation matrix is unknown in practice and estimated by the sample correlation matrix calculated from an observed \( n \times m \) \( \mathbf{Y} \) matrix with \( n \) observations of each of the \( m \) responses. The values of the PCs for each observation are here called PC scores, and the score vectors, \( \mathbf{t}_a, a = 1, 2, \ldots, m \), represent the \( n \) observed values of the PCs based on the observed \( \mathbf{Y} \) matrix.

The goal of the PCA is reduction of dimensionality, and if the variables are highly correlated, much of the systematic variation described by a correlation or covariance matrix
can be described using $A < m$ dimensions, and the remaining $m - A$ dimensions are considered to contain mostly random noise.

The loading vectors, $\mathbf{p}_a$, $a = 1, 2, \ldots, A$, define the reduced dimensional space ($A$) with respect to the original responses and the score vectors, $\mathbf{t}_a$, $a = 1, 2, \ldots, A$, are the projections of the original observations onto the $A$-dimensional reduced space.

The number of retained principal components ($A$) can be derived by several methods, see Jackson (2003). One way is to extract the number of components that are needed to reproduce a specific fraction of the variance of the original response data. When working with standardized variables, it is also common to only keep PCs with eigenvalues larger than one, so each PC explains at least as much of the total variation as one of the original variables. Cross-validation, see Wold (1978), is also frequently used to select the appropriate number of components.

If only the $A$ first PCs are used to approximate the variability in $\mathbf{Y}$, we can write:

$$\mathbf{Y} = \mathbf{T}\mathbf{P}' = \sum_{a=1}^{A} \mathbf{t}_a\mathbf{p}_a' + \mathbf{E}$$

where $\mathbf{T}$ is an $n \times m$ matrix with the score vectors as rows, $\mathbf{P}'$ is an $m \times m$ matrix with the loading vectors as columns, and the variability in the remaining $m - A$ PCs are summed up in the residual matrix $\mathbf{E}$.

### 2.2 Transfer function-noise models

This section gives a brief introduction to transfer function-noise models. For further descriptions, see, for example, Jenkins (1979), Box et al. (1994, chapter 10), Wei (2006, chapter 14), Bisgaard and Kulahci (2006), and Montgomery et al. (2008, chapter 6).

Consider the single-input time series $\mathbf{x}_t$ and the single-output time series $\mathbf{y}_t$. Assume that the input time series can be represented by a quantitative continuous variable. Further assume that the input time series has been manipulated and we want to study its effect on the output from the process. Note that the output series could be an original process response or, as proposed in this article, a linear combination of many responses represented by principal components. Assume that both $\mathbf{x}_t$ and $\mathbf{y}_t$ are zero-mean stationary time series and that they are related through the linear filter:

$$\mathbf{y}_t = \mathbf{v}(\mathbf{B})\mathbf{x}_t + \mathbf{N}_t$$

where $\mathbf{B}$ is the backshift operator, $\mathbf{v}(\mathbf{B}) = \sum_{i=0}^{s} \vartheta_i \mathbf{B}^i$ is the transfer function, and $\mathbf{N}_t$ represents the unobservable zero-mean noise. The number of coefficients in $\mathbf{v}(\mathbf{B})$ are usually assumed to be limited to a fairly small number and to follow the structure:

$$\mathbf{v}(\mathbf{B}) = \frac{\vartheta(\mathbf{B})}{\delta(\mathbf{B})} = \frac{\vartheta_0 - \vartheta_1 B - \ldots - \vartheta_s B^s}{1 - \delta_1 B - \ldots - \delta_r B^r}$$

The coefficients $\vartheta_0, \ldots, \vartheta_s$ and $\delta_1, \ldots, \delta_r$ determine the structure of the transfer function, $\mathbf{v}(\mathbf{B})$, and $s$ and $r$ are the orders of the numerator and denominator respectively. The
coefficients $\nu_i$, also called the impulse response function, can be obtained recursively from the coefficients $\omega_k, \ldots, \omega_k$ and $\delta_t, \ldots, \delta_t$, see Montgomery et al. (2008, chapter 6). Sometimes there is a delay before $x_t$ starts to affect $y_t$. If we assume that this ‘pure delay’ is $b$ time units, the transfer function-noise model can be represented by:

\[ y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + N_t \]  

Furthermore, it is assumed that $N_t$ is uncorrelated with $x_t$ and that $N_t$ can be represented by an autoregressive integrated moving average model, ARIMA($p, d, q$):

\[ \Phi(B)(1-B)^d N_t = \Theta(B)\varepsilon_t \]  

where $\Phi(B) = (1-\phi_1B-\cdots-\phi_pB^p)$, $\Theta(B) = (1-\theta_1B-\cdots-\theta_qB^q)$, and $\{\varepsilon_t\}$ represents white noise. See, for example, Montgomery et al. (2008) for the process of fitting ARIMA models.

By combining (5) and (6), the transfer function-noise model can be expressed as:

\[ y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\Theta(B)}{\Phi(B)(1-B)^d} \varepsilon_t \]  

### 2.3 Identifying transfer function-noise models

The following seven steps are taken to obtain the transfer function, $\nu(B)$, and the noise model, see also Montgomery et al. (2008). In the following descriptions we assume that the input and output series have been scaled so the mean is zero for each series.

**Step 1: Prewhiten the input series $x_t$**

If the input series $x_t$ is autocorrelated, the method of prewhitening is needed to obtain the transfer function. Prewhitenning is a procedure that transforms the input series $x_t$ into white noise. Normally, an appropriate ARIMA model is used to filter $x_t$:

\[ \alpha_t = \frac{\Phi_x(B)(1-B)^d}{\Theta_x(B)} x_t \]  

where the filtered input series, $\alpha_t$, should be white noise with zero mean and variance $\sigma^2_{\alpha}$.  

**Step 2: Apply the prewhitening filter to the output series $y_t$**

The same prewhitening filter is then applied to the output series $y_t$ to obtain

\[ \beta_t = \frac{\Phi_y(B)(1-B)^d}{\Theta_y(B)} y_t \]  

where the filtered output series has variance $\sigma^2_{\beta}$ and is not necessarily white noise. The cross correlation function between the prewhitened input series $\alpha_t$ and filtered output series $\beta_t$ is
directly proportional to the weights, $\nu_i$, in the transfer function. We have the following relation, see Montgomery et al. (2008),

$$\nu_i = \frac{\sigma_i}{\sigma_u} \rho_{i}(l)$$  \hspace{1cm} (10)

where $\rho_{i}(l)$ is the cross correlation function between $\alpha_i$ and $\beta_i$ at lag $i$, $i = 0, \pm 1, \pm 2, \ldots$.

**Step 3: Obtain initial estimates of the impulse response function $\nu_i$**

Using the sample estimates $\hat{\rho}_{i}(l)$, $\hat{\sigma}$, and $\hat{\sigma}_{\nu}$ of $\rho_{i}(l)$, $\sigma_u$, and $\sigma_{\nu}$, respectively and applying Eq. (10) we obtain the initial estimates of the impulse response function $\nu_i$ as:

$$\hat{\nu}_i = \frac{\hat{\sigma}_{\nu}}{\hat{\sigma}_u} \hat{\rho}_{i}(l)$$  \hspace{1cm} (11)

Montgomery et al. (2008) recommend using $\pm 2\sqrt{n}$, where $n$ is number of observations in the time series, as the approximate 95% confidence interval to judge the significance of the cross correlations and thus the estimated weights $\hat{\nu}_i$.

**Step 4: Specify $b$, $r$, and $s$ and obtain a preliminary estimate of the transfer function.**

The possible delay, $b$, is identified by studying $\hat{\rho}_{i}(l)$. A tentative specification of the orders $r$ and $s$ in the transfer function is made by matching the pattern of $\hat{\nu}_i$, obtained from (11), with known theoretical patterns. Examples of theoretical patterns of the impulse response function for comparison can be found in, for example, Box et al. (1994, p. 389), Wei (2006, pp. 325-326), and Montgomery et al. (2008, pp. 305-306). When $b$, $r$, and $s$ have been chosen, preliminary estimates $\hat{\phi}$ and $\hat{\delta}$ can be obtained through their relationships with $\hat{\nu}_i$. Thus, a tentative transfer function can be formed as:

$$\hat{\phi}(B) = \frac{\hat{\phi}(B)}{\hat{\delta}(B)}$$  \hspace{1cm} (12)

**Step 5: Model the noise $N_i$**

Once the preliminary transfer function has been established, the estimated noise series, $\hat{N}_i$, can be calculated as:

$$\hat{N}_i = y_i - \hat{\phi}(B) y_{i-b}$$  \hspace{1cm} (13)

By studying the time series plot, the autocorrelation function and the partial autocorrelation function of the estimated noise series in (13), an appropriate ARIMA model is chosen to model any remaining structure in the noise series.

**Step 6: Fit the overall model**

The first five steps have produced a tentative model specification:
The final estimates of the parameters $\delta = (\delta_1 \ldots \delta_T)'$, $\omega = (\omega_1 \ldots \omega_T)'$, $\phi = (\phi_1 \ldots \phi_T)'$, and $\theta = (\theta_1 \ldots \theta_T)'$ are obtained by an iterative maximum likelihood fit of the specified model to the time series.

**Step 7: Model adequacy checks**

The validity of the estimated model is studied by checking two important assumptions of the fitted model. First, the residuals from the model, $e_t$, should be white noise. Second, the independence between $x_t$ and $e_t$ should also be checked, see also Wei (2006) and Montgomery et al. (2008).

### 2.4 Intervention analysis

This section provides a brief introduction to intervention analysis and its application to study the effect of a known intervention on an output time series. For further descriptions, see Jenkins (1979), Box et al. (1994, chapter 12), and Wei (2006, chapter 10).

Assume that the single-output time series $y_t$ is affected by a known event such as a change of a qualitative treatment. For example, different input materials to a continuous process may have to be represented by a qualitative indicator variable summarizing all possible differences in the materials. Let

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-\delta_0} + \frac{\Theta(B)}{\Phi(B)(1-B)} e_t$$

where $\delta_0$ is a binary deterministic indicator variable with value 0 for nonoccurrence, and with value 1 for occurrence of the specific event and $b$ determines the possible pure delay of the intervention effect. Two common types of indicator variables are the step variable:

$$S^{(t)}_t = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases}$$

and the pulse variable:

$$P^{(t)}_t = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases}$$

where $T$ is the time of the intervention.

Due to the deterministic nature of the indicator time series, the method of prewhitening is no longer meaningful. The form of the intervention model must therefore be specified by considering the mechanisms that might cause the change and by studying the time series to suggest an appropriate model, see Jenkins (1979).

A step function is well suited for the intervention exemplified in this article, a shift of raw materials. Assuming that the intervention can be represented by the simple step variable in Eq. 16, Figure 1 shows the intervention response for different values of $\delta_0$ given a transfer function.
function on the form \( \omega_i B \left[ 1 - \delta_i B \right] \). Often a gradual response is reasonable to assume, which corresponds to the case \( 0 < \delta_i < 1 \).

\[
\begin{align*}
\delta_i &= 0 \Rightarrow \omega_i S_i^{(T)} & (a) \\
0 < \delta_i < 1 \Rightarrow \frac{\omega_i}{1 - \delta_i B} S_i^{(T)} & (b) \\
\delta_i &= 1 \Rightarrow \frac{\omega_i}{1 - B} S_i^{(T)} & (c)
\end{align*}
\]

**Figure 1.** Response to an intervention in form of a step function based on a step variable and a simple transfer function depending on different values of \( \delta_i \). The figure is adapted from Box et al. (1994, p. 464).

3. Transition time in an Experimental Blast Furnace

To illustrate the proposed approach to determine the transition time in a dynamic process we use data from an Experimental Blast Furnace (EBF). The EBF is owned and operated by Luossavaara-Kiirunavaara AB (LKAB), a Swedish producer of iron ore products (iron ore pellets in particular). The EBF is a pilot scale blast furnace, specifically designed for experimental purposes and the production capacity of the EBF is approximately thirty-five tons of hot metal per day (compared to up to 10,000 tons per day for the largest full scale furnaces). For more details about the experiments run in the EBF, see Vanhatalo and Bergquist (2007) and Vanhatalo and Vännman (2008).

Two of the most frequently used experimental factors in the EBF are the types of iron bearers (mostly iron ore pellets) and the blast parameter settings. The transition time after changes of these experimental factors is not fully known but highly important when planning the experiments in the EBF. Figure 2 presents an outline of the EBF and examples of measurement possibilities.

3.1 Transition time when changing oxygen content in the blast air

During experiments in the EBF, it is often of interest to change the production rate to test the raw materials under different process conditions. The production rate can be changed either by altering the oxygen content of the blast or by changing the blast volume. The needed
transition time when changing the oxygen content is therefore important to estimate. Process responses calculated from, for example, pressure sensors and thermocouples in the EBF can be used to study how the process reacts to changes. Table 1 presents the process responses used to analyze the transition time.

Data consisting of hourly averages for each of the variables in Table 1 from a past experimental campaign were located where the oxygen content of the blast air had been changed between two target values: 45 and 90 Nm³/h. In total, 371 observations (hours) were available for each variable.

Figure 2. Outline of the EBF process. Examples of possible responses are underlined. The two types of changes studied in this article (pellets and oxygen content in the blast air) are indicated by bold uppercase font.

Table 1. Important process responses from the EBF. The numbering is for future reference.

<table>
<thead>
<tr>
<th>Process response</th>
<th>Unit</th>
<th>Process response</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Differential pressure over furnace</td>
<td>bar</td>
<td>12. Gas speed top</td>
<td>ms⁻¹</td>
</tr>
<tr>
<td>2. Differential shaft pressure, dp5-45°</td>
<td>bar</td>
<td>13. Flame temperature</td>
<td>°C</td>
</tr>
<tr>
<td>3. Differential shaft pressure, dp5-225°</td>
<td>bar</td>
<td>14. Production rate</td>
<td>ton⁻¹</td>
</tr>
<tr>
<td>4. Top temperature</td>
<td>°C</td>
<td>15. Specific blast volume</td>
<td>Nm⁻¹ton⁻¹</td>
</tr>
<tr>
<td>5. Temperature BR 1</td>
<td>°C</td>
<td>16. Direct reduction rate, DRR</td>
<td>%</td>
</tr>
<tr>
<td>6. Temperature BR 2</td>
<td>°C</td>
<td>17. Solution loss</td>
<td>kgC⁻¹ton⁻¹</td>
</tr>
<tr>
<td>7. Burden descent rate</td>
<td>cmmin⁻¹</td>
<td>18. Wall flow index</td>
<td>°C</td>
</tr>
<tr>
<td>8. (n_{CO}) (CO gas utilization)</td>
<td>%</td>
<td>19. Center flow index</td>
<td>°C</td>
</tr>
<tr>
<td>9. Cooling effect tuyeres</td>
<td>kW</td>
<td>20. Top gas flow</td>
<td>Nm⁻¹ton⁻¹</td>
</tr>
<tr>
<td>11. Gas speed furnace</td>
<td>ms⁻¹</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PCA was conducted and the first four PCs explained 80.5% of the variation in the data, see Table 2. We found that the first PC, which explains the largest part of the variability in the process responses, separates the two oxygen contents. See Figure 3a–b, where the scores \( t_1 \) of the first PC are plotted against the scores \( t_2 \) of the second PC. We conclude that the change of oxygen content seems to explain the main variability visible in the variables in Table 1. None of the other PCs showed similar clear dependence on the oxygen content. Hence, only the scores \( t_1 \) will be studied in the following analysis. Figure 4 shows a time series plot of the 371 observations on \( t_1 \) together with the input oxygen content.

Table 2. Results from the PCA on the 371 observations of the process variables in Table 1.

<table>
<thead>
<tr>
<th>PC</th>
<th>Explained variance [%]</th>
<th>Cum. explained variance [%]</th>
<th>Eigenvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32.5</td>
<td>32.5</td>
<td>6.83</td>
</tr>
<tr>
<td>2</td>
<td>24.0</td>
<td>56.5</td>
<td>5.04</td>
</tr>
<tr>
<td>3</td>
<td>14.9</td>
<td>71.5</td>
<td>3.14</td>
</tr>
<tr>
<td>4</td>
<td>9.0</td>
<td>80.5</td>
<td>1.89</td>
</tr>
</tbody>
</table>

Figure 3a. PC score scatter plot, \( t_1 \) vs. \( t_2 \) coded according the oxygen target levels.

Figure 3b. PC loading plot \( p_1 \) vs. \( p_2 \). See Table 1 for the variable codes.

From Figure 4 we see that the oxygen content seems to affect the first principal component. When the oxygen content is increased, the first PC decreases. However, since there is autocorrelation present we need to apply transfer function-noise modeling to be able to draw any clear conclusions.
Transfer function-noise modeling

The time series of \( t_i \) may be modeled as an ordinary time series, not acknowledging the change in oxygen content, and this was done for comparative reasons. An ARIMA(0,1,1) model was fitted, where first differencing was needed to account for the nonstationary behavior (the shifts in \( t_i \)). The nonstationary behavior means that the expected value of \( t_i \) changes over time.

Since the level of \( t_i \) seems to depend on the level of the oxygen content of the blast air we try to explain the nonstationary behavior in \( t_i \) by fitting a transfer function-noise model. In the transfer function-noise model we use the oxygen content of the blast air to try to account for the shifts in \( t_i \). If the noise in the transfer function-noise model can be modeled by an ARMA model (no differencing) we assume that the shifts in \( t_i \) are mainly explained by the changes of the oxygen content. To test the explanatory performance of a transfer function-noise model for the time series in Figure 4b, models were developed in a stepwise manner following the description in section 2.2. The models found were then compared. The software JMP\textsuperscript{®} 8.0 was used for the calculations.

The oxygen content was used as the single input series \( x_i \) and the scores \( t_i \) as the single output time series. First the input and output series were prewhitened using an ARIMA(0,1,1) model to form \( \alpha_i \) and \( \beta_i \), according to Steps 1-2 in Section 2.2. In Step 3 the cross correlation function (CCF) between \( \alpha_i \) and \( \beta_i \) was estimated, see Figure 5.
Figure 5. Cross correlation function between $\alpha_i$ and $\beta_i$. The sampling interval was one hour.

By interpreting Figure 5 tentative values of the delay $b$ and the orders of $r$ and $s$ in (4) can be found as discussed in Step 4 in Section 2.2. A pure delay of one hour ($b = 1$) seems reasonable, since the lag 1 cross correlation coefficient is the first significant coefficient. We see one (possibly two) significant spikes in the CCF at lag 1 and 2. We are only interested in the cross correlations at lag 0 and positive lags to see how changes of the input is correlated to $t_i$. Large spikes at negative lags are likely due to spurious correlations.

Since the cross correlation coefficients are proportional to the impulse response function according to (10), the pattern in the CCF was compared to theoretical patterns of the impulse response function in Montgomery et al. (2008, pp. 305-306). From this comparison two tentative transfer functions were identified and fitted. The transfer functions were assumed to have denominator of degree 0, that is $r = 0$. Two possible values for the numerator in the transfer function were considered, $s = 0$ (one spike) and $s = 1$ (two spikes). The remaining correlation structure in the noise from both models was eliminated, as described in Step 5, by an ARIMA(1,0,0) model. Note that this is a AR(1) model and hence the shifts in $t_i$ can be explained by the oxygen content. Finally, the overall models were fitted as described in Step 6 in Section 2.2. Table 3 gives a summary of the two transfer functions-noise models found together with the ARIMA(0,1,1) model that does not consider changes of the input variable.

