INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI
800-521-0600
PRODUCT AND PROCESS IMPROVEMENT

USING LATENT VARIABLE METHODS

By

CHRISTIANE M. JAECKLE. DIPL. ING.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

© Copyright by Christiane M. Jaekle, May 1998
PRODUCT & PROCESS IMPROVEMENT
USING LATENT VARIABLE METHODS
Doctor of Philosophy (1998)
(Chemical Engineering)
McMaster University
Hamilton, Ontario

TITLE: Product and Process Improvement Using Latent Variable Methods

AUTHOR: Christiane M. Jaeckle, Dipl. Ing. (Universität Stuttgart, Germany)

SUPERVISOR: Professor John F. MacGregor

NUMBER OF PAGES: xiv, 173
Abstract

This thesis considers the utilization of historical process data for three different process engineering problems. Latent variable methods such as Principal Components Analysis (PCA) and Partial Least Squares (PLS) are shown to be key tools for dealing with the highly correlated variables typically found in process operating data and for ensuring feasibility of the solutions. The restrictions and limitations encountered by any databased approach are recognized and discussed.

The problem of designing process conditions that yield a new product grade quality within the range of already existing grades is addressed by latent variable regression models and their inversion. Latent variable techniques allow for models that not only describe the relationship between process conditions and quality variables but also the covariance structure within the process conditions. This leads to the design of new conditions that are consistent with the plant operations from the past. Feasibility issues of both the new quality specifications as well as of the predicted operating conditions are addressed. The approach is illustrated on a simulated process and three industrial processes.

The related topic of moving the production of a particular product grade from a plant A to another plant B when both plants have already produced a similar range of grades is treated as well. Since the two plants may differ in size, configuration etc. the process conditions for one grade may be very different in the two plants. A latent variable method is proposed which uses data from both plants to predict
process conditions for plant B for a grade previously produced only in plant A. The issue of feasibility is again addressed for both product quality and process conditions.

The last part of the thesis explores the problematic issue of utilizing large process operating databases for process performance improvement. The major problem found with using normal operating data for this purpose is correlation among manipulated variables and disturbances as caused by feedback operation or other operating strategies. Such correlation is shown to impede the extraction of causal information from the database, which is necessary in order to infer better process conditions. This implies that only in exceptional circumstances would it be feasible to use empirical databased methods for process optimization.
Acknowledgements

First of all, I would like to thank my parents Friederike and Dr. Hans-Günther Jaeckle whose continuing love and support made this thesis possible. I will never be able to thank them enough.

I would like to express my deepest gratitude to my supervisor Dr. John F. MacGregor for his kind guidance, inspiration and support throughout the course of this study. He has provided encouragement and valuable advice whenever needed. It has been a pleasure working with him for these few years.

I would also like to thank the members of my supervisory committee Dr. Andrew N. Hrymak and Dr. Román Viveros-Aguilera as well as Dr. Cameron M. Crowe and Dr. Alison J. Burnham for helpful discussions and advice.

I am grateful to Dr. Martin Navratil, Rebecca L. Durney Cronin, James J. McCrystal, and Ira Gans (all from Reichhold Chemicals) and Dr. Krishna Padayachee from 3M Canada for their help in this work.

The people in the process control group and the graduate students of the Department of Chemical Engineering have my special thanks. They have provided a friendly and enjoyable working atmosphere.

Finally, I would like to express my thanks to Fernando for his support and patience over the last two years.
# Contents

Abstract iii

Acknowledgements v

1 Introduction 1

2 Product Design through Analysis of Process Data 5
  2.1 Introduction 5
  2.2 Methodology 11
    2.2.1 Selection of Quality Variables 12
    2.2.2 Standard Regression Model and Its Generalized Inverse 14
    2.2.3 PCR-Model and Its Generalized Inverse 16
    2.2.4 Other Multivariate Latent Variable Methods 24
    2.2.5 Conditional Expectation Approach 25
  2.3 A Common Framework 27
  2.4 Example: Simulated LDPE Polymerization Process 32
    2.4.1 The Process 33
    2.4.2 Generation of Data 33
    2.4.3 Case Studies 34
  2.5 Extension to Nonlinear Systems 46
    2.5.1 Nonlinear PLS and Inversion 47
3 Industrial Applications

3.1 A Batch Solution Polymerization Process

3.1.1 The Process

3.1.2 The Data

3.1.3 Preliminary Data Analysis

3.1.4 Mahalanobis Distances Between Nominal and Predicted Process Conditions

3.1.5 Cross-Validation as a Test for Predictability

3.1.6 Model Inversions and Results

3.2 A Continuous Process

3.2.1 The Process

3.2.2 The Data

3.2.3 Preliminary Data Analysis

3.2.4 Model Inversion and Results

3.3 A Semi-Batch Emulsion Polymerization Process

3.3.1 The Process

3.3.2 The Data

3.3.3 Preliminary Data Analysis

3.3.4 Model Inversions and Results

3.4 Conclusions

4 Product Transfer from One Plant to a Second Plant

4.1 Introduction

4.2 Data Structure
4.3 Methodology ................................................. 119
  4.3.1 Selection of Grades .................................... 119
  4.3.2 Scaling of the Data .................................... 122
  4.3.3 Approach Using an Extended PCR-Model .......... 123
4.4 Simulation Example ........................................ 129
4.5 Discussion and Conclusions ............................... 134

5 Process Improvement Using Normal Operating Databases 137
  5.1 Introduction .............................................. 137
  5.2 Similarity Optimization .................................. 143
    5.2.1 An Idea for Reconstruction of Unmeasured Disturbances and
          Search for Improved Operating Conditions ............ 147
    5.2.2 Discussion .......................................... 152
  5.3 Conclusions .............................................. 156

6 Summary and Conclusions .................................. 160

A Construction of the “Null Component” $x_{null}^T$ for PCR-Inversion 165

B Data for the LDPE Example .................................. 167

References ....................................................... 169
List of Tables

2.1 Common Framework for the Methods: \( x_{new}^{T} = y_{des}^{T} \cdot M_{Method}^{T} \) .......... 28
2.2 Percentage of Variation Explained by PCR in \( X \) and \( Y \) ................. 35
2.3 Percentage of Variation Explained by NPLS in \( X \) and \( Y \) (Case 1) .... 53
2.4 Inner Relationships for the First 3 Principal Components (Case 1) .... 53
2.5 Mahalanobis Distances to Desired Grades Based on the Specified Y-Variables (Case 1): \( (y_{des}^{T} - y_{ach}^{T}) \cdot \left[ Cov(Y) \right]^{-1} \cdot (y_{des}^{T} - y_{ach}^{T})^{T} \) ........... 55
2.6 Percentage of Variation Explained by NPLS in \( X \) and \( Y \) (Case 2) ..... 57
2.7 Inner Relationships for the First 3 Principal Components (Case 2) .... 57
2.8 Mahalanobis Distances to Desired Grades Based on the Specified Y-Variables (Case 2): \( (y_{des}^{T} - y_{ach}^{T}) \cdot \left[ Cov(Y) \right]^{-1} \cdot (y_{des}^{T} - y_{ach}^{T})^{T} \) ........... 60
2.9 Mahalanobis Distances to Desired Grades Based on the Specified Y-Variables and Conv (Case 2): \( (y_{des-ext}^{T} - y_{ach-ext}^{T}) \cdot \left[ Cov(Y_{ext}) \right]^{-1} \cdot (y_{des-ext}^{T} - y_{ach-ext}^{T})^{T} \) .......... 62
3.1 Percentage of Variation Explained by PLS and PCR in \( X \) and \( Y \) ... 73
3.2 Mahalanobis Distances in the Model Space between Grades Used in the Model, Closest Prediction, and the Nominal \( x^{T} \) ................. 86
3.3 14 Variables for the Nominal Process Conditions \( x^{T} \) and Closest Predictions \( x_{closest}^{T} \) ............................................. 88
3.4 14 Variables for the Nominal Process Conditions $x^T$ and Predictions Along the Line of the $x_{null}^T$-Space for Grade 2 ......................... 89
3.5 Percentage of Variation Explained by PLS in $X$ and $Y$ Using all 6 Products ................................................................................. 92
3.6 Percentage of Variation Explained by PCA in $Y$ Using Products 1-5 . 95
3.7 Percentage of Variation Explained by PLS and PCR in $X$ and $Y = [Y_2 \ Y_4]$ Using Products 1-5 .......................................................... 95
3.8 Mahalanobis Distances in the Model Space Between Grades 1 to 5 Used in the Model. Closest Prediction, and the Actual $x^T$ of Grade 6 .... 96
3.9 Percentage of Variation Explained by PLS in $X$ and $Y$ Using all 5 Products and all 6 $Y$-Variables .......................................................... 101
3.10 Percentage of Variation Explained by PCA on all 6 $Y$-Variables ... 101
3.11 Percentage of Variation Explained by PLS and PCR in $X$ and $Y = [el \ R]$ Using Products 1, 2, 4, and 5 ................................................. 105
3.12 Mahalanobis Distances in the Model Space from Grades 1, 2, 4, 5, and Closest Prediction to the Actual $x^T$ of Grade 6 ......................... 105
3.13 $X$-Values of the 4 Model Grades, the True Grade 3, and the Predicted Grade 3 .......................................................... 107
3.14 Percentage of Variation Explained by PLS and PCR in $X$ and $Y = [el \ PS \ R]$ Using Products 1-5 ......................................................... 109
3.15 $X$-Values of the 5 Model Grades, and the Predicted Grade 3 for the Large Reactor .......................................................... 109

4.1 Percentage of Variation Explained by PCA in $X_B$, $Y_B$ and $Y_{\text{Ansat}}$ .. 130
4.2 Percentage of Variation Explained by the Extended PCR-Model in $Y_B$ Using the First Four Latent Variables in $X_B$ ......................... 131

5.1 Four Scenarios for the Example $q = MV + D$ ......................... 145
5.2 Measured Variables in a Typical Process Operating Database . . . . . . 148

B.1 Manipulated Variable Values $(X')$ for the 15 LDPE Grades . . . . . . . 167
B.2 Quality Variable Values $(Y')$ for the 15 LDPE Grades . . . . . . . . . . . . 168
List of Figures

2.1 Quality Data on 9 Grades Produced in the Past (○) and 6 New Desired Grades (*) .................................................. 6
2.2 Linear Low Density Polyethylene (LLDPE) Products .......................... 8
2.3 The Different Spaces for \( \tilde{x}_{new}^T \) and \( x_{null}^T \) when \( m = 3, A = 2 \) and \( k = 1 \) ........... 22
2.4 PCR- and OLS-Results for Number and Weight Average Molecular Weight, when \( Y = [\text{Conv MWn MWw}] \) .......................... 36
2.5 PCR- and OLS-Results for Conversion and Weight Average Molecular Weight, when \( Y = [\text{Conv MWn MWw}] \) .......................... 36
2.6 PCR- and OLS-Results for Long and Short Chain Branching, when \( Y = [\text{Conv MWn MWw}] \) .......................... 37
2.7 PCR-Inversion for \( Y_2 = [\text{MWn MWw}] \): PCR Gives Different Process Conditions for Each Grade ............................. 39
2.8 PCR-Results for Number and Weight Average Molecular Weight, when \( Y = [\text{MWn MWw}] \) ....................................................... 40
2.9 PCR-Results for Long and Short Chain Branching, when \( Y = [\text{MWn MWw}] \) ....................................................... 40
2.10 Conditional Expectation and PCR Results when \( Y = [\text{MWn MWw}] \) .......... 42
2.11 Process Operating Conditions from OLS- and PCR-Inversion
when \( Y = [\text{MWn MWw}] \) ....................................................... 43
2.12 Process Conditions from OLS- and PCR-Inversion; (OLS-values are lumped together at approximately \( P = 2750 \text{ atm}, \ Tin = 460 \text{ K} \))

2.13 NPLS and PCR Results for Number and Weight Average Molecular Weight when \( Y = [\text{Conv MWh MWh}] \)

2.14 NPLS and PCR Results for Number Average Molecular Weight and Conversion when \( Y = [\text{Conv MWh MWh}] \)

2.15 Inner Relationship Between \( X \)- and \( Y \)-Scores for the First Latent NPLS-Dimension

2.16 \( u \)-Values of Different Predictions for Grades 10-15, the Actual \( u \)-Values, and the \( u \)-Values of the Historical Grades Used for Modeling

2.17 NPLS and PCR Results for Number and Weight Average Molecular Weight when \( Y = [\text{MWh MWh}] \)

2.18 NPLS and PCR Results for Number Average Molecular Weight and Conversion when \( Y = [\text{MWh MWh}] \)

3.1 \( t \)-Scores from PCA on \( Y \): Grades Included in Model (2-8) and the New Grade (1)

3.2 \( t \)-Scores from PCA on \( Y \): Grades Included in Model (1-7) and the New Grade (8)

3.3 \( t \)-Scores of the PLS-Model on all 8 Grades

3.4 \( t \)-Scores from PLS on all 6 Products: \( t_{1-t_2} \)

3.5 \( t \)-Scores from PLS on all 6 Products: \( t_{1-t_3} \)

3.6 \( t \)-Scores from PLS on Products 1–5, Corresponding Scores from Product 6, and Closest Prediction: \( t_{1-t_2} \)

3.7 \( t \)-Scores from PLS on Products 1–5, Corresponding Scores from Product 6, and Closest Prediction: \( t_{1-t_3} \)

3.8 \( t \)-Scores from PLS on all 5 Products and 6 \( Y \)'s: \( t_{1-t_2} \)

3.9 \( t \)-Scores from PLS on all 5 Products and 6 \( Y \)'s: \( t_{2-t_3} \)
3.10 $t$-Scores from PCA on $Y = [el \ R]$; Grades 1, 2, 4, 5 and the Prediction for the New Grade 3 ................................................................. 104
3.11 $t$-Scores from PLS on Grades 1, 2, 4, 5 ($\times$), Scores for True Conditions ($\ast$), and the Prediction for the New Grade 3 ($\circ$); $Y = [el \ R]$ .................. 106
3.12 $t$-Scores from PCA on $Y = [el \ PS \ R]$; Grades 1-5 and the Prediction for Grade 3 in the Large Reactor .................................................. 108
3.13 $t_1$-$t_2$-Scores from PLS on Grades 1-5 ($\times$), and the Prediction for the New Grade 3 in the Large Reactor ($\circ$); $Y = [el \ PS \ R]$ ................. 110
3.14 $t_2$-$t_3$-Scores from PLS on Grades 1-5 ($\times$), and the Prediction for the New Grade 3 in the Large Reactor ($\circ$); $Y = [el \ PS \ R]$ ................. 110

4.1 Data Structure for Process Conditions and Resulting Quality in Two Plants A and B: Both Produce Grades Within the Same Product Family117
4.2 Squared Prediction Errors for 12 Plant B Grades ($\ast$) and 9 Plant A Grades ($\circ$): the first two grades from both plants are common .......... 131
4.3 Contributions of the $Y'$-Variables to the SPE of Grade 9 in $Y_A$ .......... 132
4.4 Qualities for $Y_B$ Grades. Desired $Y_A$ Grades, and $Y_A$ Grades Achieved in Plant B; $Y_3$-$Y_4$ ................................................................. 133
4.5 Qualities for $Y_B$ Grades. Desired $Y_A$ Grades, and $Y_A$ Grades Achieved in Plant B; $Y_3$-$Y_5$ ................................................................. 133
Chapter 1

Introduction

As computerization becomes more and more common in industry, large databases containing process operating data are readily available on many industrial processes. These are potentially very valuable resources to draw from in solving process engineering problems. In recent years much interest has been shown in using information derived from historical data for applications such as process monitoring and diagnosis, and a lot of work has been done in this area (see for e.g. Kresta et al. (1991), Nomikos and MacGregor (1994), Kourtì and MacGregor (1995)). It would be of great value if historical process data could be used also for other applications such as the design of new products, and process optimization.

In an attempt to make use of historical process operating databases this thesis deals with the following three topics:

- design of a new product grade with a specified set of quality characteristics when data on several grades of the product are already available

- transfer of the production of a product grade from one plant A to another plant B
• investigating the potential use of historical operating data for process improvement or optimization

The common theme throughout these three topics is the problem of finding new process conditions for a particular purpose. Most often empirical models are developed to predict a process outcome, e.g. product quality from given process conditions. In this thesis, however, the problems require the reverse path. Given are certain quality specifications or the desire for quality improvement, and it is the process conditions leading to such a result that need to be found.

These problems could be dealt with very efficiently if good fundamental process models were available. However, such models often do not exist or are costly and time consuming to develop. If good fundamental models are lacking, process experiments can be designed to gain information on how the process is working. Data derived from such experiments would be better suited for use in treating the above problems than the normal plant operating data. Nevertheless, it might be beneficial to try and exploit the information readily available in the plant data before one resorts to more involved studies or a full design of experiments. The results and insights gained from the historical operating data may provide potential solutions or serve as a good starting point and provide additional knowledge for a design of experiments.

Of course any databased approach faces certain limitations due to the range and the structure of the data. An empirical approach cannot and should not attempt to predict process conditions which differ in structure or lie outside the range of the process conditions present in the data. Although process conditions that differ from the conditions used in the past may exist and could possibly provide a good solution to the problem, they cannot be obtained from an approach relying solely on historical data. It is therefore important that the databased methodology restricts itself to the operating strategies that have been used in the past and are present in the available data set.
CHAPTER 1. INTRODUCTION

The structure among the process data in a particular data set arises from the fact that during normal operation process variables usually do not vary independently from each other but are rather highly correlated or even collinear. Typically the process is moving only in a low dimensional subspace of the space spanned by all process variables. This can often be seen using simple data analysis techniques on the process data. Databased approaches to the above problems should thus be able to reduce the dimensionality of the problem, and work within the lower dimensional subspace in which the operations from the past have taken place. For this reason latent variable regression methods seem a promising tool in this context. They summarize the physical process variables by projection into a much smaller number of latent variables that span the subspace in which the process has operated in the past.

Chapter 2 treats the problem of product design. It is assumed that data on several grades in the region of the new desired grade quality already exist. The design of process conditions for the new grade is performed by the inversion of a latent variable regression model. Variations on the approach using different models are developed. The use of standard regression models is also investigated and compared to the latent variable regression methods. An important issue that is dealt with in this context is the feasibility of any such process conditions obtained from such an approach. A simulated example of low density polyethylene illustrates the methodology. In the second part of the chapter the approach is extended to the use of nonlinear models. The same example is employed for illustration.

Chapter 3 contains three industrial applications of the methodology developed in chapter 2. Two batch polymerization processes and one continuous process are used to illustrate the performance of the method in "real world" scenarios.

In chapter 4 the problem of designing a new grade quality is extended to the situation where two plants (A and B) are producing various grades of a product, and the production of a particular grade has to be transferred from plant A to the
other plant B. A problem that has to be looked at is whether the production of this grade is actually feasible in plant B according to the available data. Even if this is the case the process conditions from plant A for the desired grade can in general not simply be applied in plant B. The two plants may differ in size and configuration, and therefore the process conditions resulting in the same grade in both plants will likely be different for the two plants. However, the data available from plant A where the grade is already produced may support the design of process conditions for that grade in plant B. An approach is proposed that combines the data of the two plants and predicts process conditions for plant B. A brief numerical example illustrates the methodology.

Chapter 5 is an exploratory one. It investigates the problems that occur when historical operating databases are used in attempts at finding different operating conditions that would improve the current operating performance. Several approaches that have been published in this area are discussed. From the difficulties that one faces in this context it appears that an approach that could reconstruct unmeasured disturbances from the measured process data could be helpful. Ideas for such an approach (called Similarity Optimization) using disturbance reconstruction and nearest neighbor methods to extract useful information from past operating points are presented. At the same time the limitations of such an approach and of other databased approaches, are pointed out. The chapter concludes with a summary of conditions for databases in which databased process optimization may be possible, and conditions in which any attempts at databased optimization are prone to fail.
Chapter 2

Product Design through Analysis of Process Data

2.1 Introduction

Processes can often be operated over a range of conditions to produce a variety of different grades of a product (e.g. a polymer). Each grade has a set of quality specifications such as melt flow index, tensile strength, etc. It often arises that one wants to find process operating conditions needed to achieve a new product grade having a modified set of product quality properties. An example is shown in figure (2.1) where the product quality is characterized by number and weight average molecular weight (MWN and MWw). Points 1 to 9 (○) represent average qualities of nine different grades of polymer which have been produced in the past. Operating strategies are therefore known for these grades. On the other hand, points 10 to 15 (★) mark new polymer grades which would be desirable to achieve, but where the corresponding operating conditions are still unknown.

If one has a good theoretical model of the process, then an appropriate optimization algorithm can be used to find those conditions within the system constraints
That yield the desired product. If such a model is not available, then designed experiments can be run on the process and response surface methods used to move the process to the desired conditions. However, even before one performs experiments, there exists information that is available within the historical process data obtained from past operations over the range of the existing product grades.

This chapter looks at multivariate statistical approaches to analyzing historical plant data and estimating process operating conditions which should yield a product with the desired properties. Very little has been published on this topic. One of the few publications is that of Moteki and Arai (1986). They used a combination of multivariate statistical methods (Principal Components Analysis) and theoretical models to analyze historical operating data on a low density polyethylene (LDPE) process, and inferred conditions that led to new film and injection molding grades of polyethylene. They even claimed to be able to take a competitor’s product which
their plant had not produced before by finding the operating conditions for their process that lead to the same quality. Unfortunately they give few details on their approach.

Another matter of interest in this context is one where two or more plants of a company are required to produce the same product. This product must be of identical quality in both plants so that customers anywhere in the world can accept deliveries from any plant and expect identical performance. However, these plants are often of different size and design, and have different raw material suppliers, and operating constraints such as cooling water temperatures, etc. One cannot simply specify identical operating conditions and expect to achieve the same product. On the contrary, one needs to design the specific conditions for each plant that will yield the one particular product quality. This related problem is treated in chapter 4.

It is the objective of the current chapter to examine an approach to extract the necessary information from past data, and to specify a feasible target region within which the plant operating conditions should lie in order to produce the desired quality. Since they are being derived through a database approach, any new conditions are limited by the range and structure of the historical data. It is possible that there exists a set of operating conditions which are very different from any conditions applied in the past and which would lead to the desired quality. However, such conditions cannot be found by this empirical approach as it can only return conditions having the same range and structure as past conditions.

The problem outlined above addresses both, continuous and batch processes. For continuous processes the historical database is assumed to contain data taken at steady-state operating points corresponding to different product grades. For batch processes, the historical database will contain recipes (i.e. amounts of each ingredient charged to the reactor initially), as well as feedrate profiles for various ingredients (if semi-batch operation is used), and operating profiles on process variables such as
temperature and agitator speed throughout the batch histories. In this chapter we present the methodology and a simulation example for a continuous process. Two industrial batch processes as well as one industrial continuous process are considered in chapter 3.

We assume that data are available for a) the values of the quality variables of the produced grades (e.g. molecular weights, melt flow index, etc.) and b) the operating conditions for those grades. The latter are typically characterized by the values of the manipulated variables of the process. In order to describe the relationship between operating conditions and product quality, one can fit a model \( \hat{Y} = f(X) \), where \( Y \) contains the quality measurements for the different grades and \( X \) the corresponding values of the manipulated variables.

Industrial processes are often operated over a large range of conditions to produce the full scope of products. Figure (2.2) shows a schematic for the range of Linear Low Density Polyethylene (LLDPE) products in the market.

![Figure 2.2: Linear Low Density Polyethylene (LLDPE) Products](image-url)
The behavior over this entire range of products is typically nonlinear. However, here we will mostly be concerned only with using the existing data on multiple grades of a single product type to find operating conditions for a new grade within that same product family. For example, we may be interested in producing a new film grade of LLDPE using data only from within the existing film grade region of figure (2.2). Figure (2.1) serves as an example for such data. The corresponding process behavior may be approximately linear over this region. Therefore, the first part of this chapter will exclusively deal with data that allow for a linear methodology.

On the other hand, it is possible that the data at hand are from a group of grades which range over a much larger area in the product space. Nonlinearities may be more severe in this case and linear techniques not necessarily appropriate anymore. Another situation where nonlinear methods may be considered arises when the new desired grade lies in between two groups of grades. Take as an example the injection molding grades in figure (2.2): A new grade lying between the two shaded areas would require use of data from both injection molding areas and perhaps the data from the rotational molding grades. Any databased approach in this particular situation is effectively an interpolation over a wide range of data, and therefore will be less reliable. However, if the gap between the different groups of data is not too large, nonlinear methods may capture the behavior of the process over this gap fairly well.

The basic ideas of the linear methodology are extended to nonlinear techniques at the end of the chapter. A more thorough discussion on use and abuse of these nonlinear methods is given there as well. The main underlying concepts of these approaches, however, are laid out in the greater part of the chapter dealing with linear methods.

The problem at hand is then to find new process conditions $x_{pred}^T$ for a desired product quality $y_{des}^T$. Thus, we have to use some form of model inverse to predict a
set of manipulated variables $x_{new}^T$ from a specified set of quality variables. A major problem with this approach is the fact that we have typically more manipulatable variables than independent quality variables. Most commonly, one will therefore have to predict a larger number of manipulated variables from a smaller number of quality variables. This results in an underdetermined equation system which has an infinite number of solutions $x_{new}^T$. However, not all of these solutions are acceptable. First of all, there is the requirement that the new process conditions be physically feasible. Furthermore, one would like the new conditions to be consistent with past operating procedures, which means we would like to retain the correlation structure among the manipulated variables. In summary, we want to constrain the solution to be physically feasible and consistent with the sets of process conditions from the past.

There are several reasons for demanding consistency with past process conditions: first of all, it ensures feasibility of the resulting operating strategies; secondly, any model used for this approach must be built from the existing database which has information only on past operating methods, and hence is valid only in this operating region; and thirdly, by enforcing this consistency one reduces the risk of upsetting the process or encountering unexpected problems when running the new conditions. In addition, the process conditions will not just affect the set of quality variables that are measured but also some other unmeasured ones. Even though we do not measure them, these latter quality variables may be important for the final application of the product. It is more likely that these unmeasured quality variables will remain consistent with past grades if we constrain the new conditions to follow the operating strategies from the past. This point will be illustrated in the example of section 2.4.

The following sections outline a methodology which aims at finding at least a region of process conditions that are physically feasible, and consistent with past operating strategies, and that are expected to yield the desired product qualities.
Multivariate latent variable modeling methods such as Principal Components Regression (PCR) and Partial Least Squares (PLS) provide models for both the X-space (process variables) and the Y-space (quality and productivity variables). Simultaneously, modelling both spaces is shown to be essential in this model inversion problem. The standard regression model which models the Y-space only through regression methods such as Ordinary Least Squares (OLS) is shown to be inadequate for this problem. Also contrasted with these approaches is a method based on estimating missing data using conditional expectation methods. This allows the construction of a common framework for the different approaches and provides suitable expressions that highlight the similarities and differences.

The methodology is illustrated by a simulated example of a high pressure tubular low density polyethylene (LDPE) reactor (section 2.4) and three industrial processes: two batch polymerization processes (sections 3.1 and 3.3), and one continuous process (section 3.2).

The last part of this chapter provides an extension of the concepts to nonlinear data structures, and the LDPE example is employed to illustrate performance of the nonlinear approach and compare it with the results from the linear analysis. The necessary discussion of linear versus nonlinear approaches follows. In the final section, the nature of data and potential problems are treated.

2.2 Methodology

In the following sections all the manipulated and quality variables are first mean-centered and scaled to unit variance, and then arranged as the columns in an X- and Y-matrix respectively. In general we assume that there are more manipulated variables than quality variables, and hence, that X has more columns than Y. Both matrices have the same number of rows as each row corresponds to one particular grade.
2.2.1 Selection of Quality Variables

In specifying settings for the quality variables for a new grade $y_{des}$, one first has to make sure that the production of this new grade is possible within the existing process. Although there may be several variables available to describe the quality of a grade, they are often not independent from each other and are usually highly correlated. Specifying independent values for number average and weight average molecular weights, viscosity, etc. will very likely result in a product that is physically not feasible to produce within the existing process, even if each single variable has a reasonable value. In other words, it is not possible to design just any combination of quality properties. The correlation structure among the quality variables from the existing data has to be respected. To ensure this one can build a Principal Component model (Wold et al. 1987a, Jackson 1991) using the entire set of quality data. The principal component score vectors (columns of $T$) in this model form a set of orthogonal vectors which fully describe the structure of the historical grade qualities. The number of significant components will indicate the number of independent quality variables that can be specified. $T$ can be interpreted as a transformation of all quality variables into a smaller set of independent variables. It is therefore reasonable to use the $T$-score matrix as "quality matrix" $Y$ in the design problem. (The set of physical quality variables ($Y$) could simply be replaced by the latent variables ($T$) in the methodology that is developed in this chapter without any further changes.) However, people are usually more comfortable working with "real" physical variables rather than latent variables. For this reason the following sections will use the minimum set of physical quality variables combined in a matrix $Y$ which is representative of all the quality measurements on a grade.

In order to select physical quality variables for $Y$, we still have to overcome the correlation problem for the specification of feasible new grades. Since the human mind has difficulties with thinking in terms of multiple correlated data, it is desirable to
work with a subset of significant quality variables which are nearly independent. The historical quality data serve as a basis for selecting this subset: Given the data matrix of the previously produced grades containing all the measured quality variables, one can first perform a Principal Components Analysis (PCA) to find the number of independent directions within these data. This is the maximum possible number of quality variables that can be specified independently. A properly selected subset of these quality variables would contain just this maximum number of variables and still account for most of the information present in all the quality measurements. The selective PCA algorithm as described by Roffel et al. (1989) can be applied to find such a set. By first using PCA on all available quality variables and then selective PCA (Roffel et al. 1989, Kettaneh-Wold et al. 1994) to find the most complementary subset, one should come up with a representative set of quality variables which show only little correlation. These physical variables could then be used in the quality matrix Y.

Whether Y consists of latent variables or an independent set of physical variables, one still has to make sure that the new set of quality specifications is physically feasible. But apart from physical feasibility, our approach imposes an additional restriction on the qualities that can be designed: they have to obey a covariance structure similar to the one present in the qualities from the past. Since the only information we have are the data from previously produced grades, this information can only help us in the design of new grades that are in some way consistent with these past ones. Again multivariate statistical tools such as PCA can be applied to test whether a new specified product quality is consistent with the qualities present in the data. This requires a) a small Squared Prediction Error (SPE) for the new point, and b) score values that lie within a 95% or 99% control ellipse of the PCA model plane (Kourti and MacGregor 1996). A useful consequence resulting from the described selection of quality variables is the fact that the Y-matrix will be in general
well-conditioned. The benefits of this will be seen later in the model inversion.