We present model criteria for comparison in Table 3. For details about these criteria, see, for example Montgomery et al. (2008, pp. 57-60). Generally, models with small standard deviation of the residuals, small mean absolute error, high adjusted coefficient of determination, and small values on the Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) are preferable. The AIC and SIC criteria penalize the sum of squared residuals when including additional parameters in the model. Montgomery et al. (2008) recommend using SIC over AIC.
Table 3. Comparisons of the transfer function-noise models and a univariate ARIMA(0,1,1) model for the time series of the first principal component, t₁. The models were fitted using JMP® 8.0 statistics software. The standard errors for the fitted parameters are given above or below the parameter values. The arrows next to the model criteria indicate if the corresponding criterion should be large (↑) or small (↓).

<table>
<thead>
<tr>
<th>Fitted models (hourly averages)</th>
<th>d.f.</th>
<th>s.d.</th>
<th>MAE</th>
<th>( R^2 )</th>
<th>AIC</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) ARIMA (0,1,1) ( \Delta t_1(t) = \left(1 - 0.21 B\right) \delta_t )</td>
<td>369</td>
<td>0.72</td>
<td>0.54</td>
<td>0.92</td>
<td>811</td>
<td>815</td>
</tr>
<tr>
<td>b) ARIMA (1,0,0) ( t_1(t) = 0.32 - 0.0044 B \left(1 - 0.67 B\right) \delta_t )</td>
<td>367</td>
<td>0.64</td>
<td>0.49</td>
<td>0.94</td>
<td>727</td>
<td>740</td>
</tr>
<tr>
<td>c) ARIMA (1,0,1) ( t_1(t) = 0.32 - 0.015 B \left(1 - 0.67 B\right) \delta_t )</td>
<td>365</td>
<td>0.64</td>
<td>0.49</td>
<td>0.94</td>
<td>725</td>
<td>741</td>
</tr>
</tbody>
</table>

Notes: \( \Delta \) indicates that the first difference of the time series is modeled. d.f. = degrees of freedom; s.d. = standard deviation of the residuals; MAE = Mean Absolute prediction Error; AIC = Akaike Information Criterion; SIC = Schwarz Information Criterion.

Model b) and c) perform equally well. According to model b), the gain due to the change of oxygen content is completely realized one hour after the change and according to model c) two hours after the change. However, the standard error of second parameter in the transfer function in model c) is large compared to the estimated coefficient. We therefore choose to exclude model c) from further consideration and conclude that the shift in \( t_1 \) seems to occur within the first hour after the oxygen content in the blast air has been changed.

**Increased resolution of the transition time**

Because the change was considered to be completed within the first hour after the change, we increased the resolution by analyzing the time series using ten-minute averages instead of hourly averages. We kept the PC loadings based on the PCA of the hourly averages and calculated ten-minute averages (based on minute values) for the original variables in Table 1. PC scores were then computed using the ten-minute averages and the PC loadings based on hourly averages; necessary to avoid having the extreme autocorrelation in the minute values from the process affect the calculation of the PCs.

Using the same analysis procedure as described above, transfer function-noise models were fitted to the data and the best model is given in Table 4. The parameters \( \phi_1 \) and \( \phi_2 \) in the transfer function were significant, which indicated that the change in \( t_1 \) occurred during the first 20 minutes after the change in oxygen content of the blast air. The noise was now described by an ARIMA(2,0,1) model.
Table 4. Transfer function-noise model for the first principal component, $t_1$, based on ten-minute averages. The standard errors for the fitted parameters are given above or below the parameter values.

<table>
<thead>
<tr>
<th>Fitted model (ten-minute averages)</th>
<th>d.f.</th>
<th>s.d.</th>
<th>MAE</th>
<th>$R^2_{adj}$</th>
<th>AIC</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1(t) = 7.33 - \left( 0.066 + 0.0082 B + 0.033 B^2 \right) x_t +$</td>
<td>2211</td>
<td>0.77</td>
<td>0.56</td>
<td>0.92</td>
<td>5160</td>
<td>5200</td>
</tr>
</tbody>
</table>

As $t_1$ describes the main variability of the process responses caused by the change of the oxygen content we conclude that a transition time of about 20 minutes is a reasonable estimate for the transition time of the described change in the oxygen content of the blast air.

3.2 Transition time when changing iron ore pellets

Another type of experiment common in the EBF is product development experiments, where raw materials (iron ore pellets, fluxes and fuels/reactants such as coke) with differing compositions are tested. Here we investigate the transition time when changing iron ore pellets. The pellets, together with coke and fluxes, are charged at the top of the furnace, and the burden descends through the furnace shaft. Although many surface reactions occur already in the furnace shaft, most reactions are expected to take place when the burden reaches the reaction zone, in which the pellets are reduced and molten into hot metal. The hot metal and the slag are then tapped from the bottom of the furnace. The transition time between two types of pellets in the EBF can, therefore, be estimated by studying analyses of the pig iron and the slag tapped from the furnace, which are available approximately once every hour, see Table 5.

Data from past experimental campaigns in the EBF, where changes between two types of pellets had been made and all other processing variables were held constant, were retrieved. We present the analysis of one such change-over in this article. In this example, 99 observations, approximately one hour apart, on the variables in Table 5 were available and the change of pellets occurred just before observation 38.

PCA was conducted on the variables in Table 5. It was found that the three first PCs explain 71.6% of the variation in the data, see Table 6. The first PC mainly describes the thermal state in the bottom of the furnace, where lower values on the scores $t_1$ indicate a cooler process. The second PC seems to describe a dimension that differentiates between the two pellets, see Figure 6. Figure 7 shows a time series plot of the 99 observations on $t_2$ and the time of the change of the pellets from type A to type B is indicated.
Table 5. Variables in the analysis of pig iron and slag from the EBF.

<table>
<thead>
<tr>
<th>Hot metal Unit</th>
<th>Slag Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot metal temperature (Temp) °C</td>
<td>Iron content (Fe) Wt. %</td>
</tr>
<tr>
<td>Carbon (C) Wt. %</td>
<td>Calcium oxide (CaO) Wt. %</td>
</tr>
<tr>
<td>Silicon (Si) Wt. %</td>
<td>Silicon dioxide (SiO₂) Wt. %</td>
</tr>
<tr>
<td>Manganese (Mn) Wt. %</td>
<td>Manganese oxide (MnO) Wt. %</td>
</tr>
<tr>
<td>Phosphorus (P) Wt. %</td>
<td>Sulphur in slag (S slag) Wt. %</td>
</tr>
<tr>
<td>Sulfur (S) Wt. %</td>
<td>Aluminium oxide (Al₂O₃) Wt. %</td>
</tr>
<tr>
<td>Nickel (Ni) Wt. %</td>
<td>Magnesium oxide (MgO) Wt. %</td>
</tr>
<tr>
<td>Vanadium (V) Wt. %</td>
<td>Sodium oxide (Na₂O) Wt. %</td>
</tr>
<tr>
<td>Titanium (Ti) Wt. %</td>
<td>Potassium oxide (K₂O) Wt. %</td>
</tr>
<tr>
<td>Vanadium oxide (V₂O₅) Wt. %</td>
<td>Titanium dioxide (TiO₂) Wt. %</td>
</tr>
<tr>
<td>Phosphorus oxide (P₂O₅) Wt. %</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Results from the PCA on the 99 observations of pig iron and slag variables.

<table>
<thead>
<tr>
<th>PC</th>
<th>Explained variance [%]</th>
<th>Cum. explained variance [%]</th>
<th>Eigenvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43.5</td>
<td>43.5</td>
<td>9.13</td>
</tr>
<tr>
<td>2</td>
<td>20.0</td>
<td>63.5</td>
<td>4.21</td>
</tr>
<tr>
<td>3</td>
<td>8.1</td>
<td>71.6</td>
<td>1.69</td>
</tr>
</tbody>
</table>

Figure 6a. PC score scatter plot, t₁ vs. t₂ coded according to the pellets type.

Figure 6b. PC loading plot, p₁ vs. p₂. See Table 5 for the variable codes.

Figure 7. Time series plot of t₂. Observation 38 is the first after the change of iron ore pellets has occurred. The observations are approximately one hour apart.
Intervention analysis

Since the difference between the two iron ore pellets cannot be expressed quantitatively (the pellets may differ not only in chemistry, but also in processing conditions, production time, and production plant), the transition can instead be modeled by intervention analysis. Using intervention analysis, the change of pellets can be described by a step function, \( S_i^{(T)} = 0 \) for pellets A, and \( S_i^{(T)} = 1 \) for pellets B. It is thus not possible to use prewhitening to identify the structure of the transfer function in an intervention model. Instead, the structure of the transfer function must be estimated by viewing the time series in the light of the underlying mechanisms behind the change. When new pellets are charged at the top of the furnace, they will descend for a few hours before reaching the reaction zone. It is reasonable to assume that the response will exhibit a pure delay during this descent. The newly molten material is then mixed with the remaining material from the previous burden mix in the bottom of the furnace, and it is thus likely that the chemistry of the melt will change gradually. A reasonable assumption is therefore the following transfer function (see also Figure 2)

\[
\frac{\theta_1}{1 - \delta B} S_i^{(T)}
\]

Again, models of the time series in Figure 7 were developed in a stepwise manner and compared. First an ARIMA(0,1,1) model was fitted to the time series of \( t_i \), where the differencing was needed to account for the nonstationary behavior (the shift in \( t_i \)). Thereafter the intervention variable was introduced, testing different values of the pure delay (\( b \)). The intervention variable accounts for the shift in the time series and the remaining noise was described by an ARIMA(1,0,1) model. See Table 7 for a summary of the tested models.

Table 7 shows only minor differences among the fitted models for the model criteria. It can be concluded that the intervention variable can explain the shift in \( t_i \) that otherwise warrants first differencing of the output time series. Model f) and g) in Table 7 perform similarly, and slightly better than model e) for all criteria except SIC and practically only differ in the choice of pure lag (\( b \)).

By testing other change-overs between pellet types at similar production rates in the EBF (not elaborated here) we conclude that \( b = 3 \) is probably the best choice. We chose model f) in Table 7 for calculation of the transition time.
Table 7. Comparison of intervention models and a univariate ARIMA(0,1,1) model for the second principal component, $t_2$. The standard errors for the fitted parameters are given above or below the parameter values. The arrows next to the model criteria indicate if the corresponding criterion should be large (↑) or small (↓).

<table>
<thead>
<tr>
<th>Fitted model</th>
<th>d.f.</th>
<th>s.d.</th>
<th>MAE</th>
<th>$R^2_{adj}$</th>
<th>AIC</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>e) ARIMA (0, 1, 1)</td>
<td>97</td>
<td>0.60</td>
<td>0.46</td>
<td>0.91</td>
<td>179</td>
<td>181</td>
</tr>
<tr>
<td>f) Intervention Noise ($b = 3$)</td>
<td>91</td>
<td>0.58</td>
<td>0.43</td>
<td>0.92</td>
<td>173</td>
<td>186</td>
</tr>
<tr>
<td>g) Intervention Noise ($b = 4$)</td>
<td>90</td>
<td>0.58</td>
<td>0.43</td>
<td>0.92</td>
<td>171</td>
<td>184</td>
</tr>
</tbody>
</table>

Notes: $\Delta^1$ indicates that the first difference of the time series is modeled.

Transition time

Using model f) in Table 7 we assume a pure delay of three observations (about three hours) before the intervention starts to affect the chemistry in the pig iron and slag. According to Jenkins (1979, p. 62) the estimated ‘gain’, $g$, the ultimate change of $t_2$ due to the intervention, can be calculated from the transfer function as:

$$
\hat{g} = \frac{\delta_0}{1 - \delta_1} = \frac{-0.78}{1 - 0.78} = -3.54
$$

Hence, the change of pellets will eventually cause the average of $t_2$ to decrease with 3.54 units. An estimate of the percentage of the change that has occurred after each time period (with the start in period $b = 3$) can be calculated as:

$$
\frac{\omega_0}{g}; \frac{\omega_0 \delta_1}{g}; \frac{\omega_0 \delta_1^2}{g}; \frac{\omega_0 \delta_1^3}{g}; \cdots \frac{\omega_0 \delta_1^n}{g}
$$

Figure 8 presents the estimated cumulative percentage change realized in $t_2$ as a function of the time after the intervention. After about 10 hours, about 90% of the total change has occurred which may be a reasonable cutoff to estimate the transition time. Hence we conclude that the experimenter needs to add a transition time of ten hours before measuring effects of a pellets change and probably a similar time for other raw materials charged at the top of the blast furnace, given that a comparable production rate is used in the EBF.
Validation of the transition time for the change of pellets

To examine if the estimated transition time for pellet changes in the EBF is stable, another change-over between two pellets types was analyzed. The same chemical variables of the pig iron and the slag and the same analysis procedure as described above was used. The transition pattern produced by the final model from the validation test is given in Figure 8 together with the transition pattern given by model f) in Table 7. We see that both models produce a similar transition pattern and conclude that the estimated transition time seems to be stable over time for changes of pellets in the EBF, given a specific production rate.

4. Conclusions and discussion

In this article we propose a method to determine the needed transition time between experimental runs in a continuous process. Since we often encounter a multitude of responses in a continuous process we first use PCA to create a few new, independent, linear combinations of the original variables that together summarize the main variability in the response space. We then investigate if the PCs seem to be dependent on the changes in the input time series to the process. If this is the case, we use transfer function-noise models or intervention analysis to model the dynamic relation between an input time series $x_t$ and the output time series in form of PC scores. The transition time is estimated from the fitted dynamic models. The proposed method is summarized in Figure 9.
We illustrate the method using data from an experimental blast furnace where two types of transitions were studied. The transition time after a change of a quantitative process variable was determined through transfer function-noise modeling while the change of a qualitative raw material variable was analyzed by intervention analysis. The results show that the estimated transition time for the material variable is substantially longer than for the process variable, which is important information for the experimental planning process, but also for the analysis phase. The results differed compared to the prevailing understanding among the engineers at the EBF. Previously, changes of pellets were thought to be noticeable after four hours, while changes in blast parameters were considered to take even longer time, which also affected decisions in the planning phase of experiments in the EBF. The complete transition times in the EBF process for these changes were not fully known.

The estimation of the transition time using PCs provides a summarized and manageable overview of the course of events in a multivariate response situation. However, if the change of the input affects the process in several ways with different time lags, we have to be careful. Assume a situation where our change affects response \( Y_i \) more slowly than the other responses. Then \( Y_i \) will probably be uncorrelated with most other responses and have loadings of small magnitude in the first number of PCs. That is, \( Y_i \) is correlated to the change of the input but not to the other responses due to the different lag structure. Indeed, transfer function-noise models and intervention models for single-output variables like \( Y_i \) may be needed to complement models using PCs to get the complete picture of the transition time. The experimenter may check that responses of special interest have loadings of significant magnitude in the PCs. Otherwise, single-output models should be considered.

By knowledge of the transition time we can establish the needed length of each experimental run. Each run requires enough time to include the transition time and an additional time during which responses can be sampled at the new process state. Furthermore, knowing the transition time between runs is important when selecting representative data for each run to include in the analysis of the experiment. Usually the responses during the
transition time are excluded from the analysis, which further stresses the importance to correctly determine the transition time. In addition, good estimates of the transition time in the process are useful to achieve better traceability in dynamic processes. For example, the transition time can measure the dynamic propagation of a disturbance or product change in the process output. The transition time can also be of importance for process control strategies and design of engineering control systems.

For some process variables the estimated transition may be gradual and slow (see, for example, Figure 8) and the experimenter may need to decide a reasonable cutoff. In such cases, the transition time can, for example, be defined as the time required to reach 90 percent of the total change modeled by the transfer function.

If the transition time for different experimental factors differs significantly, the experimenter may consider randomization restrictions for factors with longer transition times. Hence, split-plot designs can be arranged based on information about the transition time for the factors to minimize the required length of the whole experiment. Further descriptions about split-plot designs are given in, for example, Box and Jones (1992) and Kowalski et al. (2007). Factors with longer transition times can be natural choices for whole-plots while those with shorter transition times are potential sub-plot factors. The transition time is by no means the only consideration when deciding on the appropriate experimental design, such as a split-plot design, but since the experimental time in continuous processes normally is limited and costly the transition time is an important issue.

Acknowledgement

The financial support from the Swedish mining company LKAB and the European Union, European Regional Development Fund, Produktion Botnia is gratefully acknowledged. The authors thank all members of the LKAB ‘EBF methodology development project’ for their important contribution to the results presented here. Special thanks to Gunilla Hyllander at LKAB for valuable support.

About the authors

Erik Vanhatalo and Björn Kvarnström are PhD students at Luleå University of Technology (LTU), Luleå, Sweden. They both hold Licentiate degrees of Engineering in the subject of Quality Technology and Management and Master’s degrees in Industrial and Management Engineering from LTU. Erik Vanhatalo’s current research is focused on experimental design in continuous processes. Björn Kvarnström’s current research is focused on methods to improve traceability in continuous processes. They are both members of the European Network for Business and Industrial Statistics (ENBIS).

Bjarne Bergquist is Associate Professor of Quality Management and head of the Division of Quality Technology, Environmental Management and Social Informatics at LTU. He holds a Master’s degree in Mechanical Engineering from LTU and a PhD in Materials Science from Linköping University, Sweden. His main research interest is focused on process control and experimental design, especially for continuous process applications. He is a member of ENBIS.
Kerstin Vännman is a Professor in Statistics with special emphasis on Industrial Statistics at the Department of Mathematics, LTU. Her main research interest is currently in the field of statistical process control, including capability analysis, as well as the design of experiments and multivariate data analysis, but she also has an interest in statistical education. She is a member of ISI, ASA and ENBIS.

References


PAPER E

Analyzing Two-Level Factorial Experiments with Time Series Responses

Vanhatalo, E., Bergquist, B., and Vännman, K. (2009)

Research report to be submitted for publication:
Analyzing Two-Level Factorial Experiments with
Time Series Responses

Erik Vanhatalo¹, Bjarne Bergquist¹, and Kerstin Vännman²
¹Quality Technology and Management
²Department of Mathematics
¹²Luleå University of Technology, SE-97187, Luleå, Sweden
Correspondence to Erik Vanhatalo: E-mail: erik.vanhatalo@ltu.se, Phone: +46 920 49 17 20

Abstract: Dynamic processes exhibit a time-delay between disturbances and the resulting process response. It is therefore necessary to acknowledge process dynamics, such as transition times, when planning and analyzing experiments in such processes. In this article we explore and compare different methods to estimate location effects for two-level factorial experiments with time series responses. Particularly, we outline the use of intervention-noise modeling to estimate the effects and compare this method with averaging out the observations of the response of each run. The comparisons are made by simulated experiments using a dynamic continuous process model. The results show that the effect estimates for the different analysis methods are similar. Using the average of the response in each run, but removing the transition time, is found to be a relatively robust and straightforward method, while intervention-noise models are found to be more comprehensive, render fewer spurious effects, find more of the active effects for unreplicated experiments, and provide the possibility to model effect dynamics.

Keywords: Two-level factorial design, Time series analysis, Process dynamics, Intervention-noise model, Location effects, Industrial experiments, Simulation.

1. Introduction

Many industrial processes exhibit a dynamic behavior and combined with typical high measurement sampling frequencies, the measurement series become autocorrelated. The autocorrelation is especially evident in process industry, where process dynamics contribute to slow-moving propagations of disturbances. When experimenting on such systems, the observed responses are represented by time series. In these situations, our experience is that the time series aspects often are ignored and instead a single response value is assigned to each experimental run. Analysis procedures that ignore the dynamic nature of the responses may be ineffective or even erroneous. To disregard the time series characteristics by, for example, averaging out the entire time series including the transition periods when the process reacts to different treatments should be a poor alternative, as it likely leads to underestimation of location effects and overestimation of the variation.

We argue that a comparison and discussion of different ways to analyze experiments with time series responses may be valuable for many experimenters. The purpose of this article is hence to propose, illustrate, and compare different ways to analyze factorial experiments with time series responses. Here, standard replicated and unreplicated two-level factorials with three
experimental factors are used as examples, and we limit our study to the estimation of location effects. To compare the analysis methods, we use a simple simulation model that emulates an experiment performed on a process with dynamic behavior.

The only prior work we have found in design of experiments (DoE) literature that explicitly focuses on the analysis of experiments with time series responses is Hau et al. (1996). They use regression analysis with the response series in each run as the dependent variable and time as the independent variable. The overall average and trend for each run are estimated and these regression parameters are then used as response observations of each run.

The serial dependence between adjacent observations in many industrial processes suggests that methods such as autoregressive integrated moving average (ARIMA) models may be more effective, see Box et al. (2008). However, according to Montgomery et al. (2008) the estimation of the parameters of an ARIMA model requires at least 50 observations from a time series, implying that the response series from each run needs to include sufficiently many observations. Having 50 observations may also be inadequate. Each run needs to include as many observations as needed to capture process reactions, and the observations should be sampled close enough to capture the speed of change of relevant events.

Modeling of dynamic relations between, for example, experimental factors and responses, is possible using transfer function-noise modeling or intervention analysis. Already Jenkins (1979, p. 70) stated that “…intervention models represent generalizations of methods used for the analysis of data, usually not expressed as time series, and referred to by statisticians as ‘the design and analysis of experiments’.” See, for example, Box et al. (2008) for a discussion of intervention analysis. Transfer function-noise models also allow for modeling of the dynamic relation between experimental factors and the response, see Bisgaard and Kulahci (2006a; 2006b) and Box et al., (2008).

2. Process model under normal operation

In this article we assume that a stationary autoregressive moving average model, ARMA$(p,q)$, can be used to represent the (undisturbed) process response, $y_t$, under normal operation:

$$y_t = \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where $y_t$ is the value from the process at time $t$, $\delta$ is the constant term in the model, $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive (AR) coefficients, $p$ is the order of the AR part of the model, $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average (MA) coefficients, $q$ is the order of the MA part of the model, and $\varepsilon_t$ is Gaussian white noise. Even if we define our process to be stationary during normal operation, it may still exhibit cyclic behavior and strong autocorrelation. Here we assume that the process is operating in such a way that the process settles at a new level some time after a treatment change, and that the process can be represented by a stationary model after stabilization. The stationarity of an ARMA$(p,q)$ process is related to the AR part of the model, see, for example, Montgomery et al. (2008, p. 253). If the absolute values of the roots of the polynomial:
are all less than one, then the ARMA($p,q$) process is stationary. By choosing $\phi_1 = \phi_2 = \ldots = \phi_p = 0$ the process is reduced to a MA($q$) process, and when $\theta_1 = \theta_2 = \ldots = \theta_q = 0$ the process is reduced to an AR($p$) process. The mean of a stationary ARMA($p,q$) process is:

$$\mu = E(y_t) = \frac{\delta}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}$$

(3)

3. Dynamic simulation model

This section presents the model and assumptions to be used for simulating the process' reactions to the experimental interventions.

The underlying assumptions are the following: only the process mean is affected by the experimental treatments (not the variance or inherent process dynamics), and it is also assumed that the expected value of the process may change as a result of the experimental treatments. For some industrial experiments, this would not be true, and instead control actions may transfer process reactions to control variables. Process control is used, for instance, for quality concerns or for personal and plant safety reasons, and control may be needed during experimentation, see Vanhatalo (2009) and Vanhatalo and Bergquist (2007). Even if the assumption of open-loop operation is unrealistic for some applications, the analysis methods discussed here should be valid for situations where a control variable must be used as the actual response. See Box and MacGregor (1974; 1976) for a discussion of systems operating under closed-loop control.

For the forthcoming simulated experiments, we choose a $2^3$ factorial design, which results in seven possible effects that can affect the time series in (1). Thus, we write the time series model during the experiment as:

$$y_t = \delta + r^{(A)}_t + r^{(B)}_t + \ldots + r^{(AB)}_t + \ldots + r^{(ABC)}_t + \sum_{i=1}^{\text{n}} \phi_i y_{t-i} + \epsilon_t - \sum_{i=1}^{\text{n}} \theta_i \epsilon_{t-i}$$

(4)

where, $r^{(A)}_t, r^{(B)}_t, \ldots, r^{(AB)}_t, \ldots, r^{(ABC)}_t$ are the contributions to the mean of the time series at time $t$ due to possible main and interaction effects. Note that the effects are time-dependent. Letting the effects depend on time allows for modeling of situations where the effects gradually develop and stabilize, and modeling of different dynamic responses for the different effects. The time between the intervention until the response has stabilized on a new level is here referred to as the transition time, see also, for example, Black Nembhard and Valverde-Ventura (2003) or Vanhatalo et al. (2009).

Below, the modeling of the effect dynamics is exemplified by using the main effect of factor $A$, but all effects are modeled in the same way. As $t$ increases and $A$ is left unchanged, $r^{(A)}_t$ approaches its long-term value $r^{(A)}$. The pace may, however, differ for the different factors. The model is intended to emulate a gradual change of the response and should also
allow for a ‘pure lag’. The pure lag labels the possible initial delay before the effect starts to develop. We use transfer functions and ideas from intervention analysis to model this behavior, see, for example, Box and Tiao (1976), Jenkins (1979), and Box et al. (2008, Chapter 13). First, let a binary indicator variable, called the step variable, represent the two levels of each factor or interaction. Thus, for factor $A$:

$$S_{A,i}^{(0)} = \begin{cases} -1, & \text{for all } t \text{ when factor } A \text{ is kept at its low level} \\ 1, & \text{for all } t \text{ when factor } A \text{ is kept at its high level} \end{cases} \quad (5)$$

The dynamic response pattern of the effects is then modeled by a transfer function. This means that the contribution to the mean of the time series at time $t$ due to factor $A$ is given by:

$$Z_{i}^{(A)} = \frac{\omega_{A}}{1 - \psi_{A}B} S_{i}^{(A)}, \quad (6)$$

which corresponds to a change with the rate determined by the constant $\psi_{A}$, $0 \leq \psi_{A} < 1$, and the initial gain constant $\omega_{A}$, see Box et al. (2008, p. 531). The ‘pure lag’ is conveyed by the pure lag constant $b_{A}$, and $B$ is the backshift operator on $t$ so that $B_{t} = t$. We can thus re-write (6) on the form:

$$Z_{i}^{(A)} = \omega_{A}^{(A)} S_{i}^{(A)} + \psi_{A}^{(A)} Z_{i-1}^{(A)} \quad (7)$$

The resulting change pattern of (7) is gradual if $\psi_{A} > 0$ and eventually, given that $S_{i}^{(A)}$ remains unchanged, $Z_{i}^{(A)}$ approaches the long-term value:

$$Z_{A} = \frac{\omega_{A}}{1 - \psi_{A}} \quad (8)$$

The choice of $\psi_{A}$ determines the ‘inertia’ of the effect and a larger value of $\psi_{A}$ results in longer transition times. Letting $\psi_{A} = 0$ results in a direct response with value $\omega_{A}$ after any pure delay. The change pattern of the effect thus means that the effect is approximated through a first-order dynamic response to a step change. This means that the change rate is proportional to the difference between the effect at time $t$ and the equilibrium at the high and low level, see Box et al. (2008, pp. 442-447).

The contribution of the AR part of the undisturbed process must also be considered to obtain the expected long-term main effect of $A$. Using (3) and (8), and given that the main effect of factor $A$ is defined as the expected change in the response when factor $A$ is changed from its low to its high level, the expected long-term effect of factor $A$ is:

$$A_{\text{effect}} = \frac{2\tau_{A}}{\left(1 - \phi_{1} - \phi_{2} - \ldots - \phi_{p}\right)\left(1 - \psi_{A}\right)} = \frac{2\omega_{A}}{\left(1 - \phi_{1} - \phi_{2} - \ldots - \phi_{p}\right)\left(1 - \psi_{A}\right)} \quad (9)$$

Consider the following example as an illustration of the suggested simulation model. Let $\omega_{A} = 0.153$ and $\psi_{A} = 0.5$ (from one of the forthcoming simulations in Section 6) and let the undisturbed process be an ARMA(1,1) process with $\phi_{1} = 0.6$. Also let the undisturbed process be affected by introducing factor $A$ at its high level at time $t = 1$, and then factor $A$ is changed
to its low level at time \( t = 49 \). For ease of illustration, disregard the current mean of the process at time \( t = 0 \) and let \( y_0 = 0 \). Furthermore, disregard the Gaussian noise \( \xi \) that normally affects the process, then with a pure delay \( b_d = 1 \) the ‘deterministic part’ of the effect is seen in Figure 1.

**Figure 1.** An illustration of the dynamics (the first 97 observations) of the A effect, given \( \psi_d = 0.5 \), \( \phi_d = 0.153 \), \( \phi \) = 0.6, and \( b_d = 1 \). Here \( r_d = 0.306 \) and \( A_{f_{0.5}} = 1.53 \).

4. Five tentative methods of analysis

This section proposes and outlines five tentative methods, with increasing sophistication, to analyze a two-level factorial design with time series responses.

4.1 Method I, based on averages for each run

A naïve way to analyze an experiment with time series responses is to ignore the time series aspects of the response and calculate the averages of the observations in each run. The averages are then used as responses in a traditional analysis using analysis of variance (ANOVA) or a normal probability plot of the estimated effects. More formally, let run \( i \), \( i = 1, 2, \ldots, K \), include the response observations \( y^{(1)}_i, y^{(2)}_i, \ldots, y^{(n)}_i \), where \( n_i \) is the number of observations in run \( i \). Then the response in the \( i \)th run is the average

\[
\bar{y}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} y^{(j)}_i
\]  

(10)

4.2 Method II, based on averages after removal of the transition time

The expected consequence of including observations during the transition time in Method I is an underestimation of the location effects. A presumable improvement, given that the transition time is known or can be estimated, is to eliminate the observations during the transition time from each run. Vanhatalo *et al.* (2009) propose a formal method to determine the transition time for experimental factors in dynamic processes based on transfer function-noise modeling or intervention analysis. A less formal alternative is to use engineering judgements based on inspection of the time series. Once the transition time is estimated, the observations during the transition time are removed from each run and the adjusted averages are calculated and used as responses.
4.3 Method III, based on estimated parameters from an ARMA model

Methods I and II ignore a possible serial dependence between adjacent observations. An appropriate time series model may therefore produce better estimates of the variability and mean of each run. Practically, the proposed procedure involves dividing the time series from the experiment into separate time series for each run, and then fitting an appropriate ARMA model to the separate time series. The model building procedure we use for Method III follows recommendations given in Montgomery et al. (2008) and Box et al. (2008). Note that we disregard possible transition times between runs in Method III, just like in Method I. The appropriate model is determined by studies of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the original observations and the residuals. The estimated parameters, such as the mean, standard deviation, or even the AR and MA coefficients can then be used as new single responses in an ANOVA or a normal probability plot of the effects. Since this article focuses on location effects, only the estimated means from the ARMA models are used as the single response observation.

If the time series cannot be described by a stationary ARMA model, a stationary time series can often be created by taking the first difference of the time series:

\[ w_t = y_t - y_{t-1} = (1 - B)y_t \]  \hspace{1cm} (11)

or higher-order differences \( w_t = (1 - B)^d y_t \). The differenced time series can then be modeled by a stationary ARMA model and the model is now called an autoregressive integrated moving average model ARIMA\((p,d,q)\). However, note that if differencing is needed \((d > 0)\), Method III cannot estimate the mean of the specific run since the process then, by definition, has no fixed mean. For such runs, Method III will render a missing value for the mean. If effects are large, we also expect that some of the runs will exhibit a nonstationary behavior in the beginning of those runs where level of an active factor is changed.

4.4 Method IV, based on estimated parameters from an ARMA model after removal of the transition time

A possible solution to avoid the need for differentiation of the time series is, again, to exclude the observations during the transition time and estimate the appropriate ARMA\((p,q)\) model for the remaining observations. In a real industrial process, however, to acknowledge the transition time is not a guarantee that the rest of the time series can be assumed to be stationary as other disturbances may affect the response. Another possible complication is that the removal of response observations, when there are few observations to begin with, can result in too few response observations and uncertainty in the estimation of model parameters. Analysis methods III and IV are therefore not appropriate if too few observations are available in any of the runs.

4.5 Method V, based on intervention-noise modeling

Intervention analysis enables simultaneous analysis of the entire time series from all runs of the experiments and modeling of the effect dynamics. Because the time series is not divided among the runs, the problem of having too few observations to fit a time series model is mitigated.
Intervention analysis requires input time series for each main effect and interaction effect to model their relation to the output response series. Normally, intervention analysis uses a binary indicator variables with values 0 and 1, but here the two levels are coded as –1 (low level) and +1 (high level), following DoE conventions for two-level factorial designs.

Let $y_t$ be the time series response at time $t$ from the entire experiment. Then we assume that:

$$y_t = \frac{\omega_A(B)}{\delta_A(B)} z_{-b_A} + \frac{\omega_B(B)}{\delta_B(B)} z_{-b_B} + \ldots + \frac{\omega_{AB}(B)}{\delta_{AB}(B)} z_{-b_{AB}} + N_t$$

where $z_{-b_A}$, for example, is a binary deterministic indicator variable with value –1 when factor $A$ is on its low level, and with +1 when $A$ is on its high level, $b_A$ determines the possible pure delay of the intervention effect of the main effect of $A$, and $N_t$ is the remaining noise after the contributions from the input variables have been accounted for. An ARIMA($p,d,q$) model is used to account for any remaining structure in the noise, $N_t$, thus producing an intervention-noise model.

The general structure of the transfer function of, for example, factor $A$’s main effect is written as:

$$\frac{\omega_A(B)}{\psi_A(B)} = \frac{\omega_{b_A}-\omega_{-b}B-\ldots-\omega_{-b}B^s}{1-\psi_{b_A}B-\ldots-\psi_{b_A}B^r}$$

where $s$ and $r$ are the orders of the numerator and denominator polynomials respectively.

A possible drawback of using a binary coded variable for quantitative factors is that any deviation of the quantitative experimental factor settings from the experimental plan, such as difficulties of reaching and maintaining ±1, as well as unintended variation in the factors is disregarded. When there are only quantitative experimental factors, transfer function-noise models could be used instead of intervention analysis, since transfer function-noise models allow the use of actual factor settings, see, for example, Box et al. (2008, chapters 11-12).

Another difference between transfer function-noise models and intervention-noise models is that the analyst needs to postulate a tentative structure for the transfer functions in intervention analysis. When the inputs variables are quantitative continuous variables, the so-called ‘pre-whitening’ procedure is typically used to determine the structures, see, e.g., Jenkins (1979).

### 4.5.1 Model building procedure

To iteratively test all possible transfer functions in (13) for the model in (12) may become overwhelming and the fitting of all parameters at once can cause numerical problems. We therefore propose the following simplifications. Let $s = 0$ and $r = 1$ in (13), which limit the possible candidates for the transfer function and gives a simple transfer function that can model a gradual response, see also Box et al. (2008, p. 531):

$$\frac{\omega_A(B)}{\psi_A(B)} = \frac{\omega_{b_A}}{1-\psi_{b_A}B}$$

We also propose starting the analysis with a zero pure lag, $b = 0$, for all effects.
One way to analyze the experiment is to use backward-selection as follows. First estimate the parameters of the transfer functions, and then successively exclude nonsignificant transfer functions. However, we sometimes encounter numerical problems using this approach, such as non-convergence of the iterative maximum likelihood estimation algorithm in the software. An alternative and, in our opinion simpler way, is to iteratively fit the transfer functions for the effects, then fit an ARMA($p, q$) model for the resulting noise series. The proposed analysis procedure has four distinct steps:

1. Analyze the experiment using Method II. Estimate all effects and rank the effects based on their absolute sizes. Focus on the effects of largest absolute size that are found active or nearly active in, for example, an ANOVA.

2. Fit the transfer functions for the effects, starting with the effects that were found active and nearly active using Method II. Adjust the pure lags if appropriate. We use the model criteria described below as an aid to determine the appropriate pure lags for the factors. Study the resulting residuals from the models with the fitted transfer functions. Look for any remaining structure that is related to the remaining input variables. If no such structure is found, continue to Step 3.

3. Study the ACF and PACF of the residuals from the model from step 2. Determine the appropriate ARMA (or ARIMA) model for the noise series and then fit the overall model. Time series from the process sampled before the experiment or from stable operation in one or a few of the experimental runs can also be used to find a tentative model for the noise series.

4. Study the significance of the estimated parameters of the transfer functions in the model and make the necessary adjustments. The effects are estimated through the parameters of the different transfer functions in the final selected model. The transfer functions for non-significant effects are removed and cannot be estimated using Method V.

An effect in Method V is considered significant if the fitted parameters of its transfer function are large compared to their standard errors (the corresponding $p$-values are smaller than the chosen significance level). In steps 2-4 above, competing models are compared using model criteria such as the adjusted coefficient of determination ($R^2_{adj}$), standard deviation of the residuals, mean absolute prediction error, Akaike information criterion (AIC), and Schwarz information criterion (SIC), see Montgomery et al. (2008, pp. 57-60). Models with small standard deviation of the residuals, small mean absolute error, high adjusted coefficient of determination, and small values on the AIC and SIC are preferable. The SIC generally results in the choice of a more parsimonious model and is recommended over AIC by Montgomery et al. (2008).