Another way to select suitable quality specifications is to start from the space spanned by the \(t\)-scores of the PCA-model (\(T\)-space or score space): Any point in that space that lies within the cluster of historical scores corresponds to qualities that are physically feasible \textit{and} consistent with the available data. The PCA-model provides a one-to-one relationship between physical quality variables and latent variables (\(t\)-scores). One can therefore select a grid of score points within the historical data cluster and calculate the corresponding quality values. They serve as a first set of suggestions for possible new grades. Typically, the grades that are more interesting to a particular customer lie in a certain region of the score space. Thus, one can go back to the score space and choose a finer grid of points in only that particular region. This results in a new set of possible grades which are now closer to the customer's wishes. Iterating between scores and qualities in this way should eventually result in a grade quality that would be feasible \textit{and} desirable to the customer.

### 2.2.2 Standard Regression Model and Its Generalized Inverse

Under some suitable transformation of the \(Y\)s and \(X\)s, the relationship between the product quality (\(Y\)) and the operating conditions (\(X\)) can often be reasonably well approximated by the linear model

\[
Y = X \cdot \beta + E \quad (n \times k) \quad (n \times m) \quad (m \times k) \quad (n \times k)
\]

(2.1)

where the rows in \(X\) and \(Y\) correspond to manipulated and quality variables in meancentered and autoscaled form.

The simplest and probably best known way to estimate \(\beta\) in the linear empirical model (2.1) is Ordinary Least Squares (OLS). The regression parameter estimates \(\hat{\beta}_{ij}\) are obtained by
\[
\hat{\beta} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y.
\]

The first difficulty with OLS arises when \( X \) is ill-conditioned. This often occurs with plant operating data because many of the manipulated variables are moved together in a correlated manner. The inversion of \((X^T \cdot X)\) can then be a problem, and OLS will give estimates of \( \beta \) that have extremely large variances.

\(
\text{Var} \left( \hat{\beta}_{OLS} \right) = (X^T \cdot X)^{-1} \cdot \sigma^2 \).
\)

To help overcome this, one can use regularized least squares solutions such as Ridge Regression (Draper and Smith 1981), which give biased solutions but are much better conditioned. If the number of variables in \( X \) (\( m \)) is greater than the number of grades in the historical data (\( n \)), then an OLS solution does not exist.

Assuming the model (2.1) and parameter estimates \( \hat{\beta} \), new process conditions \( x_{new}^T \) have to be predicted from the desired quality specifications \( y_{des}^T \) such that

\[
y_{des}^T = x_{new}^T \cdot \hat{\beta} \quad (1 \times k) \quad (1 \times m) \quad (m \times k).
\]

If \( k \geq m \), \( x_{new}^T \) can be estimated by OLS or a regularized regression method such as Ridge Regression. However, in general there exist more process conditions than quality variables (\( m > k \)). In this case, since \( \hat{\beta} \) is not square and \( m > k \), solving equation (2.2) for \( x_{new}^T \) is an underdetermined equation system which has an infinite number of solutions. Therefore, a commonly used estimate \( \hat{x}_{new}^T \) is the least squares solution of equation (2.2) which has the minimum 2-norm:

\[
\hat{x}_{new}^T = y_{des}^T \cdot (\hat{\beta}^T \cdot \hat{\beta})^{-1} \cdot \hat{\beta}^T.
\]

\((\hat{\beta}^T \cdot \hat{\beta})^{-1} \cdot \hat{\beta}^T\) is the generalized inverse or pseudo-inverse of \( \hat{\beta} \). The solution (2.3) ensures that \( \hat{x}_{new}^T \) has the following properties:

\[
\min \|x_{new}^T - \hat{x}_{new}^T\|_2 \quad \text{and} \quad \min \|\hat{x}_{new}^T\|_2.
\]
CHAPTER 2. PRODUCT DESIGN

However, there are some serious problems with the use of this standard regression model (2.1), and its inversion (2.3). Since this model (2.1) does not contain any information about the structure within \( X \), the relationships among the manipulated variables that were present during the production of previous grades are ignored. As a consequence the solution given by (2.3) will not respect those previous structural relationships when solving for the new conditions \( x_{new}^T \). Therefore, this formulation based on the standard regression model (2.1) cannot guarantee consistency of the new conditions with past operating procedures. These problems are analyzed analytically in section (2.3) and are illustrated in the example (section 2.4).

2.2.3 PCR-Model and Its Generalized Inverse

Multivariate statistical projection techniques work with data matrices by extracting the main underlying orthogonal directions of variation in correlated data. They summarize the information within the large number of physical variables in a usually much smaller number \( A \) of independent latent variables (Geladi and Kowalski 1986). Principal Components Analysis (PCA) for example transforms the \( (n \times m) \) data matrix \( X \) containing \( m \) highly correlated manipulated variables into a \( (n \times A) \) matrix \( T \) containing only \( A \) independent latent variables which are linear combinations of the original manipulated variables. The weights for these linear combinations are stored in a matrix \( V_A \) which serves to transform \( T \) into \( X \) and vice versa:

\[
X = T \cdot V_A^T + E \nonumber \tag{2.4}
\]

where \( E \) is the error or residual matrix. (Note: In the PCA literature, \( \hat{X} \) is conventionally expressed as \( \hat{X} = T \cdot P^T \). In order to distinguish between this orthogonal matrix \( P \) in PCA/PCR and the \( P \) in PLS, which is not orthogonal, we replace the former matrix in PCA by \( V_A \). This indicates a \( (m \times A) \) matrix with orthonormal
CHAPTER 2. PRODUCT DESIGN

columns, where \( A \leq m \).) PCA can also be interpreted in terms of the singular value
decomposition of \( X \) where only the \( A \) dominant singular directions are retained:

\[
X = U_A \cdot \Sigma_A \cdot V_A^T + E
\]

where \( U_A \) has orthonormal columns and \( \Sigma_A \) is a diagonal matrix with the \( a \)-th element being
the \( a \)-th largest singular value of \( X \). By comparison with equation (2.4), the score
matrix \( T \) can be expressed as

\[
T = U_A \cdot \Sigma_A
\]

In effect, the columns of \( U_A \) are the normalized score vectors.

Both OLS and PCR are linear regression methods. In OLS the columns of \( Y \)
are regressed onto the columns of \( X \). In PCR the columns of \( Y \) are regressed onto
the columns of the latent variables (scores) of \( X \) (\( T \) in (2.4)), that is

\[
\hat{Y} = T \cdot B \quad \text{where} \quad B = (T^T \cdot T)^{-1} \cdot T^T \cdot Y \tag{2.5}
\]

\[
\hat{X} = T \cdot V_A^T \tag{2.6}
\]

or

\[
\hat{Y} = U_A \cdot \Sigma_A \cdot B \quad \text{where} \quad B = \Sigma_A^{-1} \cdot U_A^T \cdot Y \tag{2.7}
\]

\[
\hat{X} = U_A \cdot \Sigma_A \cdot V_A^T \tag{2.8}
\]

Equations (2.6) and (2.8) provide a model for the covariance structure of \( X \).
While the \( A \) columns of \( U_A \) are orthogonal and of unit length, the entries in \( \Sigma_A \)
and the rows in \( V_A^T \) contain the features of the variance and correlation structure
present in \( X \). In the product design problem being treated in this chapter, it would be desirable to impose this same covariance structure onto the conditions \( x_{new}^T \) for the new product. This would make these new operating conditions consistent with the past ones. Thus, the new conditions should conform with the model \((2.8)\) as

\[
x_{new}^T = u_{new}^T \cdot \Sigma_A \cdot V_A^T
\]

\((1 \times m) \quad (1 \times A) \quad (A \times A) \quad (A \times m)\)

Accordingly, it is the elements of \( u_{new}^T \) that have to be estimated. This can be done using equation \((2.7)\)

\[
y_{des}^T = u_{new}^T \cdot \Sigma_A \cdot B
\]

\((1 \times k) \quad (1 \times A) \quad (A \times A) \quad (A \times k)\) \quad (2.9)

This equation has the same structure as equation \((2.2)\). However, while in \((2.2)\) one has to find \( m \) elements for \( x_{new}^T \), here one only has to find \( A < m \) elements for \( u_{new}^T \), i.e. the equation system has been reduced in dimension. The problem is redefined to estimate \( A \) latent variables for the process conditions from \( k \) quality variables. Consequently, we can distinguish between three cases in equation \((2.9)\):

1. \( k > A \): The model inversion corresponds to a projection from a high dimensional \( (k) \) space to a lower dimensional \( (A) \) space. It can be performed as a simple least squares projection.

\[
\hat{u}_{new}^T = y_{des}^T \cdot B^T \cdot \left( B \cdot B^T \right)^{-1} \cdot \Sigma_A^{-1}
\]

\[
\hat{x}_{new}^T = \hat{u}_{new}^T \cdot \Sigma_A \cdot V_A^T
\]

\[
= y_{des}^T \cdot B^T \cdot \left( B \cdot B^T \right)^{-1} \cdot V_A^T
\]

2. \( k = A \): In this case an exact inversion from one \( k \)- or \( A \)-dimensional basis to the other is possible.

\[
\hat{u}_{new}^T = y_{des}^T \cdot B^{-1} \cdot \Sigma_A^{-1}
\]

\[
\hat{x}_{new}^T = \hat{u}_{new}^T \cdot \Sigma_A \cdot V_A^T
\]

\[
= y_{des}^T \cdot B^{-1} \cdot V_A^T
\]
3. $k < A$: This is the most common situation. Although the number of variables
to be predicted has been reduced from $m$ to $A$, still a projection from a lower
($k$) to a higher ($A$) dimensional hyperplane is required. This is obviously the
worst of the three cases. It occurs when the effective rank of the $X$-matrix ($A$)
is larger than the number of independent $Y$-variables ($k$), that is when there
are variations in some directions of the process operating space ($X$) which have
little influence on the particular set of quality variables under consideration ($Y$).

For cases 1. and 2. there exists a unique solution of minimum squared error.
In the third underdetermined case, however, the set of solutions is infinite.

In analogy to equation (2.3) we use the pseudo-inverse for this case (3.) to
invert (2.9):

$$
\hat{u}_{new}^T = y_{des}^T \cdot \left( B^T \cdot \Sigma_A^2 \cdot B \right)^{-1} \cdot B^T \cdot \Sigma_A
$$

(2.10)

The new process conditions can now be calculated as

$$
\begin{align*}
\hat{x}_{new}^T &= \hat{u}_{new}^T \cdot \Sigma_A \cdot V_A^T \\
(1 \times m) &= \left(1 \times A \right) \left(A \times A \right) \left( A \times m \right) \\
&= y_{des}^T \cdot M_{PCR}^T \\
(1 \times k) &= \left( k \times m \right)
\end{align*}
$$

(2.11)

(2.12)

where

$$
M_{PCR}^T = \left( B^T \cdot \Sigma_A^2 \cdot B \right)^{-1} \cdot B^T \cdot \Sigma_A^2 \cdot V_A^T
$$

(2.13)

$$
= \left(Y^T \cdot U_A \cdot U_A^T \cdot Y \right)^{-1} \cdot Y^T \cdot U_A \cdot U_A^T \cdot X
$$

Note that the matrix being inverted in equation (2.13) is well-conditioned
since it depends solely on $Y$ and $U_A$. The columns in $Y$ are chosen as an independent
set of the quality variables, and the columns of $U_A$ are orthogonal by definition. The
case where one or more columns of $Y$ are in the nullspace of $U_A$ is a pathological one:
It implies that none of the process conditions included in $X$ have any influence on
the particular quality variable(s) specified in these column(s) of $Y$. Consequently, an
approach based on such data cannot design process conditions which would lead to desired values of these qualities. Therefore, this case is not considered any further.

Equation (2.11) reflects the way in which PCR and its inverse impose the covariance structure of the past operating conditions onto the new conditions $\hat{x}_{new}^T$: $\tilde{u}_{new}^T$ which is obtained from equation (2.10) is multiplied by $\Sigma_A$ and $V_A^T$ which contain the variance and correlation structure of the original data $X$. The new $\hat{x}_{new}^T$ therefore follows the same covariance structure as the past operating conditions. There is no equivalence to this in the approach based on the standard regression model (2.1) using OLS (section 2.2.2). The reason for this is that - unlike the latent variable methods (PCR, PLS) - in the standard regression model (2.1) no model for $X$ is proposed, only one for the correlation structure between $X$ and $Y$. Moreover, equation (2.3) for this model estimates $\hat{x}_{new}^T$ as the minimum norm solution to equation (2.2).

The minimization of the norm of $x^T$ with respect to the Euclidean distance assumes that the elements of vector $x^T$ are independent. However, $x^T$ represents the set of manipulatable variables for which we are not assuming independent values but rather correlated ones. Thus, the minimum norm (or generalized) inverse of the standard regression model completely ignores the correlation and variance structure present in the historical data for $X$. On the other hand, in this multivariate approach using the PCR model, $\tilde{u}^T$ is calculated as the minimum norm solution to equation (2.9), and again the resulting $\tilde{u}_{new}^T$-vector is estimated having independent elements. Only here this is adequate since these $A$ elements of $\tilde{u}_{new}^T$ correspond to the $A$ columns of the matrix $U_A$ which has orthonormal and therefore independent columns. In other words, the minimum norm inverse given by equation (2.10) cannot destroy any covariance structure since the elements are supposed to be independent and have equal variance.

This enables us to impose the desired covariance structure through multiplication by $\Sigma_A$ and $V_A^T$.

Equation (2.12) shows that $\hat{x}_{new}^T$ is a linear combination of the rows of $M_{PCR}^T$. 
which itself is a linear combination of the A rows in \( V_A^T \). While the past operating conditions (rows in \( X \)) fall in the space spanned by the A rows of \( V_A^T \), the \( \hat{x}_{\text{new}}^T \) obtained from (2.12) fall only into the \( k \)-dimensional space (which in fact is a subspace of \( V_A^T \)) defined by the rows of \( M_{\text{PCR}}^T \). In other words, equation (2.12) constructs only that part of the new operating conditions which corresponds to the projection of the past process conditions \( X \) onto the \( M^T \)-space. This is the part of the new operating conditions which - according to the model - affects the product quality: According to the model \( \hat{x}_{\text{new}}^T \) is sufficient to produce the desired product quality. However, there are features in the process space \( X \) which do not affect this particular set of quality characteristics \( y_{\text{des}}^T \) and which are not accounted for in \( \hat{x}_{\text{new}}^T \). We denote this space by \( x_{\text{null}}^T \). These \( x_{\text{null}}^T \) may still be important operating features which should be respected in any new operating conditions \( x_{\text{pred}}^T \). They lie within the \( A \)-dimensional \( V_A^T \)-space that defines past process operations (\( X \)), but outside the subspace of \( M^T \), and they do not affect \( Y \).

This suggests, that any feature lying in the space of \( x_{\text{null}}^T \) can be added to \( \hat{x}_{\text{new}}^T \), where \( x_{\text{null}}^T \) lies in the space which is orthogonal to \( M \) and spans the remaining \((A-k)\) dimensions of the \( A \)-dimensional \( X \)-space. Adding this component should not alter the relationship between \( x_{\text{pred}}^T = \hat{x}_{\text{new}}^T + x_{\text{null}}^T \) and \( y_{\text{des}}^T \). In other words, \( x_{\text{null}}^T \) lies in the remaining \((A-k)\) dimensions of \( V_A^T \) that are not spanned by \( M^T \). (See appendix A for details on the calculation of \( x_{\text{null}}^T \)). Since this component \( x_{\text{null}}^T \) has no effect on \( Y \), there is some degree of freedom with respect to its magnitude and direction within that \((A-k)\)-dimensional space. The only limitation is the constraint on \( x_{\text{pred}}^T \) to be statistically consistent with past operating conditions. By choosing a whole range of different \( x_{\text{null}}^T \) within the \( 2\sigma \)-range of \( X \) in that additional space, it is possible to look at a “window” of process conditions \( x_{\text{pred}}^T = \hat{x}_{\text{new}}^T + x_{\text{null}}^T \) that would, in principle, yield the same \( y_{\text{des}}^T \). Using some process knowledge the most promising set of new process conditions \( x_{\text{pred}}^T \) can be selected. In this way, when \( k < A \), one can only present a
window of possible solutions by inverting the model from $y_{des}^T$. Figure (2.3) serves as an example to illustrate this graphically.

![Diagram illustrating possible solutions along $X_{null}$-direction](image)

**Figure 2.3:** The Different Spaces for $\hat{x}_{new}^T$ and $x_{null}^T$ when $m = 3$, $A = 2$ and $k = 1$

The $X$-matrix in this example contains three manipulated variables which are shown as coordinates. However, the $X$-data fall only into a two-dimensional space indicated by the gray plane. This plane represents the main underlying directions in $X$. The actual area the data fall into is determined by their covariance structure and may be approximated by an ellipse in that plane. The lengths of the ellipse axes are proportional to the singular values in $\Sigma_A$. In this figure, the $Y$-matrix is assumed to consist of a single quality variable only, and thus $k = 1$. Accordingly, the corresponding $M$-space is one-dimensional as well, and as a subspace of the true $X$-space, it defines a line in the plane. This line is the locus of all $\hat{x}_{new}^T$-predictions obtained from different $y_{des}^T$. Additional components $x_{null}^T$ can be added to $\hat{x}_{new}^T$ such that the whole two-dimensional $X$-space is spanned. Resulting $x_{pred}^T = \hat{x}_{new}^T + x_{null}^T$ for one $\hat{x}_{new}^T$ from different $x_{null}^T$ are shown as white dots on a second line in figure...
(2.3). Any point on this second line \( x_{null}^T \)-line corresponds to process conditions that are constructed such that their covariance structure is consistent with the historical \( X \)-data and result in the desired \( y_{des}^T \). However, we have to make sure that we restrict ourselves to \( x_{pred}^T \) that also lie within the range of past operating procedures. In other words, we want to stay within the region into which the past \( x^T \) fall. A two-step procedure leads to such an adequate “window” of process conditions:

1. Only vectors \( x_{null}^T \) within the 2\( \sigma \)-range of \( X \) along the \( x_{null}^T \)-line are added to \( x_{new}^T \). (An estimate for the variance of \( X \) along the line can be obtained by projecting the rows of the existing \( X \)-matrix onto that line in the plane.) Since for the typically small number of grades (rows in \( X \)) this \( \sigma \)-estimate has a large confidence interval, this first step is only a crude measure and needs to be refined in step 2.

2. Different \( x_{pred}^T \) along the line are transformed into the score space: \( t_{pred}^T = x_{pred}^T \cdot V_A \), and it is checked where in the \( T \)-space \( t_{pred}^T \) lies compared to the \( t_i^T \) of the past conditions. If \( t_{pred}^T \) does not fall within the cluster of past \( t_i^T \)-s, it should not be included in the set of conditions since it represents an extrapolation beyond the historical data region.

Note: For the case in figure (2.3) where \( A = 2 \), all \( x_{pred}^T \) for different \( y_{des}^T \) should lie within the ellipse which corresponds to the historical \( x^T \) (rows of \( X \)) in the plane of figure (2.3). For a case \( A \geq 3 \), \( x_{null}^T \) must be selected such that \( x_{pred}^T \) still lies within the ellipsoid or hyperellipsoid defined by the range of the historical data. Extrapolating outside of these ellipsoids cannot guarantee reasonable results. For \( A - k > 1 \), the approach will give regions of possible operation; the process conditions \( x_{pred}^T \) from those regions should all yield the specified properties \( y_{des}^T \).

Indispensable factors in the decision making on which of the conditions to apply are opinion and experience of the plant operator or engineer. Faced with the
set of suggested conditions, his/her experience may again rule out the ones that seem too different from normal operating procedures, or are undesirable in some way.

2.2.4 Other Multivariate Latent Variable Methods

Apart from PCR there exists a whole range of other multivariate statistical modeling techniques which theoretically could be used and inverted for this problem. Common to all of them is the representation of data matrices by latent variables that summarize certain types of variation. PCR focuses on the high variance directions in \( X \). Canonical Correlation Regression (CCR) focuses only on the high correlation directions between \( X \) and \( Y \), Reduced Rank Regression (RRR) focuses on high variance directions in \( Y \), and Projection to Latent Structures (PLS) lies somewhere in between by simultaneously explaining variation in \( X \) as well as the correlation with \( Y \). Burnham et al. (1996) discuss the objective function frameworks behind all these methods.

For the problem at hand which tries to construct a \( x_{new}^{T} \) from a given \( y_{des}^{T} \), the key is to have a model for the structure of the data in \( X \) as well as for the relationship between \( X \) and \( Y \). PLS and PCR are therefore preferred over methods like OLS, CCR and RRR, which provide no model for \( X \).

PLS models mainly those high variance directions of \( X \) which are predictive of \( Y \). It tends to explain \( Y \) with a smaller number of latent variables than PCR. However, there may be variations in the process conditions which do not affect the measured quality variables but which are essential in other aspects - for example for safety reasons or for quality properties which are not measured. A PLS-model may not capture these directions in the first few latent variables since these directions have little correlation with \( Y \), and hence one may have to add more latent variables in order to explain these parts of \( X \) as well.

PCR models \( X \) first - thereby capturing all the high variance directions of \( X \)
within the first latent variables - and then regresses $Y$ onto the latent variables of $X$. However, since some of the latent variables in $X$ may not be highly correlated with $Y$, PCR often needs more principal components to describe the main part of $Y$ than PLS.

In most cases PCR and PLS have been found to give similar results. In general of course, the method giving a smaller number of latent variables is preferable in order to keep $(A-k)$ as small as possible. $(A-k)$ is the number of directions in $X$ which are not captured by the $M$-space and in which $x^T_{null}$ has freedom to move.) (see appendix A). In this sense there may be cases where a PLS-model is beneficial compared to a PCR-model or vice versa, depending on which method uses fewer components to explain the greatest amount of both $X$ and $Y$.

Note: If $Y$ is close to being orthogonal to one of the columns of $T$ in PCR, then $B$ and consequently any inverse of $B$ are ill-conditioned. In such a case one should use PLS.

The methodology for the inversion of PLS-models is identical to that for PCR described earlier. In the industrial applications in chapter 3 PLS-models are employed, and in the extension to nonlinear systems (section 2.5) a nonlinear PLS formulation is used.

### 2.2.5 Conditional Expectation Approach

Although this approach looks at the problem from a quite different angle, it nevertheless leads to results that are comparable to and supporting of those obtained in the previous sections.

Past operating conditions $X$ and the corresponding product qualities $Y$ are combined into a single matrix, and are looked at as samples from a multivariate normal distribution. The unknown operating conditions $x^T_{new}$ that would correspond to the newly specified quality $y^T_{des}$ are interpreted as missing data within a new row
added to the existing $X \mid Y$ data matrix, cf.

$$
[X \mid Y] = \begin{pmatrix}
X & Y \\
\vdots & \vdots \\
x_{new}^T & y_{des}^T
\end{pmatrix}.
$$

The approach used is based on that of Nelson et al. (1996) and the Expectation-Maximization (EM) method of Little and Rubin (1987).

Here, we only use the conditional expectation step of that approach. Based on the estimates for the mean and covariance of the data from past grades, the missing values $x_{new}^T$ are calculated as the conditional expectations given the values for the "observed" variables $y_{des}^T$. (In our problem this converts to: Given the quality values $y_{des}^T$, what is the conditional expectation for $x_{new}^T$ - based on the mean and the covariance matrix of the combined data set $[X \mid Y]$?) For mean-centered $X$ and $Y$, the covariance matrix can be written as

$$
cov([X \mid Y]) = \begin{pmatrix}
X^T \cdot X & X^T \cdot Y \\
Y^T \cdot X & Y^T \cdot Y
\end{pmatrix}
$$

and the conditional expectation for $x_{new}^T$ given the values of $y_{des}^T$ is (Johnson and Wichern (1988), p.131; Nelson et al. (1996))

$$
\hat{x}_{new}^T = \frac{y_{des}^T \cdot (Y^T \cdot Y)^{-1} \cdot Y^T \cdot X}{(1 \times m)}
= \frac{y_{des}^T \cdot M_{cond}^T}{(1 \times k)}
= \frac{y_{des}^T}{(k \times m)}.
$$

(2.14)

This expression is equivalent to an inverse regression where $X$ is regressed onto the columns of $Y$ rather than vice versa. We can see that conceptually this expression is similar to the one for $\hat{x}_{new}^T$ from PCR-inversion (see equation (2.12) in section (2.2.3)). The only difference is that in PCR $Y$ and $X$ are replaced by their projections onto $U_A$.

This approach could be beneficial in cases where it is desired to preset certain values for one or more manipulated variables. In such a case these fixed manipulated
variables would be included with the desired qualities \( y_{\text{des}}^T \), and \( x_{\text{new}}^T \) would consist only of the remaining still unknown manipulated variables. However, as discussed in section (2.2.1), one has to make sure that the values selected for this set of quality and prespecified \( X \)-variables are consistent with the past data for these variables. Again, this can be done by performing PCA on this particular set of variables in the historical data, and then applying the model to the new specified values of the fixed manipulated and quality variables. The corresponding Squared Prediction Error has to be small and the \( t \)-scores have to lie within the cluster of \( t \)-scores from the historical data in order for the set of specified values to be applicable.

Equation (2.14) shows that again \( x_{\text{new}}^T \) can fall only into a \( k \)-dimensional subspace of the space spanned by the past operating conditions \( X \). However, in contrast to the PCR and PLS model inversions of the previous sections, we again do not have a model for the structure in \( X \) from this approach. It is therefore not possible to find that \( (A - k) \)-dimensional hyperplane for \( x_{\text{null}}^T \) in the \( X \)-space within which alternative choices of \( x_{\text{pred}}^T \) will have the same predicted quality \( y_{\text{des}}^T \).

## 2.3 A Common Framework

For the methods described above the new operating conditions \( x_{\text{pred}}^T \) consist of two parts:

- \( \hat{x}_{\text{new}}^T \) which is derived using the relationship between \( X \) and \( Y \), and the new values \( y_{\text{des}}^T \)

- and an additional component \( x_{\text{null}}^T \) which complements \( \hat{x}_{\text{new}}^T \) and which is derived solely from information contained within \( X \) and does not affect \( Y \).

The calculations of \( x_{\text{null}}^T \) for different methods depend strongly on the way \( \hat{x}_{\text{new}}^T \) has been calculated as can be seen in appendix A. However, \( \hat{x}_{\text{new}}^T \) for all the
different methods can be expressed as
\[ x_{new}^T = y_{des}^T \cdot M_{Method}^T \]

where \( M_{Method}^T = \left( \hat{Y}^T \cdot \hat{Y} \right)^{-1} \cdot \hat{Y}^T \cdot \hat{X} \).

\( \hat{X} \) and \( \hat{Y} \) are different for each method, but they can always be interpreted as a projection of \( X \) and \( Y \) onto some space. Table 2.1 gives the \( \hat{X} \) and \( \hat{Y} \) for each method.

Table 2.1: Common Framework for the Methods: \( x_{new}^T = y_{des}^T \cdot M_{Method}^T \)

<table>
<thead>
<tr>
<th>Method</th>
<th>( M_{Method}^T )</th>
<th>( \hat{Y} )</th>
<th>( \hat{X} )</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{PCR} )</td>
<td>((y^T \cdot u_A \cdot u_A^T \cdot y)^{-1} \cdot y^T \cdot u_A \cdot u_A^T \cdot x)</td>
<td>( u_A^T Y )</td>
<td>( u_A^T X )</td>
<td>- similar to PCR</td>
</tr>
<tr>
<td>( M_{PLS} )</td>
<td>((y^T \cdot u_A \cdot u_A^T \cdot x)^{-1} \cdot y^T \cdot u_A \cdot u_A^T \cdot x)</td>
<td>( u_A^T )</td>
<td>( u_A^T X )</td>
<td>- similar to PCR</td>
</tr>
<tr>
<td>( M_{OLS} )</td>
<td>((y^T \cdot u \cdot u^T \cdot y)^{-1} \cdot y^T \cdot u \cdot u^T \cdot x)</td>
<td>( \Sigma^{-1} u^T \cdot y )</td>
<td>( \Sigma^{-1} u^T \cdot x )</td>
<td>- does not maintain the covariance structure of ( X ) - no ( x_{null}^T ) since no model for ( X )</td>
</tr>
<tr>
<td>( M_{Cond} )</td>
<td>((y^T \cdot y)^{-1} \cdot y^T \cdot x)</td>
<td>( y )</td>
<td>( x )</td>
<td>- inverse regression - no ( x_{null}^T ) since no model for ( X )</td>
</tr>
</tbody>
</table>

The nomenclature refers to the following relationships:

**PCR:**
PCA on \( X \):
\[ \hat{X} = u_A \cdot \Sigma_A \cdot v_A^T \]
CHAPTER 2. PRODUCT DESIGN

PLS:

$$
\hat{X} = U_A \cdot S_A \cdot P^T
$$

$$
\hat{Y} = U_A \cdot S_A \cdot Q^T
$$

OLS:

$$
X = U \cdot \Sigma \cdot V^T
$$

$U$, $U_A$ and $U_A$ have orthonormal columns
$V$, $V_A$ have orthonormal columns
$P$, $Q$ have full column rank
$\Sigma$, $\Sigma_A$, $S_A$ are diagonal matrices.

The similarity of PCR and PLS is obvious in the expressions: In both cases $X$ and $Y$ are projected onto a $(n \times A)$ matrix ($U_A$ and $U_A$) with orthonormal columns which span the column space of $\hat{X}$. In PCR these columns stem from the singular value decomposition in $X$, in PLS they are derived from eigenvalue decompositions on $X \cdot X^T \cdot Y \cdot Y^T$ and the residual covariance matrices in both spaces (Höskuldsson 1988). However, as discussed in section 2.2.4 in general the space covered by PCR and the one covered by PLS are often almost the same as soon as both $X$ and $Y$ are sufficiently explained.

Fitting OLS inversion into this framework allows some greater insights into why inversion of the standard regression model does not preserve the covariance structure of the data. A first issue to address is the ill-conditioning of $X$: It can be shown that with one (or more) singular values of $X$ going towards zero, the rows in $M_{OLS}^T$ approach zero. Since $X$ is meancentered and autoscaled, $x_{new}^T$ will also approach zero. Therefore, when $X$ is very ill-conditioned, the final resulting operating conditions will be very close to the mean values of the conditions from the past - no matter what has been specified for $y_{des}^T$. 
CHAPTER 2. PRODUCT DESIGN

But even if $X$ is not highly correlated, the standard regression model and its inversion are not able to respect the covariance structure present in $X$. Relating OLS to PCR sheds some light onto this problem: OLS is identical to PCR when all $(m)$ principal components are retained.

PCR on $X$ and $Y$ with the maximum number of principal components can be written as:

$$X = T \cdot V^T \quad \text{with} \quad V \cdot V^T = V^T \cdot V = I$$

$$\hat{Y} = T \cdot B$$  \hspace{1cm} (2.15)

Note, that $X$ is explained completely by this model and use of $\hat{X}$ is therefore unnecessary.

The OLS model is denoted by

$$\hat{Y} = X \cdot \hat{\beta}$$

Since $X = T \cdot V^T$, we have $B = V^T \cdot \hat{\beta}$ in equation (2.15). We expand the OLS-expression for $\hat{x}_{new}^T$ (see equation (2.3)) by multiplying by the identity matrix $I = V \cdot V^T$:

$$\hat{x}_{new}^T = y_{des}^T \cdot (\hat{\beta}^T \cdot \hat{\beta})^{-1} \cdot \hat{\beta}^T$$

$$= y_{des}^T \cdot (\hat{\beta}^T \cdot V \cdot V^T \cdot \hat{\beta})^{-1} \cdot \hat{\beta}^T \cdot V \cdot V^T$$

$$= y_{des}^T \cdot (B^T \cdot B)^{-1} \cdot B^T \cdot V^T$$

The first part of this expression $(y_{des}^T \cdot (B^T \cdot B)^{-1} \cdot B^T)$ is the generalized inverse of equation (2.15) applied to $y_{des}^T$. Since the term is then multiplied by $V^T$ we could define: $\hat{x}_{new}^T = y_{des}^T \cdot (B^T \cdot B)^{-1} \cdot B^T$ in order to be able to write $\hat{x}_{new}^T = \hat{t}_{new}^T \cdot V^T$. $V^T$ is simply an orthogonal matrix which rotates $t^T$ into $x^T$. The crucial part is the
CHAPTER 2. PRODUCT DESIGN

generalized inverse \((B^T \cdot B)^{-1} \cdot B^T\). According to equation (2.15) this inverse gives the minimum norm \(t^T\) in a Euclidean sense that solves the underdetermined equation \(y_{des}^T = \tilde{t}_{new}^T \cdot B\). This minimum norm in a Euclidean sense allows for equal variance for each element in \(t^T\). The columns in \(T\), however, have different variances. In other words, while in PCR the \(t^T\)-elements are weighed according to the variation explained by the corresponding principal component, a minimum Euclidean norm in OLS inversion gives equal weight to the \(t^T\)-elements.

The \(X\)-data plotted in the \(T\)-space lie within an ellipsoid or hyperellipsoid where the lengths of the axes correspond to the standard deviation of the columns in \(T\) (singular values in \(X\)). Since \(M^T\) from each method has only \(k\) rows, \(\tilde{x}_{new}^T\) lies only in a subspace (spanned by the rows of \(M^T\)) of the \(T\)-space. (This has been discussed in section 2.2.3.) In order to maintain the covariance structure, the \(\tilde{x}_{new}^T\) values should lie within the intersection of the hyperellipsoid and that subspace.

The expressions from OLS for \(\tilde{X}\) and \(\tilde{Y}\) are the projections of \(X\) and \(Y\) onto \(U\) (see table (2.1)). These projections are then scaled by \(\Sigma^{-1}\), i.e. the projection along each \(U\)-axis is scaled by the inverse of the singular value of \(X\) (which is the standard deviation of the corresponding latent variable \(t\)) along that axis. Consequently, the variation of \(Y\) in the dominant direction of \(X\) (first column of \(U\)) is weighed less than the variation of \(Y\) in the second direction. This results in equally weighed \(t^T\)-elements (i.e. unit variance of each \(t^T\)-element), and in terms of the data it is equivalent to squeezing the hyperellipsoid into the shape of a hypersphere. By eliminating the variances of the principal components in this way, of course one removes an essential part of the covariance information within \(X\). The resulting \(\tilde{x}_{new}^T\) now lie in a subspace intersection of the hypersphere and possibly outside the original hyperellipsoid, i.e. the OLS-predictions do not exhibit the proper variance, and as such they are not consistent with the past data anymore.
CHAPTER 2. PRODUCT DESIGN

In the above paragraph we have shown that the OLS-inversion is identical to building a PCR-model with the maximum number of principal components and inverting it such that the new $t^T$-vector has minimum Euclidean norm. The difference in the PCR-inversion which was described in section 2.2.3 is that there we solve for the Euclidean minimum norm $u^T$-vectors, and then let $t^T_{new} = \hat{u}^T_{new} \cdot \Sigma$ (e.g. for $A = m$). This is equivalent to finding the minimum length $t^T$ with respect to the Mahalanobis distance of $t^T$. It allows elements in $t^T_{new}$ which correspond to principal components with greater variability to assume larger values than elements corresponding to principal components with less variation.

By including only $A < m$ principal components in the model, one essentially assigns a zero variance to the scores for principal components $A + 1$ to $m$. On the other hand, including these principal components in the model would restrict the corresponding scores to very small values since they would be scaled by their small variance. However, it is preferable to discard principal components with small variance in order to keep $A$ and therefore $A - k$ small, so that the additional component $x^T_{null}$ does not need to be calculated in too many directions.

2.4 Example: Simulated LDPE Polymerization Process

The following sections illustrate these concepts using as a process a fundamental simulation model of a high pressure tubular reactor system for production of low density polyethylene (LDPE) (Kiparissides et al. 1993, MacGregor et al. 1994). The performance of OLS and PCR will be compared and the findings discussed.
2.4.1 The Process

The process used in the following study is the production of low density polyethylene (LDPE). The polymerization takes place in a high pressure tubular reactor with a cooling jacket. Simulated are two zones of the reactor with feed to the first zone and a quenching stream to the second zone. (The qualities are therefore not end product qualities.) The five measured quality and productivity variables are:

- Conversion (Conv)
- Number Average Molecular Weight (MWN)
- Weight Average Molecular Weight (MWW)
- Long Chain Branching (LCB)
- Short Chain Branching (SCB).

Six manipulated variables are available:

- Inlet Temperature of the Feed (Tin)
- Pressure in the Reactor (P)
- Initiator Feed Rate to both Zones (Fi1, Fi2)
- Solvent Flow to both Zones as % of Ethylene (Cs1, Cs2).

More details on this process and its simulation can be found in (Kiparissides et al. 1993, MacGregor et al. 1994).

2.4.2 Generation of Data

Although six manipulated variables are available, they are not set independently. Except for some added random error, Fi2 and Cs2 (initiator flow rate and solvent
concentration in the feed to zone 2) are kept at the same levels as their counterparts in the feed to zone 1. This leaves only four variables to be moved. Their settings have been chosen such that they are varied in a correlated manner, and the resulting X-matrix has only three (two major and one minor) independent underlying directions. 15 points representing different grades of LDPE have been simulated in this way. They are recorded in appendix B. Out of these 15 grades 9 have been chosen to represent the average process conditions and quality of previously produced grades. These 9 points (1–9 in figure (2.1)) form the historical data base for the multivariate statistical model. The quality values of the remaining 6 grades serve as quality specifications that one may want to produce. Hence, we pretend not to know the settings for the manipulated variables of these 6 points.

2.4.3 Case Studies

The method of data generation described above ensures that the new quality specifications are a) physically feasible and b) consistent with grades produced in the past. We start by finding out how many and which quality variables we should include in the Y-matrix. PCA on the matrix containing all five quality variables shows that after the first three principal components 99% of the overall variation is explained. Consequently, we are looking for the three most significant quality and productivity variables. Selective PCA (Roffel et al. 1989) determines these as MWw, MWn, Conv - in this order. These three variables account for 96% of the variation in all five variables. MWn and MWw alone account for 76.25% of the total variation.

In the following studies, X consists of the values for the six manipulated variables for the nine grades that have been produced in the past. Two different Y-matrices are considered: Y3 = [Conv MWn MWw] and Y2 = [MWn MWw] for each of the nine grades.

Table 2.2 shows how much variation is explained by PCR in two models: X
and Y3, and X and Y2. In both cases, the number of principal components in the
PCR-model is chosen as \( A = 3 \).

Table 2.2: Percentage of Variation Explained by PCR in X and Y

<table>
<thead>
<tr>
<th>No. of Principal Components</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>63.70</td>
<td>96.36</td>
<td>99.15</td>
<td>99.24</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>case 1: Y3</td>
<td>23.51</td>
<td>68.07</td>
<td>94.05</td>
<td>96.22</td>
<td>97.58</td>
<td>99.10</td>
</tr>
<tr>
<td>case 2: Y2</td>
<td>29.03</td>
<td>95.35</td>
<td>97.04</td>
<td>98.51</td>
<td>98.74</td>
<td>99.36</td>
</tr>
</tbody>
</table>

Case 1: Y3 = [Conv MWhn MWw]

The use of the standard regression model and OLS is problematic even in this small example since we have to invert an ill-conditioned \( X^T \cdot X \). However, OLS was performed using MATLAB to obtain values for \( \hat{\beta} \). The variances of these OLS estimates were very large.

In PCR on the other hand, we encounter here the special case where the number of Y-variables equals the number of principal components in the model \( k = A \). For the inversion of equation (2.9) this means that we can directly invert \( B \) and calculate \( x_{new}^T = y_{des}^T \cdot B^{-1} \cdot V \). Since for \( k = A \) the M-space spans all the main directions in X, there are no orthogonal directions \( x_{null}^T \). We therefore come up with one unique set of manipulated variables for each desired new grade.

The resulting process conditions from PCR- and OLS-inversion were applied to the full nonlinear simulated process and the corresponding quality values obtained. The results are shown in figures (2.4), (2.5), and (2.6). The points denoted by a (*) are the desired qualities, (o) denotes the qualities achieved from OLS-inversion, and (x) the qualities achieved from PCR-inversion.

If one compares the estimates of the quality and productivity variables obtained from OLS- and PCR-inversion, apart from one exception (grade 13) PCR
Figure 2.4: PCR- and OLS-Results for Number and Weight Average Molecular Weight, when $Y = [\text{Conv MWN MWw}]$

Figure 2.5: PCR- and OLS-Results for Conversion and Weight Average Molecular Weight, when $Y = [\text{Conv MWN MWw}]$
achieves products much closer to the desired ones than OLS. (As mentioned in section 2.3 OLS for ill-conditioned $X$ tends to move the new conditions towards the mean of the historical conditions. The numerical values of grade 13 happen to lie close to that mean, and therefore OLS achieves a good result.) Whereas figures (2.4) and (2.5) show the three measured responses (Conv, MWN and MWw) that are included in $Y$, figure (2.6) displays the two quality variables which are not present in $Y$ (LCB and SCB), but which nevertheless are very important quality characteristics for the product. Since there is no immediate information about them in the model data $Y$ their results are not as good as for MWN and MWw. However, we can see that PCR-qualities are much closer to the LCB and SCB-qualities of the desired products than OLS-qualities. This is one of the benefits of constructing process conditions which are consistent with past operating strategies. Other unmeasured quality variables will maintain their same historical relationship with those that are measured.
CHAPTER 2. PRODUCT DESIGN

Remark: In this case, LCB and SCB are not specified in the Y-matrix, so realistically there are no desired values for them. However, we did simulate grades 10-15 in a way consistent with the grades 1-9 which are used for the model building, and therefore, we have values for LCB and SCB for these grades. If PCR returns qualities close to the ones of 10-15, this indicates that the new products obtained from PCR-inversion are consistent with the grades produced in the past. This is not generally true for OLS, as can be seen in figure (2.6).

Case 2: \( Y_2 = [M\text{Wn M}\text{Ww}] \)

Some Results from PCR Inversion

As discussed earlier, a PCA analysis on all five quality variables indicated there are three underlying directions among these variables which account for the total variation within them. Thus, ideally one would want to specify three independent quality variables. However, in many processes only a few correlated properties are actually measured on a product, and these may be insufficient to span the entire quality space of a product. In this case, by specifying MWn and MWw only two out of the three directions are being set, and about 24% of the total quality variation is not accounted for. This example will highlight some of the issues that arise in the case of a Y-matrix that does not account for the total quality variation of a product.

For the PCR-inversion this case provides a different situation from case 1: the number of \( Y \)-variables \( (k = 2) \) is less than the number of independent directions in \( X \) \( (A = 3) \). Consequently, the resulting \( \hat{x}_{new}^T \) conditions will lie only in a 2-dimensional subspace of \( X \). The additional direction, which is complementing the \( M \)-space (see figure (2.3)), must be calculated. Projecting the \( X \)-data onto this direction gives an estimate of the amount of variation in \( X \) that exists along that direction. Five different vectors \( x_{null}^T \) (within \( 2\sigma \) of the rows of \( X \) along this direction) were added to each \( \hat{x}_{new}^T \). This gives a set of five suggestions \( x_{pred}^T = \hat{x}_{new}^T + x_{null}^T \) for each new grade. From the
five suggestions for each grade a subset was selected such that their $t_{pred}^T = x_{pred}^T \cdot V_A$ lay within the cluster of $t_i^T$-s from the past grades. For example, for grade 11 only two of the five suggestions had appropriate $t$-values and were implemented.

In figure (2.7) two of the resulting manipulated variables (Cs1 and P, inlet solvent concentration to zone 1 and inlet pressure) are plotted versus each other for the various sets of conditions. The points corresponding to the same grade lie on a line, and the lines for all the grades are parallel to each other. This is not surprising since it reflects the fact that the additional component $x_{null}^T$ is free only to move along $(A - k = 1)$ directions in this case. The lines in figure (2.7) are the projection of this direction onto the Cs1-P-plane.

Again the new process conditions were implemented using the detailed fundamental simulation process, and the resulting qualities plotted in figures (2.8) and (2.9).
Figure 2.8: PCR-Results for Number and Weight Average Molecular Weight, when $Y = [\text{MWN} \ \text{MWw}]$

Figure 2.9: PCR-Results for Long and Short Chain Branching, when $Y = [\text{MWN} \ \text{MWw}]$
In general, the grades obtained by implementing the new process conditions come close to the desired ones as in case 1. Looking at these PCR-results in figures (2.8) and (2.9), we see another interesting but expected feature: While measured qualities (MWN and MWw) resulting from different $x^{T}_{pred}$ suggestions for one grade (figure (2.8)) are almost the same, the resulting unmeasured qualities (LCB and SCB in figure (2.9)) display much more variation. (For each grade the resulting unmeasured qualities are connected by dashed lines in figure (2.9).) This can be explained by the random component $x^{T}_{null}$ in $x^{T}_{pred}$. It was shown that according to the model this component does not affect the measured qualities $Y$ (MWN and MWw), however, there is no guarantee that it does not affect the unmeasured qualities (LCB and SCB). According to figure (2.9), $x^{T}_{null}$ has a larger effect on SCB than on LCB.

The LCB- and SCB-values for each new grade fall almost on a line as well. However, this is not necessarily expected since here we only see the effect of changing the operating conditions for one grade along a line. The fact that for this process these qualities also change in an almost one-dimensional way indicates that at least in that range the process is adequately described by a linear model.

Since in this example it was not necessary to specify a fixed value for any of the manipulated variables in $X$, there is no incentive to apply the Conditional Expectation approach. After all, as mentioned in section 2.2.5, the lack of a model for $X$ leaves us without the benefits of a solution window from this method. We only obtain a single set of process conditions for each grade. However, for illustrative purposes figure (2.10) shows the Conditional Expectation results for the case when $Y = \{MWN \ MWw\}$.

Comparison of Conditional Expectation and PCR results shows that the grades obtained from the Conditional Expectation approach may even be better than the ones from PCR. However, there is no way to predict which method will perform better for a certain process.
Figure 2.10: Conditional Expectation and PCR Results when $Y = [\text{MWn MWw}]$

The score space obtained from a PLS model on this data set is nearly the same as the score space obtained from PCR, and the results are similar. Resulting plots from PLS are therefore not shown.

**Some Results from Inversion of the Standard Regression Model Based on OLS**

This section shows some results from inversion of the standard regression model (2.1) with parameters estimated by OLS, although the discussion in sections 2.2.2 and 2.3 advise against the use of such a model in this context.

Inversion of the standard regression model (2.1) fitted by OLS again gives very poor results in this case. One of the problems with this OLS approach is illustrated in figure (2.11). This figure shows the values for the two initiator flow rates (Fl1 and Fl2) to the two zones as they have been designed by OLS-inversion (○) and by PCR-inversion (×) for the different grades.

The way the process has been run in the past, the two rates are nearly identical
at all times. However, since the OLS regression model (2.1) does not provide a model for the $X$-space, it cannot and does not respect the existing correlation between the two rates, whereas it can be seen that the results from PCR-inversion continue to obey the correlation structure of the past data. The numerical values from OLS-inversion for these data show clearly that the correlation structure is not respected. The general expression for $\hat{x}_{\text{new}}^T$ is $\hat{x}_{\text{new}}^T = y_{\text{des}}^T \cdot M^T$; in other words, $\hat{x}_{\text{new}}^T$ is a linear combination of the rows of $M^T$ which themselves are linear combinations of the rows of $V^T$. For OLS-inversion

$$M_{\text{OLS}}^T = \left( Y^T \cdot U \cdot \Sigma^{-2} \cdot U^T \cdot Y \right)^{-1} \cdot Y^T \cdot U \cdot \Sigma^{-1} \cdot V^T .$$

Thus, the $i$-th column in

$$\left( Y^T \cdot U \cdot \Sigma^{-2} \cdot U^T \cdot Y \right)^{-1} \cdot Y^T \cdot U \cdot \Sigma^{-1}
$$

contains the weights for the $i$-th principal component (row $i$ in $V^T$):
In the historical data the rows in $V^T$ are weighed in decreasing order, here however, we see that OLS-inversion gives the largest weights to components 2 and 5, and almost no weight to the first component which actually is the most important. In comparison, weights for $V^T$ in

$$M_{PCR}^T = \left( Y^T \cdot U_A \cdot L_A^T \cdot Y \right)^{-1} \cdot Y^T \cdot U_A \cdot \Sigma \cdot V_A^T$$

- even when the maximum number of principal components is retained and $U_A = U$ and $V_A = V$, are

<table>
<thead>
<tr>
<th></th>
<th>1. PC</th>
<th>2. PC</th>
<th>3. PC</th>
<th>4. PC</th>
<th>5. PC</th>
<th>6. PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>row 1</td>
<td>-1.3954</td>
<td>-1.0746</td>
<td>-0.0959</td>
<td>0.0186</td>
<td>0.0021</td>
<td>-0.0010</td>
</tr>
<tr>
<td>row 2</td>
<td>1.8630</td>
<td>-0.5726</td>
<td>0.0833</td>
<td>-0.0362</td>
<td>0.0021</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

One can see that the absolute values of the coefficients decrease from left to right, thereby assigning the most weight to the first principal component and increasingly less weight to later ones.

The effect of ill-conditioning in $X$ can be seen in figure (2.12) where the newly designed values for the inlet pressure and inlet temperature ($P$ and $T_{in}$) are plotted versus each other.

While the values from PCR are spread out, the values from OLS lie almost on one spot. This can be explained by the fact that one of the two rows in $M^T_{OLS}$ is more than 50% shorter than the other, i.e. one of the rows is tending towards zero as explained in section 2.3. Accordingly, the $\hat{x}_{new}^T$ move almost exclusively in the single direction of the dominant row in $M_{OLS}^T$. Projection of this row onto the P/T-in-plane results almost in a point since the OLS-conditions fall closely together in this plane.
Figure 2.12: Process Conditions from OLS- and PCR-Inversion; (OLS-values are lumped together at approximately $P = 2750$ atm, $T_{in} = 46$ K)

In other words OLS-inversion gives approximately the same inlet temperature and pressure for each new desired grade.

The example shows that it is possible to infer process conditions on the basis of historical data which yield new product grades that are close to the desired ones. If one is not satisfied with the result, one can use the information from the actual implementation of the new conditions and add it to the data set of existing grades. Continued iteration of model building and inversion with updated data sets should further refine the process conditions in order to finally achieve a quality which is sufficiently close to the desired quality.
2.5 Extension to Nonlinear Systems

The previous sections of the chapter are based on the assumption that the process can be adequately described by linear behavior over the range of product grades used in the model. As the new desired grade must lie within that same range the same linear behavior can be expected. However, there exist cases where nonlinear methods would be more appropriate. The main features of the previously introduced linear model inversion are therefore extended to nonlinear model inversion.

Since the concepts for the design of new process conditions remain the same in the nonlinear case an extension to nonlinear methodology is straightforward. Several issues and cautions involved with the use of nonlinear models and techniques versus linear ones are raised at the end of this section. A more general discussion on areas of linear and nonlinear process behavior is given in the introduction to this chapter.

There are several ways of introducing nonlinearities into latent variable projection methods:

1. As long as the nature of the nonlinearities is known, it is often possible to transform the variables (e.g. by taking logarithms) and thereby convert the problem into a linear one. (The inversion methodology for linear models could then be followed, and in a final step the obtained process conditions $x_{pred}^T$ would have to be transformed back to their original units.)

2. The $X$-matrix can be augmented by adding additional columns containing the squared ($x_i^2$) and cross product terms ($x_i \cdot x_j$) of other columns. However, Berglund and Wold (1997) show that fairly general quadratic relationships can be treated by augmenting the $X$-matrix with columns that contain only the squared values ($x_i^2$) of the variables. (For the current inversion problem, however, this would cause difficulties since we cannot expect the new process conditions to be able to reconstruct the exact relationship between a $X$-variable...
and its squared term - unless additional restrictions among the columns of the augmented \( X \)-matrix are imposed during the optimization.\)

3. Nonlinear PLS methods (Wold et al. 1989, Wold 1992) retain linear models for the \( X \)- and \( Y \)-spaces, but use nonlinear quadratic, cubic or spline inner relationships between the latent variables of the \( X \)- and \( Y \)-spaces.

In this section we only consider the inversion of Nonlinear PLS models with a quadratic inner relationship between \( X \)- and \( Y \)-scores. Inversion of models with more complex nonlinear inner relationships follows logically from there. However, as discussed later in this section it is often not advisable to use any more complex nonlinear relationships than quadratics.

### 2.5.1 Nonlinear PLS and Inversion

As indicated above, a Nonlinear PLS (NPLS) model represents \( X \) and \( Y \) as linear latent variable spaces

\[
\hat{X} = T \cdot P^T
\]
\[
\hat{Y} = Z \cdot Q^T
\]

and introduces a nonlinear inner relationship between the latent variables \( z_a \) of the \( Y \)-space and \( t_a \) of the \( X \)-space such as the quadratic relationship:

\[
z_a = c_{0a} + c_{1a} \cdot t_a + c_{2a} \cdot t_a^2 \quad a = 1, \ldots, A
\]

where \( t_a \) and \( z_a \) are the score vectors for the \( a \)-th dimension, and \( c_{ia} \) are scalar coefficients. The algorithm and more details on the NPLS-model can be found in (Wold et al. 1989).

The requirements for the design of new product quality specifications are the same as in the linear case: For a desired new grade \( y_{des}^T \) we want to obtain new
process conditions $x^T_{new}$ such that the established nonlinear model still holds. Since consistency with past operating strategies is required, the new process conditions must again maintain the same covariance structure as the data from the past.

To obtain this covariance structure from the model we have to decompose $T$. Since the columns in $T$ are orthogonal we can write

$$
T = U_A \cdot \Sigma_A \quad (n \times A) \quad (A \times A) \quad (n \times A)
$$

(2.16)

where $U_A$ has orthonormal columns: $U^T_A \cdot U_A = I$ (identity matrix). and $\Sigma_A$ is diagonal and contains the standard deviations of the different score vectors in $T$.

From here $\dot{X}$ can be rewritten as

$$
\dot{X} = U_A \cdot \Sigma_A \cdot P^T
$$

where $\Sigma_A$ and $P^T$ contain the variance and correlation information of the past data $X$. We therefore postulate for the new process conditions:

$$
\dot{x}^T_{new} = \dot{u}^T_{new} \cdot \Sigma_A \cdot P^T \quad (1 \times m) \quad (1 \times A) \quad (A \times A) \quad (A \times m)
$$

(2.17)

and

$$
\ddot{t}^T_{new} = \dot{u}^T_{new} \cdot \Sigma_A,

= \begin{bmatrix}
\dot{u}_{new1} \cdot \sigma_1 & \dot{u}_{new2} \cdot \sigma_2 & \ldots & \dot{u}_{newA} \cdot \sigma_A
\end{bmatrix}
$$

In this way we ensure that the new process conditions indeed follow the same covariance structure as the past ones.

This enforcement of consistency, i.e. of maintaining the historical covariance structure, reduces the problem from estimating $m$ correlated process variables $x_j$ ($j = 1, \ldots, m$) to the estimation of $A$ independent latent variables $\hat{u}_{newa}$ ($a = 1, \ldots, A$), where $A < m$. The problem that remains is to find these latent values $\hat{u}_{newa}$ such that
the resulting product quality is the desired \( y^T_{des} \). However, \( y^T_{des} \) has to fit the model as well, that is:

\[
y^T_{des} = z^T_{new} \cdot Q^T
\]

\((1 \times k) \cdot (A \times k)\)

(2.18)

where for each element \( a \) in \( z^T_{new} \):

\[
z_{new_a} = c_{0a} + c_{1a} \cdot \hat{u}_{new_a} \cdot \sigma_a + c_{2a} \cdot \hat{u}_{new_a}^2 \cdot \sigma_a^2.
\]

(2.19)

In order to combine equations (2.18) and (2.19), we first rewrite the vector \( y_{j,des} \) in terms of single elements \( y_{j,des} \):

\[
y_{j,des} = \sum_{a=1}^{A} z_{new_a} \cdot q_{ja} \quad \text{for } j = 1, \ldots, k
\]

where \( q_{ja} \) is the \((j, a)\)-th element of \( Q \). \( y_{j,des} \) can now be expressed as

\[
y_{j,des} = \sum_{a=1}^{A} \left( c_{0a} + c_{1a} \cdot \hat{u}_{new_a} \cdot \sigma_a + c_{2a} \cdot \hat{u}_{new_a}^2 \cdot \sigma_a^2 \right) \cdot q_{ja} \quad \text{for } j = 1, \ldots, k.
\]

(2.20)

This finally states our problem: We have to solve a set of \( k \) equations (2.20) to obtain \( A \) values \( \hat{u}_{new_a} \). They can then be substituted into equation (2.17):

\[
z^T_{new} = \hat{u}_{new}^T \cdot \Sigma_A \cdot P^T
\]

to give process conditions that are:

1. designed to give the desired product properties \( y^T_{des} \)

2. consistent with past operating strategies.

For one new grade \( y^T_{des} \) (2.20) is a set of \( k \) quadratic equations in \( A \) unknowns \( \hat{u}_{new_a} \). Depending on the number \((k)\) of equations and \((A)\) of variables, a solution can be found using different numerical techniques. In the case of a square system \((A = k)\) we can apply the Multivariate Newton’s Method to find roots of the system.
An overdetermined system where \( A < k \) is not very likely in this type of problem. It would imply that there is significant variation in the product quality \( Y \) that cannot be explained by the process variables present in \( X \) - either because \( X \) does not contain all the variables that affect \( Y \) or because the noise ratio in \( Y \) is very high. Either case would make it fairly impossible to "design" the desired quality \( y_{des}^T \) by estimating the proper values for the process variables in \( X \). However, a mathematical solution for this overdetermined system can be found e.g. by using a least squares technique to find the \( A \) unknowns minimizing the error sum of squares in equations (2.20).

As before, the most common case is one where the process variables vary in more independent directions than the number of quality variables specified in \( Y \): \( A > k \). Such an underdetermined nonlinear equation system can again have an infinite number of solutions which - according to the model - will all result in the desired \( y_{des}^T \).

Solving equation (2.20) is an optimization problem and many algorithms are available to find solutions. One possible way is to minimize

\[
\mathbf{f} (\hat{\mathbf{u}}_{\text{new}})^T \cdot \mathbf{f} (\hat{\mathbf{u}}_{\text{new}})
\]

using different starting values. where

\[
\mathbf{f}^T \cdot \mathbf{f} = \sum_{j=1}^{k} \left[ y_{j,\text{des}} - \sum_{a=1}^{A} \left( c_{0a} + c_{1a} \cdot \hat{u}_{\text{new}_a} \cdot \sigma_a + c_{2a} \cdot \hat{u}_{\text{new}_a}^2 \cdot \sigma_a^2 \right) \cdot q_{ja} \right]^2 \quad (2.21)
\]

An underdetermined system \( (A > k) \) can be accommodated by fixing a grid of values for \( A - k \) unknowns \( \hat{u}_{\text{new}_a} \), and solving through optimization for the remaining \( k \) \( \hat{u}_{\text{new}_a} \). Connecting the solutions defines a curve or surface of possible solutions.

We are only interested in solutions that are consistent with past operating strategies. Maintaining the covariance structure by multiplying the estimates \( \hat{u}_{\text{new}}^T \) with \( \Sigma_A \) and \( P^T \) addresses the structural part of this consistency. We also have to make sure that the new conditions are in the same range as the past ones. In the historical data set the \( A \) columns in \( U_A \) are orthonormal. If all \( A = m \) latent
dimensions were included in the model, the matrix $U_A$ would have orthonormal rows as well. In other words, the 2-norm of a row $u^T$ would be $\|u^T\|_2 = 1$ in that case. Typically however, the number of latent dimensions used in the model is far less than the maximum possible number ($A < m$) and therefore $\|u^T\|_2 < 1$. Thus, a quite generous upper limit for all the $U_{\text{new}}^T$ obtained from equations (2.20) is $\|U_{\text{new}}^T\|_2 < 1.0$. This provides us with an appropriate area over which we should conduct the search for solutions for the nonlinear equations (2.20). The locus of solutions may be a nonlinear curve itself, however, we are only interested in that part of the solution curve that intersects the historical $u$-cluster. Following from the discussion on $\|U_{\text{new}}^T\|_2$ above, a generous limit for this cluster is generated by the bounds of $-1$ and $1$ for each value $U_{\text{new}}^T$. Therefore, the search for solutions to equations (2.20) can be limited to the unit hypersphere defined by these bounds. For systems with $A - k = i$ and $i > 0$, one can select a fine grid of fixed values within the $i$-dimensional hypersphere for $i$ of the $A$ elements $U_{\text{new}}$, and solve the resulting square system for the remaining $k$ elements of $U_{\text{new}}$. Different starting values at each grid point can help finding multiple solutions. However, for only slight nonlinear process behavior we do not expect multiple solutions to fall within the historical data cluster; and solutions outside of this cluster are not viable in our case.

Out of the $U_{\text{new}}$-vectors obtained from the optimization, we retain those which

1. result in zero residuals $f^T \cdot f$ and

2. fall within the region of historical $u$-vectors.

In order to fulfill the second condition, we plot the historical $u$-values in the $u_i/u_j$ planes and check whether the new $U_{\text{new}}^T$-values fall inside the cluster of the historical ones. A new prediction $U_{\text{new}}^T$ could actually have $\|U_{\text{new}}^T\|_2 < 1.0$ but still fall outside of the cluster. However, this visual test becomes rather unwieldy for large $A$, and one may choose to plot the $u$-planes only for the first few latent dimensions and
compute $\|\hat{u}^T_{\text{new}}\|_2$ instead.

After a valid set of solutions $\hat{u}_{\text{new}}$ has been found, the corresponding process conditions can be calculated using equation (2.17)

$$\hat{x}^T_{\text{new}} = \hat{u}^T_{\text{new}} \cdot \Sigma_A \cdot P^T.$$  

Similar to the linear case, for $A > k$ the method for nonlinear model inversion results in a whole window of operating conditions. Process knowledge has to be combined with the methods described above to select the most promising of the conditions.