5. Choice of process model for the simulations

We now return to explain how the experiments are simulated by choosing the underlying process model. The choice of the process dynamics, including the dynamics of active factors is inspired by the authors’ work with an experimental blast furnace, see Vanhatalo and Bergquist.
(2007) and Vanhatalo and Vännman (2008). The simulated response we use to exemplify our work in this article is carbon monoxide (CO) efficiency (hereafter $\eta_{\text{CO}}$); an important response for the blast furnace where higher values generally are preferred as they indicate a more energy-efficient process. Throughout this article, $\eta_{\text{CO}}$ is measured in percent units and it is assumed that a new observation on $\eta_{\text{CO}}$ is available each hour. Hourly data correspond to the sampling frequency of other important responses in the studied blast furnace, such as chemical analysis of the pig iron and the slag. Using hourly observations, the gas efficiency response, under normal operation, can be described by an ARMA(1,1) process:

$$\eta_{\text{CO}} = \delta + \phi \eta_{\text{CO},t-1} + \epsilon_t - \theta \epsilon_{t-1}$$  \hfill (15)$$

The ARMA(1,1) process is stationary if $|\phi| < 1$. The mean of a stationary ARMA(1,1) process is:

$$\mu = E(y_t) = \frac{\delta}{1-\phi}$$  \hfill (16)$$

and the variance is, see Box et al. (2008, p. 82):

$$\sigma^2_{\text{ARMA}(1,1)} = \sigma^2_{\epsilon} \frac{1 + \theta^2 - 2\phi\theta}{1-\phi^2}$$  \hfill (17)$$

Based on stable furnace operation, the following parameter values are assigned to the ARMA(1,1) model in (15): $\delta = 19.2$, $\phi = 0.6$, and $\theta = -0.28$. Hence, the model used for simulating the response under normal operation is:

$$\eta_{\text{CO}} = 19.2 + 0.6\eta_{\text{CO},t-1} + \epsilon_t + 0.28\epsilon_{t-1},$$  \hfill (18)$$

with $\sigma^2_{\epsilon} = 0.36$. This implies a process with the mean $19.2/(1-0.6) = 48$ and the standard deviation:

$$\sigma_{\text{ARMA}(1,1)} = \sqrt{0.36 \cdot \frac{1 + 0.28^2 + 2 \cdot 0.6 \cdot 0.28}{1-0.36^2}} = \sqrt{0.585} = 0.765$$  \hfill (19)$$

Figure 2 presents a simulated time series with 100 observations from the model in (18), which emulates to a process under normal, undisturbed, operation.

![Figure 2. Simulated time series with 100 observations from the ARMA(1,1) model in Eq. (18).](image-url)
5.1 Simulation of experiments using the process model

To compare the analysis methods, described in Section 4, two-level factorial experiments are used. A $2^3$ fully randomized factorial design with and without replicates is chosen. This design allows for tests of how the analysis methods perform for effects of different sizes and for multiple effects, including interaction effects.

Each simulated run of the design lasts for 48 hours, and a new observation of the response is obtained each hour. The length of the run is limited to 48 hours to avoid an unrealistically lengthy experiment, whereas shorter runs make tests of methods III and IV difficult. Even 48 hours for each run results in $2 \times 16 = 32$ days of operation for an experiment with one replication. Each simulated experiment produces a time series, $\gamma_t$, that may be divided into separate time series for each run, see Table 1.

We argue that it is reasonable to relate the size of the simulated effects to the standard deviation of the ARMA(1,1) process under normal operation. Using (9) and (19), $SN_A$ is defined as the signal-to-noise ratio for the main effect of factor $A$ in relation to the standard deviation of the process response under normal operation:

$$SN_A = \frac{2\omega_3}{(1 - \psi_A)(1 - \phi_1)\sqrt{\sigma_e^2 + \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{(1 - \phi_1^2)}}}$$

The larger $SN_A$ is, the easier it will be to detect the effect through the noise of the process.

6. Illustration and a first comparison of analysis methods

For the first sets of simulations, assume that only the main effect of factor $A$ is significant with an amplitude of two standard deviations of the process response under normal operation, that is, $SN_A = 2$. In the simulations, the size and dynamics of the effects are specified through $\omega_3$, $\psi_A$, and $b_1$. Here we keep $\psi_A$ and $b_1$ constant and vary $\omega_3$ to change the size of the simulated effect. The pure lag is arbitrarily chosen to $b_1 = 1$, mostly to illustrate its effect during the simulations. Furthermore, letting $\psi_A = 0.5$ the model will produce a response that gradually stabilizes in each run, see Figure 1.
A choice of $\omega_A = 0.153$ and $\psi_A = 0.5$ will result in the (long-term) effect (in percent units):

$$A_{\text{rel}} = \frac{2 \cdot 0.153}{(1 - 0.5)(1 - 0.6)} = 1.53$$

(21)

and the signal-to-noise ratio:

$$\text{SN}_A = \frac{2 \cdot 0.153}{(1 - 0.5)(1 - 0.6) \cdot 0.765} = 2$$

(22)

Ten randomized $2^3$ factorial experiments are simulated using $A_{\text{rel}}$ in (21). All other effects are set to 0. Each simulation uses a new randomization order of the 16 runs in the design. We are aware that ten simulations are few, but time series model building, especially Method V, is an iterative approach where the analyst must take active part in each step. Therefore, we cannot (and do not want to) automate the analysis of the simulated time series. This makes the analysis time consuming.

The proposed analysis methods are illustrated by outlining the analysis procedure for the first simulated experiment below. The illustrated experiment has the randomized run order: $\text{AbC}$, $\text{ABc}$, $\text{abc}$, $\text{ABC}$, $\text{ABc}$, $\text{aBc}$, $\text{Abc}$, $\text{aBC}$, $\text{Abc}$, $\text{AbC}$, $\text{abc}$, $\text{aBC}$, $\text{abC}$, $\text{aBc}$, $\text{abC}$, $\text{ABC}$, and the corresponding time series from the experiment is given in Figure 3.

![Figure 3. Simulated time series: $\eta = 19.2 + 0.66\theta_{c0}, + \varepsilon_i + 0.28\theta_{c1} + \varepsilon_{c1}$, with $\omega_A = 0.153$, $\psi_A = 0.5$, $b_A = 1$, $\text{SN}_A = 2$, and a run length of 48 hours.](image)

6.1 Analysis using Method I and II

Using Method I and II, the averages for each run of the time series in Figure 3 are calculated both when the transition time is included (Method I) and after it is removed (Method II). We
thus need to estimate the transition time. Here the transition time is estimated by visual
inspection of the time series and we conclude that the transition time is approximately ten
hours. Consequently, the first ten observations in each run are disregarded. Another way to
estimate the transition time is to consider those runs that produce a change of the different
factors and assign individual transition times for the runs accordingly. However, such an
approach requires the analyst to speculate about which of the effects that are active before
doing the analysis. Therefore, here the robust choice is made to eliminate the same transition
time for all runs, long enough to assume that even the slowest of effects are fully developed.

Table 2 presents the averages using Methods I and II. Table 3 presents an ANOVA table based
on the averages in Table 2. The significance level 0.05 is used in all analyses.

### Table 2. Averages for the runs of the simulated 2^3 factorial experiment in Figure 3 using Method
I (M-I) and Method II (M-II). In Method II, the first ten hours of each run are excluded.

<table>
<thead>
<tr>
<th>Std. order</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>47.59</td>
<td>47.29</td>
</tr>
<tr>
<td>abc</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>48.67</td>
<td>48.79</td>
</tr>
<tr>
<td>abc</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>47.13</td>
<td>47.38</td>
</tr>
<tr>
<td>abc</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>48.89</td>
<td>48.66</td>
</tr>
<tr>
<td>abc</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>47.31</td>
<td>47.45</td>
</tr>
<tr>
<td>abc</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>48.36</td>
<td>48.46</td>
</tr>
<tr>
<td>abc</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>47.15</td>
<td>47.16</td>
</tr>
<tr>
<td>abc</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>48.86</td>
<td>48.27</td>
</tr>
</tbody>
</table>

### Table 3. ANOVA and estimated effects for Method I (M-I) and Method II (M-II). In Method II
the first ten hours of each run is excluded.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares M-I</th>
<th>Mean square M-I</th>
<th>F value M-I</th>
<th>Prob&gt;F M-I</th>
<th>Estimated effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>8.167</td>
<td>1.167</td>
<td>24.16</td>
<td>&lt; .0001</td>
<td>1.388</td>
</tr>
<tr>
<td>A</td>
<td>7.710</td>
<td>7.710</td>
<td>159.6</td>
<td>&lt; .0001</td>
<td>1.388</td>
</tr>
<tr>
<td>B</td>
<td>0.001</td>
<td>0.001</td>
<td>0.030</td>
<td>&lt; .0001</td>
<td>0.019</td>
</tr>
<tr>
<td>C</td>
<td>0.335</td>
<td>0.335</td>
<td>6.938</td>
<td>&lt; .0001</td>
<td>0.289</td>
</tr>
<tr>
<td>AB</td>
<td>0.002</td>
<td>0.002</td>
<td>0.042</td>
<td>&lt; .0001</td>
<td>0.033</td>
</tr>
<tr>
<td>AC</td>
<td>0.001</td>
<td>0.001</td>
<td>0.023</td>
<td>&lt; .0001</td>
<td>0.017</td>
</tr>
<tr>
<td>BC</td>
<td>0.008</td>
<td>0.008</td>
<td>0.152</td>
<td>&lt; .0001</td>
<td>0.043</td>
</tr>
<tr>
<td>ABC</td>
<td>.008</td>
<td>.008</td>
<td>.152</td>
<td>&lt; .0001</td>
<td>.101</td>
</tr>
<tr>
<td>Pure error</td>
<td>.386</td>
<td>.805</td>
<td>8.048</td>
<td>.101</td>
<td></td>
</tr>
<tr>
<td>Cor. Total</td>
<td>8.553</td>
<td>9.817</td>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From Table 3, the main effect of factor $A$ is found significant for the average response
whether or not the transition time is excluded from the runs. In both cases, the main effect
of factor $C$ is significant. The estimated effect for $A$ is slightly larger when the transition time is
removed, and so is the absolute value of the $C$ effect.
6.2 Analysis using Method III and IV

Using Methods III and IV, the appropriate ARMA model is fitted to each run. Time series models are fitted both including (Method III) and excluding (Method IV) the observations from the estimated ten-hour transition time. The time series used in Method IV include the last 38 observations of each run. Thirty-eight observations are rather few to fit a time series model, but they are sufficiently many to illustrate the analysis method. The estimated mean of each run is then used as the single response in an ANOVA.

Table 4 gives the estimated parameters from the appropriate ARMA model fitted to each run using Minitab® 15 software. Table 5 presents an ANOVA table based on the estimated mean of each run.

Table 4. The estimated mean ($\hat{\mu}$), the estimated AR(1) coefficient ($\hat{\phi}_1$), and/or MA(1) coefficient ($\hat{\theta}_1$) from the best fitting ARMA model to each run for Method III (M-III) and IV (M-IV). In Method IV the first ten hours of each run are excluded.

<table>
<thead>
<tr>
<th>Std. order</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M-III</td>
<td>M-IV</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>47.44</td>
<td>.786</td>
</tr>
<tr>
<td>$\hat{\phi}_1$</td>
<td>.4876</td>
<td>-.534</td>
</tr>
<tr>
<td>$\hat{\theta}_1$</td>
<td>.4742</td>
<td>.379</td>
</tr>
<tr>
<td>$\hat{\mu}abc$</td>
<td>.4869</td>
<td>-.519</td>
</tr>
<tr>
<td>$\hat{\mu}Abc$</td>
<td>.4732</td>
<td>.627</td>
</tr>
<tr>
<td>$\hat{\mu}aBc$</td>
<td>.4838</td>
<td>.401</td>
</tr>
<tr>
<td>$\hat{\mu}abC$</td>
<td>.4697</td>
<td>.808</td>
</tr>
<tr>
<td>$\hat{\mu}A$</td>
<td>.4832</td>
<td>.740</td>
</tr>
</tbody>
</table>

Table 5. ANOVA table and the estimated effects for Method III (M-III) and Method IV (M-IV). In Method IV the first ten hours of each run are excluded.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>D.f.</th>
<th>Mean square</th>
<th>F value</th>
<th>Prob&gt;F</th>
<th>Estimated effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>M-III</td>
<td>M-IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>.0076</td>
<td>.0014</td>
<td>.0076</td>
<td>.0014</td>
<td>.289</td>
<td>.0159 .605</td>
</tr>
<tr>
<td>C</td>
<td>.244</td>
<td>.514</td>
<td>.244</td>
<td>.514</td>
<td>9.296</td>
<td>5.574 .0159 .0459</td>
</tr>
<tr>
<td>AB</td>
<td>.0947</td>
<td>.077</td>
<td>.0947</td>
<td>.077</td>
<td>3.603</td>
<td>.838 .0942 .387</td>
</tr>
<tr>
<td>AC</td>
<td>.0310^a</td>
<td>.0105</td>
<td>.0310</td>
<td>.0105</td>
<td>.0110</td>
<td>.0002 .0020 .919</td>
</tr>
<tr>
<td>BC</td>
<td>.0410^b</td>
<td>.0042</td>
<td>.0410</td>
<td>.0042</td>
<td>.0016</td>
<td>.0456 .969 .836</td>
</tr>
<tr>
<td>ABC</td>
<td>.0042</td>
<td>.146</td>
<td>.0042</td>
<td>.146</td>
<td>1.580</td>
<td>.698 .244</td>
</tr>
<tr>
<td>Pure error</td>
<td>.210</td>
<td>.738</td>
<td>.0263</td>
<td>8</td>
<td>.092</td>
<td></td>
</tr>
<tr>
<td>Cor. Total</td>
<td>7.946</td>
<td>9.763</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By comparing Table 5 and Table 3 we see that using the estimated mean from the ARMA models gives similar results as using the arithmetic run average. Again, it seems reasonable to remove the estimated transition time before fitting the appropriate ARMA models.
model, since the estimated effect gets closer to the true value of 1.53. The estimated main effect of factor $A$ for Method IV is slightly larger ($1.439 > 1.426$) than the estimated effect produced by Method II. The main effect of factor $C$ is still significant on the significance level 0.05.

6.3 Analysis using Method V

The results of using Method II show that the effects of largest absolute sizes are $A$ and $C$, and this finding is used as input to Method V according to the procedure in Section 4.5.1. Table 6 provides a comparison of different models in this model building procedure. All models were fitted using JMP® 8 software.

We start by fitting an intervention model using only $A$ and $C$. Different pure lags are then tested before studying the resulting noise series. Model b) using a pure lag of 3 for $A$ seems slightly better than the model with no pure lags in a). There is no dramatic improvement of the model criteria by lagging $A$, and it is hence possible to argue in favor for model a) due to simplicity. However, we choose a pure lag of 3 for $A$. No further improvement is found by testing different pure lags for factor $C$. Note that the residuals from model b) are autocorrelated but at this point they seem unrelated to any of the remaining main and interaction effects. An ARMA(1,1) model is found appropriate for the noise series from model b). A model including $A$, $C$, and the ARMA(1,1) model for the noise is then fitted and the significance of the parameters in the transfer functions and noise model are evaluated.

As expected, adding the ARMA(1,1) model for the noise series in model c) gives a large improvement of the model criteria. The numerator in the transfer function for $C$ in model c) is no longer significant on the 0.05 level ($p$-value 0.074). Model d), where $C$ has been dropped, is used for comparison to see if $C$ significantly improves the model. The model criteria for models c) and d) are similar, and the conclusion is therefore that factor $C$ does not improve the model and is hence not included in the final model.

Using model d) in Table 6, the estimated (long-term) effect of $A$ is calculated as:

$$A_{\text{long}} = \frac{2\hat{\omega}_{0.A}}{1-\hat{\psi}_{1.A}} = \frac{2 \cdot 0.265}{1 - 0.641} = 1.476$$

(23)

Figure 4 shows a time series plot of the fitted values of model d) versus the observed time series. The nice convergence of the fitted time series to the observed values indicates that model d) seems to follow the process’ behavior well.
Table 6. Comparison of intervention-noise models for the time series in Figure 3. The \( p \)-values for the estimated parameters are given above or below the parameter values. The arrows next to the model criteria indicate if the corresponding criterion should be large (↑) or small (↓). The models are fitted using JMP® 8.0 statistics software.

<table>
<thead>
<tr>
<th>Fitted model</th>
<th>d.f.</th>
<th>s.d.</th>
<th>MAE</th>
<th>( R^2_{\text{adj}} )</th>
<th>AIC</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>a). A, C, ( r_{AC} = 0 ) ( r_{AC} = 1 ) ( b_1 = 0 ) ( b_2 = 0 )</td>
<td>762</td>
<td>0.85</td>
<td>0.68</td>
<td>0.440</td>
<td>1926</td>
<td>1950</td>
</tr>
<tr>
<td>b). A, C, ( r_{AC} = 0 ) ( r_{AC} = 1 ) ( b_1 = 3 ) ( b_2 = 0 )</td>
<td>760</td>
<td>0.84</td>
<td>0.68</td>
<td>0.445</td>
<td>1915</td>
<td>1939</td>
</tr>
<tr>
<td>c). A, C + ARMA(1,1) for the noise ( r_{AC} = 0 ) ( r_{AC} = 1 ) ( b_1 = 3 ) ( b_2 = 0 )</td>
<td>758</td>
<td>0.57</td>
<td>0.45</td>
<td>0.744</td>
<td>1327</td>
<td>1360</td>
</tr>
<tr>
<td>d). A + ARMA(1,1) for the noise ( r_j = 0 ) ( r_j = 1 ) ( b_1 = 3 )</td>
<td>760</td>
<td>0.58</td>
<td>0.46</td>
<td>0.742</td>
<td>1331</td>
<td>1354</td>
</tr>
</tbody>
</table>

Notes: d.f. = degrees of freedom; s.d. = standard deviation of the residuals; MAE = Mean Absolute prediction Error; AIC = Akaike Information Criterion; SIC = Schwarz Information Criterion.
6.4 Results from ten simulations with SN$_A = 2$

Tables 7-8 presents the results from Methods I-V based on ten simulations using $\omega_A = 0.153$, $\psi_A = 0.5$, SN$_A = 2$. The run order is randomized for each simulation, and each run is 48 hours. The analyses are made according to the procedures illustrated above. To compare the analysis methods, we use the average estimated $A$ effect, the number of true effects found active, and the number of false effects found active.