An application of the approach is shown in the following section.

### 2.5.2 Case Study on LDPE-Polymerization

This study deals with the exact same process and data sets as described in section 2.4. The matrix $X$ consists of the six manipulatable variables (inlet temperature of the feed ($T_{\text{in}}$), pressure in the reactor ($P$), initiator feed rate to both zones ($F_{i1}$, $F_{i2}$) and solvent flow to both zones ($C_{s1}$, $C_{s2}$)). They have been varied such that three latent directions prevail in $X$. Again we will look at the two cases when $Y$ contains three or two quality variables respectively:

1. $Y_3 = [\text{Conv MWn MWw}]$, $k = 3$

2. $Y_2 = [\text{MWn MWw}]$, $k = 2$

where Conv stands for conversion, and MWn, MWw for number and weight average molecular weight.

**Case 1: $Y_3 = [\text{Conv MWn MWw}]$**

A nonlinear PLS-model with quadratic inner relation was calculated for the 9 historical grades. Results are shown in table 2.3:
Table 2.3: Percentage of Variation Explained by NPLS in $X$ and $Y$ (Case 1)

<table>
<thead>
<tr>
<th>No. of Principal Components</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>55.73</td>
<td>96.28</td>
<td>99.44</td>
<td>99.93</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>case 1: $Y_3$</td>
<td>35.93</td>
<td>70.65</td>
<td>95.80</td>
<td>96.73</td>
<td>98.63</td>
<td>99.00</td>
</tr>
</tbody>
</table>

Comparison to the linear PLS-model (not displayed) shows that the two models explain about the same amount of variation in each dimension. Furthermore, the loading vectors of both models are very similar which suggests that the individual latent variables describe the same type of variation in both models. Consequently, the NPLS-model will use three latent variables as well. Compared to the linear PCR-model (see table 2.2 in section 2.4.3), NPLS explains the same amount of variation in $X$ and a little bit more in $Y$. In table 2.4 the coefficients for the first three inner relations of the NPLS-model suggest mostly linear behavior for the first two dimensions. Only the third dimension has a significant square term. We may conclude that the data show only slightly nonlinear behavior, and that a linear model is perhaps adequate. However, the nonlinear approach using the NPLS-model introduced above, is carried out for illustrative purposes.

Table 2.4: Inner Relationships for the First 3 Principal Components (Case 1)

<table>
<thead>
<tr>
<th>Principal Components No.</th>
<th>Quadratic Inner Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\hat{z}_1 = -0.25 + 0.56 \cdot t_1 + 0.09 \cdot t_1^2$</td>
</tr>
<tr>
<td>2</td>
<td>$\hat{z}_2 = -0.16 + 0.66 \cdot t_2 + 0.07 \cdot t_2^2$</td>
</tr>
<tr>
<td>3</td>
<td>$\hat{z}_3 = -0.12 + 2.12 \cdot t_3 + 0.71 \cdot t_3^2$</td>
</tr>
</tbody>
</table>

In order to obtain process conditions for six new grades (see section 2.4.2) the $3 \times 3$ nonlinear equation system was inverted using a least squares algorithm to minimize $f^T \cdot f$ according to equation (2.21). No multiple solutions were found. For each of the six new desired grades the resulting estimated process conditions were
implemented on the fundamental simulation and the corresponding product qualities obtained. Plots on the desired and achieved product qualities can be seen in figures (2.13) and (2.14).

![Graph](image)

Figure 2.13: NPLS and PCR Results for Number and Weight Average Molecular Weight when \( Y = [\text{Conv MWN MWW}] \)

In terms of the two molecular weight properties MWN and MWw (figure (2.13)), NPLS achieves qualities that are closer to the desired values in all but one case (grade 10). However, NPLS does not perform as well with respect to conversion (figure (2.14)), where PCR-grades are closer to the desired values in most cases.

To obtain a better feel for the overall closeness to the desired results for all six grades, table (2.5) shows the Mahalanobis distances between desired \( y_{des}^T \) and actually achieved grade quality \( y_{ach}^T \):

\[
(y_{des}^T - y_{ach}^T) \cdot [\text{Cov}(Y)]^{-1} \cdot (y_{des}^T - y_{ach}^T)^T
\]

where \([\text{Cov}(Y)]^{-1}\) is the inverse of the covariance matrix of the historical \( Y \)-data. Except for grades 11 and 14, PCR-Inversion shows better results than NPLS- Inver-
Table 2.5: Mahalanobis Distances to Desired Grades Based on the Specified Y-Variables (Case 1): \( (y_{des}^T - y_{ach}^T) \cdot [Cov(Y)]^{-1} \cdot (y_{des}^T - y_{ach}^T)^T \)

<table>
<thead>
<tr>
<th></th>
<th>grade 10</th>
<th>grade 11</th>
<th>grade 12</th>
<th>grade 13</th>
<th>grade 14</th>
<th>grade 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR-Inversion</td>
<td>0.096</td>
<td>0.149</td>
<td>0.086</td>
<td>0.277</td>
<td>0.162</td>
<td>0.061</td>
</tr>
<tr>
<td>NPLS-Inversion</td>
<td>0.199</td>
<td>0.144</td>
<td>0.151</td>
<td>0.414</td>
<td>0.024</td>
<td>0.094</td>
</tr>
</tbody>
</table>
sion. Apparently, the good performance of NPLS with respect to MWn and MWw does not outweigh its poorer performance in conversion. It may be interesting to investigate a bit further why the NPLS-inversion achieves good results for MWn and MWw but not for conversion. If we were to look at the Q-loading plots for $Y'$ of the NPLS-model (not shown), we would find that it is the third dimension which focuses on explaining conversion. However, the variance in $X$ for this component amounts only to 1.48 while the variance for the first and second dimensions are 24.30 and 19.22. (These variances can be found from the diagonal elements of $T^T \cdot T$). In other words, there is little information in $X$ with respect to changes in conversion as compared to changes in molecular weight properties. The same is true for the linear PCR-model. Therefore, we conclude that although the nonlinear model may provide a slightly better fit for the historical data set than the linear model, this does not necessarily lead to better inversion results, as we can see in this example.

This is an indication that there is no advantage gained by a nonlinear approach for this particular data set. The coefficients for the inner relationships in the NPLS-model support this by giving very low weight to the quadratic term in the first two dimensions (see table 2.4).

**Case 2: $Y_2 = [\text{MWn MWw}]$**

In this case only two of the previous three quality variables are assumed to be available for inclusion in $Y$: $Y_2 = [\text{MWn MWw}]$. Apart from this, the same data are used as above. Table 2.6 shows the amount of variation explained in the resulting NPLS-model.

Again, linear and nonlinear PLS explain almost the same amounts of variation (linear PLS results are not shown). The loadings are very similar in both models as well, which indicates that the individual principal components describe the same type of variation. With respect to the corresponding linear PCR-model (refer to table 2.2),
Table 2.6: Percentage of Variation Explained by NPLS in X and Y (Case 2)

<table>
<thead>
<tr>
<th>No. of Principal Components</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>46.49</td>
<td>96.35</td>
<td>99.44</td>
<td>99.93</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>case 2: Y2</td>
<td>58.93</td>
<td>97.53</td>
<td>99.29</td>
<td>99.63</td>
<td>99.84</td>
<td>99.95</td>
</tr>
</tbody>
</table>

A slight improvement can be observed for the amount of Y explained by the NPLS-model.

The coefficients for the inner relationships of the three dominant principal components can be seen in Table 2.7.

Table 2.7: Inner Relationships for the First 3 Principal Components (Case 2)

<table>
<thead>
<tr>
<th>Principal Components No.</th>
<th>Quadratic Inner Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\hat{z}_1 = -0.23 + 0.64 \cdot t_1 + 0.10 \cdot t_1^2$</td>
</tr>
<tr>
<td>2</td>
<td>$\hat{z}_2 = -0.11 + 0.52 \cdot t_2 + 0.04 \cdot t_2^2$</td>
</tr>
<tr>
<td>3</td>
<td>$\hat{z}_3 = -0.01 + 0.43 \cdot t_3 + 0.04 \cdot t_3^2$</td>
</tr>
</tbody>
</table>

The first dimension shows a slight curvature between the latent vectors $t_1$ and $\hat{z}_1$ as illustrated in Figure 2.15. This is ideal for the application of a nonlinear PLS model.

For this case of only $k = 2$ Y-variables we have two nonlinear equations but $A = 3$ latent variables $\hat{u}_{new}$. Defining a grid from $-1$ to $1$ (stepwidth 0.1) for $\hat{u}_{new1}$, at each grid point the resulting minimization problem $f^T \cdot f$ for $\hat{u}_{new2}$ and $\hat{u}_{new3}$ was solved according to equation (2.21). Since the grid consists of 21 different values for $\hat{u}_{new1}$, we obtained 21 $\hat{u}_{new}$-vectors for each of the six new desired grades. Out of those, the ones with $\|\hat{u}_{new}\|_2 < 1.0$ were selected. This resulted in eight $\hat{u}_{new}$-vectors for grade 10, two for grade 14 and three for each of the remaining four grades. (Note: In order to be complete in terms of possible multiple solutions for equation system (2.20), we would have to repeat the procedure and search for solutions along a grid for $\hat{u}_{new2}$ and $\hat{u}_{new3}$ as well.) Figure 2.16 shows historical and predicted $u$-values in
the $u_1 - u_3$-plane. Indicated by (×) are the $u$-values of the actual grades 10–15 as they were simulated. (In a “real” example we would not know these latter values.) Different predictions $u$ for each grade are denoted by (o), and connected by a dashed line to indicate solution curves. The $u$-values of the historical grades used in the model are shown as (∗).

For the new and desired grades 11 to 15 the loci of solutions follow slight curves. Those curves for the different grades are parallel. This is expected, since from grade to grade the coefficients of the nonlinear equation system stay the same; only the constants $y_{j, des}$ change (see equations (2.20)). However, the eight solutions for grade 10 show very clearly that the locus of solutions is parabolic, suggesting at least two solutions for some $\tilde{u}_{new3}$ values. On the other hand, the strongly curved part of the solution locus for grade 10 is outside of the range of the historical $u$-data, and should not be considered for possible implementation. We suspect that for models
with only slightly curved inner relations, the inversion should also result in only one slightly curved branch of solutions inside of the range of the data. If there are multiple solutions for some of the grid points, they will most likely lie outside of the historical data range.

As before, the corresponding process conditions for these \( \hat{u}_{new} \)-vectors were computed, implemented and the results compared to the results from the linear PCR-inversion. For this underdetermined case we have several process conditions for each grade, all of which should yield the same values of \( MWn \) and \( MWw \), but in practice lead to slightly varying values (see e.g. figure (2.8) for results from the linear PCR approach). The reason for this is mismatch between the empirical model (here PCR or NPLS) and the fundamental simulation model. Therefore, we compare for each grade the results that came closest to the desired \( MWn \) and \( MWw \) values in each model inversion. The Mahalanobis distances between desired and achieved grades

Figure 2.16: \( u \)-Values of Different Predictions for Grades 10-15, the Actual \( u \)-Values, and the \( u \)-Values of the Historical Grades Used for Modeling
are calculated to find these “closest” grades for each method. The results in table 2.8 reveal that NPLS performs much better than linear PCR for this Y-space.

Table 2.8: Mahalanobis Distances to Desired Grades Based on the Specified Y-Variables (Case 2): \( (y_{des}^T - y_{ach}^T) \cdot [Cov(Y)]^{-1} \cdot (y_{des}^T - y_{ach}^T)^T \)

<table>
<thead>
<tr>
<th></th>
<th>grade 10</th>
<th>grade 11</th>
<th>grade 12</th>
<th>grade 13</th>
<th>grade 14</th>
<th>grade 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR-Inversion</td>
<td>0.0082</td>
<td>0.0144</td>
<td>0.0169</td>
<td>0.0434</td>
<td>0.0133</td>
<td>0.0159</td>
</tr>
<tr>
<td>NPLS-Inversion</td>
<td>0.0004</td>
<td>0.0114</td>
<td>0.0002</td>
<td>0.0029</td>
<td>0.0004</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

The same results are displayed in figure (2.17). We can see that the grades obtained from NPLS inversion are indeed very close to the desired ones.

![Figure 2.17: NPLS and PCR Results for Number and Weight Average Molecular Weight when Y = [MWN MWw]](image)

Figure 2.17: NPLS and PCR Results for Number and Weight Average Molecular Weight when Y = [MWN MWw]

Already in the previous case which was set up to achieve pre-specified conversion values at the same time as the desired molecular weight properties, NPLS showed good performance with respect to the molecular weights. However, in the
present case where the model aims only at achieving the molecular weight properties, it performs better. Obviously, the relationship between X and MWN and MWw is suited very well for a nonlinear approach.

Although in this case only MWN and MWw are assumed to be available as quality measurements, we know that they account for only 76% of the total variation in all five product properties. Conversion is mostly uncorrelated with MWN and MWw, and accounts for most of the remaining variation. We pretend conversion measurements are unavailable in this case, however, since we are dealing with a simulated process, we can obtain the values for conversion. In order to compare this case (2) of only two Y-variables to the previous case (1) which also included conversion in Y, we should look at the resulting conversion in case 2 as well (figure (2.18)).

![Figure 2.18: NPLS and PCR Results for Number Average Molecular Weight and Conversion when Y = [MWN MWw]](image)

Even though the modeled nonlinearities are not tailored towards achieving
certain conversion specifications, NPLS grades come fairly close to the actual conversion of the desired grades. Except for grade 11, the conversions obtained from the nonlinear approach are closer to the “desired” conversion values than their counterparts from linear PCR. Of course, if one compares the conversion results between the two NPLS-models from case 1 and 2, the model of case 1, where Conv is a specified \( Y \)-variable, achieves better results - with the exceptions of grades 10 and 13, though.

We can also include the unmeasured (but yet simulated) conversion values in the distance measure between desired and achieved grades. In this way we can compare NPLS to PCR with respect to all three qualities.

Table 2.9 shows the distances for the six grades from the desired grades when \( y^T_{des}, y^T_{ach} \) and \( Y=[\text{MWN MWw}] \) are augmented with conversion measurements:

\[
y^T_{des-ext} = [y^T_{des} \quad \text{Conv}\_{des}]
\]

\[
y^T_{ach-ext} = [y^T_{ach} \quad \text{Conv}\_{ach}]
\]

\[
Y_{ext} = [Y \quad \text{Conv}]
\]

Table 2.9: Mahalanobis Distances to Desired Grades Based on the Specified \( Y \)-Variables and Conv (Case 2): \( (y^T_{des-ext} - y^T_{ach-ext}) \cdot [\text{Cov}(Y_{ext})]^{-1} \cdot (y^T_{des-ext} - y^T_{ach-ext})^T \)

<table>
<thead>
<tr>
<th></th>
<th>grade 10</th>
<th>grade 11</th>
<th>grade 12</th>
<th>grade 13</th>
<th>grade 14</th>
<th>grade 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR-Inversion</td>
<td>3.3149</td>
<td>0.2436</td>
<td>7.5285</td>
<td>0.0731</td>
<td>6.2851</td>
<td>0.8607</td>
</tr>
<tr>
<td>NPLS-Inversion</td>
<td>0.0293</td>
<td>0.6927</td>
<td>0.2006</td>
<td>0.0301</td>
<td>0.0432</td>
<td>0.4496</td>
</tr>
</tbody>
</table>

As expected, we see smaller distances for NPLS grades than for PCR. An exception is grade 11, where NPLS results in a much larger conversion than the actual desired grade. Thus, in this case that does not specify conversion in \( Y \), the nonlinearities in NPLS which are geared towards modeling the relationship between \( X \) and the molecular weights, seem to capture the relationship between \( X \) and the unmeasured conversion as well, and far better than the linear PCR model.
Of course one has to keep in mind that in this case we have chosen the process conditions for each grade shown in the plots from a whole window of possible solutions. The inversion forces these solutions to result in the desired $Y'$-variables, however, any unspecified properties can still change for the different process conditions. This is due to the fact that the space of all five quality measures is really 3-dimensional (see section 2.4.3), and although the two molecular weights are specified, the remaining properties are not fully determined. Since these latter properties change with the different process conditions for one grade, the distance of the resulting grade to the desired one changes as well, if unspecified properties are included in the distance measure. In other words, choosing a different set of operating conditions for one grade results in a change of the distance recorded in table (2.9). (The distances in table (2.8) remain invariant.)

In general it is not possible to predict the effect that the process conditions obtained from NPLS-inversion will have on the quality properties not specified in $Y$. The nonlinearities described by the model between $X$ and the variables in $Y$ may in some cases serve as a fairly good model for the unspecified quality properties as well, in which case those other qualities will be achieved quite consistently with the past grades. This is true for the present example. However, it is also possible that the nonlinear model of the relationship between $X$ and $Y$ is a very poor predictor of the relationship between $X$ and the unspecified qualities. In this case these unspecified qualities may turn out as quite different from the values in the past.

In summary, we can conclude the following: while in the previous case with three $Y$-variables the linear approach should be preferred, in the case with two $Y$-variables the NPLS approach has definitely advantages over the linear PCR method. As mentioned before, the nature of the inner relationships of the two NPLS-models already suggested this. In case 1, the first two inner relations were almost linear with a significant quadratic term showing only in the third dimension, whereas in case
2 the slight curvature in the inner relation of the first dimension made a nonlinear approach much more promising.

The following section contains further discussion on the use of nonlinear versus linear model inversion.

2.5.3 Discussion on the Use of Nonlinear Model Inversion

Especially in the second case where only two quality variables (MWn and MWw) were used, the nonlinear model inversion worked well. However, as mentioned in the introduction to this chapter one should be very careful when using nonlinear methods. A first and very rough indication when and when not to use a nonlinear method is the size of the domain covered by the grades used in the modeling. In section 2.1 it was discussed that data which either stem from a very large domain of grades or even from two different classes of grades may invite use of a nonlinear model. Data that are spread over a smaller area can typically be described by linear models, although one may find a case where the process behaves in a nonlinear manner even throughout a small range of product grades.

Many risks are involved in the context of nonlinear model inversion. Therefore, before resorting to nonlinear techniques, one should investigate whether a linear approach is indeed unreasonable. Since nonlinear models need more parameters than a linear model we have to be careful not to overfit the data. For the inversion this would lead to the same type of problems as for prediction. Especially when looking at this problem of product design, we typically do not have that many historical grades available. Nine grades - as we had in our example - is already a lot. This relatively small number of grades leads to low residual degrees of freedom. In linear PLS, the residual degrees of freedom are often calculated as \((n - A - 1) \cdot (m - A)\) (SIMCA 1996), where \(n\) is the number of observations (grades), \(A\) is the number of latent variables and \(m\) is the number of X-variables. There is no discussion in the
literature about degrees of freedom for NPLS. However, since for a quadratic inner relationship one is fitting two parameters more in each dimension than for a linear inner relationship, one possibility for estimating the degrees of freedom would be: \((n - A - 1) \cdot (m - A) - 2A\). In the case of nine grades, six \(X\)-variables and three principal components this results in only nine degrees of freedom left in the residuals. An additional fourth component would exhaust all degrees of freedom.

If the process behavior is very nonlinear, more parameters are needed to describe these nonlinearities, and most likely there will not be enough data to allow prediction of new process conditions. We therefore recommend not to go beyond a quadratic inner relationship. In (Wold et al. 1989, Wold 1992) the use of cubic inner relationships and cubic splines is described. However, these nonlinear inner functions are highly sensitive to slight changes in the data, and inversion of such a model may result in poor predictions.

We recommend that one builds a linear PLS-model first. If the inner relationships between the scores of the dominant latent vectors display a slight curvature, a NPLS-model with quadratic inner relations can be calculated. Furthermore, if the NPLS-model explains about the same amount of variation as the linear PLS-model but with fewer latent variables, then both models can be considered equivalent. However, for inversion one would prefer the nonlinear model with fewer latent variables.

As mentioned before, nonlinear modeling requires a larger number of data points than linear modeling. If only a few grades are available nonlinear techniques become more dangerous in terms of overfitting, and very sensitive to one extreme grade. At the very least one would want the data points to be spread out well in order to increase the predictability of the model. On the other hand, if we have many grades available we may want to select only those surrounding the desired new grade and build a model on grades in the immediate area around that new grade. Such a local model may describe the process behavior in that particular region better than
a global model based on all available data. Furthermore, we may be able to describe the process in that smaller region by a linear model which is more robust than a nonlinear one for reasons discussed above.

As a rule of thumb we think that as long as a linear model describes the relationship between $X$ and $Y$ sufficiently (i.e. not much worse than a nonlinear model) and with a small number of principal components ($A - k \leq 2$) one should choose this linear model over a nonlinear one.

2.6 Conclusions

In this chapter, we have developed a methodology based on latent variable techniques that uses historical process data to determine a window of process operating conditions within which a product with desired quality specifications can be produced. It is important not only to design process conditions that will achieve the specified qualities $y^*_{des}$, but also to ensure that these process conditions retain the same covariance structure as the historical data. This is the key point which leads to preference of multivariate methods such as PLS and PCR over other methods. The advantage of these multivariate latent variable models arises from the fact that unlike other methods they model the $X$-space (process conditions) as well as the $Y$-space (product quality). In particular, the advantages of PCR and PLS over OLS have been outlined: Multivariate techniques reduce the dimensionality of the variable space and at the same time lead to better conditioned model inverses, which respect past operating strategies. This ensures that these new conditions maintain the covariance structure of past operating conditions, as well as giving a product quality (including unmeasured properties) that is in some way consistent with the other grades.

Any approach based solely on empirical data entails certain limitations: Not every set of process and quality variables is equally suitable for this purpose of product
development. Analyzing the model can provide insight into how much information exists in the data. If only about 50% of $Y$ or less is explained due to the lack of information in $X$, of course any predictions of process conditions from that model will not be very reliable. Such data do not contain the necessary information for a databased approach to this problem.

Since we are using empirical models which rely on data from past production, the new quality specifications should lie within the range of past ones, although in some cases slight extrapolations may be possible. For the same reason it is not possible to predict process conditions which are completely different from any others used before. Even though there may exist other sets of conditions which achieve the desired quality, this methodology cannot find these conditions if they are outside the range and correlation structure of the data and therefore the validity of the model.

The approach discussed here does not necessarily lead to the desired quality in one step. Repeatedly following the sequence of data analysis, running the suggested process conditions, adding the result to the data set, rebuilding the model and re-inverting it should lead to the desired result in a few iterations. It is expected that some fine tuning of the resulting process conditions may require a few additional designed experiments. The process conditions obtained from the inversion should serve as a very good starting point for the experimental design.
Chapter 3

Industrial Applications

The databased approach for product design as developed in the previous chapter is applied to three different examples from industry: two batch processes (sections 3.1 and 3.3) and one continuous process (section 3.2). Each example involves data from a real industrial plant, and discusses some of the difficulties encountered in the particular data set.

This chapter also introduces the use of Mahalanobis distances to evaluate the prediction results, and the use of cross-validation to estimate the potential predictability of the method for a given data set. The different steps of the methodology are outlined in more detail in the first application, while application sections 3.2 and 3.3 are kept shorter in terms of theoretical explanations. The figures and tables shown are not necessarily of the same type in each example. They were selected depending on what appeared to be most illustrative in each case.
3.1 A Batch Solution Polymerization Process

3.1.1 The Process

The process is an operator controlled batch polymerization producing several different polymer grades. Differences in the amounts and sometimes in the types of the ingredients together with variations in the temperature profile are used to obtain the different grades of the polymer.

A process cycle consists of two stages: In the first stage three (in some products four) preliminary ingredients are mixed and heated. The operator ensures that sufficient time is allowed for most of the material to react to generate one of the prime components of the polymer. After cooling down, more ingredients are added for the second stage of the cycle. The batch is heated up to a top temperature, possibly cooled down a bit again and then held at constant temperature. Heating and cooling follow a certain pattern which is unique to the specific grade. While the batch is held at constant temperature the operator takes frequent samples to make sure that certain quality measurements follow a desired trend and that all specifications are finally met.

One of the major problems with a databased approach in this example arises from the fact that the process is not computerized. The temperature trajectories are implemented by operators through successive heating and cooling of the reactor. Of course, this leads to a far less consistent way of operation than an automation of the temperature profile would provide.

3.1.2 The Data

The data available to us from this process are on eight different grades of the polymer. For each grade we have the production run sheets of four batches.

In a preliminary step the information on the production run sheets needs to
be converted to an $X$-matrix consisting of the variables that are manipulated in the process and a $Y$-matrix containing the final quality measurements on the respective batches.

As indicators of product quality we have measurements on three variables for the final polymer from each batch run: the viscosity of the polymer solution at 50% non volatiles, the acid value, and the color.

There are two groups of process variables that can be varied from grade to grade: The first group is composed of the amounts of different ingredients charged to the batch for the eight different grades. Ten ingredient variables listed in the production sheets are used in this study.

The second group consists of selected characteristics of the temperature profile of the batches for each grade. The time dependent information in these profiles is converted into a number of variables that are common to all grades. In general, these variables describe the important features of the profiles such as certain slopes, maximum or minimum temperature, times for certain phases of the batch, etc.

The time allowed for the first stage of a batch cycle depends on how fast the ingredients are being converted. The second stage is initiated only after sufficient material has been converted in the first stage. In this sense the reaction temperature-time information of the first stage is under closed loop conditions and cannot be used as manipulatable variables in the context of this problem. One simply has to assume that for any future new grade sufficient time will also be allowed to achieve the proper conversion in the first stage of the process.

In its final stage each batch is held at constant temperature until all three quality variables are within certain specification limits. For some grades the batch is held at the top temperature (maximum temperature), for others the temperature is lowered after a certain amount of time and then held at the new temperature. In spite of these differences the trajectories have to be summarized by a set of variables
that are common to all available grades. For this study four variables were found that are used to represent the temperature trajectories of the different grades: time at top temperature, top temperature, the weighted average of time at top temperature and holding time, and the weighted average of top and holding temperature.

After selecting the $X$- and $Y$-variables that are to be used in the study, the "nominal" variable values for each grade have to be determined. Data on four batch runs are available for each grade. However, some of these batches encountered some problems during production and have to be considered as outliers. After eliminating such problem batches, the variable values of the remaining batches within each grade are averaged. This results in a $(8 \times 14)$ $X$-matrix, and a $(8 \times 3)$ $Y$-matrix for the eight grades. The eight averaged sets of data may not yield the exact operating specifications for the eight grades, but they are representative data for different product grades from this plant.

Remark: On a general note it may be worthwhile mentioning the case where a single batch has resulted in out of specification product. This batch may be a very valuable one and should not be discarded before a closer investigation. If the cause for the deviation is a mistake of running the process, i.e. a wrong se: point value or an ingredient charged by the wrong amount, this batch could be used to represent an additional grade. This grade would be represented only by this single batch as compared to an average over several batches for the other grades, but the benefit would be the enriched information content in the data by the additional point.

Unfortunately, this data set is far from the ideal situation envisioned in our approach. We are faced with the following difficulties:

1. A closer look at the temperature information of the second stage reveals that the batches are of very different run lengths even within one product grade. The heating and cooling patterns differ a lot for different products. Even within one product the suggested trajectories are not always closely followed in all
the batches due to unavailability of tanks or other problems with a batch. Consequently, the batch to batch variation within a product grade is often as large as the grade to grade variation for the temperature data.

2. With respect to the ingredient data comes another major difficulty: Not all the ingredients are used for all the product grades. In fact, two ingredient variables display zeros for four out of the eight grades. Another ingredient variable shows only two non-zero values out of eight. This causes the following problem for the databased approach: the actual variation between the non-zero values (which contains the important information) is overpowered by the much larger difference from zeros to non-zero values. This poses a problem for this study as will become apparent in the results section 3.1.6. Ideally the method was developed for data on products that all use the same set of ingredients, and vary in temperature settings and feedrates etc., but not in the type of ingredients.

3. Out of the three quality variables, color does not have very narrow specifications. The color of a grade is acceptable as long as it is not too dark, i.e. as long as the color value does not exceed a certain limit. Consequently, the color measurements show large variation even within a single grade. However, since we only have two other quality properties, we retain color as a variable and in the inversion pretend that a certain color value would be desirable.

The above difficulties with the data provide a very severe test of the methodology proposed in chapter 2.

3.1.3 Preliminary Data Analysis

In order to specify possible new product quality we can only specify independent quality characteristics. A PCA-model on all three quality measurements (viscosity,
acid value and color) results in a two component model that explains 97\% of the total variation. This means that although we have three quality measurements, only two of them are independent of each other. Selective PCA (Roffel et al. 1989) is used to choose an appropriate pair of $Y$-variables that best defines this two dimensional quality space. In this case selective PCA selects acid value and color as the most complementary set of $Y$-variables which account for 94.1\% of the total variation.

Both a PCR- and a PLS-model were developed using this new $Y$-matrix. The $X$-matrix consists of ten variables that denote ingredient amounts and the four variables that characterize the temperature profile as described above. The sum of squares explained by these models is shown in table 3.1:

<table>
<thead>
<tr>
<th>Latent Variable No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$-PLS</td>
<td>45.70</td>
<td>75.59</td>
<td>83.72</td>
<td>93.63</td>
</tr>
<tr>
<td>$Y$-PLS</td>
<td>44.54</td>
<td>78.26</td>
<td>94.26</td>
<td>97.07</td>
</tr>
<tr>
<td>$X$-PCR</td>
<td>49.69</td>
<td>76.54</td>
<td>89.00</td>
<td>93.95</td>
</tr>
<tr>
<td>$Y$-PCR</td>
<td>30.20</td>
<td>65.24</td>
<td>69.44</td>
<td>92.46</td>
</tr>
</tbody>
</table>

The fourth PLS dimension is insignificant since there are insufficient degrees of freedom in the data. In three components PCR explains far less of $Y$ than PLS and only a bit more of $X$. This suggests that even though the amount of variation explained in $X$ is not that different for the two models, the type of variation used in $X$ to explain $Y$ must be different. An important requirement when choosing a model for the inversion is to keep \( A - k \) as small as possible in order to have a small number of $x_{null}^T$-directions in the $x_{pred}^T$-space that do not affect $Y$. Under these premises we are obviously dealing with a data set for which a PLS-model must be preferred over a PCR-model. 94\% of $Y$ is explained by 84\% of $X$. Since the objective of this data analysis is to construct a new $x_{pred}^T$ from a desired $y_{des}^T$ we would like to explain as
much as possible of the data matrix $X$ in the model. However, considering the large uncertainty in this particular data set, 84% is quite good.

The PLS-model approximates $X$ by $A = 3$ principal components. The number of independent variables in $Y$ is $k = 2$. Therefore, the first inversion step

$$u_{new}^T = y_{des}^T \cdot \left( Q \cdot S_A^2 \cdot Q^T \right)^{-1} \cdot Q \cdot S_A$$

where $T = U_A \cdot S_A$

predicts only that part of the new $x_{pred}^T$ which is predictive of $Y$. It lies in a 2-dimensional subspace of the 3-dimensional $X$-space. Only when the additional direction $x_{null}^T$ (one dimensional) is added to $x_{new}^T$ the complete solution space is obtained. Any point $x_{pred}^T$ in that space results - according to the model - in the same $y_{des}^T$.

Within the scope of this work it was not feasible to predict process conditions for a completely new grade and have the company run those new conditions in the plant. Therefore, in order to get an idea of how well the approach would work for a new grade, we have to test it on each of the eight grades provided in the data set. We pretend not to know the $x$ values of one of the grades, build the model on the remaining seven grades, and using the quality of the grade that was left out as $y_{des}^T$ we predict the required process conditions $x_{pred}^T$. The whole procedure is repeated for each of the eight grades.

### 3.1.4 Mahalanobis Distances Between Nominal and Predicted Process Conditions

By “nominal” process conditions we mean here the values for the original $(8 \times 14)$ $X$-matrix obtained by building averages over the batches for each grade. The term “nominal" is used here rather than “true" or “actual" to stress the fact that the values recorded in the original $X$-matrix are only estimates of the true grade conditions based on averages on only two to four somewhat variable batches. Therefore, these nominal
values may have error in that they are only rough estimates of the conditions for the "true" grades in this example.

For each grade that we try to predict we obtain a solution space along a line which is defined by the point

\[ x_{\text{new}}^T = u_{\text{new}}^T \cdot S_A \cdot P^T \]

and the direction \( x_{\text{null}}^T \).

If we want to compare the nominal conditions to the new predicted ones, we should do that in the \( X \)-space spanned by the model since this is the space where new process conditions are predicted. We therefore have to project the nominal process conditions onto the model space of \( X \), and compare this projection to the predicted conditions given by the solution space of the inversion.

It is possible to check how close the projection of the recorded or nominal \( x \) values lies to this solution space. If the projection of the nominal \( x \) values lies within the solution space, then of course we can select \( x_{\text{pred}}^T \) such that the distance between projection of \( x^T \) and \( x_{\text{pred}}^T \) is zero. If, however, the solution space does not contain the projection of the nominal \( x^T \), the closest we can get to those projected nominal conditions is determined by the distance of projected \( x^T \) to that solution space. Since the \( X \)-variables are not independent of each other and not of equal variance the distance measure between two vectors (projected \( x^T \) and \( x_{\text{pred}}^T \)) should take that into account. In other words, instead of the Euclidean distance we should use the Mahalanobis distance. The Mahalanobis distance between two points \( x_1 \) and \( x_2 \) in the model space of \( X \) is defined by

\[
dist_{\text{Mah}} = (x_1^T - x_2^T) \cdot [Cov(\hat{X})]^{-1} \cdot (x_1 - x_2)
\]

The difference \( x_1 - x_2 \) is scaled by the inverse covariance matrix of \( \hat{X} = T \cdot P^T \) as described by the PLS-model since it is this approximation \( \hat{X} \) of \( X \) that we try to construct from the inversion.
We now show that the above Mahalanobis distance in the $X$-space is equivalent to the Euclidean distance in the "U"-space. (Note: $U_A$ here refers to the scaled $T$-matrix ($T = U_A \cdot S_A$).)

$$\text{dist}_{Eucl} = (u_1^T - u_2^T) \cdot (u_1 - u_2) \quad .$$  \hfill (3.2)

If we invert the transformation between $u^T$ and $x^T$: ($x^T = u^T \cdot S_A \cdot P^T + \text{error}$) we obtain:

$$u^T = x^T \cdot P \cdot (P^T \cdot P)^{-1} \cdot S_A^{-1} \quad .$$

Substituting this into equation (3.2) we find

$$\text{dist}_{Eucl} = (x_1^T - x_2^T) \cdot \left( P \cdot (P^T \cdot P)^{-1} \cdot S_A^{-2} (P^T \cdot P)^{-1} \cdot P^T \right) \cdot (x_1 - x_2) \quad .$$  \hfill (3.3)

Taking into account that $\hat{X}$ is already mean-centered the covariance of $\hat{X}$ is proportional to $\hat{X}^T \cdot \hat{X}$. It can be shown that equations (3.1) and (3.3) are equivalent since

$$\frac{1}{n-1} \left[ \text{Cov}(\hat{X}) \right]^{-1} = (\hat{X}^T \cdot \hat{X})^{-1} = \left( P \cdot S_A \cdot U_A^T \cdot U_A \cdot S_A \cdot P^T \right)^{-1} = \left( P \cdot S_A^2 \cdot P^T \right)^{-1} \quad \text{since} \quad U_A^T \cdot U_A = I \quad .$$  \hfill (3.4)

Since $P$ is a ($m \times A$) matrix (in this case ($14 \times 3$)) and $m > A$, the inverse in equation (3.4) is singular. The pseudoinverse (Moore-Penrose inverse) of the expression, however, is equal to the middle term in equation (3.3).

This shows that the Mahalanobis distance between the "closest" $x_{\text{pred}}^T$ and the projection of the nominal $x^T$ is equivalent to the Euclidean distance between the corresponding $u_{\text{pred}}^T$ and $u^T$. The "closest" point in the solution space to the projected nominal conditions is conveniently found in the "U"-space: One transforms $x^T$ into $u^T$ and projects $u^T$ onto the solution space. This projection onto the solution space,
the closest point $u_{\text{closest}}^T$, or its equivalent in the $X$-space ($x_{\text{closest}}^T$), is the best we can do in terms of reconstructing the nominal process conditions.

If the Euclidean distance between $u_{\text{closest}}^T$ and the nominal point $u^T$ is smaller than the distances of all the other grades used in the model to $u^T$, then the inversion approach has provided an improvement in the sense that selecting a feasible point from the solution space and implementing it should - according to the model - get us closer to the desired quality of the new grade than any of the process conditions used in the past.

This comparison of distances to the projected nominal values of process conditions for the grade that we consider our "new" grade is the best we can do in this example to evaluate the performance of the method. Considering the large uncertainty in the data due to the large variation in operating procedures and the measurement errors, we do not expect the solution space to indeed contain the projection of the nominal process conditions $x^T$. (Even for a more ideal data set the model error would very likely cause the solution space to "pass" the projected nominal point rather than contain it.) Comparing the distances though should give an idea of the benefits derived from this methodology for a particular data set. In fact it is suggested that this cross-validation procedure (where each of the available grades is left out of the data set and redesigned using only the remaining grades in the model) always be carried out before one takes upon the design of a completely new grade. The results of such a cross-validation study will provide a feel for the ability of the methodology to achieve reliable inversion results for the particular data set being used. This is discussed in more detail in the next section.

3.1.5 Cross-Validation as a Test for Predictability

The procedure described in chapter 2 of this thesis provides an estimate of the process conditions ($\hat{u}^T_{\text{new}}, \hat{x}_{\text{new}}^T$) or the region of equivalent conditions if $A > k$. It would also
be desirable if we could get some measure of the goodness of these inversions or some meaningful measure of the uncertainty in the predictions and the uncertainty in the quality that might be obtained by implementing the predicted conditions. This section discusses some of the difficulties in getting a closed form solution for these uncertainty regions, and then proposes a modified cross-validation approach which can be used to provide a measure of the predictability of the inversion procedure based on the existing grade data.

In most cases each of the existing grades has been produced on several different occasions, yielding a range of process conditions \( x_i^T \), and product quality measures \( y_i^T \) for each of the \( i = 1, \ldots, n \) grades. In the inversion procedure presented in chapter 2 the average conditions \( \bar{x}_i^T \) and \( y_i^T \) obtained for each grade \( i \) are used in the matrices \( X \) and \( Y \) on which the PCR- or PLS-model is built. This use of the average conditions rather than the individual realizations within each grade is preferred since it is the covariance structure among the grades that one wants to describe with the model rather than the usually quite different covariance structure that is present in the data within a given grade. The former arises from the way the plant conditions are changed by the operators to achieve the desired grade changes, while the covariance structure within grades arises from the naturally occurring disturbances (e.g. impurities, heat transfer variations, etc.) and variations in the plant operation that may change slightly each time one returns to produce the same grade. However, the several realizations of each grade can be used to obtain an estimate of the error covariance matrices for the means of the grades, i.e. \( \Sigma_X, \Sigma_Y \). Then, for \( A = k \) in principle one could derive an expression for the uncertainty in \( \hat{u}_{new}^T \) and the uncertainty in the quality \( y_{pred}^T \) that arises due to this uncertainty in the original data base. \( y_{prod}^T \) denotes the predictions for the quality that results when the predicted conditions \( \hat{u}_{new}^T \) are implemented.

The model inversion leads to expressions for \( \hat{u}_{new}^T \) and \( y_{pred}^T \) which are nonlinear
functions in the matrices of grade means $X$, $Y$:

$$
\hat{u}_{new}^{T} = y_{des}^{T} \cdot \left(U_{A}^{T} \cdot Y\right)^{-1} \quad \text{and} \quad U_{A} = f(X, Y)
$$

$$
y_{pred}^{T} = \hat{u}_{new}^{T} \cdot U_{A}^{T} \cdot Y.
$$

One could use Taylor Series expansions of these models to obtain approximate expressions for the error covariance matrices of $\hat{u}_{new}^{T}$ and $y_{pred}^{T}$ in terms of $\Sigma_{X}$ and $\Sigma_{Y}$. If one then assumed a distribution form for $\hat{u}_{new}^{T}$ and $y_{pred}^{T}$ (e.g., multivariate normal), then approximate confidence regions could be established for these predictions. However, given the often small number of grade replications available to estimate $\Sigma_{X}$, $\Sigma_{Y}$, the Taylor Series linearizations, and the unknown distributional properties for $\hat{u}_{new}^{T}$ and $y_{pred}^{T}$, this approach is not considered further in this thesis. Instead a data-driven cross-validation approach which is described below has been developed to provide some assessment of the predictability of the inversion methodology.

Cross-validation is often used as a procedure for assessing the predictability of models based on the existing data (e.g. Wold (1978)). The multiple correlation coefficient $R^2$ indicates for a given model how well the model fits the data. $R^2$ gives the percentage of variation in $Y$ fitted by the model relative to the total variation in $Y$:

$$
R^2 = 1 - \frac{SS[Y - \hat{Y}]}{SS[Y]},
$$

where "SS" stands for the sum of squares of the matrix elements of the argument. The limits on $R^2$ are $0 \leq R^2 \leq 1$. A $R^2$-value close to one indicates a good fit of the model to the given data.

In order to estimate the predictive power of the model cross-validation is used, and a $Q^2$-value corresponding to $R^2$ can be calculated. For this purpose one or more data points are left out of the data set, a model is built on the remaining data and then used to predict $y_{pred}^{T}$ for the points that were left out. This is done until each of the data points has been left out and predicted once. Using the matrix of predictions
\( \hat{Y} \) the Prediction Error Sum of Squares (PRESS) is calculated as:

\[
\text{PRESS} = \sum_{i=1}^{n^*} \text{SS}[Y_i - \hat{Y}_i] .
\]

The \( Q^2 \)-value

\[
Q^2 = 1 - \frac{\text{PRESS}}{\text{SS}[Y]}
\]

is a measure for the predictability of the model. It gives the percentage of variation in \( Y \) that can be predicted by cross-validation relative to the total variation in \( Y \). A high \( Q^2 \) indicates good predictive power of the model. Usually \( Q^2 \) is less than \( R^2 \). A \( Q^2 \)-value close to \( R^2 \) says that the model is nearly as good in predicting new points as it is in fitting the given data.

In the context of model inversion the interest is not in predicting \( Y \) but in assessing the quality of the inversion for a given data set. One wants to measure how reproducible the inversion process is, and what predictability one can expect when inverting for a new grade. A modification of the cross-validation approach described above for the inversion procedure is given in the following few paragraphs.

Each of the grades is in turn eliminated from the data set, a model is built on the remaining \( n - 1 \) grades and inverted to predict process conditions for the grade that was left out. It may not make sense to do this for all \( n \) available grades, since some of them might require extrapolation when predicted from the other grades, and hence this would lead to a large prediction error. It is therefore recommended to do the inversion only for those \( n^* \) grades that lie within the range of the remaining \( n - 1 \) grades in each case (i.e. ones for which one might consider using this inversion procedure). Based on the predicted \( u^T \)-values obtained from the inversion in that way \( (u^T_{\text{predicted}}) \), and the \( u^T \)-values of the nominal process conditions when projected onto the \( u \)-space of the model \( (u^T_{\text{nominal}}) \), one can calculate a measure for goodness of prediction \( (\bar{Q}^2) \) similar to the \( Q^2 \)-value described above:

\[
\bar{Q}^2 = 1 - \frac{\sum_{i=1}^{n^*} \text{SS}[u^T_{\text{nominal}_i} - u^T_{\text{predicted}_i}]}{\sum_{i=1}^{n^*} \text{SS}[u^T_{\text{nominal}_i} - \bar{u}^T_{\text{nominal}}]} ,
\]
CHAPTER 3. INDUSTRIAL APPLICATIONS

where the summation is carried out over those grades \( i \) for which new process conditions are predicted (i.e. which do not require extrapolation). \( \bar{u}_{n_{\text{nominal}}}^T \) is the mean of the projected \( u \)-values of the nominal conditions \( (u_{n_{\text{nominal}}}^T) \) for the \( n^* \) grades.

\( \hat{Q}^2 \) is a measure for how well a model inversion on the given data set is able to predict process conditions for a new grade - provided this new grade does not require extrapolation out of the range of the given data. It represents the fraction of the total sum of squares of the \( u^T \)-values of the nominal process conditions in the data set that can be predicted by the inversion method. Of course in cases where the inversion solution consists of a whole space of solution points \( (\lambda > k) \) such a measure becomes less meaningful. However, in this case one could use as \( u_{\text{predicted}}^T \) the prediction for each grade that comes closest to the projected nominal conditions (in Mahalanobis sense).

In a similar way, the \( R^2 \)-value can be adapted to measure the goodness of the inversion when process conditions \( \hat{u}_{\text{new}}^T \) are predicted for the points included in the model. All \( n \) data points are used in this model, however, the inversion is again carried out only for those \( n^* \) grades that are used in \( \hat{Q}^2 \). A value \( \hat{R}^2 \) can be calculated as

\[
\hat{R}^2 = 1 - \frac{\sum_{i=1}^{n^*} SS(u_{\text{model}}^T - u_{\text{predicted}}^T)}{\sum_{i=1}^{n^*} SS(u_{\text{model}}^T - \bar{u}_{\text{model}}^T)},
\]

where \( u_{\text{model}}^T \) is the \( u^T \)-vector for grade \( i \) obtained from the model on all \( n \) grades, and \( \bar{u}_{\text{model}}^T \) is the average of the \( n^* \) values of \( u_{\text{model}}^T \) used in the comparison. The \( n^* \) values of \( u_{\text{predicted}}^T \) are the values predicted from inversions of the model based on all \( n \) data points. In order to allow for a valid comparison the number of latent variables included in the model used in \( \hat{R}^2 \) should be the same as the number of latent variables used in the models for \( \hat{Q}^2 \).

A value of \( \hat{R}^2 \) close to one implies a good fit of the model inversion on the given data, and a value of \( \hat{Q}^2 \) close to one implies good predictive power of the model inversion for a new grade. Usually, one would expect \( \hat{Q}^2 < \hat{R}^2 \). A \( \hat{Q}^2 \)-value close
to $\hat{R}^2$ indicates that the prediction of new grades is as reliable as the prediction of grades used in the model.

The described cross-validation procedure can assist in developing some feel for the performance of the method for a given data set. If $\hat{Q}^2$ is considered too small then the predictability of the method for that particular data set is dubious, and the inversion should be used with caution.

However, even this cross-validation approach is questionable if only a small number of data points is available. In such a case the best one can do is to perform the model inversion on maybe one of the given grades that does not require extrapolation when predicted from the other grades. This result together with the predicted process conditions for the actual new desired grade may assist in assessing the potential of the methodology in a small data set.

### 3.1.6 Model Inversions and Results

As indicated in section 2.2.1 we have to check in the $Y$-space if the new $y_{des}^T$ is consistent with the grades that are used in the model. Only if it is consistent does it make sense to attempt the design of process conditions for the new grade. Although in this study we only test the approach on the already available grades, we follow this procedure in order to see whether a prediction of process conditions for each grade from a model on the remaining grades looks to be feasible.

For that purpose a PCA-model is calculated on $Y$. This $Y$-matrix consists of acid values and color values for the seven grades included in the model. These two $Y$-variables have been selected as the most complementary pair that spans the full 2-dimensional quality space. Therefore, any new $(1 \times 2)$ $y_{des}^T$ falls into that PCA-plane giving a squared prediction error of zero. However, the locus of the predicted $t$-scores for $y_{des}^T$ compared to the $ts$ of the other grades gives an indication whether the new grade is within the range spanned by the "historical" data. If $y_{des}^T$ falls outside
the cluster of the other $y$s in the score space, then an inversion of a model on the $X/Y$-data will be an extrapolation for this new grade and should not be trusted.

Figure 3.1 illustrates such a case where the predicted $t$-scores of the new grade (grade 1) are far from the $t$-scores of the other seven grades (2–8) used to build the model. Prediction of process conditions for this grade would thus require extrapolation outside of the range of the model. Design of process conditions corresponding to this grade 1 should not be attempted based on the given data set.

![Figure 3.1: t-Scores from PCA on $Y$; Grades Included in Model (2–8) and the New Grade (1)](image)

Using grade 8 as the "new" grade for which process conditions have to be found gives a completely different picture. As shown in figure 3.2 the predicted $t$-scores for this grade lie right in the center of the other $t$-scores. From this aspect a decent prediction of new process conditions for grade 8 can be expected.

The corresponding tests on $Y$-spaces for all the grades were carried out, the models were inverted and the solution spaces of possible process conditions calculated.
Two questions are of interest:

1. Are the new process conditions $x_{pred}^T$ closer to the nominal $x^T$ than the process conditions for any of the other grades?

2. How close to the nominal variable values do the predicted values get?

To answer question 1 we employ Mahalanobis distances between the nominal and the predicted process conditions as well as Mahalanobis distances between nominal conditions and conditions used in the model. Since we cannot account for the error between the model and the nominal conditions, these Mahalanobis distances are calculated within the model space of $X$. For the nominal conditions and the conditions of the model grades we therefore use their projections onto the model space of $X$ in these distances. The predictions of new process conditions lie in this model.
space by default. In the following sections we imply that the Mahalanobis distances are calculated within the space spanned by the model.

Table 3.2 shows the Mahalanobis distances of the closest predictions $x_{closest}^T$ to the projected nominal conditions for grades 2 to 8. The table contains seven double rows for each of the seven grades that were predicted. The first of the two rows for each grade lists the numbers of the grades used in the respective model. The second row displays the distance between the projection of the nominal $x^T$ of the new grade, and the point in the solution space closest to it, as well as the distances from the projected nominal $x^T$ to the projections of all other grades. If for a new grade the projected process conditions of another grade are closer to the projected $x^T$ than the closest prediction, this is indicated by showing the number of the grade that was closer in boldface and with three asterixes (*). Predictions for grade 1 are not included as they were very poor, which was expected from the preliminary analysis on the PCA-scores for $Y'$. That same PCA analysis of $Y'$ for the other grades indicated slight extrapolations for grades 2, 3, 4, and a somewhat larger extrapolation for 7, but not severe enough to render the method inapplicable.

For four out of the seven grades under consideration (4, 5, 6, and 8) the inversion results in a solution space that comes closer to the projection of the actually used process conditions than any other grade. As mentioned before, this shows that the method does indeed provide an improved guess of conditions for those new grades. As for the other three grades, there exist one or two historical grades with process conditions closer to the projected $x^T$ than the inversion could provide. For the prediction of grades 2 and 3 this is not very surprising since the $t_1-t_2$ plot of the PLS-model on all eight grades (see figure 3.3) shows that grades 2, 3 and 5 are very close in terms of the characteristics of their process conditions. It is therefore not surprising that some of these three grades are as close or closer to the projected nominal $x^T$ than the closest prediction. For grade 7 it is a bit more surprising that grade 6 has process
Table 3.2: Mahalanobis Distances in the Model Space between Grades Used in the Model, Closest Prediction, and the Nominal $x^T$

<table>
<thead>
<tr>
<th>No.</th>
<th>closest</th>
<th>1</th>
<th>3 ...</th>
<th>4</th>
<th>5 ...</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>closest</td>
<td>1</td>
<td>1.6085</td>
<td>0.1182</td>
<td>0.8406</td>
<td>0.0308</td>
<td>1.0294</td>
<td>1.0147</td>
</tr>
<tr>
<td>3</td>
<td>closest</td>
<td>1</td>
<td>2 ...</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.4488</td>
<td>0.5747</td>
<td>0.2692</td>
<td>0.8054</td>
<td>0.7025</td>
<td>1.6190</td>
<td>1.7475</td>
<td>1.5522</td>
</tr>
<tr>
<td>4</td>
<td>closest</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.0052</td>
<td>1.9334</td>
<td>0.5368</td>
<td>0.9391</td>
<td>0.3304</td>
<td>0.5713</td>
<td>0.1571</td>
<td>1.3789</td>
</tr>
<tr>
<td>5</td>
<td>closest</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.0594</td>
<td>2.2079</td>
<td>0.2719</td>
<td>0.2483</td>
<td>0.8205</td>
<td>1.0468</td>
<td>0.8828</td>
<td>0.4986</td>
</tr>
<tr>
<td>6</td>
<td>closest</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.1713</td>
<td>0.9146</td>
<td>1.2973</td>
<td>1.1826</td>
<td>1.0219</td>
<td>1.1836</td>
<td>0.1773</td>
<td>0.9463</td>
</tr>
<tr>
<td>7</td>
<td>closest</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6 ...</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.1421</td>
<td>1.9500</td>
<td>1.3963</td>
<td>1.3788</td>
<td>0.5840</td>
<td>1.0752</td>
<td>0.1310</td>
<td>1.2187</td>
</tr>
<tr>
<td>8</td>
<td>closest</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.0751</td>
<td>1.4885</td>
<td>0.7078</td>
<td>0.2699</td>
<td>2.8407</td>
<td>0.9026</td>
<td>0.6861</td>
<td>1.1455</td>
</tr>
</tbody>
</table>

conditions slightly closer to those of 7 than any predictions. However, the plot of the PCA-scores on $Y$ (not shown) indicates that in order to predict conditions for grade 7, the extrapolation outside of the model region is larger than for the other grades (except for grade 1). Also in the PLS score space shown in figure 3.3, grade 6 is the model grade which is closest to grade 7.

To answer the second question on the performance of the inversion approach for this data set, table 3.3 shows the nominal $x^T$-values for grades 2–8 and below each of these are shown the values for the predictions that according to the Mahalanobis distances came closest to the projection of these $x^T$-values.

The results, as we can observe, are not all that favorable. Of major concern are the negative values for some of the variables. It is impossible to implement values less than zero for any of the variables. However, a closer look at these variables for
which some predictions are negative reveals that those are the variables which have zero values in the data set. As discussed in section 3.1.2, these variables pose less than ideal conditions on a databased approach.

In general, one may resort to the fact that the solutions displayed in table 3.3 are not the only possible process conditions for the grades. For each grade we have a whole space of solutions along the line of the respective $x_{null}^T$. The particular points shown in table 3.3 are the ones which are the closest (in the sense of the Mahalanobis distance in the $X$-space) to the nominal process conditions that we know. It is thus possible to search along this line of solutions for points that look more favorable, i.e. have no negative values for any of the variables. Table 3.4 shows different process conditions for grade 2 as they are changed along that line without affecting the qualities in $Y$. Unfortunately, in this case it is not possible e.g. for grade 2 to find a point with only positive variable entries. Variable 3 reaches a positive value only in the last row of the table. However, at this point variable 5 has assumed
Table 3.3: 14 Variables for the Nominal Process Conditions $x^T$ and Closest Predictions $x^T_{closest}$

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>V10</th>
<th>V11</th>
<th>V12</th>
<th>V13</th>
<th>V14</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.61</td>
<td>.11</td>
<td>0</td>
<td>.0019</td>
<td>.00</td>
<td>.014</td>
<td>.064</td>
<td>.008</td>
<td>.23</td>
<td>.0003</td>
<td>.56</td>
<td>302</td>
<td>250</td>
<td>302</td>
</tr>
<tr>
<td></td>
<td>.63</td>
<td>.12</td>
<td>-.003</td>
<td>.0019</td>
<td>-.004</td>
<td>.020</td>
<td>.010</td>
<td>.22</td>
<td>.0001</td>
<td>.49</td>
<td>278</td>
<td>250</td>
<td>377</td>
<td>244</td>
</tr>
<tr>
<td>3</td>
<td>.58</td>
<td>.09</td>
<td>-.006</td>
<td>.018</td>
<td>.00</td>
<td>.014</td>
<td>.21</td>
<td>.0002</td>
<td>.60</td>
<td>125</td>
<td>250</td>
<td>399</td>
<td>234</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.61</td>
<td>.09</td>
<td>-.006</td>
<td>.0029</td>
<td>.00</td>
<td>.007</td>
<td>.25</td>
<td>.0003</td>
<td>.62</td>
<td>102</td>
<td>254</td>
<td>252</td>
<td>244</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.51</td>
<td>.08</td>
<td>0</td>
<td>.0015</td>
<td>.00</td>
<td>.025</td>
<td>.34</td>
<td>.0002</td>
<td>.89</td>
<td>83</td>
<td>250</td>
<td>402</td>
<td>228</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.51</td>
<td>.10</td>
<td>.011</td>
<td>.0015</td>
<td>.004</td>
<td>.029</td>
<td>.30</td>
<td>.0003</td>
<td>.79</td>
<td>311</td>
<td>245</td>
<td>320</td>
<td>245</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.58</td>
<td>.11</td>
<td>0</td>
<td>.0017</td>
<td>.00</td>
<td>.047</td>
<td>.26</td>
<td>.0002</td>
<td>.58</td>
<td>259</td>
<td>250</td>
<td>259</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.62</td>
<td>.12</td>
<td>-.002</td>
<td>.0002</td>
<td>.00</td>
<td>.099</td>
<td>.23</td>
<td>.0002</td>
<td>.60</td>
<td>186</td>
<td>253</td>
<td>327</td>
<td>244</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.49</td>
<td>.11</td>
<td>.015</td>
<td>.0015</td>
<td>.059</td>
<td>.003</td>
<td>.32</td>
<td>0</td>
<td>.84</td>
<td>508</td>
<td>240</td>
<td>508</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.52</td>
<td>.11</td>
<td>.008</td>
<td>.0014</td>
<td>.042</td>
<td>.005</td>
<td>.30</td>
<td>.0002</td>
<td>.74</td>
<td>348</td>
<td>244</td>
<td>402</td>
<td>241</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.47</td>
<td>.11</td>
<td>.017</td>
<td>.0012</td>
<td>.062</td>
<td>0</td>
<td>.34</td>
<td>.003</td>
<td>.85</td>
<td>401</td>
<td>240</td>
<td>401</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.43</td>
<td>.10</td>
<td>.014</td>
<td>.0013</td>
<td>.093</td>
<td>-.01</td>
<td>.38</td>
<td>.0000</td>
<td>.99</td>
<td>437</td>
<td>240</td>
<td>536</td>
<td>232</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.61</td>
<td>.14</td>
<td>0</td>
<td>.0019</td>
<td>.00</td>
<td>.006</td>
<td>.24</td>
<td>0</td>
<td>.54</td>
<td>381</td>
<td>250</td>
<td>381</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.58</td>
<td>.13</td>
<td>.011</td>
<td>.0017</td>
<td>.014</td>
<td>.009</td>
<td>.24</td>
<td>.0002</td>
<td>.62</td>
<td>444</td>
<td>242</td>
<td>423</td>
<td>245</td>
<td></td>
</tr>
</tbody>
</table>

A negative value. Furthermore, we have moved already outside of the range of the historical data: e.g. the value of variable 1 in this last row is larger than any value of variable 1 in the data set (compare table 3.3). It would be possible to extend the table going upwards and still be within the range of the historical grades, but the value of variable 3 which is negative at the top of the table would grow even more negative. If one wanted to consider only process conditions that are in the range of those applied before, one would probably have to draw the line in table 3.4 after the 8th prediction, after which variable 1 starts exceeding its values from the historical data set. The other variables also become smaller or larger than their historical counterparts. Only points above this line can be considered within the range of previous conditions.