**Table 7.** The estimated main effects of factor $A$ using Methods I-V (M-I to M-V) for the ten simulations. The mean squared error is calculated as the mean squared deviation of the estimated effects from the ‘true’ effect of 1.53.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>M-I</th>
<th>M-II</th>
<th>M-III</th>
<th>M-IV</th>
<th>M-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (the illustrated example)</td>
<td>1.388</td>
<td>1.426</td>
<td>1.317</td>
<td>1.439</td>
<td>1.476</td>
</tr>
<tr>
<td>2</td>
<td>1.740</td>
<td>1.672</td>
<td>1.750</td>
<td>1.720</td>
<td>1.776</td>
</tr>
<tr>
<td>3</td>
<td>1.347</td>
<td>1.547</td>
<td>1.247</td>
<td>1.476</td>
<td>1.598</td>
</tr>
<tr>
<td>4</td>
<td>1.577</td>
<td>1.745</td>
<td>1.443</td>
<td>1.792</td>
<td>1.772</td>
</tr>
<tr>
<td>5</td>
<td>1.366</td>
<td>1.502</td>
<td>1.313</td>
<td>1.505</td>
<td>1.495</td>
</tr>
<tr>
<td>6</td>
<td>1.406</td>
<td>1.410</td>
<td>1.425</td>
<td>1.465</td>
<td>1.484</td>
</tr>
<tr>
<td>7</td>
<td>1.276</td>
<td>1.416</td>
<td>1.159</td>
<td>1.293</td>
<td>1.421</td>
</tr>
<tr>
<td>8</td>
<td>1.187</td>
<td>1.277</td>
<td>1.211</td>
<td>1.336</td>
<td>1.274</td>
</tr>
<tr>
<td>9</td>
<td>1.409</td>
<td>1.534</td>
<td>1.302</td>
<td>1.539</td>
<td>1.539</td>
</tr>
<tr>
<td>10</td>
<td>1.359</td>
<td>1.422</td>
<td>1.261</td>
<td>1.357</td>
<td>1.428</td>
</tr>
<tr>
<td><strong>Average estimated effect</strong></td>
<td>1.406</td>
<td>1.495</td>
<td>1.343</td>
<td>1.492</td>
<td>1.528</td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
<td>0.154</td>
<td>0.137</td>
<td>0.168</td>
<td>0.160</td>
<td>0.155</td>
</tr>
<tr>
<td><strong>Mean squared error</strong></td>
<td>0.0368</td>
<td>0.0181</td>
<td>0.0603</td>
<td>0.0245</td>
<td>0.0218</td>
</tr>
</tbody>
</table>

**Figure 4.** Fitted values using model d) in Table 6 versus the observations from the simulated experiment.
Table 8. The number of active effects found for the analysis methods I-V (M I to M V). The significance level 0.05 is used during the analyses.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>M-I</th>
<th>M-II</th>
<th>M-III</th>
<th>M-IV</th>
<th>M-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (the illustrated example)</td>
<td>A, C</td>
<td>A, C</td>
<td>A, C</td>
<td>A, C</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>A, B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A, ABC</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>A, B</td>
<td>A</td>
<td>A, B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Tot. number of active effects 13 11 12 12 10

Tot. number of false effects 3 1 2 2 0

Based on these initial simulations with one active effect some tentative conclusions are drawn before doing further simulations and analyses. The results in Table 7 show that the effects tend to be underestimated by Method I and III and we conclude that the observations during the transition time should be removed before calculating averages or fitting an ARMA model to each run. We also note that similar results are obtained using the adjusted average (Method II) of each run or the estimated mean from the ARMA models (Method IV). Furthermore, the average of effect estimates from Method V is somewhat closer to the true simulated effect of 1.53 than the other methods. The analysis methods using ARMA models for each run (Methods III and IV) do not seem to produce better location effect estimates than using averages. Fitting ARMA models to each run is also more dependent on the number of observations in each run and more time-consuming, making it a less attractive method. Methods I, III, and IV are therefore excluded from further comparisons.

7. Further simulations with $A$ as the only active effect

In this section, further simulations varying the size of effect $A$ is performed to compare the performance of Methods II and V. The analysis, especially using intervention-noise models, cannot be automated and requires manual manipulations by the analyst. Hence, the number of simulations and cases are limited due to time concerns. We choose to keep the $2^3$ factorial design constant, the run length constant at 48 observations (hours), and the dynamics of the effect constant, $\psi_A = 0.5$ and $b_A = 1$. The same model for the process under normal operation is also used for all simulations, see (18).

The analysis methods are compared for different effect sizes by varying $\alpha_A$. Tables 9-12 give the results from ten simulations for the cases with $SN_A = 0$ (no effect), $SN_A = 0.5$, and $SN_A = 1$ (the results for $SN_A = 2.0$ are presented in Section 6). Method II is based on an estimate of the transition time of ten hours for all cases.
Table 9. Results from Methods II and V for ten simulations in the case with SN\textsubscript{1} = 0 (no effect). The simulated effect, \( A_{\text{no}} = 0 \), is determined by: \( \alpha = 0 \), \( \psi = 0 \), and \( b = 0 \). For Method II the first ten observations in each run are excluded. For Method V the time series are adequately modeled by an ARMA(1,1) model for all ten simulations. Therefore, effect estimates of \( A \) are not available (n.a.) using Method V. Method II provides an effect estimate for each simulation.

<table>
<thead>
<tr>
<th>Performance of analysis method</th>
<th>Method II</th>
<th>Method V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average estimated effect of ( A )</td>
<td>0.00800</td>
<td>n.a.</td>
</tr>
<tr>
<td>Standard deviation of estimated effects</td>
<td>0.149</td>
<td>n.a.</td>
</tr>
<tr>
<td>Mean squared error</td>
<td>0.0202</td>
<td>n.a.</td>
</tr>
<tr>
<td>Tot. number of effects falsely declared active</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 10. Results from Methods II and V for ten simulations in the case with SN\textsubscript{1} = 0.5. The simulated effect, \( A_{\text{0.5}} = 0.382 \), is determined by: \( \alpha = 0.0382 \), \( \psi = 0.5 \), and \( b = 1 \). For Method II the first ten observations in each run are excluded. For Method V the noise series are adequately modeled by an ARMA(1,1) model for all ten simulations.

<table>
<thead>
<tr>
<th>Performance of analysis method</th>
<th>Method II*</th>
<th>Method V*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average estimated effect of ( A )</td>
<td>0.319</td>
<td>0.374</td>
</tr>
<tr>
<td>Standard deviation of estimated effects</td>
<td>0.107</td>
<td>0.059</td>
</tr>
<tr>
<td>Mean squared error</td>
<td>0.014</td>
<td>0.0004</td>
</tr>
<tr>
<td>Tot. number of effects declared active</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Tot. number of active effects not found</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>
| Tot. number of effects falsely declared active | 2  | 0  *The performance of Method V is based on two estimated effects for \( A \), since only significant effects can be estimated. Method II provides an effect estimate for all ten simulations.

Table 11. Results from Methods II and V for ten simulations in the case with SN\textsubscript{1} = 1.0. The simulated effect, \( A_{\text{1.0}} = 0.765 \), is determined by: \( \alpha = 0.0765 \), \( \psi = 0.5 \), and \( b = 1 \). For Method II the first ten observations in each run are excluded. For Method V the noise series are adequately modeled by an ARMA(1,1) model for all ten simulations.

<table>
<thead>
<tr>
<th>Performance of analysis method</th>
<th>Method II</th>
<th>Method V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average estimated effect of ( A )</td>
<td>0.719</td>
<td>0.721</td>
</tr>
<tr>
<td>Standard deviation of estimated effects</td>
<td>0.167</td>
<td>0.188</td>
</tr>
<tr>
<td>Mean squared error</td>
<td>0.027</td>
<td>0.034</td>
</tr>
<tr>
<td>Tot. number of effects declared active</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Tot. number of active effects not found</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tot. number of effects falsely declared active</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

By studying Tables 7-11 the following can be noticed. Method II and Method V seem to produce similar average effect estimates, standard deviations of the estimated effects, and mean squared errors. However, the number of significant effects found and false active effects differ. False effects are declared active more frequently using Method II, especially for small effects. A possible explanation can be that the ARMA noise model in Method V manages to
adjust for the smaller shifts and cyclical behavior of the time series caused by the random variation in the process. For Method V in Table 10 we see that only two out of ten transfer functions for $A$ are found significant. One possible explanation is that the ARMA model may account for the added variation caused by a small effect ($SN_A = 0.5$).

8. A $2^3$ experiment with three active effects

A replicated $2^3$ factorial experiment with three active effects is simulated to compare the performance of the analysis methods in a situation with more than one active effect. Again (18) describes the behavior of the response during normal operation. This time the choices for the simulated effects are motivated by outlining a fictitious blast furnace example. The three experimental factors are given in Table 12.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Explanation</th>
<th>Low level (–)</th>
<th>High level (+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Iron ore pellet type</td>
<td>Type 1</td>
<td>Type 2</td>
</tr>
<tr>
<td>B</td>
<td>Blast volume, Nm$^3$/h</td>
<td>1600</td>
<td>1800</td>
</tr>
<tr>
<td>C</td>
<td>Moisture content of blast air, g/ Nm$^3$</td>
<td>15</td>
<td>30</td>
</tr>
</tbody>
</table>

Assume that the pellets of Type 2 are better in the sense that their performance in the blast furnace result in a higher carbon monoxide efficiency ($\eta_{CO}$). Also assume that an increased blast volume has a negative effect on $\eta_{CO}$, and at the same time increases the production rate. The moisture content of the blast air does not affect $\eta_{CO}$. We also assume that the Type 2 pellets perform better on the low level of the blast volume and that the Type 1 pellets perform better on the high level. That is, we have a positive main effect of factor $A$, a negative main effect of factor $B$, and a negative interaction effect $AB$. Table 13 gives the parameters used for the simulations.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$\omega$</th>
<th>$\psi$</th>
<th>$\beta$</th>
<th>Effect (long-term)</th>
<th>$SN_{p,e}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.0612</td>
<td>.5</td>
<td>.612</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-.2754</td>
<td>0.1</td>
<td>-1.53</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>-.0918</td>
<td>0.5</td>
<td>-.918</td>
<td>1.2</td>
<td></td>
</tr>
</tbody>
</table>

Again, ten simulated experiments are performed and the resulting time series responses are analyzed using Methods II and V. By visual inspection of the time series we estimate the transition time to ten hours and, consequently, the first ten hours of each run are disregarded using Method II.

The results from the analysis of the ten simulated experiments are given in Tables 14 and 15. Method II and Method V produce similar estimates of the effects, standard deviations, and mean squared errors for the effect estimates. Again, Method II seems to generate more false active effects. We also note that the pure lags, $(\hat{b}_A, \hat{b}_B, \text{ and } \hat{b}_{AB})$, for the transfer functions in the intervention-noise models are not consistently estimated throughout the ten simulations. A possible explanation could be that the pure lags used in the simulations are small compared to
the total run length and that a slower response can be modeled either by adding a pure lag or by increasing the denominator in the transfer function.

Table 14. The estimated effects using analysis methods II (M-II) and V (M-V) for the ten simulations. The mean squared error is calculated as the mean squared deviation of the estimated effects from the ‘true’ simulated effects. The pure lags for the estimated effects for Method V are given in brackets after each estimated effect. For Method V the noise series are adequately modeled by an ARMA(1,1) model for all ten simulations.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>M-II</th>
<th>M-V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>.463</td>
<td>-1.483</td>
</tr>
<tr>
<td>2</td>
<td>.455</td>
<td>-1.626</td>
</tr>
<tr>
<td>3</td>
<td>.653</td>
<td>-1.553</td>
</tr>
<tr>
<td>4</td>
<td>.625</td>
<td>-1.876</td>
</tr>
<tr>
<td>5</td>
<td>.653</td>
<td>-1.383</td>
</tr>
<tr>
<td>6</td>
<td>.431</td>
<td>-1.507</td>
</tr>
<tr>
<td>7</td>
<td>.604</td>
<td>-1.789</td>
</tr>
<tr>
<td>8</td>
<td>.487</td>
<td>-1.457</td>
</tr>
<tr>
<td>9</td>
<td>.890</td>
<td>-1.413</td>
</tr>
<tr>
<td>10</td>
<td>.444</td>
<td>-1.454</td>
</tr>
<tr>
<td>True effect</td>
<td>.612</td>
<td>-1.53</td>
</tr>
<tr>
<td>Average estimated effect</td>
<td>.571</td>
<td>-1.352</td>
</tr>
<tr>
<td>Standard deviation of estimated effects</td>
<td>.144</td>
<td>.163</td>
</tr>
<tr>
<td>Mean squared error</td>
<td>.0204</td>
<td>.0245</td>
</tr>
</tbody>
</table>

Table 15. The number of significant effects found using Methods II (M-II) and V (M-V). The significance level 0.05 is used during the analyses.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>M-II</th>
<th>M-V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Significant effects</td>
<td>‘False’ significant effects</td>
</tr>
<tr>
<td>1</td>
<td>A, B, AB, AC</td>
<td>AC</td>
</tr>
<tr>
<td>2</td>
<td>A, B, AB, ABC</td>
<td>ABC</td>
</tr>
<tr>
<td>3</td>
<td>A, B, AB</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>A, B, AB, BC</td>
<td>BC</td>
</tr>
<tr>
<td>5</td>
<td>A, B, AB</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>A, B, AB</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>A, B, AB</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>A, B, AB</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>A, B, AB</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>A, B, AB</td>
<td>-</td>
</tr>
<tr>
<td>Tot. number of effects</td>
<td>33</td>
<td>3</td>
</tr>
</tbody>
</table>

9. An unreplicated $2^3$ experiment

Unreplicated experiments are often used in industry to generate information at a low cost, but they lack an independent estimate of the experimental error. Analysis of unreplicated two-level factorials is traditionally made by studying a normal (or half-normal) probability plot of the effects (Daniel, 1959). In the case of a $2^3$ design, however, there are only 7 effects to plot and
determination of the reference distribution of the ‘inert’ effects is difficult. To reduce the subjectivity of the normal probability plotting technique, several formal methods to analyze unreplicated factorials have been proposed in the literature, see Hamada and Balakrishnan et al. (1998) and Chen and Kunert (2004) for a review and comparison of important methods.

Here we simulate an unreplicated $2^3$ factorial and then analyze the experiment using Method II and Method V. The background and parameters used in the replicated $2^3$ design (Section 8) are also used in this example, except for the omission of the replicate. For Method II the effects are calculated after removing the first ten observations in each run due to the transition time, and then Lenth’s method (Lenth, 1989) and the Box and Meyer method (Box and Meyer, 1986) are used for formal analysis of the estimated effects. We also use the method outlined in Bergquist et al. (2009), which builds on the Box and Meyer method.

Using Lenth’s (1989) method we define an effect as likely active if the absolute value of the effect is larger than Lenth’s margin of error (ME), and clearly active if larger than Lenth’s simultaneous margin of error (SME). Let $\hat{\epsilon}_1, \hat{\epsilon}_2, \ldots, \hat{\epsilon}_n$ be the effect estimates. Then the Lenth’s pseudo standard error (PSE) of the effects is:

$$\text{PSE} = \frac{1.5 \times \text{median}\left|\hat{\epsilon}\right|}{\left|\hat{\epsilon}_{10}\right|}$$

(24)

where $s_0 = 1.5 \times \text{median}\left|\hat{\epsilon}\right|$. Two 95 percent confidence intervals for the effects are:

$$\text{ME} = t_{.975,d} \times \text{PSE} \quad \text{and} \quad \text{SME} = t_{.95,d} \times \text{PSE},$$

(25)

respectively. In (25) $t_{.975,d}$ denotes .975th quantile of the $t$ distribution with $d$ degrees of freedom, $d = m/3$, and $\gamma = \left(1 + 95^{1/n}\right)/2$.

Box and Meyer (1986) recommend using the prior probability $\alpha = 0.2$ that an effect is active, and $k = 10$, which determines the ‘inflation factor’ for the standard deviation of an active effect, as recommended by Box and Meyer. Under the assumption that the effects $\epsilon_i$ are independent and identically distributed from the Gaussian mixture $(1-\alpha)N\left(0, \sigma^2\right) + \alpha N\left(0, k^2\sigma^2\right)$, the posterior probability, $P$, that effect $i$ is active given $\hat{\epsilon}_i$ and $\sigma$ is:

$$P\left(\text{active} \mid \hat{\epsilon}_i, \sigma\right) = \frac{\alpha}{k} \exp\left(\frac{-\hat{\epsilon}_i^2}{2k^2\sigma^2}\right) + \frac{1-\alpha}{k} \exp\left(\frac{-\hat{\epsilon}_i^2}{2\sigma^2}\right)$$

(26)

The posterior probabilities that each effect is active are calculated using a numerical integration procedure described in Bergquist et al. (2009). Effects are considered active if their posterior probability is $\geq 0.5$. We also investigate the performance of the adjusted version of the Box and Meyer method using a three-step procedure outlined in Bergquist et al. (2009). The analysis principles, frequently used in the analysis by normal probability plots, of hierarchy and heredity are included through allowing individual prior probabilities of the effects. These prior probabilities are: 0.5 for main effects, 0.3 for two-factor interactions exhibiting strong heredity
PAPER E

(both main factors are active), 0.02 for other two-factor interactions, and 0.01 for the three-factor interaction.

From Table 16, it can be concluded that the effect estimates are comparable using Method II and V. Although Method V produces an average estimated effect of $B$ somewhat further from the true simulated effect than Method II, the opposite is true for the estimates of the $A$ and $AB$ effects. Hence, there seems to be no clear difference between the methods regarding the effect estimates. The standard deviations and the mean squared errors for the estimated effects are also comparable for the two methods.