Apart from the variables with negative entries, this table shows another interesting feature. The conditions have been changed along the $x_{null}$-axis where - according to the model - the changes in $x$ should not affect the quality of the product. The way in which the variables change with respect to each other indicates that
Table 3.4: 14 Variables for the Nominal Process Conditions $z^T$ and Predictions Along the Line of the $z_{null}^T$-Space for Grade 2

<table>
<thead>
<tr>
<th></th>
<th>Var 1</th>
<th>Var 2</th>
<th>Var 3</th>
<th>Var 4</th>
<th>Var 5</th>
<th>Var 6</th>
<th>Var 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominal</td>
<td>0.6071</td>
<td>0.1094</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0437</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>predicted</td>
<td>0.5960</td>
<td>0.1003</td>
<td>-0.0050</td>
<td>0.0017</td>
<td>0.0178</td>
<td>0.0320</td>
<td>0.0106</td>
</tr>
<tr>
<td></td>
<td>0.6013</td>
<td>0.1041</td>
<td>-0.0047</td>
<td>0.0018</td>
<td>0.0144</td>
<td>0.0302</td>
<td>0.0106</td>
</tr>
<tr>
<td></td>
<td>0.6065</td>
<td>0.1080</td>
<td>-0.0043</td>
<td>0.0018</td>
<td>0.0110</td>
<td>0.0284</td>
<td>0.0105</td>
</tr>
<tr>
<td></td>
<td>0.6117</td>
<td>0.1118</td>
<td>-0.0040</td>
<td>0.0018</td>
<td>0.0077</td>
<td>0.0266</td>
<td>0.0104</td>
</tr>
<tr>
<td></td>
<td>0.6170</td>
<td>0.1156</td>
<td>-0.0036</td>
<td>0.0018</td>
<td>0.0043</td>
<td>0.0249</td>
<td>0.0103</td>
</tr>
<tr>
<td></td>
<td>0.6222</td>
<td>0.1194</td>
<td>-0.0033</td>
<td>0.0019</td>
<td>0.0009</td>
<td>0.0231</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>0.6275</td>
<td>0.1232</td>
<td>-0.0029</td>
<td>0.0019</td>
<td>-0.0025</td>
<td>0.0213</td>
<td>0.0101</td>
</tr>
<tr>
<td></td>
<td>0.6298</td>
<td>0.1249</td>
<td>-0.0028</td>
<td>0.0019</td>
<td>-0.0040</td>
<td>0.0205</td>
<td>0.0101</td>
</tr>
<tr>
<td></td>
<td>0.6327</td>
<td>0.1271</td>
<td>-0.0026</td>
<td>0.0019</td>
<td>-0.0059</td>
<td>0.0195</td>
<td>0.0100</td>
</tr>
<tr>
<td></td>
<td>0.6379</td>
<td>0.1309</td>
<td>-0.0022</td>
<td>0.0019</td>
<td>-0.0093</td>
<td>0.0177</td>
<td>0.0100</td>
</tr>
<tr>
<td></td>
<td>0.6432</td>
<td>0.1347</td>
<td>-0.0019</td>
<td>0.0020</td>
<td>-0.0127</td>
<td>0.0159</td>
<td>0.0099</td>
</tr>
<tr>
<td></td>
<td>0.6484</td>
<td>0.1385</td>
<td>-0.0015</td>
<td>0.0020</td>
<td>-0.0161</td>
<td>0.0142</td>
<td>0.0098</td>
</tr>
<tr>
<td></td>
<td>0.6537</td>
<td>0.1424</td>
<td>-0.0012</td>
<td>0.0020</td>
<td>-0.0194</td>
<td>0.0124</td>
<td>0.0097</td>
</tr>
<tr>
<td></td>
<td>0.6589</td>
<td>0.1462</td>
<td>-0.0008</td>
<td>0.0020</td>
<td>-0.0228</td>
<td>0.0106</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>0.6641</td>
<td>0.1500</td>
<td>-0.0005</td>
<td>0.0021</td>
<td>-0.0262</td>
<td>0.0088</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>0.6694</td>
<td>0.1538</td>
<td>-0.0002</td>
<td>0.0021</td>
<td>-0.0296</td>
<td>0.0070</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>0.6746</td>
<td>0.1576</td>
<td>0.0002</td>
<td>0.0021</td>
<td>-0.0330</td>
<td>0.0052</td>
<td>0.0094</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Var 8</th>
<th>Var 9</th>
<th>Var 10</th>
<th>Var 11</th>
<th>Var 12</th>
<th>Var 13</th>
<th>Var 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominal</td>
<td>0.2294</td>
<td>0.0003</td>
<td>0.5569</td>
<td>302.5000</td>
<td>250.0000</td>
<td>302.5000</td>
<td>250.0000</td>
</tr>
<tr>
<td>predicted</td>
<td>0.2462</td>
<td>0.0002</td>
<td>0.5629</td>
<td>152.0245</td>
<td>249.5616</td>
<td>390.1427</td>
<td>254.2133</td>
</tr>
<tr>
<td></td>
<td>0.2420</td>
<td>0.0002</td>
<td>0.5510</td>
<td>171.6289</td>
<td>249.5650</td>
<td>388.1023</td>
<td>253.7118</td>
</tr>
<tr>
<td></td>
<td>0.2378</td>
<td>0.0002</td>
<td>0.5390</td>
<td>191.2334</td>
<td>249.5684</td>
<td>386.0620</td>
<td>253.7102</td>
</tr>
<tr>
<td></td>
<td>0.2336</td>
<td>0.0002</td>
<td>0.5271</td>
<td>210.8378</td>
<td>249.5718</td>
<td>384.0216</td>
<td>253.7087</td>
</tr>
<tr>
<td></td>
<td>0.2294</td>
<td>0.0002</td>
<td>0.5152</td>
<td>230.4422</td>
<td>249.5752</td>
<td>381.9813</td>
<td>250.2071</td>
</tr>
<tr>
<td></td>
<td>0.2253</td>
<td>0.0001</td>
<td>0.5032</td>
<td>250.0466</td>
<td>249.5786</td>
<td>379.9409</td>
<td>247.0556</td>
</tr>
<tr>
<td></td>
<td>0.2211</td>
<td>0.0001</td>
<td>0.4913</td>
<td>269.6510</td>
<td>249.5820</td>
<td>377.9006</td>
<td>243.2041</td>
</tr>
<tr>
<td></td>
<td>0.2193</td>
<td>0.0001</td>
<td>0.4861</td>
<td>278.2128</td>
<td>249.5834</td>
<td>376.0095</td>
<td>243.8585</td>
</tr>
<tr>
<td></td>
<td>0.2169</td>
<td>0.0001</td>
<td>0.4794</td>
<td>289.2555</td>
<td>249.5854</td>
<td>375.8603</td>
<td>244.7025</td>
</tr>
<tr>
<td></td>
<td>0.2127</td>
<td>0.0001</td>
<td>0.4674</td>
<td>308.8599</td>
<td>249.5888</td>
<td>373.8199</td>
<td>246.2010</td>
</tr>
<tr>
<td></td>
<td>0.2086</td>
<td>0.0001</td>
<td>0.4555</td>
<td>328.4643</td>
<td>249.5921</td>
<td>371.7796</td>
<td>247.6994</td>
</tr>
<tr>
<td></td>
<td>0.2044</td>
<td>0.0001</td>
<td>0.4436</td>
<td>348.0687</td>
<td>249.5955</td>
<td>369.7392</td>
<td>249.1979</td>
</tr>
<tr>
<td></td>
<td>0.2002</td>
<td>0.0000</td>
<td>0.4317</td>
<td>367.6731</td>
<td>249.5989</td>
<td>367.6989</td>
<td>250.6963</td>
</tr>
<tr>
<td></td>
<td>0.1960</td>
<td>0.0000</td>
<td>0.4197</td>
<td>387.2776</td>
<td>249.6023</td>
<td>365.6585</td>
<td>252.1948</td>
</tr>
<tr>
<td></td>
<td>0.1918</td>
<td>0.0000</td>
<td>0.4078</td>
<td>406.8820</td>
<td>249.6057</td>
<td>363.6182</td>
<td>253.6932</td>
</tr>
<tr>
<td></td>
<td>0.1877</td>
<td>0.0000</td>
<td>0.3959</td>
<td>426.4864</td>
<td>249.6091</td>
<td>361.5779</td>
<td>255.1917</td>
</tr>
<tr>
<td></td>
<td>0.1835</td>
<td>0.0000</td>
<td>0.3839</td>
<td>446.0908</td>
<td>249.6125</td>
<td>359.5375</td>
<td>256.6901</td>
</tr>
</tbody>
</table>
part of the covariance in $X$ that is uncorrelated with the variation in $Y$.

After having predicted process conditions for seven out of the eight grades in the data set, we can calculate the measures for the fit and predictability of the inversion for this data set ($\tilde{R}^2$ and $\tilde{Q}^2$) as suggested in section 3.1.5. In this case the latent variable dimension of the $X$-space is greater than the number of variables in $Y$ ($A - k = 1$), and hence we obtain a line of possible solutions. To calculate $\tilde{Q}^2$ we use for each grade the $u_{predicted}^T$ that comes closest to the nominal $u^T$ for that grade. When including only grades 2–8 in the summation $i = 1, \ldots, n^*$, the resulting values are $\tilde{R}^2 = 85\%$, and $\tilde{Q}^2 = 68\%$. Taking into consideration all the problems with the data set, these results are surprisingly good. Since 68% of the variability among the $u_{nominal}^T$ could be predicted, it might be reasonable to use the procedure to predict starting conditions for a new grade, although one would probably have to iterate on the procedure or perform designed experiments around the new point.

The results in this section are for a process which is mostly operator controlled. Trajectories are not automated and the batch to batch variation within each grade is very high. The largest difficulty with this data set, however, lies in the fact that several ingredients are used only in some of the product grades. In principle one might then argue that with different ingredients these are not simply different grades within the same product. As discussed before, this poses extremely tough conditions onto the method. On top of that it appears possible that some of the ingredients are changed from grade to grade in order to affect some product properties that are important only to the customer using the grade. However, we do not have measurements of those customer related properties. The available properties (viscosity, acid value, color) may not really be the important variables affected by some of the ingredient changes.

This study is shown as an example of a case where the methodology reaches its limits. Nevertheless, even in this "bad case scenario" the method does show
reasonable performance for some of the grades: For grades 4, 5, 6 and 8 a solution could be found closer to the nominal conditions in the model space than any of the other grades (see table 3.2). The variable values for these grades are all physically feasible when looking at the closest predictions in table 3.3. Other feasible process conditions for these grades can be found along the $x^T_{null}$-axis. For the remaining grades (2, 3, and 7) solutions were found which were not as close to the nominal conditions in the model space as one or more of the existing grades. However, in each of these cases those existing grades were very close to the nominal grade. In summary, it can be said that the method performed reasonably well on a data set which was used as a tough test on the method.

3.2 A Continuous Process

3.2.1 The Process

This data set is treated in a very confidential manner by the company. Therefore, not much information on the process is revealed. The exact type of product or any specifics on the process are unknown, except that it is from a continuous process, where the independent ($X$-) variables are monitored by an automatic data collection facility. Each time a sample of the product is taken and quality variables ($Y$) are measured, the corresponding process variables are recorded automatically. The data represent steady state operation.

3.2.2 The Data

Data are available for six different product grades. The process conditions for the grades differ from each other in the amounts of ingredients that are used, as well as in operating conditions (e.g. temperatures, etc.). Unfortunately, the data set does not
CHAPTER 3. INDUSTRIAL APPLICATIONS

contain the ingredient amounts themselves. However, the company indicated that the effects of these ingredient amounts show up in the variables that are available. Recorded are eighteen independent variables which we assume to be the X- or manipulatable variables. Variables 1-6 are flow rates, variables 7, 9, 11, and 17 are weights, variables 8 and 10 are temperatures, variables 12-16 are flow rates, and variable 18 is proportional to voltage. Product quality is characterized by four different Y-variables Y₁-Y₄. All we know is that they represent various measured quality characteristics deemed important for the process.

For each product we received operating conditions and corresponding quality variables repeatedly collected over a period of often several months. These data sets for the six different products were screened independently by PCA analyses and subsequently outliers were removed. From those pruned data sets the average variable values were calculated within each grade. The resulting set of data consists of six observations for the six products, eighteen X- and four Y-variables. In all of the models discussed further below, the data used in the models are meancentered and scaled to unit variance.

3.2.3 Preliminary Data Analysis

A PLS-model on all six products fully explains in five dimensions the (6 × 18) X-matrix, and the (6 × 4) Y-matrix as is shown in table (3.5).

Table 3.5: Percentage of Variation Explained by PLS in X and Y Using all 6 Products

<table>
<thead>
<tr>
<th>Latent Variable No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>34.7</td>
<td>48.2</td>
<td>76.8</td>
<td>88.1</td>
<td>100</td>
</tr>
<tr>
<td>Y</td>
<td>47.4</td>
<td>81.7</td>
<td>87.6</td>
<td>96.2</td>
<td>100</td>
</tr>
</tbody>
</table>
Figures 3.4 and 3.5 display the t-score values for the first three latent dimensions of this model.

![Graph showing t-scores from PLS on all 6 Products: t1-t2](image)

Figure 3.4: t-Scores from PLS on all 6 Products: $t_1$-$t_2$

The scores of all six products are well spread out over the space. We were hoping that one of the products might have scores that are surrounded by the scores of the other five. This would provide us with a perfect candidate for a product grade that we could leave out of the data set, and try to re-design its process conditions using a model based on the remaining five products. In spite of the lack of such an "ideal" point, product 6 was selected as the one where we pretend not to know the process conditions ($x^T$ for grade 6), and we attempt to predict such process conditions for grade 6 from the remaining grades 1–5. This selection of grade 6 was also guided by the fact that we obtained data for grade 6 much later than the data for the other grades. Apart from that, figures 3.4 and 3.5 show that product 6 does not seem to be a bad choice, considering the fact that none of the other grades has scores that lie more centered in the cluster. Figure 3.5 illustrates that certain features of grades 5
and 6 must be very similar since these two grades lie extremely close to each other in the $t_1$-$t_3$-plane.

The following analysis uses products 1–5 as "historical" grades, and the quality data for product 6 as a new grade for which process conditions need to be designed. The nominal process conditions for grade 6 are used later on to evaluate the design results.

The first step is to decide how many and which quality variables ($Y$) of products 1–5 are the most complementary subset to cover the quality characteristics of this product family. Table (3.6) shows the results of a PCA analysis on the $Y$-matrix of grades 1–5. Selective PCA on this same data set reveals that $Y_4$ and $Y_2$ together account for 83.3%, and $Y_4$, $Y_2$, and $Y_3$ account for 97.8% of the variation in the data.

In this situation it is very hard to decide whether to use two or three $Y$-variables for the model and the inversion. Considering the fact that the above PCA-model (table (3.6)) is based on only five data points, the third principal component
Table 3.6: Percentage of Variation Explained by PCA in Y Using Products 1–5

<table>
<thead>
<tr>
<th>No. of Principal Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>55.5</td>
<td>90.3</td>
<td>99.4</td>
<td>100</td>
</tr>
</tbody>
</table>

leading to over 99% of the total variation may just be fitting error. Cross-validation does not work very well with such few data points. According to the SIMCA-software (see (SIMCA 1996)), none of the components passes the cross-validation test, but for the third component there are not even enough degrees of freedom in the data. For this study it was decided to select only $Y_2$ and $Y_4$ as the $Y$-space for the model and its inversion.

Remark: The model using $Y_2$, $Y_3$, and $Y_4$ as the $Y$-space has also been inverted as a test, but the obtained process conditions result in large extrapolations out of the range of the conditions used in any of the products. This indicates that the data are probably overfitted in this case, and therefore using only two $Y$-variables seems a better choice.

### 3.2.4 Model Inversion and Results

Table (3.7) shows the results of PLS and PCR performed on product grades 1–5 using $Y = [Y_2 \ Y_4]$.

Table 3.7: Percentage of Variation Explained by PLS and PCR in $X$ and $Y = [Y_2 \ Y_4]$ Using Products 1–5

<table>
<thead>
<tr>
<th>Latent Variable No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-PLS</td>
<td>41.2</td>
<td>60.7</td>
<td>85.3</td>
<td>100</td>
</tr>
<tr>
<td>Y-PLS</td>
<td>65.9</td>
<td>94.6</td>
<td>97.3</td>
<td>100</td>
</tr>
<tr>
<td>X-PCR</td>
<td>41.7</td>
<td>72.4</td>
<td>87.3</td>
<td>100</td>
</tr>
<tr>
<td>Y-PCR</td>
<td>60.6</td>
<td>54.1</td>
<td>90.3</td>
<td>100</td>
</tr>
</tbody>
</table>
Both, PLS and PCR perform similarly for three components, and one could probably use either model for an inversion. In this study the PLS-model with three latent variables is chosen for the inversion. SIMCA indicates that there may not be enough degrees of freedom left for the third dimension, but since it explains only about 3% in $Y$ but another 24% in $X$ we include it in the model, hoping that not too much overfit is introduced by this third component.

As there are only two $Y$-variables specified, but three latent variables used in the model, the inversion results again in a whole line as solution space for process conditions for product 6. Along this line of solutions the set of process conditions closest (in Mahalanobis sense) to the projection of the “true” process conditions given by the data for product 6 is selected. The Mahalanobis distance between the closest prediction and the projection of the true process conditions onto the model space for grade 6 is calculated, and compared to the Mahalanobis distances between the projections of each of the five model grades and the projection of the true point in the $X$-space. (Details on the calculation of such Mahalanobis distances are given in section 3.1.4.) The distances are recorded in table (3.8):

Table 3.8: Mahalanobis Distances in the Model Space Between Grades 1 to 5 Used in the Model, Closest Prediction, and the Actual $x^T$ of Grade 6

<table>
<thead>
<tr>
<th>Inversion</th>
<th>Grades Used in the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>closest</td>
<td>1</td>
</tr>
<tr>
<td>0.1963</td>
<td>1.8073</td>
</tr>
</tbody>
</table>

The point in the solution space that comes closest to the projected true process conditions according to the Mahalanobis distance measure, i.e. our best prediction, is closer to these projected true conditions of grade 6 than the projected conditions of grades 1–4. However, product grade 5 seems to have $x^T$-values which are even closer in the model space to grade 6 than this best prediction. As has been mentioned before,
when looking back at figures 3.4 and 3.5 this is not very surprising. According to the PLS-model on all six grades, the first and third $t$-scores values for grades 5 and 6 are almost identical. The $t_2$-values are different but not too far apart. This proximity of scores for grade 5 and 6 ("true") can also be seen in figures 3.6 and 3.7, where the $t$-scores from the PLS model with $Y = [Y_2 \ Y_4]$ on grades 1-5 ($\times$) are displayed together with the projection of the true $x^T$-values for grade 6 ($\ast$) into that score space, and the resulting scores of the prediction that comes closest to the true values ($\circ$).

![Graph showing t-Scores from PLS on Products 1-5, Corresponding Scores from Product 6, and Closest Prediction: $t_1$-$t_2$](image)

Figure 3.6: $t$-Scores from PLS on Products 1-5, Corresponding Scores from Product 6, and Closest Prediction: $t_1$-$t_2$

The values for the $X$-variables as predicted by the point closest to the true $x^T$-values for grade 6, lie well within the range of variable values given in the data set. For reasons of confidentiality the $X$-values are not shown here.

Taking into account that the predictions for product grade 6 have been estimated from a model that is based on five data points only, the results are reasonable.
Chapter 3. Industrial Applications

Figure 3.7: $t$-Scores from PLS on Products 1–5, Corresponding Scores from Product 6, and Closest Prediction: $t_1$-$t_3$

Although the small number of data points does not allow for the cross-validation approach described in section 3.1.5, the obtained predictions could be used as a starting point for some experiments that would result in more data which in turn could be used in a new model.

3.3 A Semi-Batch Emulsion Polymerization Process

3.3.1 The Process

The data are from an emulsion polymerization process that is carried out in semi-batch fashion. At a first stage some of the ingredients are fed into the reactor. The reactor is then heated to a certain temperature at which time the feed of further
CHAPTER 3. INDUSTRIAL APPLICATIONS

ingredients starts. Feeding takes place over a prescribed period of time after which the reaction is left to continue until it reaches a certain conversion. Subsequently, the reactor is emptied and its contents are stripped of unreacted monomer. The product is then put into storage tanks that usually contain the compounded material of several batches. It is from these storage tanks that the samples for the obtained product quality measurements are taken. Therefore, it is not possible to attribute individual quality values to the different batches.

3.3.2 The Data

Data from five different products are available for this study. For reasons mentioned above, the quality data consist of only one set of quality values for each product, that is, those values had to be used "as is". No average qualities over several batches could be calculated. However, the fact that the quality measurements were taken on a sample of the compounded product from several batches amounts to some kind of averaging over several batches. The five available quality variables are: gel content, elongation at break, tensile strength, glass transition temperature, and particle size.

In terms of manipulatable variables only variables pertaining to the batch reaction are considered. The effect of the stripping is neglected in this study. Several recordings of batch runs together with the "cook sheets" are available for each product. A "cook sheet" contains the detailed instructions on how to fill and run the reactor for a specific product. Since in every case the implemented process conditions are very close to those listed on the cook sheets, these latter values (given in the cook sheets) are taken as the "true" process conditions for each grade.

Although many factors contribute to the final quality of the product, only variables that indeed are set differently in the five products are included in the $X$-matrix. Twelve such variables constitute this $X$-matrix. Variables 1–3 are the charge of a reactant, the charge of the seed latex, and the amount of catalyst initially fed
into the reactor during the first stage. Variables 4–8 denote feedrates for different reactants, variable 9 is the set point temperature of the reactor jacket, and variables 10–12 contain the total times over which the different feed streams are fed into the reactor.

This particular data set poses a problem not just because of the extremely small number of data points (five product grades), but also because of the fact that two out of these five polymers (1 and 2) are produced in a large size reactor, while the other three grades (3, 4, and 5) are produced in a much smaller sized reactor.

Two concepts have been investigated in handling this difference in reactor size:

1. The data for the two different reactor groups (large and small) are first mean-centered individually for each reactor group, and then combined and scaled to unit variance. Of course this reduces the degrees of freedom in the data. In this particular data set, however, some of the twelve variables do not change for products within one or the other reactor group. Thus, this form of mean centering results in some of the columns in $X$ having zeros either for products 1 and 2, or for products 3, 4, and 5. It would be preferable to also scale the data using the standard deviation of each reactor group, but since some variables are constant within a reactor group, this is not possible here.

2. Another way to account for the two reactor sizes is to introduce another $Y$-variable that indicates the reactor group by assuming either 1 for the large size reactor or 0 for the small one.

PLS-models that follow concept 2 have one latent dimension more than models that employ concept 1. The additional dimension (usually the first one) serves to explain the variation in the data that is introduced by the differences in reactor size. This variation is eliminated in concept 1 by mean centering within each reactor group.
In fact, the second and third dimensions of a “concept 2 model” are quite similar to the first and second dimensions of a “concept 1 model” in this example.

3.3.3 Preliminary Data Analysis

The present study follows concept 2. The initial ingredient charges and feedrate variables in $X$ are scaled in proportion to the final amount of material charged in the reactor for each grade. Afterwards, the $(5 \times 12)$ $X$-matrix and the $(5 \times 6)$ $Y$-matrix are mean-centered and scaled to unit variance. (Remember, that $Y$ contains now five quality variables plus the additional sixth variable $R$ which indicates the reactor group.) The results of a PLS-model on this data set can be seen in table (3.9).

Table 3.9: Percentage of Variation Explained by PLS in $X$ and $Y$ Using all 5 Products and all 6 $Y$-Variables

<table>
<thead>
<tr>
<th>Latent Variable No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>85.7</td>
<td>91.4</td>
<td>97.4</td>
<td>100</td>
</tr>
<tr>
<td>$Y$</td>
<td>26.1</td>
<td>75.8</td>
<td>96.5</td>
<td>100</td>
</tr>
</tbody>
</table>

The first step now is again to select a complementary subset of quality variables for the $Y$-space in the inversion model. PCA on all six $Y$-variables explains the data as shown in table (3.10). It seems questionable whether the third component really helps predict the data or whether it is already overfitting the $5 \times 6$ data points.

Table 3.10: Percentage of Variation Explained by PCA on all 6 $Y$-Variables

<table>
<thead>
<tr>
<th>Principal Component No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>59.9</td>
<td>86.3</td>
<td>98.4</td>
<td>100</td>
</tr>
</tbody>
</table>

Selective PCA on these data finds elongation ($el$), reactor size ($R$), and particle size ($PS$) as the most complementary subset containing only three variables. It
accounts for 98.2% of the variation. If only two variables are selected, the most complementary ones are elongation and reactor size, and account for 83.8% of the variation in the data.

Obviously, for the specification of a new product grade in this case one would want to specify the reactor in which the grade should be produced. Of course this implies that when only two $Y$-variables are to be specified, only one "real" quality characteristic (such as e.g. elongation) can be chosen. (If more grades were available, specification of a reactor group variable $R$ may still allow for the design of more than only one other quality characteristic.)

Having only five product grades available also makes it difficult to eliminate one and re-design it from the remaining four. Of course any model and its inversion based on only four points can at most give crude starting values in the design of a new grade quality. Figures 3.8 and 3.9 show the $t$-scores for the PLS-model on all five grades and six $Y$-variables that was already displayed in table (3.9). It can be seen that product 3 is the one that is more contained within the remaining four products than any of the other grades. It belongs to the group of three grades that are produced in the smaller reactor. The model used to design grade 3 therefore still contains two grades from the large reactor, and two grades from the smaller size reactor.

Two studies are described in the following sections:

1. Product 3 is taken out of the data set and its process conditions assumed unknown. It is then re-designed by inverting a model based on the remaining four grades.

2. So far product 3 has been produced only in the small reactor. The information on all five product grades can be used to design a production for grade 3 that could be carried out in the big reactor.
Figure 3.8: t-Scores from PLS on all 5 Products and 6 Y's: $t_1-t_2$

Figure 3.9: t-Scores from PLS on all 5 Products and 6 Y's: $t_2-t_3$
3.3.4 Model Inversions and Results

Re-Designing Product Grade 3

A PLS-model using all six $Y$-variables but leaving out product 3, explains 100% of the variation in $X$ and $Y$ in three latent dimensions. Therefore, it is clear that for the inversion only two latent dimensions should be used. But with only two dimensions in the $X$-space no more than two independent $Y$-variables can be explained. In fact, PCA on the $Y$-matrix containing only products 1, 2, 4, and 5 results in only two significant components explaining 87.3% of the variation in $Y$ (three components explain 100%). Selective PCA on the same matrix also indicates elongation and reactor size as the two $Y$-variables accounting for most of the variation (85.1%). The $Y$-space in this study consists therefore of elongation and reactor size values only.

Figure 3.10 shows where in the $t$-space of the PCA-model on $Y = [e_l \ R]$ the scores for a "new" product 3 are predicted.

![Figure 3.10: $t$-Scores from PCA on $Y = [e_l \ R]$; Grades 1,2,4,5 and the Prediction for the New Grade 3](image_url)
CHAPTER 3. INDUSTRIAL APPLICATIONS

Table (3.11) displays the sum of squares explained by PLS and PCR in $X$ and $Y' = [e_1 R]$ when grade 3 is left out of the data set. The PLS-model seems far more appropriate to use since it serves so much better in explaining the $Y$-space than the PCR-model. It is clear that only the first two dimensions can be considered significant, and since $Y$ also consists of two variables, the inversion will result in a single solution point.

Table 3.11: Percentage of Variation Explained by PLS and PCR in $X$ and $Y' = [e_1 R]$
Using Products 1, 2, 4, and 5

<table>
<thead>
<tr>
<th>Latent Variable No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$-PLS</td>
<td>86.2</td>
<td>92.0</td>
<td>100</td>
</tr>
<tr>
<td>$Y$-PLS</td>
<td>52.6</td>
<td>98.8</td>
<td>100</td>
</tr>
<tr>
<td>$X$-PCR</td>
<td>86.3</td>
<td>94.9</td>
<td>100</td>
</tr>
<tr>
<td>$Y$-PCR</td>
<td>52.2</td>
<td>55.4</td>
<td>100</td>
</tr>
</tbody>
</table>

The model is inverted, and the Mahalanobis distances between the projections of the true process conditions for grade 3 and of the four model grades onto the model space are calculated as laid out in section 3.1.4. The distance between the true and the newly predicted conditions in the model space is obtained in the same way. Table (3.12) shows these distances. Considering the few data points this study is based on, this result is extremely encouraging. The predicted process conditions are closer to the true ones in the model space than any of the grades used in the model.

Table 3.12: Mahalanobis Distances in the Model Space from Grades 1, 2, 4, 5, and Closest Prediction to the Actual $x^*$ of Grade 6

<table>
<thead>
<tr>
<th>Inversion</th>
<th>Grades Used in the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>closest</td>
<td></td>
</tr>
<tr>
<td>0.1395</td>
<td>1.6399 1.4340 0.2985 0.6547</td>
</tr>
</tbody>
</table>

Figure 3.11 shows where the newly predicted and the true process conditions
fall in the $t_1$-$t_2$-plane of the PLS-model used in the inversion. Remark: In this figure the $t$-scores of product grade 4 seem actually closer to the projection of the "true" conditions than the predicted $t$-scores, although the Mahalanobis distances as displayed in table (3.12) indicate otherwise. This discrepancy is due to the difference between the term $P : (P^T \cdot P)^{-1}$ in the calculation of the Mahalanobis distance and the term $W : (P^T \cdot W)^{-1}$ used in the calculation of the $t$-scores from a PLS-model for a given $x$.

![Graph](image)

Figure 3.11: $t$-Scores from PLS on Grades 1,2,4,5 ($\times$), Scores for True Conditions ($\ast$), and the Prediction for the New Grade 3 (○); $Y = [el \: R]$

Table (3.13) lists the mean-centered and scaled values of the twelve $X$-variables (V1 to V12) for the model grades 1,2,4,5, the true grade 3, and predicted conditions for grade 3. Only for variable 2 does the predicted value (last row in the table) lie outside of the range of the other grades. The predicted values of variable 4, a feedrate, seems quite small compared to its "true" value. However, this smaller feedrate is compensated by a longer feed time (variable 11) than in the true conditions. The
product of feedrate and feed time (in their original units) results in a slightly smaller value than what the product of the true variable values gives. However, this product (between this feedrate and feed time) from the predicted values is well within the range of the corresponding products for the existing grades produced in the smaller reactor. In other words, the amount of material fed is still predicted within the range of the existing data.

Another "handicap" in using this data set for predicting grade 3 can be seen from the table: Variables 9–12 are constant within each reactor group for the model grades. However, grade 3 which belongs to the small reactor group of grades 4 and 5, has true values that differ from those constant for grades 4 and 5. Of course, the model does not have any information on such a variation, since it contains only grades 1, 2, 4, and 5. It therefore must predict these variable values for grade 3 to be the same values present in grades 4 and 5, as table (3.13) shows.

Table 3.13: X-Values of the 4 Model Grades, the True Grade 3, and the Predicted Grade 3

<table>
<thead>
<tr>
<th>Gr.</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>V10</th>
<th>V11</th>
<th>V12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.41</td>
<td>0.18</td>
<td>-1.29</td>
<td>-0.87</td>
<td>-0.59</td>
<td>-0.49</td>
<td>-0.88</td>
<td>-1.03</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>-1.46</td>
<td>-0.28</td>
<td>-0.68</td>
<td>-0.83</td>
<td>-1.09</td>
<td>-0.85</td>
<td>-0.68</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>4</td>
<td>-0.71</td>
<td>0.64</td>
<td>0.79</td>
<td>1.31</td>
<td>0.02</td>
<td>1.20</td>
<td>0.86</td>
<td>0.85</td>
<td>-0.87</td>
<td>-0.87</td>
<td>-0.87</td>
<td>-0.87</td>
</tr>
<tr>
<td>5</td>
<td>-0.71</td>
<td>0.64</td>
<td>0.79</td>
<td>0.24</td>
<td>1.40</td>
<td>0.38</td>
<td>0.87</td>
<td>0.86</td>
<td>-0.87</td>
<td>-0.87</td>
<td>-0.87</td>
<td>-0.87</td>
</tr>
<tr>
<td>true</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.72</td>
<td>0.64</td>
<td>0.78</td>
<td>1.39</td>
<td>0.83</td>
<td>0.86</td>
<td>0.67</td>
<td>1.78</td>
<td>-2.60</td>
<td>-0.76</td>
<td>-1.13</td>
<td>-1.18</td>
</tr>
<tr>
<td>pred</td>
<td>-0.65</td>
<td>0.72</td>
<td>0.74</td>
<td>0.64</td>
<td>0.88</td>
<td>0.72</td>
<td>0.86</td>
<td>0.84</td>
<td>-0.87</td>
<td>-0.87</td>
<td>-0.87</td>
<td>-0.87</td>
</tr>
</tbody>
</table>

Considering the strong reservations one must have towards this data set because of the small number of data points, and the fact that the products were produced in two different sized reactors the methodology appears to have worked very well. These shortcomings also prevent the use of cross-validation in this example as described in section 3.1.5.
Designing Product Grade 3 for the Large Reactor

For this study the information on all five grades including product 3 (which has so far been produced only in the small reactor) can be used. As shown in table (3.9), the PLS-model on all five grades and six quality variables explains 100% in $X$ and $Y$ in four latent dimensions. It is therefore possible to use the first three latent dimensions of the model and specify three $Y$-variables for the inversion. Since the $Y$-values for product grade 3 in the small reactor are already known to be feasible, the recorded quality values from grade 3 for elongation and particle size can be used as they are. Only the value for the reactor group changes from 0 to 1 to indicate the larger size reactor.

Figure 3.12 shows the $t_1$-$t_2$-plane of a PCA-model on $Y = [el \ PS \ R]$, and where it predicts the scores for a grade 3 that is to be produced in the large reactor.

![Figure 3.12: t-Scores from PCA on $Y = [el \ PS \ R]$; Grades 1–5 and the Prediction for Grade 3 in the Large Reactor](image)

The variation explained by PLS and PCR on all five grades and using $Y =$
[el PS R] is given in table (3.14). For three dimensions one could probably employ either model; here the PLS-model is chosen for inversion.

Table 3.14: Percentage of Variation Explained by PLS and PCR in X and Y = [el PS R] Using Products 1–5

<table>
<thead>
<tr>
<th>Latent Variable No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-PLS</td>
<td>85.7</td>
<td>93.3</td>
<td>97.4</td>
<td>100</td>
</tr>
<tr>
<td>Y-PLS</td>
<td>45.8</td>
<td>70.2</td>
<td>99.0</td>
<td>100</td>
</tr>
<tr>
<td>X-PCR</td>
<td>85.8</td>
<td>93.4</td>
<td>97.5</td>
<td>100</td>
</tr>
<tr>
<td>Y-PCR</td>
<td>44.5</td>
<td>69.8</td>
<td>95.4</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 3.13 shows that the new process conditions fall on the same side of the PLS score space where grades 1 and 2 (both produced in the large reactor) are. It also lies fairly close to grade 1 (see figures 3.13 and 3.14), which comes closer to grade 3 in terms of elongation and particle size than grade 2.

The meancentered and scaled X-values of grades 1–5 as well as the X-values for the prediction of grade 3 in the large reactor are listed in table (3.15). No true process conditions are available for grade 3 produced in the large reactor, but as the table shows, the values lie within or not far from the X-values given for the five grades (especially grade 1).

Table 3.15: X-Values of the 5 Model Grades, and the Predicted Grade 3 for the Large Reactor

<table>
<thead>
<tr>
<th>Gr.</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>V10</th>
<th>V11</th>
<th>V12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.68</td>
<td>0.05</td>
<td>-1.55</td>
<td>-1.08</td>
<td>-0.80</td>
<td>-0.70</td>
<td>-1.11</td>
<td>-1.18</td>
<td>0.96</td>
<td>1.09</td>
<td>1.09</td>
<td>1.09</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>-1.74</td>
<td>-0.47</td>
<td>-0.90</td>
<td>-1.06</td>
<td>-1.33</td>
<td>-1.07</td>
<td>-0.88</td>
<td>0.96</td>
<td>1.09</td>
<td>1.09</td>
<td>1.09</td>
</tr>
<tr>
<td>3</td>
<td>-0.62</td>
<td>0.56</td>
<td>0.67</td>
<td>1.04</td>
<td>0.70</td>
<td>0.72</td>
<td>0.58</td>
<td>1.21</td>
<td>-1.43</td>
<td>-0.65</td>
<td>-0.90</td>
<td>-0.93</td>
</tr>
<tr>
<td>4</td>
<td>-0.62</td>
<td>0.56</td>
<td>0.68</td>
<td>0.97</td>
<td>-0.15</td>
<td>1.08</td>
<td>0.80</td>
<td>0.42</td>
<td>-0.24</td>
<td>-0.77</td>
<td>-0.64</td>
<td>-0.62</td>
</tr>
<tr>
<td>5</td>
<td>-0.61</td>
<td>0.56</td>
<td>0.68</td>
<td>-0.04</td>
<td>1.31</td>
<td>0.22</td>
<td>0.80</td>
<td>0.43</td>
<td>-0.24</td>
<td>-0.77</td>
<td>-0.64</td>
<td>-0.62</td>
</tr>
<tr>
<td>pred</td>
<td>1.59</td>
<td>-0.06</td>
<td>-1.48</td>
<td>-1.32</td>
<td>-0.52</td>
<td>-0.92</td>
<td>-1.09</td>
<td>-1.22</td>
<td>1.04</td>
<td>1.08</td>
<td>1.11</td>
<td>1.11</td>
</tr>
</tbody>
</table>
Figure 3.13: $t_1$-$t_2$-Scores from PLS on Grades 1-5 ($\times$), and the Prediction for the New Grade 3 in the Large Reactor ($\circ$); $Y = [el \ PS \ R]$

Figure 3.14: $t_2$-$t_3$-Scores from PLS on Grades 1-5 ($\times$), and the Prediction for the New Grade 3 in the Large Reactor ($\circ$); $Y = [el \ PS \ R]$
CHAPTER 3. INDUSTRIAL APPLICATIONS

The analogous study has been carried out using $Y = [e_l \ R]$ in a PLS-model with only two latent dimensions. The inversion results are similar to those shown above using three $Y$-variables, and are therefore not displayed here.

This industrial example, although providing only a minimal number of data points, illustrates the potential of the methodology that has been developed in chapter 2. The example is also of interest from another aspect: The five grades come from two differently sized reactors, thereby posing an additional challenge. Two possible ways of treating this situation have been indicated. Furthermore, a strategy has been outlined how this inversion problem could be used to move the production of a grade from one reactor to a different one. It is of course necessary for this approach that the two reactors have process conditions that are expressed by the same set of manipulatable variables. Otherwise, the methodology outlined in the next chapter would have to be used.

3.4 Conclusions

The three examples show the potential of the methodology described in the previous chapter 2 for different industrial processes. Each data set has its own shortcomings and problems, such as very small number of data points (grades), grades produced in differently sized reactors, or even changes in types of ingredients from grade to grade. These difficulties have been discussed and where possible ways of handling them have been indicated. The effects of such problems on the databased approach are shown within the results sections of the examples.

Although none of the industrial data sets represents the “ideal” example for this databased approach, the results obtained in these examples are promising. Unfortunately, it was not possible to really implement any of the predicted process conditions for the “new” grades. However, the predicted conditions are compared to
those of the existing grades. The results show that in cases where the new grade does not require an extrapolation outside of the range of the grades used in the model, the new conditions lie within the range of the "historical" conditions, and appear to provide reasonable process conditions. They may serve as a starting point for an experimental design aimed at optimizing these conditions.
Chapter 4

Product Transfer from One Plant to a Second Plant

4.1 Introduction

The problem treated in this chapter can be seen as an extension to the product design problem. Consider two similar plants, A and B, owned by same company at two different locations. Both plants produce products out of the same product family (e.g. LDPE film grades). They even may have some overlap in their product lines, i.e. some of the grades may be produced in both plants. The necessity often arises for a grade $y_{A_{des}}^T$ which is currently only manufactured in plant A to be produced also in plant B - of course with the same quality specifications.

Although making products from the same family and based on the same physical principles, the two plants might be of different size and configuration, or use different technology. They also might have different raw material suppliers, cooling water temperatures, etc. Therefore, in order to produce identical quality on a particular grade, the operating strategies and conditions for production of this grade will look different for the two plants. The question that is dealt with in this chapter is: how
can we use the information about production in plant A to find the corresponding process conditions in plant B that will lead to the same grade quality?

Since the grade that one wants to produce in plant B is already being produced by another plant, there is no question of physical feasibility of the desired grade as we are faced with in chapter 2. Therefore, it is not necessary here to find the most complementary subset of independent quality variables for \( Y \). On the contrary, all available quality variables (whether they are correlated or not) can be used as long as they are available for all product grades from both plants.

The different data matrices that are assumed to be available for this problem are introduced in section 4.2. This leads to the question of which product grades should be included when building a model that is used for the current purpose. Section 4.3.1 addresses this issue and describes a strategy in order to select the grades that are beneficial to solving the problem. In 4.3.2 the pretreatment of the data is outlined. This is followed in section 4.3.3 by the description of an approach that can be applied to combine the data from two plants in order to derive plant B process conditions for a grade so far produced only in plant A. This approach is illustrated by a simple MATLAB example in 4.4. A few additional issues and ideas are discussed in the final section 4.5.

Although a few key points are repeated here, the reader is referred back to the introduction and discussion sections of chapter 2 for a general philosophy on databased approaches to this type of problem.

### 4.2 Data Structure

Before looking at possible solutions to this problem, the structure of the data that one would typically expect in this type of problem must be investigated. In general, the type of data is the same as in the product design problem (chapter 2), except
that we assume data for a series of grades belonging to the same type of product to be available for both plants.

Each plant is characterized by a matrix $X$ containing the individual process conditions of the plant as they are changed from grade to grade, and a matrix $Y$ that contains the corresponding quality measurements for those grades. Obviously each plant may have produced a different number of grades out of that product family. This number is denoted by $n_A$ and $n_B$ respectively, and marks the number of rows in the data matrices $X$ and $Y$ for the two plants. In the matrices for one plant each row corresponds to a particular grade.

In case there are common grades between the two plants, we denote the matrix which contains the quality specifications for those common grades by $Y_c$ to contrast with grades that are produced either only in plant A, $Y_{Au}$, or only in plant B, $Y_{Bu}$, where "c" stands for "common", and "u" for "unique". (Grades that are supposed to be identical in both plants, but do not turn out that way belong to matrices $Y_{Au}$ and $Y_{Bu}$.)

The $(n_A \times k)$ matrix $Y_A$ is constructed as $Y_A = \left[ \begin{array}{c} Y_{Au} \\ Y_c \\ Y_{Bu} \end{array} \right]$ and contains for now all the grades that have been produced in plant A. (In section 4.3.1 we test whether some grades may have to be eliminated from $Y_A$ in order to benefit the databased approach to this problem.) In the same way matrix $Y_B$ contains all the grades that are available from plant B, where $Y_B = \left[ \begin{array}{c} Y_c \\ Y_{Bu} \end{array} \right]$. These $Y$-matrices are assumed to contain the same number of columns ($k$) with each of the $k$ quality variables corresponding to the same columns in $Y_A$ and $Y_B$.

The process conditions for both plants resulting in the respective grades that are unique to each plant are denoted by $X_{Au}$, $X_{Bu}$. Their rows relate to those of $Y_{Au}$ and $Y_{Bu}$. The process conditions for the common grades are summarized in matrix $X_{Ac}$ for plant A, in $X_{Bc}$ for plant B such that their respective rows correspond to the grades in $Y_c$. 


Within each plant $P$ the columns in $X_{PA}$ have to correspond to the same process variables as the columns in $X_{PC}$ so that they can be combined to a $(n_A \times m_A)$ matrix $X_A = \begin{bmatrix} X_{Ae} \\ X_{Ac} \end{bmatrix}$ and a $(n_B \times m_B)$ matrix $X_B = \begin{bmatrix} X_{Be} \\ X_{Bc} \end{bmatrix}$. However, since the two plants are expected to be dissimilar in some ways, some of the process variables used in $X_A$ may differ from those used in $X_B$. One can also not expect the number of process variables in the two plants to be same ($m_A \neq m_B$).

Figure (4.1) shows these different data blocks as a schematic. The question that arises is: how can the different blocks be used in predicting plant $B$ process conditions for the desired grade $y_{A_{des}}^T$ which so far has only been produced in plant $A$?

Obviously, the main information on how to find such conditions for plant $B$ is contained in $X_B = \begin{bmatrix} X_{Be} \\ X_{Bc} \end{bmatrix}$ and $Y_B = \begin{bmatrix} Y_c \\ Y_{Bu} \end{bmatrix}$. However, we already know how to produce $y_{A_{des}}^T$ in plant $A$, and how the process conditions in both plants ($X_{Ac}, X_{Be}$) are related to the quality measurements of prior common grades ($Y_c$). Adding this information on $X_{Ac}, x_{A_{des}}^T$, and $y_{A_{des}}^T$ to the information on plant $B$ may be beneficial in dealing with this problem.

The data on the “unique” grades in plant $A$ ($X_{Au}, Y_{Au}$) are useful in helping to better define the latent variable space of $X_A, Y_A$. However, to help with prediction of plant $B$ conditions, only those grades of $Y_{Au}$ that are consistent with the covariance structure of the plant $B$ grades ($Y_B = \begin{bmatrix} Y_c \\ Y_{Bu} \end{bmatrix}$) should be retained. This is discussed further in section 4.3.1.

Ideally one would prefer a method that employs all the relevant information in the data blocks. However, there are some simple approaches that use only a subset of information:

1. $X_B \rightarrow Y_B$

Using PCR or PLS to model the relationship between the plant $B$ process conditions and the product quality in plant $B$, the inversion approach of chapter
Figure 4.1: Data Structure for Process Conditions and Resulting Quality in Two Plants A and B; Both Produce Grades Within the Same Product Family
2 could be applied, and process conditions for the new grade predicted.

2. $X_{Ac} \rightarrow X_{Bc}$

If there are enough common grades between the two plants, and especially if at the same time the two plants are very similar, then a latent variable model (e.g. PLS) relating $X_{Ac}$ and $X_{Bc}$ could be used to predict a new set of process conditions for $y^T_{A_{des}}$ for plant B from the corresponding process conditions in plant A. But especially when parts of the equipment in the two plants or the operating strategies are very different it is likely that process conditions in plant A can predict only some features of the corresponding process conditions in plant B. In other words, $X_{Ac}$ will most likely not project into all of the directions spanned by $X_B$.

3. Using the PLS-model of 2. to update the inversion approach of 1. is another option. One possible way to achieve this would be to use the predictions $x^T_{B_{predPLS}}$ from the PLS-model (2.) to determine the point in the nullspace (obtained from the model inversion (1.)) that comes closest to $x^T_{B_{predPLS}}$. On the other hand, one might reject this option and be in favor of retaining that window of operating conditions which is provided in the nullspace from the inversion.

\[
\begin{pmatrix}
X_{Ac} & X_{Bc} & y^c \\
\vdots & \vdots & \vdots \\
x^T_{A_{des}} & ??? & y^T_{A_{des}}
\end{pmatrix}
\]

Process conditions from both plants for only the common grades and the corresponding quality measurements can be combined in one matrix. A model can be fit to describe the variation within this data structure. The plant B process conditions for $y^T_{A_{des}}$ can then be treated as missing data, and their expected values given the values for the quality measurements $y^T_{A_{des}}$ and the process conditions for plant A $x^T_{A_{des}}$ can be calculated based on the model. This approach
would be analogous to the Conditional Expectation approach (section 2.2.5 in chapter 2).

Since the approaches indicated above use only part of the information available, they will not be discussed further. The approach outlined in section 4.3.3 of this chapter is an attempt to use as much of the relevant information as possible, however, it does not include the data on process conditions in plant A (\(X_A\)) at all. In particular it is the latent space of plant B process conditions, \(X_B\), and that of the grade quality in \(Y_B\), and selected grades from plant A, \(Y_A\), that is used to build an extended PCR-model. The part of this model that corresponds to plant B is then inverted on a desired grade from plant A, \(y^{T}_{A_{des}}\), to predict plant B process conditions for this grade.

4.3 Methodology

The previous section has described the problem under investigation in terms of the type of data that we assume to be available, and the structure we expect in such data. Based on these premises, sections 4.3.1 and 4.3.2 suggest a pretreatment of the data, while section 4.3.3 outlines a model and its inversion for this problem.

4.3.1 Selection of Grades

In this work we are looking for process conditions for plant B that can produce a grade \(y^{T}_{A_{des}}\), which so far has only been produced in plant A. For this purpose it makes sense to not only use plant B data but also data available from plant A. However, only grades from plant A that conform with the covariance structure of plant B grades will be of support in designing process conditions in plant B. There is no guarantee that all the grades that have been produced by plant A can be produced by plant B, even
if there exist some common grades between the plants. The quality matrices $Y_A$ and $Y_B$ each span a space within which the respective plant has been producing so far. A databased approach should only attempt to predict conditions for a new grade if it lies within the same space spanned by the $Y$-matrix for that plant.

In order to select the grades from plant A that might be produced in plant B, and that may prove useful in building a model, we need to compare the spaces that are spanned by $Y_A$ and $Y_B$ respectively. Several situations can be distinguished:

1. Both spaces, $Y_A$ and $Y_B$, have the same latent variable dimension and overlap completely. In this case all the grades in $Y_A$ can be included in the model. Any grade contained in $Y_A$ should also be producable in plant B except perhaps for grades that lie far from any of the $Y_B$ grades. It is possible that hard constraints on the manipulated variables in plant B may not permit production of a grade from the A plant that lies in the $Y_B$-space but far from existing $Y_B$ grades. However, it may still be worthwhile to look at the predictions for process conditions for such a grade. It should be apparent from the set of solutions obtained for $x_{B_{pred}}^T$ whether they look feasible, i.e. whether the model still has validity in that area, and whether an attempt to implement them seems promising or not.

2. $Y_B$ shows a higher dimensionality than $Y_A$ but the space of $Y_A$ is contained within the $Y_B$-space. In this case the same comments made above for case 1 will again hold.

3. A more general case is one where the spaces of $Y_A$ and $Y_B$ intersect, but where part of the $Y_A$-space is not contained within $Y_B$. This implies that there are some grades in $Y_A$ that have a quality structure which is different from that of $Y_B$. Obviously, the available data for plant B do not allow the prediction of process conditions for production of such grades in plant B. Even if it were possible,
plant B would have to be operated differently from the way all the existing $Y_B$ grades are produced, and therefore such an operating strategy cannot be deduced from the structure of the available data. Hence it is necessary to exclude those grades in $Y_A$ that have this different structure. Once this is done we will again be in the situation described under case 1 or 2.

4. There is no overlap between the two spaces $Y_A$ and $Y_B$. In this case no solution is possible based only on the empirical data.

One way to test for similarity between spaces is described in (Krzanowski 1979). Orthogonal basis vectors are calculated by PCA for each space. The two basis systems are then rotated until their directions are most similar. The angles between those most similar basis systems are a measure of spatial dissimilarity between the two spaces.

In this work we use another method to compare the two spaces and to select suitable plant A grades that lie within the space of $Y_B$. A PCA-model is built on $Y_B$ to describe the quality space spanned by plant B. All the plant A grades have to be tested against this model, i.e. they are projected onto the model space of $Y_B$ and their Squared Prediction Error (SPE) is calculated. By comparing those SPEs to the SPEs of the model grades $Y_B$, one can detect the grades in $Y_A$ with a much larger SPE. These grades have a different latent structure from the $Y_B$ grades, and therefore should be excluded from the $Y_A$ data used for the modeling.

Note: Unfortunately, in most data sets there will not be enough grades in plant B available to allow for estimation of a confidence limit for the SPE, as is done in Nomikos and MacGregor (1995). Therefore, the decision whether or not to include a certain grade in $Y_A$ might have to be based on one's own judgment on the acceptable size of the SPE for the $Y_A$ grades.

Consider the case where a $Y_A$ grade that one wants to produce in plant B has
an SPE slightly larger than all the $Y_B$ grades. It may be worthwhile to investigate the reasons for this larger SPE. This can be done by looking at the contribution of each quality variable to the SPE-value. Contribution plots as described in (MacGregor et al. 1994) can be employed for this purpose. They display how much each $Y$-variable contributes to the SPE. If it turns out that it is mostly one variable that causes the somewhat large SPE, and at the same time this variable is not of high importance (an example could be color which might not be vital in terms of application of the product compared to other quality variables), then this $Y_A$ grade might be included in the model in spite of its slightly different structure. In addition, it is important to look at the projection of such a grade onto the PCA-space of $Y_B$. Since we are probably unable to achieve the exact same quality as in plant A, the PCA-model can be used to give a prediction of the actual quality that is achievable in plant B.

Although a $Y_A$ grade may have a small SPE, its projection into the $t$-score space of the PCA-model may still fall far from the $t$-scores of the $Y_B$ grades. In fact, it is possible that all the selected $Y_A$ grades form a cluster in the $t$-space which is separate from the cluster of $Y_B$ grades. This may happen when there are no common grades between the plants, and when the means of $Y_A$ and $Y_B$ differ a lot from each other.

### 4.3.2 Scaling of the Data

The columns in $X_B$ and $Y_B$ are meancentered and scaled to unit variance. Since selected grades from plant A are supposed to be lined up with $Y_B$ later on, $Y_A$ is also centered and scaled using the mean and the standard deviation of the $Y_B$ data: the means of the variables in $Y_B$ are subtracted from the columns in $Y_A$, which are then scaled by the standard deviation of the columns in $Y_B$. Since the approach discussed in 4.3.3 does not use the data in $X_A$ at all, this eliminates any concerns about centering and scaling the $X_A$ data. Scaling both $Y_A$ and $Y_B$ with the mean and
standard deviation of $Y_B$ offers the following benefits:

1. The common grades between $Y_A$ and $Y_B$ still have equal values after scaling and adjusting for the mean, and therefore have the same $t$-scores as well. A PCA-model on the scaled and mean adjusted \([Y_A, Y_B]\) will thus be more focused on the common space.

2. This common mean-centering and scaling of $Y_A$ and $Y_B$ ensures that the relationship between the two spaces is not altered. In other words, if the unscaled $Y_A, Y_B$ fall into the same space or intersect in a certain way, then so will the scaled matrices when scaled and mean adjusted as suggested above.

### 4.3.3 Approach Using an Extended PCR-Model

In this section first the concepts and motivation for the approach are laid out before the mathematical equations for it are given in detail. In order to keep things simple, some additional issues are presented only at the end of the section.

The suggested approach to this problem is closely related to the PCR inversion approach in chapter 2. It is based on two data matrices: $X_B$ ($n_B \times m_B$) and \([Y_A, Y_B]\) ($(n_A + n_B) \times k$).

$X_B$ contains the adjustable variables of plant B that make up the process conditions for the different grades. As far as the quality matrix \([Y_A, Y_B]\) is concerned there is no requirement that it should consist of an independent subset of quality variables. Since plant A already produces the desired product grade $y_{A_{des}}^T$, its physical feasibility is ensured. We also assume that it is possible to produce $y_{A_{des}}^T$ in plant B, as it belongs to the grades that have a reasonably small SPE with respect to a model on historical plant B grades, and $t$-score values within the range of the $Y_B$ grades (see section 4.3.1). Thus, the quality matrix \([Y_A, Y_B]\) may contain any available quality
measures that are considered vital for the final product, and are available for both plants.

Obviously, any latent variable model (e.g. PCA) on the augmented matrix $[Y_A^T, Y_B^T]$ will change the directions of the principal components one would get from a model on $Y_B$ alone. However, since the grades in $Y_A$ have been selected to lie mostly in the same space as the $Y_B$ grades, the actual latent space that is obtained for $Y_B$ by the model on $[Y_A^T, Y_B^T]$ should be effectively the same as the space captured by e.g. a PCA-model on $Y_B$ alone. The benefit of using the augmented matrix is that through the increase in data points it leads to a more solid model, in particular in the area that is common to $Y_A$ and $Y_B$ grades.

The purpose of any model in this context is to eventually estimate process conditions $x_{B_{pred}}^T$ for plant B that result in $y_{A_{des}}^T$. With the same arguments as laid out in the chapter on product design, we shift our interest from estimating the correlated values for the manipulated variables directly to estimating the independent and scaled (to equal variance) underlying latent variables of $X_B$. Since the $Y$ data here are correlated and possibly rank deficient, it makes sense to relate the latent variables of $X_B$ not to $Y_B$ directly but to the latent variables of $Y_B$. These latent variable spaces that underlie $X_B$ and $[Y_A^T, Y_B^T]$ can be estimated by PCA or singular value decomposition (SVD) where only the first few significant components are retained.

We assume that the space spanned by $Y_B$ is basically contained within the space spanned by $X_B$, that is that all important operating features that give rise to the quality features of the different grades are contained within $X_B$. However, estimating the latent spaces for $X_B$ and $Y_B$ separately will result in two different representations of the common space. Furthermore, the process conditions typically span a larger space (dimension $B$) than the quality variables (dimension $C$) as already discussed in chapter 2 ($B \geq C$, where $B$ denotes the number of latent directions in $X_B$, and $C$ the number of latent directions in $Y_B$). The key is to transform the latent
space of $X_B$, $(U_B)$, such that it lines up as closely as possible with the latent space formulation of $Y_B$, $(U_{yB})$. This transformation can then be inverted in order to derive an estimate for the normalized $X$-scores $(u_{Bpred}^T)$ from the $Y$-scores of $y_{A_{des}}^T$, $(u_{yB_{des}}^T)$. These $Y$-scores are already known from the model, since $y_{A_{des}}^T$ is a data point in $[Y_A^T, Y_B^T]$. As before, the covariance structure from the historical data $X_B$ is then imposed onto the latent values $u_{Bpred}^T$ and the corresponding values for the manipulated variables obtained.

The ideas that have been outlined so far are shown in more detail in the mathematical formulation that follows. Since the grades in $Y_A$ have been selected to lie within the space spanned by $Y_B$, we assume $\text{rank} \left(\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}\right) = \text{rank}(Y_B) = C$.

We begin with singular value decompositions on $X_B$ and $[Y_A^T, Y_B^T]$ where the first $B$ and $C$ significant components respectively are retained. They serve as approximations for $X_B$ and $[Y_A^T, Y_B^T]$:

$$
\hat{X}_B = U_B \cdot \Sigma_B \cdot V_B^T \\
(n_B \times m_B) \quad (n_B \times B) \quad (B \times B) \quad (B \times m_B)
$$

$$
\begin{bmatrix} y_A \\ y_B \end{bmatrix} = U_y \cdot \Sigma_y \cdot V_y^T \\
(n \times k) \quad (n \times C) \quad (C \times C) \quad (C \times k)
$$

(4.1)

where $B$ is the number of latent directions in $X_B$, $C$ is the number of latent directions in $[Y_A^T, Y_B^T]$, and $n = n_A + n_B$. $U_B$ and $U_y$ can be interpreted as the latent variables that span $X_B$ and $[Y_A^T, Y_B^T]$ respectively.

To relate the latent variables of $X_B$ to the latent variables of $Y_B$, we first need to separate the matrix $U_y$ into the two parts that correspond to $Y_A$ and $Y_B$ respectively:

$$
U_y = \begin{bmatrix} U_{yA} \\ U_{yB} \end{bmatrix}
$$

The latent structure of the grades in $Y_B$ is represented by $U_{yB}$ which is of dimension
(n_B \times C).

In order to express \( U_{yB} \) in terms of the latent variables of \( X_B \), we must rotate and stretch the latent variables \( U_B \) until they line up as closely as possible with the latent variables \( U_{yB} \). In other words, we are looking for a matrix \( R_B \) such that:

\[
U_{yB} = U_B \cdot R_B
\]

\( (n_B \times C) \quad (n_B \times B) \quad (B \times C) \quad .
\]

An estimate for \( R_B \) can be obtained from the least squares solution to equation (4.2):

\[
\hat{R}_B = U_B^T \cdot U_{yB}
\]

\( (B \times C) \quad (B \times n_B) \quad (n_B \times C) \quad .
\]

The above is basically Principal Components Regression (PCR) where the principal components of the \( Y_B \)-space are regressed onto the principal components of the \( X_B \)-space.

The \( Y \)-scores for the grade \( y_{A_{des}} \), that is to be produced in plant B are already available as a row in \( U_{y_A} \), since \( y_{A_{des}}^T \) is a row in the \( Y_A \)-matrix and as such part of the model (4.1). An advantage of using the combined matrix \( \left[ Y_A \right] \) is that \( Y \)-scores for a grade are the same whether a grade is produced in plant A or B, i.e. \( u_{yB_{des}}^T = u_{yA_{des}}^T \).

The problem that remains is to transform these values from the latent space of \( Y_B \) to the latent space of \( X_B \). The key here is equation (4.2) which can be written for the desired grade as:

\[
u_{yB_{des}}^T = u_{B_{pred}}^T \cdot \hat{R}_B
\]

\( (1 \times C) \quad (1 \times B) \quad (B \times C) \quad .
\]

As in chapter 2, for \( B > C \) we have an underdetermined equation system with \( C \) equations and \( B \) unknowns which has an infinite set of solutions \( u_{B_{pred}}^T \) with minimum error. Each of the solutions can be decomposed into the minimum norm solution

\[
\hat{u}_{B_{new}}^T = u_{yB_{des}}^T \cdot (\hat{R}_B^T \cdot \hat{R}_B)^{-1} \cdot \hat{R}_B^T
\]

(4.4)
and a null component $u_{B_{null}}^T$ so that $u_{B_{pred}}^T = \tilde{u}_{B_{new}}^T + u_{B_{null}}^T$. Whereas the above inversion of the PCR-model provides the best solution that lies in that subspace of $\hat{X}_B$ which according to the model affects the product quality, the null components lie in the remaining subspace of $\hat{X}_B$ that does not affect the product quality. This latter subspace is characterized by the following relationship:

$$u_{B_{null}}^T \cdot \hat{R}_B \cdot \Sigma_y \cdot V_y^T = 0 \quad (4.5)$$

In other words, any vector $u_{B_{null}}^T$ that lies in the left nullspace of $\hat{R}_B$ fulfills equation (4.5). This nullspace is defined with help of the singular value decomposition of $\hat{R}_B$ by:

$$\hat{R}_B = [G_1 : G_2] \cdot \Sigma_{\hat{R}_B} \cdot V_{\hat{R}_B}^T$$

$$(B \times C) \quad (B \times C) \quad (B \times (B-C)) \quad (B \times C) \quad (C \times C)$$

$G_2$ in this equation denotes the left nullspace of $\hat{R}_B$. Hence, $u_{B_{null}}^T$ can be any vector in the space spanned by the rows of $G_2^T$:

$$u_{B_{null}}^T = \lambda^T \cdot G_2^T$$

$$(1 \times B) \quad (1 \times (B-C)) \quad ((B-C) \times B)$$

where $\lambda^T$ is an arbitrary row vector of the proper dimension.

Although mathematically $u_{B_{null}}^T$ is not only of arbitrary direction within $G_2^T$, but also of arbitrary magnitude, for the current problem only those solutions are acceptable that are within the range of the historical data. In other words any $u_{B_{pred}}^T = \tilde{u}_{B_{new}}^T + u_{B_{null}}^T$ must fall within the cluster of the existing $U_B$-rows in order to guarantee a solution that is compatible with the past plant B operating strategies that are represented in $X_B$.

The reconstruction of values for the manipulated variables is achieved in the same fashion as in chapter 2. By imposing the variance and covariance information contained in $\Sigma_B$ and $V_B^T$ the values for the scaled variables are obtained:

$$\hat{x}_{B_{pred}}^T = (\tilde{u}_{B_{new}}^T + u_{B_{null}}^T) \cdot \Sigma_B \cdot V_B^T$$
Multiplying with the standard deviation and adding the mean of $X_B$ will give the actual values for the manipulated $X$-variables.