Table 16. The estimated effects using analysis methods II (M-II) and V (M-V) for the ten simulations. The mean squared error is calculated as the mean squared deviation of the estimated effects from the ‘true’ simulated effects. The pure lags for the estimated effects for Method V are given in brackets. “n.a.” means that the effect is non-significant and cannot be estimated. For Method V the noise series are adequately modeled by an ARMA(1,1) model for all ten simulations.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>M-II</th>
<th>M-V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>.388</td>
<td>-1.110</td>
</tr>
<tr>
<td>2</td>
<td>.433</td>
<td>-1.879</td>
</tr>
<tr>
<td>3</td>
<td>.581</td>
<td>-1.474</td>
</tr>
<tr>
<td>4</td>
<td>.617</td>
<td>-1.379</td>
</tr>
<tr>
<td>5</td>
<td>.423</td>
<td>-1.511</td>
</tr>
<tr>
<td>6</td>
<td>.712</td>
<td>-1.436</td>
</tr>
<tr>
<td>7</td>
<td>.508</td>
<td>-1.306</td>
</tr>
<tr>
<td>8</td>
<td>.856</td>
<td>-1.385</td>
</tr>
<tr>
<td>9</td>
<td>.644</td>
<td>-1.652</td>
</tr>
<tr>
<td>10</td>
<td>.611</td>
<td>-1.543</td>
</tr>
<tr>
<td>True effect</td>
<td>.612</td>
<td>-1.33</td>
</tr>
<tr>
<td>Average estimated effect</td>
<td>.577</td>
<td>-1.468</td>
</tr>
<tr>
<td>Standard deviation of estimated effects</td>
<td>.143</td>
<td>.296</td>
</tr>
<tr>
<td>Mean squared error</td>
<td>.020</td>
<td>.042</td>
</tr>
</tbody>
</table>

*Based on six estimated effects for $A$ and nine estimated effects for $AB$

Table 17 reveals the difference among the methods. Method II combined with Lenth’s method seems to be the most conservative, declaring only five effects larger than Lenth’s ME (likely active), $B$ in all cases. The Box and Meyer (1986) method results in several additional effects being declared active. The selected prior probability of active effects, $\alpha = 0.2$, is conservative, given that we know that three out of seven effects are active. By using the three-step procedure outlined in Bergquist et al. (2009), we find as many true effects as with Method V but also two false effects. By using Method V and multiple intervention-noise models, 25 out of the 30 simulated active effects were considered significant. No effects are falsely declared active. Method V thus appears to perform better than the other methods when there are no replicates.
Table 17. The number of significant effects found using Method II (M-II) combined with Lenth's (1989) [LE89], Box and Meyer's (1986) [BM86], Bergquist et al. (2009) [BE09] methods, and Method V (M-V). For Lenth's ME, SME, and Method V, the significance level 0.05 is used in the analyses. The posterior probabilities for the effects are given as superscripts. False effects are underlined.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>LE89-ME</th>
<th>LE89-SME</th>
<th>BM86</th>
<th>BE09</th>
<th>Significant effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.81 (B)</td>
<td>1.94 (-)</td>
<td>B50, AB66</td>
<td>A37, B39, C50, AB69</td>
<td>B, AB</td>
</tr>
<tr>
<td>2</td>
<td>1.31 (B)</td>
<td>3.14 (-)</td>
<td>B50, AB66</td>
<td>B50</td>
<td>B, AB</td>
</tr>
<tr>
<td>3</td>
<td>1.61 (-)</td>
<td>3.86 (-)</td>
<td>B51</td>
<td>B57</td>
<td>B, AB</td>
</tr>
<tr>
<td>4</td>
<td>1.09 (B)</td>
<td>2.60 (-)</td>
<td>B55</td>
<td>A53, B57, AB52</td>
<td>A, B, AB</td>
</tr>
<tr>
<td>5</td>
<td>2.37 (-)</td>
<td>5.68 (-)</td>
<td>B56, AB58</td>
<td>A57, B57, C50, AB69</td>
<td>B, AB</td>
</tr>
<tr>
<td>6</td>
<td>1.50 (-)</td>
<td>3.59 (-)</td>
<td>B56</td>
<td>A51, B53, AB50</td>
<td>A, B, AB</td>
</tr>
<tr>
<td>7</td>
<td>1.32 (-)</td>
<td>3.15 (-)</td>
<td>B54</td>
<td>A55, B55, AB53</td>
<td>A, B, AB</td>
</tr>
<tr>
<td>8</td>
<td>2.39 (-)</td>
<td>5.71 (-)</td>
<td>B50</td>
<td>A53, B53, AB72</td>
<td>A, B, AB</td>
</tr>
<tr>
<td>9</td>
<td>1.23 (B)</td>
<td>2.94 (-)</td>
<td>B56</td>
<td>A54, B54</td>
<td>A, B</td>
</tr>
<tr>
<td>10</td>
<td>.89 (B)</td>
<td>2.12 (-)</td>
<td>A52, B50, AB52</td>
<td>A52, B59, AB62</td>
<td>A, B, AB</td>
</tr>
</tbody>
</table>

| Tot. number of active effects | 5 | 0 | 15 | 27 | 25 |
| Tot. number of false effects  | 0 | 0 | 0 | 2 | 0 |

10. Conclusions and Discussion

This article outlines and compares five methods of analysis to estimate location effects for two-level factorial experiments with time series responses. These are:

- **Method I**: using averages for each run as the response in, for example, an ANOVA,
- **Method II**: using averages for each run as the response, but with the observations during an estimated transition times removed,
- **Method III**: using the estimated mean from an ARMA model fitted to each run as the response,
- **Method IV**: using the estimated mean from an ARMA model fitted to each run, but with the observations during the estimated transition time removed, and
- **Method V**: using a multiple-intervention-noise model and estimate the effects through the estimated parameters of the significant transfer functions.

The analysis methods are compared by simulations (using Matlab® scripts, available through the corresponding author) of a dynamic continuous process under the assumption that the effects only affect the mean of the process, not process dynamics or variability. In Methods III and IV, we fit an appropriate ARMA model to each run. That is, it is not assumed that most runs should follow an ARMA(1,1) process, which might be expected knowing the background of the simulation. With experience of the process, an engineer may have enough
process knowledge to assume that the dynamics of the process will not change due to the experimental factors and hence fit, for example, the same ARMA\((p,q)\) model to each run. We chose to be more general and do not make such an assumption although it would provide the possibility to further automate the analysis of the time series from each run.

Although we believe that the assumption of unchanged dynamics of the process often is valid in practice (given that basic process setups and processing steps are left unchanged), effects on the variability of the process are probably more frequent. A study of analysis methods to estimate dispersion effects from experiments with time series responses is therefore motivated.

Due to our limited number of simulations, our conclusions are tentative. The process of building intervention-noise models requires input and iterative evaluation by the analyst in the different model building steps. The analysis process is therefore difficult to automate, which is needed in a large simulation study. Time concerns also explain why we choose to keep a number of parameters constant during the simulations. This includes the choice of the \(2^3\) design, the run lengths, and the dynamics of the effects. In a future study we aim to investigate how the analysis methods perform when the run lengths become small compared to effect dynamics. Other types of response dynamics will also be tested. Only first order dynamic responses are modeled in this study, but higher order dynamic responses may occur. Method V may then perform worse, as the simple transfer function in (14) specifically emulates a first order dynamics response.

First we conclude that observations from the transition time in the beginning of each run should be removed to avoid underestimation of location effects before using the averages of each run or the estimated means from ARMA models. Consequently, the estimation of the transition time becomes important for analysis of experiments in dynamic processes. Furthermore, using the estimated average from an ARMA model for each run of the experiment does not seem to improve estimates of location effects compared to using the run average. Moreover, splitting the entire time series among the runs and then removing the transition times can result in too few observations for reliable estimations of the time series models.

Based in the initial results and the observed drawbacks, we disregard analysis methods I, III, and IV early on in the simulation study and focus on Method II and V. Fitting time series models to each run is probably a more attractive method if dispersion effects should be estimated, as they provide an unbiased estimate of process variability in the presence of autocorrelation.

We conclude that Methods II and V produce comparable effect estimates (given that a reasonable estimate of the transition time is made using Method II). However, Method II seems to produce more false active effects than Method V, especially when effects are small. For unreplicated experiments, the results indicate that Method V find more of the active effects than Method II combined with Lenth’s (1989) and Box and Meyer’s (1986) methods for analysis of unreplicated experiments. The Bergquist et al. (2009) method finds as many active effects as Method V, but also two false effects. We view this result to be of importance for
industrial experiments in, for example, continuous processes, where replication often is difficult due to cost concerns.

We find Method II, using averages adjusted for the transition time, to be a robust and rather easy way to analyze experiments with time series responses. However, an intervention-noise model constitutes a more comprehensive method that seems to produce fewer spurious effects when the effects are small and also seems to find more of the truly active effects when there are no replications made. Another advantage with intervention-noise models is that the entire time series from the experiment is used. The models also provide means to model the dynamics of the effects – something that is ignored entirely by Method II. The estimated intervention-noise model can also be used to create an estimated time series that can be compared to the original response observations (see Figure 4). We believe that comparing a fitted time series with the actual observations can be helpful during the analysis of time series responses. Method V also allows for adding more intervention variables that can be used to model the effect of known disturbances that occur during the experiment.

We are aware of that the use of time series analysis methods to analyze industrial experiments may suffer from further complications such as critical disturbances during the experiment and missing observations that break up the time series. It is probably common that time series from industrial experiments have periods with missing observations, making straightforward use of Methods II, IV and V difficult. If the missing observations cannot be re-created by, for example, interpolation, the analyst may have to use averages from each run as the response – a method shown to work quite well.

Finally, our recommendations to the analyst interested in location effects in two-level factorials with time series responses are:

- **Step 1:** Estimate the transition time and remove it from the time series of each run. Calculate the run average and use the average as the single response in an ANOVA or in a normal probability plot of the estimated effects. This is a robust way to analyze the experiment and estimate location effects.

- **Step 2:** Fit intervention-noise models to the resulting time series from the experiment and estimate the effects through the estimated transfer functions. Use the results from Step 1 and the information about the largest effects as a guide in the model building process.

**Acknowledgment**

The financial support from the Swedish mining company LKAB and the European Union, European Regional Development Fund, Produktion Botnia is gratefully acknowledged. The authors thank LKAB and the members of the LKAB ‘EBF methodology development project’, especially Gunilla Hyllander, for the contributions to the results presented here. The authors thank Dr. Murat Kulahci for valuable feedback on the work presented in this article.
References


A Bayesian Analysis of Unreplicated Two-Level Factorials Using Effects Sparsity, Hierarchy, and Heredity


Submitted for publication
A Bayesian Analysis of Unreplicated Two-Level Factorials Using Effects Sparsity, Hierarchy, and Heredity

Bjarne Bergquist¹, Erik Vanhatalo¹, and Magnus Lundberg Nordenvaad²

¹Division of Quality Technology, Environmental Management, and Social Informatics
²Department of Computer Science and Electrical Engineering

Luleå University of Technology, SE-97187, Luleå, Sweden

Correspondence to Bjarne Bergquist: E-mail: bjarne_b@ltu.se, Phone: +46 920 49 21 37

Abstract: This article studies the viability and estimates the strengths of the sparsity, heredity, and hierarchy principles using metadata. The results from the metastudy are used for prior probability assessment in a Bayesian procedure to calculate posterior probabilities of active effects for unreplicated two-level factorials. We specify individual prior probabilities for each effect based on the results from the metastudy and the posterior probabilities are then calculated in a three-step procedure where the principles of effects sparsity, hierarchy, and heredity are successively considered. We illustrate our approach by reanalyzing experiments found in the literature.

Keywords: Unreplicated factorials, Prior information, Bayesian analysis, Posterior probability of active effects, Markov chain Monte Carlo integration, Engineering judgments.

1. Introduction

Experiments are usually expensive but often the only viable way to create process knowledge. The area of Design of Experiments (DoE) was developed in the twentieth century to increase the effectiveness and efficiency of experimentation, and DoE is now, in various forms, frequently used in applications such as research, engineering and economics.

Commonly, only two levels of the factors are tested to reduce the experimental effort, but the experimental venture may be large even so. Unreplicated factorials are therefore often used to generate information at lower experimental cost and powerful analysis methods for unreplicated factorials are always sought for. This article discusses analysis methods for unreplicated two-level factorials.

Analysis of unreplicated experiments often rests on three implicit hypotheses. The first hypothesis, the effects sparsity principle, is used by almost all methods. According to the sparsity principle, only a few of the estimated effects are likely to be active. The rest of the tested main or interaction effects have no practical influence on the measured responses, and the contrasts could thus be used as estimates of the experimental noise.

The second hypothesis is that active interactions are less likely than active main factors, and the higher the order of the interaction, the less likely it is that it is active. This principle is usually called the effects hierarchy principle, and is used, for example, to plan screening experiments. Since the main purpose of screening experiments is to investigate activity of
many factors rather than to obtain precise cause and effect relations, the possibility to separate active aliased effects are often sacrificed arguing that active interactions are less common than active main effects. The effects hierarchy principle is also often used during analysis of unreplicated experiments, where higher order effects are considered less likely than, for example, main effects although they are of similar size.

The third and final hypothesis used for separation of active and inert contrasts is the effect heredity principle, which states that an interaction is more likely to be active if its parent factors are active. In this article, we refer to these three principles as the governing principles.

A standard way to analyze unreplicated two-level factorials is to study a normal probability plot of the effects. The normal probability plot lets the analyst boost analysis power of experiments lacking independent variation estimates, through use of the heredity and hierarchy principles, and the effects sparsity is the general assumption on which the analysis method rests. Normal probability plotting, or half-normal probability plotting, see Daniel (1959) and Daniel (1976), lets the analyst pinpoint outliers deviating from a distribution estimate based on the contrasts closest to zero. However, the normal probability plot is an analysis tool where the result of the analysis is highly dependent on the analytical skills and judgment of the user. Two skilled analysts could come to different conclusions, as the selection procedure includes a series of subjective classifications and considerations. The analyst must, for instance, select factors not likely active and individually weigh the hierarchical and heredity principles. According to our experience, many find analysis by normal probability plots difficult; in particular to incorporate the hierarchy and heredity principles in the analysis requires skill and experience.

More formal tests to assess the activity of effects from unreplicated factorials have been proposed in the literature. Finney (1945) proposed to use the hierarchical principle to select contrasts that a priori were deemed unlikely and to use these for error estimation, which may work well when there are such contrasts. More recent methods are based on the effects sparsity principle and start by sorting the contrasts based on their absolute sizes. Some fraction of the smallest contrasts are then used to calculate the reference distribution of inert contrasts, see, for example, Voss (1988), Lenth (1989), Berk and Picard (1991), Dong (1993), and Schneider et al. (1993). All these methods subjectively include some contrasts likely to be inactive in the reference distribution estimate; a procedure shown to have low power, see Haaland and O’Connel (1995). Other procedures based on distribution tests of all contrasts have been proposed. Venter and Steel (1996), formed a null hypothesis that all contrasts were inert and came from the same reference distribution. The normality assumption of the reference distribution was then tested and contrasts causing rejections of the null hypothesis (and thus being outliers) were classified as active. Similar approaches have been used by Le and Zamar (1992) and Sandvik-Wiklund and Bergman (1999), Hamada and Balakrishnan (1998) provide a review and comparison of the above mentioned and other methods. Effect hierarchy and heredity are seldom regarded in the methods referred to above. Hamada and Wu (1992) presented a method complementing sparsity selection with an iterative comparison of effect heredity.
We believe that formal methods, like the ones above, have the advantage that they may be automated and do not rely as heavily on the skills of the analyst. On the other hand, they usually lack the possibility for the analyst to value the likelihood of certain contrasts being active based on prior information such as the governing principles. The Bayesian approach, introduced by Box and Meyer (1986), is an exception. Box and Meyer used a prior probability, $\alpha$, that the effects are active in the experiment and a parameter, $k$, expressing the larger standard deviation generated by active effects as to that of inert effects. Using the Bayes theorem, posterior probabilities of each contrasts being active are then calculated and contrasts having a posterior probability greater than 0.5 are considered more likely active than inert.

To use the Bayesian approach, the analyst must specify model parameters before the analysis. Prior estimates of, for instance, how many times more variation that may be attributed to active factors than inert factors, and the estimated probability that a factor or interaction is active should be specified. Setting priors is a task for which the result is sensitive. The use of Bayesian algorithms by practitioners is further limited since these methods are not available in standard statistical computer packages.

In this article we look for ways to formalize the analysis procedure for unreplicated two-level experiments that can incorporate prior knowledge or knowledge about the governing principles to assess the activity of individual effects. We believe that the Bayesian approach is suitable in this respect, since it allows the incorporation of the experimenter’s prior knowledge about the activity of effects. In absence of prior knowledge about specific effects, the governing principles can also provide valuable decision support. We argue that since the Bayesian approach is easier to formalize, it can be made less sensitive to individual preferences and analysis experience than, for example, the result from an analysis using a normal probability plot. Indeed, the results will still depend heavily on sound reasoning by the engineer when deciding on, for example, the prior probabilities.

Effect hierarchy or heredity were not incorporated in the method by Box and Meyer (1986), but these principles can and have been implemented in Bayesian approaches. For example, Chipman (1996) and Chipman et al. (1997) incorporated effect hierarchy and heredity in a Bayesian variable-selection algorithm to search the model space of competing models and calculate their posterior probabilities. Their approach specifically targeted complex aliasing situations including screening, mixed level, and supersaturated designs. In order to account for the governing principles, the methodology was constructed using hierarchical priors and the stochastic search variable-selection algorithm presented by George and McCulloch (1993). Some simplifications, compared to the approach taken by Box and Meyer (1986), were made in order to achieve reasonable complexity. These include an informative Gamma distribution for the noise variance, rendering analytical posteriors, and additional design parameters that might be difficult to set for an analyst.

To limit complexity we build on the ideas in Chipman (1996) and Chipman et al. (1997) to consider hierarchy and heredity principles, however, we choose to incorporate them into the less parameterized method presented by Box and Meyer (1986).
The purpose of this paper is to first study the viability of the governing principles of sparsity, hierarchy, and heredity in experiments found in literature. Second, we extend the Box and Meyer (1986) approach for analysis of unreplicated factorials to a three-step analysis procedure that successively considers effect sparsity, hierarchy, and heredity and calculates posterior probabilities of effects being active and provide a numerical integration procedure. We use individual prior probabilities for the effects being active, which provides the possibility to incorporate prior knowledge. Results from the literature study are then used to illustrate how prior probabilities for the effects can be set up reflecting effect sparsity, hierarchy, and heredity. The approach is exemplified by analyzing experiments found in the literature.

2. Viability of the governing principles

A literature survey was performed to determine how often different types of effects were found active and if the governing principles did hold. To reduce risks of the original authors wrongly classifying inert effects as active due to confounding of aliased higher order interactions, only full factorials or reduced experiments with resolution of at least IV were selected, that is, two-factor interactions were allowed to be aliased with other two-factor effects, but not with main factor effects.

Experiments were located by database searches using search strings including “full factorial” and “Design of Experiments.” Experiments from research papers and experiments described in Design of Experiments textbooks such as Box, Hunter and Hunter (1978) and Montgomery (2005) were also selected. For simplicity, only work where either hard or electronic copies could be easily accessed were chosen. The conclusions of the original authors were not questioned if we did not consider the original statistical analysis dubious. This choice was made so that engineering competence was allowed to strengthen the analysis of the original authors. Papers with unclear or questionable analysis methods were directly disregarded or reanalyzed. The experiment by Tan and Tan (2005) was recalculated and since we came to the same conclusions as the original authors, it was added to the other experiments. In Carter Jr. et al. (1994), as well as in Pei et al. (2002), the examination of the papers suggested that these experiments had not been replicated, but that duplicate measurements had been taken. Duplicate measurements do not measure the total error associated with performing the experiment. Duplicates would have led to different conclusions regarding the activity of several effects, and these papers were excluded.

In total, this left 22 studies for the analysis of the strengths of the governing principles. In experiments where several responses were tested, all responses were included and treated as individual experiments, rendering 35 different experiments, see Table 1. For each full factorial, the number of active and inert main factors as well as two-factor interactions and three-factor interactions were counted. For reduced experiments, inert and active main factor contrasts as well as two-factor interactions were counted and the suggestions of the original authors concerning which of the aliased factors or interactions were active or inert was not questioned. Note, however, that for fractional factorials of resolution IV and V, active three-factor interactions may bias the interpretation.
Table 1. The selected experiments from the literature search. The column $y$ is the number of responses in the experiment. The column $k$ is the variance inflation factor (calculated only for unreplicated experiments) according to Box and Meyer (1986), see also Section 3 below.

<table>
<thead>
<tr>
<th>Author</th>
<th>Experiment Design</th>
<th>$y$</th>
<th>$k$</th>
<th>Author</th>
<th>Experiment Design</th>
<th>$y$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box et al. (1978) (p. 307)</td>
<td>Pilot plant experiment</td>
<td>2$^2$ $^1$ 1</td>
<td>-</td>
<td>Montgomery (2005) (p. 242)</td>
<td>Oxide thickness</td>
<td>2$^1$ 1</td>
<td>11.06</td>
</tr>
<tr>
<td>Box et al. (1978) (p. 378)</td>
<td>Reactor example</td>
<td>2$^1$ 1</td>
<td>11</td>
<td>Montgomery (2005) (p. 290)</td>
<td>IC-yield</td>
<td>2$^6$-$^1$ 1</td>
<td>23.67</td>
</tr>
<tr>
<td>Box et al. (1978) (p. 326)</td>
<td>Process development example</td>
<td>2$^1$ 1</td>
<td>13.9</td>
<td>Montgomery (2005) (p. 298)</td>
<td>Injection mould shrinkage</td>
<td>2$^6$-$^2$ 1</td>
<td>10.16</td>
</tr>
<tr>
<td>Gürses et al. (2002)</td>
<td>Electrocoagulation</td>
<td>2$^1$ 1</td>
<td>1</td>
<td>Pedersen and Ramulu (2006)</td>
<td>Cutting force experiment</td>
<td>2$^2$-$^1$ 1</td>
<td>-</td>
</tr>
<tr>
<td>Laus et al. (1997)</td>
<td>Polymerization</td>
<td>2$^1$ 1</td>
<td>5</td>
<td>Poon and Williams (1999)</td>
<td>Solder printing process</td>
<td>2$^8$-$^3$ 2</td>
<td>5.90</td>
</tr>
<tr>
<td>Lundquist et al. (2004)</td>
<td>Pulp-reinforced thermoplastics</td>
<td>2$^1$ 3</td>
<td>7.3 3.4</td>
<td>Reche et al. (2008)</td>
<td>Formaldehyde extraction</td>
<td>2$^5$-$^1$ 2</td>
<td>3.71</td>
</tr>
<tr>
<td>Montgomery (2005) (p. 218)</td>
<td>Plasma etch experiment</td>
<td>2$^1$ 1</td>
<td>-</td>
<td>Silva et al. (2003)</td>
<td>Serine protease coupling</td>
<td>2$^1$ 3</td>
<td>-</td>
</tr>
<tr>
<td>Montgomery (2005) (p. 228)</td>
<td>Filtration rate</td>
<td>2$^1$ 1</td>
<td>6.88</td>
<td>Smith et al. (1995)</td>
<td>Earth-moving systems</td>
<td>2$^1$ 2</td>
<td>-</td>
</tr>
<tr>
<td>Montgomery (2005) (p. 239)</td>
<td>Aircraft panel defects</td>
<td>2$^1$ 1</td>
<td>4.72</td>
<td>Tan and Tan (2005)</td>
<td>Epitaxial growth of Si/SiGe</td>
<td>2$^5$-$^1$ 3</td>
<td>-</td>
</tr>
</tbody>
</table>

2.1 Results of the literature survey

The studied experiments tested 637 effects in total including effects related to 160 main factors, 320 two-factor interactions as well as 100 three-factor interactions, see Table 2. From Table 2 it can be concluded that the effects sparsity principle holds since only 19 % (118 of 637) of the tested contrasts were active. Box and Meyer (1986) suggested a prior probability that an effect is active to 0.2, close to our result. However, the probability of main effects being active is much higher. Considering all experiments, the probability of a main factor being active is larger than 50 %, whereas a two-factor interaction is active in about 10 % of the cases. Consequently, many or most contrasts may be active in highly fractionated experiments such as a $2^{7-4}$ experiment. To use the sparsity of effects principle to separate active from inert factors for resolution III experiments without independent error estimates must thus be done with
care. On the other hand, our metastudy is based on experiments including both full and fractional factorials. If the full factorials are compared with the reduced ones, the probability of factors being active was higher in full factorials than in fractional factorials. In the fractional factorial case, 44% of tested main factors were active.

Table 2. Rate of active contrasts in experiments found in the literature.

<table>
<thead>
<tr>
<th>All experiments</th>
<th>Tested in total</th>
<th>Number of active effects</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Contrasts or effect</td>
<td>637</td>
<td>118</td>
<td>0.19</td>
</tr>
<tr>
<td>- Main factor</td>
<td>160</td>
<td>85</td>
<td>0.53</td>
</tr>
<tr>
<td>- Two-factor interaction</td>
<td>320</td>
<td>31</td>
<td>0.10</td>
</tr>
<tr>
<td>- Three-factor interaction</td>
<td>100</td>
<td>2</td>
<td>0.02</td>
</tr>
<tr>
<td>Only full factorials</td>
<td>Tested in total</td>
<td>Number of active effects</td>
<td>Rate</td>
</tr>
<tr>
<td>- Main factor</td>
<td>94</td>
<td>56</td>
<td>0.60</td>
</tr>
<tr>
<td>- Two-factor interaction</td>
<td>146</td>
<td>24</td>
<td>0.16</td>
</tr>
<tr>
<td>- Three-factor interaction</td>
<td>96</td>
<td>2</td>
<td>0.02</td>
</tr>
<tr>
<td>Only fractional factorials</td>
<td>Tested in total</td>
<td>Number of active effects</td>
<td>Rate</td>
</tr>
<tr>
<td>- Main factor</td>
<td>66</td>
<td>29</td>
<td>0.44</td>
</tr>
<tr>
<td>- Two-factor interaction</td>
<td>174</td>
<td>7</td>
<td>0.04</td>
</tr>
<tr>
<td>- Three-factor interaction</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

Reflecting on the results presented in Table 2, the effects hierarchy principle receives strong support. For full factorials, it was 3.4 times as likely for a main factor to be active compared to a two-factor interaction; ten times as likely if only the fractional factorials are considered. A likely explanation can be that full factorials often are run with a special purpose of modeling interactions, whereas reduced experiments may be run to detect whether any of many possible factors or interactions are active.

The effect hierarchy principle may also be used for other predictions besides what types of factors are likely to be active. In Table 3, the number of times an effect is of largest, second largest and third largest magnitude is displayed. It was more than 30 times as common for the largest effect to be related to a main factor than to an interaction. Hence, the results show that main effects are not only more frequent, but also generally of larger magnitude than interaction effects.

Table 3. Occurrence of largest effects. Note that two of the experiments did not generate active effects.

<table>
<thead>
<tr>
<th>Order of the largest effect in the experiments</th>
<th>Occurrence</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A main factor has the largest effect in the experiment</td>
<td>31</td>
<td>0.97</td>
</tr>
<tr>
<td>A two-factor interaction has the largest effect in the experiment</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td>A main factor has the 2nd largest effect in the experiment</td>
<td>24</td>
<td>0.75</td>
</tr>
<tr>
<td>A two-factor interaction has the 2nd largest effect in the experiment</td>
<td>8</td>
<td>0.25</td>
</tr>
<tr>
<td>A main factor has the 3rd largest effect in the experiment</td>
<td>14</td>
<td>0.56</td>
</tr>
<tr>
<td>A two-factor interaction has the 3rd largest effect in the experiment</td>
<td>11</td>
<td>0.44</td>
</tr>
</tbody>
</table>

The results also support the effects hierarchy principle. When all experiments are included, main factor effects were almost six times as common as two-factor interactions. If only full factorials are considered, the frequency of active main factors being active is about four times as high as the frequency of active two-factor interactions. For fractional factorials,
the frequency of active main factors is around ten times as high. Only full factorials had resolution large enough to test activity of three-factor interactions, where only two were active of the 100 tested.

The effect heredity principle also appears useful for evaluating unreplicated factorials, see Table 4.

<table>
<thead>
<tr>
<th>Type of heredity</th>
<th>Possibilities</th>
<th>No. of occurrences</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong heredity</td>
<td>79</td>
<td>26</td>
<td>0.33</td>
</tr>
<tr>
<td>Weak heredity</td>
<td>163</td>
<td>4</td>
<td>0.02</td>
</tr>
<tr>
<td>No heredity</td>
<td>78</td>
<td>1</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Of the 31 active two-factor interactions, 26 showed strong heredity implying that both main effects would have been selected as active on their own merits. One third of all two-factor interactions with strong heredity were active. Weak heredity (one of the two parent factors was active) was present in four cases out of 163 possibilities. Only once (Reche et al., 2000) was a two-factor interaction active without any of its parent factors being active. However, the two-factor interaction was small and a reanalysis of the experiment indicates a need for transformation as well as strong curvature tendencies. If the response is transformed, the size of the non-hereditary interaction effect is comparable to the inert effects.

3. Bayesian analysis of unreplicated experiments

Box and Meyer (1986) proposed a Bayesian approach to calculate posterior probabilities of active effects as an adjunct to graphical analysis. They refer to the effects sparsity principle and use the same prior probability for all effects. In what follows, we extend the Box and Meyer approach to allow individual prior probabilities for the effects.

Let \( T = (T_1, \ldots, T_v) \) be a vector of \( v \) estimated effects. Assume that that an “active” effect is distributed \( N(0, k^2\sigma^2) \) while an “inert” effect is distributed \( N(0, \sigma^2) \). That is, \( \sigma \) is the standard deviation of an inert effect and \( k \) is the “inflation factor” for the standard deviation of an active effect. To allow for individual prior probabilities for each effect we let \( \alpha = (\alpha_1, \ldots, \alpha_v) \) be a vector of prior probabilities that effects \( T_i \) to \( T_v \) are active. Under the assumption that the effects \( T_i, \ i = 1, 2, \ldots, v \), are independent and identically distributed from the Gaussian mixture \( (1-\alpha_i)N(0,\sigma^2)+\alpha_iN(0,k^2\sigma^2) \), the posterior probability that effect \( i \) is active, given \( \sigma \) and \( T_i \), is:

\[
Pr(\text{i active} \mid T_i, \sigma) = \frac{\alpha_i \left[ \frac{1}{k\sigma\sqrt{2\pi}} \exp\left(\frac{-T_i^2}{2k^2\sigma^2}\right) \right]}{\alpha_i \left[ \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-T_i^2}{2\sigma^2}\right) \right] + (1-\alpha_i) \left[ \frac{1}{k\sigma\sqrt{2\pi}} \exp\left(\frac{-T_i^2}{2k^2\sigma^2}\right) \right]}
\]

(1)

The conditioning on \( \sigma \) in (1) can be removed by marginalization, that is, by integrating (1) over the posterior distribution of \( \sigma \), \( p(\sigma \mid T) \). The distribution of \( \sigma \) given the estimated
effects $T$ is therefore required to finalize our analysis. Since the effects are assumed to be independently distributed, we obtain:

$$f(T|\sigma) = \prod_{i=1}^{n} f(T_i|\sigma)$$

Box and Meyer (1986) use a non-informative prior distribution for $\sigma$ (Jeffrey's prior):

$$f(\sigma) \propto \frac{1}{\sigma}$$

Using this approach for the situation with individual prior probabilities we get the following conditional distribution for $\sigma$ given $T$:

$$p(\sigma|T) \propto f(T|\sigma)f(\sigma) = \frac{1}{\sigma^{n+1}} \prod_{i=1}^{n} \left[ \frac{\alpha_i}{k} \exp \left( \frac{-T_i^2}{2k\sigma^2} \right) + (1-\alpha_i) \exp \left( \frac{-T_i^2}{2\sigma^2} \right) \right]$$

The posterior probability for an effect that we are interested in is then:

$$\Pr(i \text{ active} | T) = \int_0^\infty \Pr(i \text{ active} | T, \sigma)p(\sigma | T)d\sigma$$

The integral in (5) can be computed by numerical integration as recommended by Box and Meyer (1986), see also Stephenson et al. (1989). As outlined in the Appendix, we use the Markov chain Monte Carlo (MCMC) approach and the Metropolis algorithm to perform this task. MatLab® code to perform all calculations of the posterior probabilities is available by contacting the corresponding author.

### 3.1 The sparsity, hierarchy, and heredity principles in three rounds

Our proposed Bayesian approach allows us to specify individual prior probabilities for all the effects (or contrasts) in an experiment and hence to consider prior knowledge in the analysis. Here we specifically focus on the situation when we lack prior knowledge about which effects are more likely to be active. Instead, we incorporate the results from the literature study about the sparsity, hierarchy, and heredity principles. Specific prior knowledge about factors can of course easily be combined with the governing principles.

The results concerning the sparsity principle suggest that in general about 20 percent of the effects in an experiment are active. For comparative reasons, we set the prior probability of activity for factors and interactions to 0.2 in a first round of the analysis procedure, equivalent to the recommendations in Box and Meyer (1986). We also used $k = 10$, although this number can be seen as conservative. The average value of $k$ for the studies using normal probability plots during analysis in Table 1 was 7.63, and values as low as 2.83 have been used. A lower value on $k$ would increase the posterior probability of activity.

However, the results from the literature study indicate that main effects are active more often than in 20 percent of the cases. We therefore incorporated the results regarding the hierarchy principle in the second analysis round. Prior probabilities of activity selected to reflect the hierarchy principle, for example, 0.5 for the main effects, 0.1 for the two-factor
interaction effects, and 0.01 for three-factor and higher order interactions. The posterior probabilities are then calculated adjusted for hierarchy. For a two-level factorial with 15 estimated effects, the selected priors sum to an average prior probability of \( \frac{4 \times 0.5 + 6 \times 0.1 + 5 \times 0.01}{15} = 0.177 \), and for a design with 7 effects a slightly "optimistic" average prior probability of 0.259. We note here that our proposed prior probabilities should be considered as guidelines. The engineer may select appropriate prior probabilities for the effects reflecting the prior knowledge for each unique experiment but we argue that the average prior probability should be calculated in this step to consider effects sparsity.

In the third round, the prior probabilities were adjusted to reflect the heredity principle. From the posterior probabilities in the second round it is possible to determine which main effects that seem to be active (posterior probability larger than 0.5). The prior probabilities for two-factor interactions exhibiting strong heredity are then increased to 0.3, and the prior probabilities for two-factor interactions with weak or no heredity are reduced to 0.02. The posterior probabilities are then recalculated to reflect effect heredity. Notice that changing the prior probability of an effect also affects the posterior probabilities of all other effects, as the estimate of the standard deviation of random effects depends on the priors.

This three-step procedure produces posterior probabilities for all effects reflecting, in turn, effect sparsity, hierarchy, and heredity. The three posterior probabilities for each effect can then be compared when deciding on whether the effect is active or not. We argue that the three steps is a formalization of the thinking process when analyzing a normal probability plot, successively considering the three governing principles. Figure 1 gives a summary of the procedure.

Figure 1. An outline of the three-step Bayesian analysis procedure in which the sparsity, hierarchy, and heredity principles are incorporated in the prior probabilities for the effects.

### 4. Examples

To illustrate our proposed Bayesian approach using the sparsity, hierarchy, and heredity principles we choose to analyze experiments from the literature where we do not have any prior knowledge about the activity of effects.

#### 4.1 The spray coating experiment

Consider the article by Saravanan et al. (2001) which describes an unreplicated \( 2^4 \) spray coating application experiment where the effects of altered fuel ratio (A), carrier gas rate (B),
frequency of detonations (C) and spray distance (D) were measured on six responses. Here we choose to only analyze the porosity response (vol.% of the Al₂O₃ coating).

For the porosity response, the original authors concluded that factors A, B, and D were active and that all interactions including two-factor interactions were only measuring noise. Indeed, A, B, and D are the largest effects but it is unclear from the article whether the original authors had some prior knowledge that was used during the analysis. We now reanalyze the experiment assuming lack of prior knowledge and using our proposed three-step Bayesian approach.

The effects and the three sets of prior probabilities, $\alpha$, used in our analysis, as well as the posterior probabilities, $Pr$, are given in Table 5. Figure 2 shows a normal probability plot over the effects in Table 5. The posterior probabilities from the three steps in the analysis procedure can be viewed and compared in Figure 3.

From our analysis we also conclude that factors A, B, and D are active, and this is true for all rounds. Furthermore it seems likely that factor C as well as the two-factor interaction BC is active, which becomes even more prominent after considering the hierarchy principle (round 2).

![Figure 2. Normal probability plot for the effects in the experiment in Saravanan et al. (2001).](image)

Effect heredity is incorporated in round 3, and this further increases the posterior probability of BC. In addition, the posterior probability of AC increases from about 0.4 to 0.8 and the interaction may be considered active due to heredity. We conclude that it is likely that A, B, C, D, BC, and AC are active. The spray coating example illustrates how the posterior probabilities in the three-step procedure produce information about the activity of effects and the effects of the consideration of the governing principles become transparent. Furthermore it illustrates how the consideration of effects hierarchy and heredity can increase the posterior probability of main effects and interactions exhibiting strong heredity (like the AC interaction above).
Table 5. Effects, prior and posterior probabilities for the analysis of the porosity response for the experiment in Saravanam et al. (2001). \( k = 10 \). \( \alpha \) for Round 1 is 0.2 for all effects. The posterior probabilities with values of 0.5 or larger are underlined.

<table>
<thead>
<tr>
<th>Model term (effect)</th>
<th>Estimated effect</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (fuel ratio)</td>
<td>-2.048</td>
<td>0.5</td>
<td>0.5</td>
<td>0.969</td>
<td>0.992</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>B (gas rate)</td>
<td>0.929</td>
<td>0.5</td>
<td>0.5</td>
<td>0.699</td>
<td>0.946</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>C (detonations)</td>
<td>0.787</td>
<td>0.5</td>
<td>0.5</td>
<td>0.654</td>
<td>0.914</td>
<td>0.984</td>
<td></td>
</tr>
<tr>
<td>D (spray distance)</td>
<td>1.022</td>
<td>0.5</td>
<td>0.5</td>
<td>0.728</td>
<td>0.960</td>
<td>0.994</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>-0.193</td>
<td>0.1</td>
<td>0.3</td>
<td>0.049</td>
<td>0.023</td>
<td>0.102</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>-0.488</td>
<td>0.1</td>
<td>0.3</td>
<td>0.456</td>
<td>0.416</td>
<td>0.720</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>-0.0510</td>
<td>0.1</td>
<td>0.3</td>
<td>0.025</td>
<td>0.012</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>0.884</td>
<td>0.1</td>
<td>0.3</td>
<td>0.685</td>
<td>0.864</td>
<td>0.985</td>
<td></td>
</tr>
<tr>
<td>BD</td>
<td>-0.196</td>
<td>0.1</td>
<td>0.3</td>
<td>0.050</td>
<td>0.023</td>
<td>0.104</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>0.237</td>
<td>0.1</td>
<td>0.3</td>
<td>0.074</td>
<td>0.034</td>
<td>0.159</td>
<td></td>
</tr>
<tr>
<td>ABC</td>
<td>-0.103</td>
<td>0.01</td>
<td>0.01</td>
<td>0.029</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>ABD</td>
<td>0.034</td>
<td>0.01</td>
<td>0.01</td>
<td>0.025</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>ACD</td>
<td>-0.261</td>
<td>0.01</td>
<td>0.01</td>
<td>0.095</td>
<td>0.005</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>BCD</td>
<td>-0.041</td>
<td>0.01</td>
<td>0.01</td>
<td>0.025</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>ABCD</td>
<td>-0.056</td>
<td>0.01</td>
<td>0.01</td>
<td>0.026</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td>0.18</td>
<td>0.26</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. The posterior probabilities of effects after round 1, 2 and 3 for the spray coating experiment.

4.2 The leaf spring heat treatment experiment

In an experiment originally presented by Pignatiello and Ramberg (1985), leaf springs for trucks were heat treated, and the free height of the leaf spring was studied in a two-level 16 run factorial design. The five studied factors were furnace temperature in °F (B), heating time
in seconds (C), transfer time in seconds to the camber former (D), hold time in seconds for the part in the camber former (E), and quench oil temperature in °F (O). The experiment was replicated three times, and the authors assumed in their analysis that the replicates were true replicates, rather than duplicate measures, see Table 6.

Table 6. The design used in leaf spring experiment and the corresponding data.

<table>
<thead>
<tr>
<th>Furnace temp.</th>
<th>Heating time</th>
<th>Transfer time</th>
<th>Hold time</th>
<th>Oil temp.</th>
<th>Y = Free height</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>Rep. 1</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>7.78</td>
</tr>
<tr>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>8.15</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>7.50</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>7.59</td>
</tr>
<tr>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>7.94</td>
</tr>
<tr>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>8.69</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>7.56</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>7.50</td>
</tr>
<tr>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>7.88</td>
</tr>
<tr>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>7.50</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>7.63</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>7.32</td>
</tr>
<tr>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>7.56</td>
</tr>
<tr>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>7.18</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>7.81</td>
</tr>
</tbody>
</table>

The analysis of the experiment followed standard analysis of variance, see Table 7 for the sum of squares and p-values. It was concluded that the contrasts including the main effects and interactions $O$, $B$, $C$, $CO$ and $E$ were active. The $p$-value of the $BO$ interaction was comparatively low, but it was not concluded to be active and this conclusion was not further discussed by Pignatiello and Ramberg.

Assume now that only one unreplicated experiment had been performed, how likely is it that the Bayesian approach depicted here would point to the same conclusions, based on single-replicate experiment? To try to answer this question using the assumption of true replicates, one of the three replicate measurements were drawn from each run, to generate a sample of 200 single-replicate experiments drawn from the $3^6 = 43,046,721$ possible combinations, (for instance [7.78, 8.15, 7.50, 7.75, 8.00, 7.69, ... , 7.59]). Each of the 200 samples thus represented one possible outcome if the experiment had not been replicated. Contrasts were calculated for each sample, and then the posterior probabilities for each contrast were calculated, thus providing a sample of 200 posterior probability vectors. The average posterior probabilities, $\bar{P}$, and the standard deviation of the posterior probabilities, $s_p$, for round 1 to 3 are given in Table 7.
Table 7. Extract of original ANOVA analysis based on three replicates (Pignatiello and Ramberg, 2000), and posterior calculations based on 200 single replicate runs for leaf spring experiment. The contrasts considered active in the original analysis are underlined and so are the average posterior probabilities with values of 0.5 or larger in the three rounds.

<table>
<thead>
<tr>
<th>Model term (contrast)</th>
<th>Original analysis</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum of Squares</td>
<td>p-value</td>
<td>$\bar{P}$</td>
<td>$\sigma_p$</td>
</tr>
<tr>
<td>B+CDE</td>
<td>0.587</td>
<td>&lt;0.001</td>
<td>0.52</td>
<td>0.36</td>
</tr>
<tr>
<td>C+BDE</td>
<td>0.373</td>
<td>&lt;0.001</td>
<td>0.40</td>
<td>0.27</td>
</tr>
<tr>
<td>D+BCE</td>
<td>0.010</td>
<td>0.44</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>O+BCDEEO</td>
<td>0.809</td>
<td>&lt;0.001</td>
<td>0.65</td>
<td>0.24</td>
</tr>
<tr>
<td>BC+DE</td>
<td>0.064</td>
<td>0.63</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>BD+CE</td>
<td>0.065</td>
<td>0.69</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>BO+CDEO</td>
<td>0.086</td>
<td>0.03</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>CD+BE</td>
<td>0.015</td>
<td>0.35</td>
<td>0.10</td>
<td>0.12</td>
</tr>
<tr>
<td>CO+BDEO</td>
<td>0.328</td>
<td>&lt;0.001</td>
<td>0.37</td>
<td>0.24</td>
</tr>
<tr>
<td>DO+BCEO</td>
<td>0.035</td>
<td>0.16</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>E+BCD</td>
<td>0.129</td>
<td>0.01</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>BKO+DEO</td>
<td>0.001</td>
<td>0.81</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>BDO+CEO</td>
<td>0.020</td>
<td>0.26</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>CDO+BEO</td>
<td>0.027</td>
<td>0.21</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>E0+BCDO</td>
<td>0.009</td>
<td>0.47</td>
<td>0.07</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Note: $\bar{P}$ is the average posterior probability for the contrast based on 200 randomly drawn single-replicate samples from the data in Table 6. $\sigma_p$ is the standard deviation for the 200 posterior probabilities for each round.

If we assume that the original analysis of the activity of effects by Pignatiello and Ramberg was correct, that analysis can be used as a benchmark. If we were to base our decisions on the Box-Meyer approach (round 1), we would, on average conclude that O and B were the only active contrasts. Using the hierarchy principles (round 2), one would also include C. With all governing principles (round 3), CO would also on average be considered active approximately as often as not. The posterior probability of the E contrast were for seven samples (of the 200) larger than 0.5 after round 1, a figure that increased to 59 samples after round 2 and 71 samples after round 3. Note also that the average posterior probabilities of active effects increase from round 1 to 3. As an illustration of the simulation results, a 3D histogram of the posterior probabilities of the CO contrast is given in Figure 4.

The leaf spring example shows how the power of the analysis can be increased by consecutively considering the sparsity, hierarchy and heredity principles in our Bayesian approach. In fact, for four out of five contrasts in this example we did on the average arrive at the same conclusion of activity using an unreplicated experiment as the original authors did using three replicates of the design. We do not mean that the replication of the design was unnecessary, only that with a powerful analysis method we would draw approximately the same conclusions even if replication was not possible.
5. Conclusions and discussion

We found all the governing principles of effect sparsity, hierarchy, and heredity to be viable from our literature review. These principles are valuable and may improve the strength of the analysis of unreplicated experiments. The principles are not novel, but their power have seldom been discussed or validated in the literature. The paper by Li et al. (2006) is an exception where the power of the three principles are investigated using metadata, and our results largely agree with those presented by Li et al. However, a major difference is that potential three-factor interaction effects were more frequent in the study by Li et al. (4.5-9 percent active) than we found in our study. This difference may partly be explained by including experiments with indications of curvature and transformation problems in the Li et al. study – such experiments that we chose to exclude.

Indeed, there are reasons to reflect on the numbers given for rates of active factors. In the literature survey, main factor effects were more often active than inert. A built-in problem of making a literature survey is that experiments rendering no active effects are probably less likely to appear in journal articles and book examples, and thus we argue that the frequency of activity presented in this paper, as well as in the Li et al. (2006) study is probably somewhat overestimated.

If main effects would be active at such high frequency, as reported here, in highly fractionated experiments where main effects are heavily aliased, it would be difficult for the experimenter to draw conclusions. However, experimenters performing experiments with a screening purpose often add factors with less information regarding their activity than in full factorials often seen later when process knowledge is better. In fact, many full factorials selected
in the literature survey were initiated since previous experiments had indicated that some of the main factors and interactions were active. Hence, the results regarding the frequency of active main effects should probably be seen as an upper limit in a screening experiment. A trend supporting this notion can be noted in Table 2, where main effects are more seldom active in the fractional factorials than in the full factorials.

In this article we outline a Bayesian approach for analysis of unreplicated two-level factorials, which extends the Box and Meyer (1986) method to allow for individual prior probabilities of all effects. Process knowledge can therefore be considered when determining the prior probabilities for each effect. In this article we focus on the situation when process knowledge is limited and instead we illustrate how knowledge about the governing principles of effect sparsity, hierarchy, and heredity can be used to increase the power of the analysis. We use the results from our literature study to illustrate how the principles can be used to increase the power of the Box and Meyer method. Indeed, process knowledge and knowledge about the principles can be combined to further increase analysis power.

Our proposed method uses three steps where effect sparsity, hierarchy, and heredity is successively added by adjusting the prior probabilities for the effects. The posterior probabilities for activity of each effect are then calculated using MCMC integration. The advantage of performing all three steps is that the experimenter can compare the three posterior probabilities for each effect in a formal procedure that successively considers the governing principles. We agree with Box and Meyer (1986) that the Bayesian approach should be considered complementary to a normal probability plot. Together, the approaches provide a strong tool for deciding on the activity of effects of an unreplicated experiment.

Note that if the proposed three-step approach changes the prior probability for a specific effect, this also results in slight changes of the posterior probability of all other effects. For “borderline effects” (posterior probabilities around 0.5), this may push the posterior probability past 0.5, as the distribution of random effects is updated. The step-wise approach updating the prior probabilities does, therefore, raise the question of the assumption of independency of effects is stretched too far, see also the discussion in Chipman et al. (1997) concerning the choice of hierarchical priors. Since the prior probabilities of the effects in step 3 in our method are based on the posterior probabilities in step 2, the priors depend on previous results of the analysis. However, note that the selection procedure is based on the accumulated knowledge of the likelihood of the effects being active. We thus argue that the three-step approach is a better way to formally consider effect the hierarchy and heredity principles that alternatively must be considered in a one-shot analysis of a normal probability plot.

As industrial experimentation is expensive, most experimenters will not settle for conclusions such as “the effect of treatment A was not quite significant on the 5 % significance level”. According to our experience, the experimenter will usually include process knowledge in determining which factors should be considered active. If an effect corresponds to what was expected, large trust may be put classifying an effect as active although the effect is not statistically significant. However, there is a risk that experimenters go too far in this direction.
A formal elicitation method like the one we propose is a way to make engineering reasoning and analysis results more transparent for unreplicated experiments.

Bayesian analysis methods provide excellent possibilities to incorporate prior knowledge of different kinds for decision-making, but currently require mathematical and statistical knowledge above that of the average user of designed experiments. To facilitate the use of Bayesian analysis methods for unreplicated factorials, the methods probably need to be made available in common statistical analysis software.

Acknowledgement

We gratefully acknowledge the financial support from the Swedish mining company LKAB, as well as the County Administrative Board under grant 303-02863-2008, and the Regional Development Fund of the European Union, grant 43206 which made this research possible.

About the authors

Bjarne Bergquist is Associate Professor of Quality Management and head of the Division of Quality Technology, Environmental Management and Social Informatics at Luleå University of Technology (LTU), Sweden. He holds a M.Sc. degree in Mechanical Engineering from LTU (1991) and a Ph.D. in Materials Science from Linköping University, Sweden (1999) where he performed experiments on batch and continuous processes in the powder metallurgical industry. He also has a background as a process engineer, doing experiments in a paper mill. His main research interest is focused on process control and experimental design, especially for continuous process applications. He is a member of the European Network for Business and Industrial Statistics (ENBIS).

Erik Vanhatalo is a Ph.D. student in the subject of Quality Technology and Management at LTU and holds a Licentiate degree of Engineering in the subject of Quality Technology and management (2007) and a M.Sc. degree in Industrial and Management Engineering (2004) from LTU. His current research is focused on the use of experimental design and multivariate statistical methods especially in continuous processes. He is a member of ENBIS.

Magnus Lundberg Nordenvaad is a Senior Lecturer at the Department of Computer Science and Electrical Engineering, LTU, and holds a senior research position at the Swedish Defence Research Agency. He holds a Ph.D. degree in Signal Processing from the School of Electrical and Computer Engineering, Chalmers University of Technology, Gothenburg, Sweden (2003) and a M.Sc. degree in Computer Science and Engineering from LTU (1998). He has held visiting positions at Purdue University, Colorado State University, and the University of Florida. His research interests primarily lie in statistical signal processing and how it applies to digital communications, radar, sonar, navigation, process diagnostics, land-mine detection, and high-level power estimation in CMOS architectures.

References:


Appendix: calculating posterior probabilities using Markov chain Monte Carlo (MCMC) integration

To remove the conditioning on $\sigma$ in Eq. (1) we need to integrate over the posterior distribution of $\sigma^*$, $p(\sigma | T)$. We use MCMC integration where $N$ samples $\sigma_{n^*}$, $n=1,\ldots,N$, are drawn from $p(\sigma | T)$. Then Eq. (5) can be approximated through:

$$\Pr\{i \text{ active} \mid T\} \approx \frac{1}{N} \sum_{n=1}^{N} \Pr\{i \text{ active} \mid T, \sigma_{n^*}\}$$ (A.1)

The samples, $\sigma_{n^*}$, are generated by creating a Markov Chain with stationary distribution $p(\sigma | T)$. We apply the Metropolis algorithm (with a symmetric jumping distribution) which in turn is a special case of the more general Metropolis-Hastings algorithm, see, for example, Gelman et al. (2004):

- initialize $\sigma_1$ (see below)
- for $n = 2 \ldots N$
  - propose an update of $\sigma_{n^*}$, $\sigma'$, by adding a symmetrically distributed random variable (we use a normally distributed variable), that is,
    - $\sigma' = \sigma_{n-1} + \tau$, where $\tau \in N(0, \gamma)$.
    - If $\sigma' > 0$, calculate $q = p(\sigma' | T) / p(\sigma_{n-1} | T)$, else $q = 0$. Note that since $q$ is a ratio, it is sufficient to know $p(\sigma | T)$ up to proportionality in Eq. (4).
  - Draw a uniformly distributed random variable $a$ between zero and one, that is, $a \in U(0,1)$.
    - If $a < q$, (keep the new sample $\sigma'$)
      - $\sigma_n = \sigma'$
    - else (keep the old sample $\sigma_{n-1}$)
      - $\sigma_n = \sigma_{n-1}$
- end.

The procedure will converge under fairly weak conditions, see Gelman et al. (2004). Furthermore, Gelman et al. (2004) recommends an acceptance rate of 0.44 for new samples, $\sigma'$, for a one-dimensional problem. To achieve the wanted acceptance rate we first calculate a reasonable starting value for $\sigma_1$ (or $\phi$ after reparameterization, see below). We also continuously adjust the standard deviation, $\gamma$, of the symmetrical jumping variable, $\tau$, to avoid the algorithm to get caught in, for example, heavy tails of the posterior distribution of $\sigma$.

To obtain a reasonable start value $\sigma_1$, some of the effects with the smallest absolute values are selected. $\sigma_1$ is then calculated as the standard deviation of these effects. The effects are sorted based on their absolute value and then the smallest half (rounded down to the nearest integer) are selected and the standard deviation of the effects are calculated. For example, the 3, 7, and 15 effects of smallest absolute value are selected for the cases with 7, 15, and 31 effects (contrasts) respectively. Furthermore, an initial setting of $\gamma$ is required, here chosen as $0.2 \times \sigma_1$. After every 100 samples of the Metropolis algorithm, the acceptance rate is
adjusted. If the acceptance rate is smaller (larger) than 0.44, \( \gamma \) is decreased (increased) with 5 percent. The acceptance rate is recalculated after another 100 samples. This procedure automatically calibrates the standard deviation of the symmetrical jumping variable \( \gamma \). We have also found that a burn-in period of 1,000 samples before starting to sum the posterior probabilities in (A.1) is useful to reduce the possible bias from the starting values of \( \sigma_i \) and \( \gamma \).

The total number of samples, \( N \), required for stable approximations of the posterior probabilities varies among examples we have tested. Using \( N = 100,000 \) has produced posterior probability estimates stable down to the third decimals of the posterior probabilities. Using these settings, the calculation time for each round in our proposed method is a few seconds on a PC with a 1.7 GHz processor. Higher precision is achieved by increasing \( N \).

For some parameter settings, the distribution in (4) becomes challenging to integrate. Particular concerns include near-singularities and heavy tails. In these cases, a large number of samples, \( N \), are generally required to ensure proper convergence of (A.1). To limit these effects, the problem was reparameterized using:

\[
\omega = \frac{1}{\sigma}
\]

Hence, using this variable change in Eq. (5), (1), and (4) we now have:

\[
\Pr(i \text{ active } | T) = \int_{0}^{\infty} \Pr(i \text{ active } | T, \omega) p(\omega | T) d\omega
\]

\[
\Pr(i \text{ active } | T, \omega) = \frac{\alpha \left[ \omega \exp\left(\frac{-T_i^2 \omega^2}{2k^2}\right)\right]}{\alpha \left[ \frac{\omega}{k\sqrt{2\pi}} \exp\left(\frac{-T_i^2 \omega^2}{2k^2}\right)\right] + (1 - \alpha) \left[ \frac{\omega}{\sqrt{2\pi}} \exp\left(\frac{-T_i^2 \omega^2}{2}\right) \right]}
\]

\[
p(\omega | T) \propto \omega^{x+y-1} \prod_{i=1}^{\nu} \left[ \frac{\alpha}{k} \exp\left(\frac{-T_i^2 \omega^2}{2k^2}\right) + (1 - \alpha) \exp\left(\frac{-T_i^2 \omega^2}{2}\right) \right]
\]

We can now calculate the posterior probability for the effects by generating samples from \( p(\omega | T) \) with the Metropolis algorithm and approximate the integral in (5) by:

\[
\Pr(i \text{ active } | T) \approx \frac{1}{N} \sum_{i=1}^{N} \Pr(i \text{ active } | T, \omega_i)
\]

However, for cases with a large number of effects (large \( \nu \)) we encounter another difficulty. The product of a large number of expressions of the kind \( a \exp(x) + b \exp(y) \) in (A.5) may be small and cause numerical problems, especially when calculating the ratio, \( q_i \), in the Metropolis algorithm. To solve this problem, \( p(\omega | T) \) was rewritten to:
\[ p(\omega \mid \mathbf{T}) \propto \omega^{r-1} \exp \left\{ \log \left( \prod_{i=1}^{n} \left[ \frac{\omega_{i}}{k} \exp \left( \frac{-T_{i}^2 \omega_{i}^2}{2k^2} \right) + (1 - \alpha_{i}) \exp \left( \frac{-T_{i}^2 \omega_{i}^2}{2} \right) \right] \right) \right\} \]  \hspace{1cm} (A.7)  

\[ \propto \omega^{r-1} \exp \left\{ \sum_{i=1}^{n} \log \left[ \frac{\omega_{i}}{k} \exp \left( \frac{-T_{i}^2 \omega_{i}^2}{2k^2} \right) + (1 - \alpha_{i}) \exp \left( \frac{-T_{i}^2 \omega_{i}^2}{2} \right) \right] \right\}. \]  

We then use

\[ l_{1}(i, \omega) = \log\left( \alpha_{i}/k \right) \frac{-T_{i}^2 \omega_{i}^2}{2k^2} \quad \text{and} \quad l_{2}(i, \omega) = \log\left( 1 - \alpha_{i} \right) \frac{-T_{i}^2 \omega_{i}^2}{2} \]  \hspace{1cm} (A.8)  

We can now write

\[ p(\omega \mid \mathbf{T}) \propto \omega^{r-1} \exp \left\{ \sum_{i=1}^{n} \log \left( \exp(l_{1}(i, \omega)) + \exp(l_{2}(i, \omega)) \right) \right\} \]  \hspace{1cm} (A.10)  

Formula A.10 does not solve the problem completely but allows us to use the following equality:

\[ \log(\exp(x) + \exp(y)) = \max \{x, y\} + \log(1 + \exp(-|x - y|)) = g(x, y) \]  \hspace{1cm} (A.11)  

We can use this when calculating the ratio, \( q \), in the Metropolis algorithm and thereby create a robust implementation. That is,

\[ \frac{p(\omega^{\prime} \mid \mathbf{T})}{p(\omega_{\text{old}} \mid \mathbf{T})} = \frac{\omega^{r-1}}{\omega_{\text{old}}^{r-1}} \exp \left\{ \sum_{i=1}^{n} \left[ g(l_{1}(i, \omega), l_{1}(i, \omega^{\prime})) - g(l_{1}(i, \omega_{\text{old}}), l_{1}(i, \omega_{\text{old}})) \right] \right\} \]  \hspace{1cm} (A.12)