As in the chapter on product design we again obtain not just a single solution, but a whole region of operating conditions for plant B which according to the model should result in the same grade quality as $y_{A_{ext}}^T$ from plant A. Again it is also left up to the operator or plant engineer to decide upon the most promising conditions that one may want to implement as a first step. Iterations between implementing the predicted conditions, re-building the model with the newly obtained data point, and re-inverting may also be necessary to fine tune the quality values.

To keep the mathematical outline of the approach simple two issues have been left aside so far:

1. The inversion in equation (4.4) requires the $(B \times C)$ matrix $\hat{R}_B$ to be of full rank $C$ ($C \leq B$). $\hat{R}_B$ is calculated as

   \[
   \hat{R}_B = U_B^T \cdot U_{y_B} 
   \]

   \[
   (B \times C) \quad (B \times n_B) \quad (n_B \times C)
   \]

   and therefore, $\hat{R}_B$ is of rank $C$ unless at least one of the columns in $U_{y_B}$ lies in the nullspace of $U_B$. However, this would suggest that some of the variation in $Y_B$ is caused by an “event” that is not present in the variation in $X_B$. In other words, there is variation in the quality of the grades that is not caused by the manipulated variables contained in $X_B$. Obviously it is not possible in such a case to change the grade quality in the desired way by only changing the manipulated variables that are present in $X_B$. Such a data set would not be suitable for a databased approach, and hence is not considered any further.

2. The second issue addresses the assumption made earlier that \( \text{rank} \left( \begin{bmatrix} Y_A \\ Y_B \end{bmatrix} \right) = \text{rank}(Y_B) \). This should always be the case if only $Y_A$ grades have been selected that fall into the $Y_B$-space. If however one or more grades are included in $Y_A$
that have a large SPE then this assumption that rank \( \left( \begin{bmatrix} Y_A^a \\ Y_B^a \end{bmatrix} \right) = \text{rank}(Y_B) \) might be violated. It is then possible that \( Y_B \) is effectively of a rank \( C_B \) that is less than the rank of the combined matrix \( \begin{bmatrix} Y_A^a \\ Y_B^a \end{bmatrix} \) \( (C) \). This happens if \( Y_A \) contains one or more grades which display variation in a direction that is not contained within \( Y_B \), and thus add another dimension to the augmented matrix. In this case also \( U_{yB} \) is effectively of rank \( C_B \) with \( C_B < C \). In order to maintain well conditioned matrices for later inversion, \( U_{yB} \) must then contain only those \( C_B \) columns that actually span the space of \( Y_B \).

### 4.4 Simulation Example

To illustrate the procedure of the approach described in the previous section a simple simulation in MATLAB is used.

A \((12 \times 9)\) orthonormal matrix \( Z = [z_1 \ldots z_9] \) is used to generate data on twelve grades for plant B as follows. A linear combination of \([z_1 \ldots z_5]\) is used to create a \((12 \times 5)\) \( Y_B \)-matrix. Linear combinations of \([z_1 \ z_2 \ z_3 \ z_6 \ldots z_9]\) are used to create the \((12 \times 8)\) \( X_B \)-matrix. In both cases \( z_1, \ z_2, \ z_3 \) have larger weights in the linear combination than the other \( z_i \). This ensures that the dominant three latent dimensions in \( X_B \) and \( Y_B \) overlap. The remaining dimensions are supposed to account for noise, and in \( X_B \) also for directions that have no effect on the quality space \( Y_B \).

The first two grades of \( Y_B \) are common to both plants. Additional plant A grades have to be generated such that at least some of them fall within the \( Y_B \)-space. This space is spanned by the first three principal components of a PCA-model on \( Y_B \). Linear combinations of these three components result in grades that lie within the \( Y_B \)-space. To some of these grades a small amount of independent variation is added. Including the common grades, nine grades for plant A are generated. They are collected in a \((9 \times 5)\) matrix \( Y_{A\text{initial}} \) to indicate, that the grades have not been
tested yet whether they have a structure similar to that displayed by $Y_B$ or not.

Table 4.1 shows the cumulative variance explained by separate PCA-models on $X_B$, $Y_B$, and $Y_{A_{initial}}$. (All matrices are are mean-centered and scaled to unit variance.)

Table 4.1: Percentage of Variation Explained by PCA in $X_B$, $Y_B$ and $Y_{A_{initial}}$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA on $X_B$</td>
<td>35.96</td>
<td>63.59</td>
<td>85.93</td>
<td>95.44</td>
<td>99.87</td>
<td>99.99</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>PCA on $Y_B$</td>
<td>46.70</td>
<td>80.90</td>
<td>99.08</td>
<td>99.82</td>
<td>100.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCA on $Y_{A_{initial}}$</td>
<td>51.83</td>
<td>76.77</td>
<td>99.97</td>
<td>99.99</td>
<td>100.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

According to this table, the matrices $X_B$, $Y_B$, and $Y_{A_{initial}}$ are spanned by four, three, and three latent variables respectively. The first step is to test which of the grades in $Y_{A_{initial}}$ lie in the same space as the grades of $Y_B$. As we see from table 4.1, the first three principal components span 99% of the $Y_B$-space. Therefore, it is the first three latent vectors of $Y_B$ onto which $Y_{A_{initial}}$ must be projected. For that purpose the mean of $Y_B$ has to be subtracted from $Y_{A_{initial}}$, and the resulting matrix must be scaled by the standard deviations of the columns of $Y_B$. For the mean adjusted and scaled matrix $Y_{A_{initial}}$ the squared prediction error to the PCA-model on $Y_B$ is calculated. Figure (4.2) shows the SPE-values for the grades in $Y_B$ denoted by *(*) and the SPE-values for $Y_{A_{initial}}$ grades denoted by (o).

The SPEs of $Y_{A_{initial}}$ grades 1, 2, 4, 5, 6, 7, and 8 are within the range of those obtained for the $Y_B$ grades. (Remember that 1 and 2 are the common grades and hence are expected to have the same SPE in both matrices.) Grade 3 of plant A with its large SPE should obviously not be used in $Y_A$. The SPE for grade 9 is a bit larger than any SPE of the $Y_B$ grades. The contribution plot for this SPE (figure (4.3)) shows that most of this error is due to variables 2 and 3. We pretend that we are not extremely concerned with these particular two variables and include grade 9
in $Y_A$, as well as the selected grades (1, 2, 4, 5, 6, 7, 8) with smaller SPEs.

With the combined matrix $[Y_A \ Y_B]$ and $X_B$ the extended PCR-model is built as described in section 4.3.3. Table (4.2) shows the variance explained in $Y_B$ by the model, when four latent directions are used in $X_B$ for different numbers of latent directions in $Y_B$. (Note, that $X_B$ is modeled by the same principal components already displayed in table (4.1).)

Table 4.2: Percentage of Variation Explained by the Extended PCR-Model in $Y_B$
Using the First Four Latent Variables in $X_B$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_B$</td>
<td>41.64</td>
<td>61.93</td>
<td>91.39</td>
<td>91.62</td>
<td>91.63</td>
</tr>
</tbody>
</table>

The plant B process conditions for the six grades in $Y_A$ that are not common with $Y_B$ (4, 5, 6, 7, 8, and 9) are obtained through inversion of the model as explained...
Figure 4.3: Contributions of the Y'-Variables to the SPE of Grade 9 in $Y_A$

in section 4.3.3. Since there is no real process behind this simulation, "implementation" of these conditions can only be simulated by multiplying the new process conditions with the coefficient matrix that represents the linear relationship between $X_B$ and $Y_B$. This is done for all seven conditions $\hat{x}_B^{\text{new}}$, where no null component is added. A null component would according to the model not affect the quality anyhow. After this "implementation" of $\hat{x}_B^{\text{new}}$ for the six $Y_A$ grades the achieved $Y_{B_{\text{achieved}}}$ can be compared to the desired qualities in $Y_A$. Results are shown in figures 4.4 and 4.5. A selective PCA analysis (Roffel et al. 1989) on $Y_B$ indicates that the variables $Y_4$, $Y_3$, and $Y_5$ are the most significant variables of all five, and account for 98% of the variability.

Obviously, the achieved results are very close to the desired ones which is probably also due to the simulated nature of the data. However, we do not claim that this example proves much in terms of performance of the method, in fact, the performance depends on the particular data set in a similar way as discussed in the
Figure 4.4: Qualities for $Y_B$ Grades, Desired $Y_A$ Grades, and $Y_A$ Grades Achieved in Plant B: $Y_3-Y_4$

Figure 4.5: Qualities for $Y_B$ Grades, Desired $Y_A$ Grades, and $Y_A$ Grades Achieved in Plant B: $Y_3-Y_5$
chapter on product design. The example here serves merely to illustrate application of the methodology.

4.5 Discussion and Conclusions

Based on the problem of product design treated in chapter 2, this chapter addresses the related problem of transferring production of a product grade from one plant A to another plant B.

Since for this problem data sets from two plants are available, the structure of these data and the relationships between them are very important. It is of particular interest to compare the quality spaces from the two plants. Whether process conditions in plant B for a desirable plant A grade $y_{A_{des}}^T$ can indeed be derived through a databased approach hinges largely on the relationship between the qualities produced in the two plants so far:

1. The model for plant B must still be valid in that area of the score space into which $y_{A_{des}}^T$ falls, and

2. the resulting process conditions for plant B must still be feasible and must not hit any hard constraints.

Of course there are cases where due to the different equipment and configuration of the two plants production of $y_{A_{des}}^T$ may not be possible in plant B at all. On the other hand, even if theoretically it was possible to produce $y_{A_{des}}^T$ in plant B, a databased approach cannot foresee this unless $y_{A_{des}}^T$ lies within the structure and range of the plant B grades present in the historical data. In other words, the typical limitations inherent to any databased approach apply here as well. Although the production of a particular grade $y_{A_{des}}^T$, may be possible in plant B, the historical data do not contain the necessary information to convey this possibility, if the grades so far
produced in plant B are very different from $y_{A_{des}}^T$. In the same way it is only feasible to predict process conditions that are consistent with the operating strategies that have been recorded for this plant in $X_B$.

The approach described in section 4.3.3 is only one of many possible ones. It can be seen as an extended PCR-model with inversion, where the quality space of plant B is supported by suitable plant A grades. A test has been developed that selects those plant A grades that have a latent structure similar to the grades in $Y_B$. Combining those plant A grades with $Y_B$ enhances the model on the quality structure especially in the quality area that is common to both plants.

The inversion of the model follows the same concepts as in chapter 2, where first predictions for the normalized latent variables of the new process conditions are obtained. Imposing the variance and correlation structure of the plant B process variables ($X_B$) onto these values then results in the settings for the physical manipulated variables.

If the process conditions span a higher dimensional space than the quality measurements, as often is the case, the solution can again be presented as a whole window of operating conditions which - according to the model - all lead to the same desired quality $y_{A_{des}}^T$.

The information used in the approach consists of all the plant B data and the quality data of plant A grades which have a covariance structure similar to the plant B grades. No use at all is made of the process conditions in plant A ($X_A$). However, one can think of approaches which also include this information. In the course of this work, two models and their inversions were developed that use all the data on the two plants (see figure (4.1)). They are based on multiblock latent variable methods as described in (Wangen and Kowalski 1988, Wold et al. 1996, Wold et al. 1987b). Since they are mainly based on intuition, however, they are not included here.
A more extensive study may apply different types of models and their inversions and compare their performance with respect to different data sets.
Chapter 5

Process Improvement Using Normal Operating Databases

5.1 Introduction

In the past years many industrial processes have become increasingly computerized. As a result hundreds of process measurements are often routinely collected over long periods of time, so that there exist large databases containing process operating data. At the same time, it is of great interest in industry to further improve and optimize process operating performance. When fundamental process models are not available, as often is the case, the question arises whether anything can be done by exploiting these readily available databases. This chapter attempts to explore the possibilities and limits of the use of historical databases in the context of process or quality optimization. Of course any databased approach is limited by the range and scope of the available data, and therefore one cannot expect to optimize the process in the sense of conventional optimization techniques which are based on fundamental models. The term "optimization" is used in this chapter in the sense of improvement that can be achieved given only the available process data in a region about the current operating
point. Such an improvement or optimization requires that causal information can be extracted from the data, so that a change in the current operational settings can be inferred that would lead to a better product quality. The problem which often is ignored is that such operational databases typically contain mostly non-causal data. Fisher recognized the problems inherent to such databases in his famous statement: "All one can do with such happenstance data is a postmortem to see what they died of" (Fisher n.d.). Therefore, one should be cautious towards any attempt at utilizing databases for purposes where causal information is needed, as e.g. when process conditions for improved product quality must be inferred.

Instead of using historical databases Fisher proposed experimental design as a means of obtaining suitable data. The manipulatable process variables of interest are changed independently according to a preconceived concept in order to provide data that allow for extraction of causal relationships between manipulated and response or - as in our case - quality variables. Response surface methods (RSM) (Box et al. 1978) use this concept in finding the local gradients at the current point of operation, and then give the directions of steepest ascent in which to move the process in order to get an improvement in quality in the region of the current point.

Evolutionary Operation (EVOP) as described by Box and Draper (1969) consists of a series of such experiments on-line. It is an approach to implement a simplified version of RSM as a method of routine plant operation. Simple factorial designs in only two or three variables are carried out about the current operating point until a direction of improvement is evident. The process is then moved a short distance in this direction and once a new steady-state has been reached where further improvement along the current direction of steepest ascent is not possible, another set of experiments is designed. The changes in manipulated variables are kept purposely small to avoid process upsets.

In general, such data that are obtained through independent changes in the
manipulated variables $x_i$ and the corresponding measurements of the resulting response or quality variable $y$ can be used to fit a regression model

$$y = b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + e$$

where $e$ accounts for the model error, and the data have been mean-centered. The regression coefficients $b_i$ give account of the causal relationship between manipulated variables $x_i$ and the response variable $y$.

In the case of normal operating data, however, we typically do not encounter this case of experimentally designed data. In normal process operation the manipulated variables are often moved in a highly correlated fashion. Furthermore, for some time during data collection the process is likely to be affected by unmeasured disturbances that cause highly correlated variation in some of the process variables. If operators or automated control systems react by changing the manipulated variables in order to counteract the effect of the disturbance, then feedback correlation is introduced into the data.

Of course it is possible to fit regression models to this kind of operating data. But in the face of these problems (correlation, unmeasured disturbances, feedback), it becomes clear that the regression coefficients obtained from such a model cannot be interpreted as describing the causal relationship between the variables anymore. MacGregor et al. (1991) discuss in detail the problems with empirical modeling of non-causal data. They start with a simple example of five manipulated variables $x_1 - x_5$ and one response variable $y$. A causal relationship, however, exists only between $y$ and $x_2$:

$$y = x_2 + e$$

All five $x$-variables change in a perfectly collinear manner. Consequently, an Ordinary Least Squares model is not possible here due to ill-conditioning of the data. However, other methods such as stepwise regression, PLS or Neural Networks can be used to
obtain estimates of the coefficients in the model:

\[ y = b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + b_4 \cdot x_4 + b_5 \cdot x_5 + e \]

Different estimation methods will provide models with sometimes greatly different estimates of the coefficients \( b_i \), however, all these models predict \( y \) equally well provided that the correlation structure among the data remains the same, i.e. as long as \( x_1 - x_5 \) are still perfectly collinear, and nothing else that affects \( y \) changes. It is obvious that even though the data could be called open-loop, the ill-conditioning prevents the precise estimation of the regression coefficients.

In a second example MacGregor et al. (1991) address the problems that are introduced when feedback conditions are present during data collection. The causal relationship in this second example is between overhead purity and reflux ratio of a distillation column. Theoretically, purity should rise with increased reflux ratio, thereby giving a positive coefficient for the relationship between purity and reflux. A regression model on the process data from the column, however, indicates a negative correlation between the two variables. As it turned out, an operator had been adjusting the reflux in order to keep the purity at constant level: whenever purity was low due to some disturbance, he increased the reflux, and vice versa. This introduced the negative correlation in the data. We know already from dynamic non-parametric identification that dynamic models built on data that have been collected under feedback conditions identify instead of the process model the negative inverse of the controller (Box and MacGregor 1974). The example above shows that even when building a model from steady-state data under feedback conditions, the model coefficients cannot be interpreted as causal. Moreover, they usually will show the opposite sign as compared to the true causal relationship.

A complex formal methodology based on machine learning techniques, decision trees etc. has been suggested by Saraiva and Stephanopoulos (1992) for continuous process improvement on the basis of historical operating data. (See also Saraiva
(1995), and Saraiva and Stephanopoulos (1994).) However, they do not at all address the difficulties that are inherent to such data, such as correlated operating variables, unmeasured disturbances, or feedback correlation. The presented case studies assume that the operating or manipulated variables show independent variation over the whole operating space. In this sense such databases would provide an even more ideal situation than one assumes for the application of RSM or EVOP. Unfortunately, real process data are typically of a much more unforgiving nature, and the suggested methods cannot be applied to such more realistic data.

De Smet (1993) recognizes the problems that come with typical process data, and has developed a way of collecting and organizing data in order to avoid them. She assumes that any disturbances remain constant over the period in which the operating variables are changed. The database consists of these changes in the manipulated variables and the corresponding changes in the response variables. Under the assumption that the disturbances were constant during that time, the data contains open-loop information which allows for retrieval of gradients that describe the cause and effect relationship between operating and response variables. She then uses local models built from the data to perform process optimization.

Chen and McAvoy (1996) address the problem of improving operating performance through feedback control of the system. Since they base their control algorithms on multivariate statistical models built from a historical process database, it is worthwhile mentioning their work here. They start out from an earlier approach by Piovoso and Kosanovich (1992), point out the errors in that first approach, and move on to suggest significant modifications. Being aware of problems caused by unmeasured disturbances, they assume that operator feedback actions are recorded in the historical database. Two control algorithms are suggested in their paper, both of which have very strong underlying assumptions: the first one requires a set of process variables whose variation is caused mostly by the unmeasured disturbances and not
by changes in the manipulated variables. In other words, they assume the existence of variables which basically indicate the unmeasured disturbances. The second algorithm is based on gradients of the measured process variables versus the manipulated variables. Again it is questionable how causal derivatives can be built from highly correlated data including feedback correlation. A closer look at the examples in the paper reveals that the databases that are actually used are open-loop in nature: at constant levels of disturbance the changes in manipulated variables are recorded together with the resulting responses in process variables. The authors are basically using the same type of data as De Smet (1993), only in a slightly different context: they are aiming at controlling the process rather than optimizing it. However, as discussed before, data collected during routine operation typically do not consist of independent open-loop variation in the variables and therefore will not give causal models.

In face of these examples it seems important to investigate the problem of what information can be extracted from realistic process operating data, and what can possibly be achieved by relying only on such data. The only assumptions on the database are that the recorded points represent steady-state operation, and that the process operating mode is consistent throughout the database and continues to be so at the present time. Both unmeasured disturbances and feedback responses may be present. The work discussed here is by no means complete; it is rather an attempt to shed some light onto these very subtle problems that come with process data. It is hoped that this will lead others to continue and improve the work presented here.

In the next section some ideas are developed that may serve as a way to work with these problems of correlation, unmeasured disturbances, and feedback. However, in outlining these concepts we also encounter the strict limitations of such historical process operating data. In discussing the proposed method the difficulties and hurdles of any approach using routine operating data for process optimization become more apparent.
5.2 Similarity Optimization

In this section we take a thorough look at the difficulties inherent to typical historical databases with respect to the extraction of causal information. The investigation motivates the attempt to reconstruct unmeasured disturbances which is outlined in section 5.2.1. As the discussion section 5.2.2 reveals, however, this approach is also severely limited by the ill-conditioned nature of normal operating data. Again it should be pointed out that the term "optimization" in this context should not be confused with its conventional use, but rather taken as synonymous with the term "improvement".

Initial work in this area has been conducted by De Smet (1993) as mentioned in the previous section. Recognizing the problems involved in such a databased approach to process optimization, she resorts to the reorganization of the database to collect any data that involve causal effects. In particular, all the changes in the manipulated variables and the resulting changes in the steady-state responses are saved. Under the assumption that the process disturbances did not change over this period these data could be used to provide causal gradients for the manipulated variables in the region in which the data are collected. However, the question remains whether anything can be done - using the readily available data - that does not require these assumptions and reorganizations.

There are basically two sources of variation in operating data: one source obviously comes from changes in the manipulated variables which in turn cause variation in intermediate process variables and in product quality. The changes in the manipulated variables are assumed to be recorded in the database. However, a second source of variation lies in disturbances which also affect intermediate and quality variables. Unfortunately, this second source of variation is not directly measurable. Moreover, it often causes another pitfall in databases: some operating variable(s)
may have been moved deliberately in order to lessen or completely counteract the
effect of a disturbance on the product quality. This is what is referred to as feedback
control. In the course of feedback the variation caused initially by the disturbance
is transferred from the intermediate process variables and quality variables to one or
more manipulated variables that are changed in response to the disturbance effect.

Consider a simple example of one manipulated variable \( MV \), one quality or
performance variable \( q \), and one unmeasured disturbance \( D \). Let the causal relation-
ship between these variables be

\[
q = MV + D
\]

A database for this example process could arise from any of the following four
different scenarios:

1. No disturbance is present \((D = 0)\), and the quality is affected only by a change
   in the manipulated variable \( MV \). This is a deterministic open-loop process.

2. Both, disturbance and manipulated variable assume values different from zero
   \((D \neq 0 \text{ and } MV \neq 0)\). Thus, the quality \( q \) is the result of the sum of both. The
   open-loop effect of the relationship between \( q \) and \( MV \) is compounded with the
   additional effect of the disturbance \( D \) on \( q \).

3. The disturbance is present \((D \neq 0)\), but the manipulated variable is moved in
   feedback manner to counteract the disturbance effect on \( q \). As as result \( q \) does
   not change at all. This is the case of perfect control.

4. The last case falls in between the cases 2 and 3: the manipulated variable
   \( MV \) is moved in correlation with the disturbance, but the control is imperfect,
   i.e. the quality \( q \) is still changing due to the combined effect of disturbance
   \( D \) and the manipulated variable \( MV \). With \( MV = -K \cdot q \), where \( K \) is the
“controller gain”, we can have positive feedback \((K < 0)\) resulting in positive correlation between \(MV\) and \(D\), or negative feedback \((K > 0)\) resulting in negative correlation between \(MV\) and \(D\).

These four cases are illustrated in table (5.1).

Table 5.1: Four Scenarios for the Example \(q = MV + D\)

<table>
<thead>
<tr>
<th>Effect</th>
<th>(D)</th>
<th>(MV)</th>
<th>(q)</th>
<th>Expected Value of (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 deterministic open-loop</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>(b = 1)</td>
</tr>
<tr>
<td>2 open-loop + disturbance</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>(b = 1)</td>
</tr>
<tr>
<td>3 perfect control</td>
<td>4</td>
<td>-4</td>
<td>0</td>
<td>(b = 0)</td>
</tr>
<tr>
<td>4 imperfect control: (MV = -K \cdot q)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) positive correlation btw. (MV) and (D)</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>(\hat{b} = 2)</td>
</tr>
<tr>
<td>b) negative correlation btw. (MV) and (D)</td>
<td>4</td>
<td>-2</td>
<td>2</td>
<td>(\hat{b} = -1)</td>
</tr>
</tbody>
</table>

Under the assumption that the values of the disturbance \(D\) are not available, a database would contain only columns 4 and 5 of table (5.1) - the values for the manipulated variable \(MV\) and the quality \(q\). A regression coefficient \(b\) between \(q\) and \(MV\): \(q = b \cdot MV + e\) can be estimated by fitting a model to the data. The expected values for such a regression coefficient (\(\hat{b}\)) in each of the four cases are given in the last column. The true causal coefficient is \(b = 1\). It is obvious from table (5.1) that the presence of feedback correlation (cases 3 and 4) entails very misleading results when the estimated regression coefficient \(\hat{b}\) is interpreted as a causal relationship. In the more general case of several \(MV\)'s the difficulties caused by correlation with disturbances are even more complex.

On the other hand, even if there exists no correlation between the manipulated variables and the disturbances, if the manipulated variables are highly correlated among each other, then the least squares parameter estimates in a regression model
CHAPTER 5. PROCESS OPERATING IMPROVEMENT

will be highly correlated, and it is difficult to estimate the individual causal coefficients. This has been shown in section 5.1 by the example of MacGregor et al. (1991). However, with the same argument that led to the databased design of process conditions for a new product quality in chapter 2, one could in this case extract the independent latent directions in which to move the manipulated variables while keeping the same covariance structure as in the historical database. In that sense under the constraint of a fixed covariance structure among the manipulated variables, some causal information can be obtained (when there is no correlation between the manipulated variables and the disturbances). Therefore, the element that really results in the inability to extract causal information is the presence of unmeasured disturbances that are correlated with manipulated variables. This may include the possibility of feedback actions as well as cases where certain operating strategies happen to cause correlation between one or more manipulated variables and a disturbance.

This suggests that much of the problem hinges on whether or not it is possible to reconstruct unmeasured disturbances - solely from the information that is present in the database. It is not necessary to reconstruct them as physical variables; it should be sufficient to reconstruct a summary of their effect on the measured variables.

Consider the situation where a process is operating at steady-state. Measurements of operating variables, intermediate process variables as well as some measure of process performance are assumed to be available. The idea of “Similarity Optimization” is as follows: if the disturbances that currently affect the process could be reconstructed, it should be possible to search the historical database for points where similar disturbance situations occurred. Among those points from the past with disturbance reconstructions similar to the current ones, one only has to identify those that resulted in a better quality or performance. From them we can learn how to possibly improve the performance under the current circumstances.

For this approach several assumptions need to be made about the database.
CHAPTER 5. PROCESS OPERATING IMPROVEMENT

First of all, data spanning the entire operating region of a plant are likely to be highly nonlinear. In this approach we employ local models on selected data from only that region of the operating space where the plant is currently operating. This should allow for the use of linear models. We also assume that the database contains steady-state data from one consistent mode of operation only, so that the covariance structure of the data is consistent throughout the database. Another necessary assumption for the approach described above is that the disturbances are constant over the periods of time where the search for nearest neighbors takes place and new operating conditions are implemented.

5.2.1 An Idea for Reconstruction of Unmeasured Disturbances and Search for Improved Operating Conditions

The basic ideas of Similarity Optimization are presented throughout the next few paragraphs. However, the subsequent section will show that there are severe limitations with the approach and indeed with any approach to process improvement based on realistic operating data.

Although the issue of defining process improvement in terms of a quality or performance measure is important, we do not address it here any further. In the following sections we assume for simplicity that quality or performance is defined by a single variable \( q \) and can be measured.

The concept of an approach for the reconstruction of disturbances from a process operating database hinges on the fact that the sources for variation in the data are the changes in the manipulated variables \( MV_i \) and the disturbances \( D_j \). A database on normal operating data will typically consist of

- process variables that are manipulated independently of any process disturbances: \( MV_i, i = 1, \ldots, I \),
CHAPTER 5. PROCESS OPERATING IMPROVEMENT

- manipulated variables that are used to counteract disturbances (or are in some other way correlated with disturbances): $MV_i^D$, $i = 1, \ldots, J$,

- measured intermediate process variables which can also include some “end properties” other than the quality: $X_0^k$, $k = 1, \ldots, K$.

- and a quality or performance variable: $q$.

Table (5.2) illustrates a possible layout of such a database. Each row in table (5.2) corresponds to a particular steady-state operating point.

Table 5.2: Measured Variables in a Typical Process Operating Database

| Independent Manipulated Variables $MV_i$ | Disturbance Correlated Manipulated Variables $MV_i^D$ | Intermediate Process Variables $X_0^k$ | Quality $q$
|-----------------------------------------|--------------------------------------------------|--------------------------------------|------|
| $MV_{i_1}$, $MV_{i_2}, \ldots, MV_{i_7}$ | $MV_{i_1}^D$, $MV_{i_2}^D, \ldots, MV_{i_7}^D$ | $X_{0_1}$, $X_{0_2}, \ldots, X_{0_{K_1}}$ | $q_1$
| $MV_{i_2}$, $MV_{i_3}, \ldots, MV_{i_7}$ | $MV_{i_2}^D$, $MV_{i_3}^D, \ldots, MV_{i_7}^D$ | $X_{0_{i_2}}$, $X_{0_{i_2}}, \ldots, X_{0_{i_{K_2}}}$ | $q_2$

Although disturbances are likely to have occurred during the collection of these data, they are not measured. In order to reconstruct the disturbance effects we have to capture the variation that originates from the disturbances, and separate it from the variation that is caused by independent changes in the manipulated variables $MV_i$.

We assume that changes in the $MV_i$ are not due to any disturbance reactions, and are uncorrelated with disturbances. The settings of the operating conditions ($MV_i$) are expected to fluctuate somewhat, resulting in slightly different steady-states after process shut-downs and start-ups, or because different operators may adjust certain set points differently. Although the $MV_i$ vary independently from the
disturbances, they nevertheless may be correlated among each other due to certain operating strategies or constraints.

Whenever it is known that a particular manipulated variable is correlated with a disturbance, it must be assigned to the group of $MV_j^D$-variables. These $MV_j^D$-variables are not free to be used later for adjustments to improve operating performance, since doing so would alter the correlation structure among the variables, which could lead to very unexpected results since the database does not provide any information on this new structure. The distinction between $MV_i$ and $MV_j^D$ may seem fuzzy at this point, but in the course of the chapter it will become apparent that it is an important feature.

The variables $(MV_j^D, X0_k, q)$ respond to the changes in the disturbances and the $MV_i$. They are therefore correlated with both, $MV_i$ and unmeasured disturbances.

The concept for the reconstruction of disturbances is based on separating the database variables (see table (5.2)) into two groups and regressing one onto the other: The data matrix $X$ collects those variables that only contain the "known" variation, the $MV_i$, whereas $Y$ consists of the variables that also vary due to the "unknown" effects, the disturbances.

$$X = \begin{bmatrix} MV_1 & MV_2 & \ldots & MV_i \end{bmatrix}$$

$$= \begin{bmatrix} MV_i \end{bmatrix}$$

$$Y = \begin{bmatrix} MV_1^D & MV_2^D & \ldots & MV_j^D & X0_1 & X0_2 & \ldots & X0_k & q \end{bmatrix}$$

$$= \begin{bmatrix} MV_j^D & X0_k & q \end{bmatrix},$$

where $MV_i$, $MV_j^D$, $X0_k$, and $q$ stand here for the mean-centered and scaled columns of the measurements.

A regression model between $[MV_j^D X0_k q]$ and $[MV_i]$ will capture the common
variation between the two matrices:

\[
\hat{Y} = f(\hat{X})
\]

\[
\begin{bmatrix} MV_j^D X_0_k q \end{bmatrix} = f \left( \begin{bmatrix} MV_i \end{bmatrix} \right) .
\]

(5.1)

This variation contains only the open-loop effects of the changes in \( MV_i \) on the process variables and the quality. Since the \( MV_i \) are typically correlated among each other, a PLS-model is more appropriate than an OLS-model. The residuals of \( Y = [MV_j^D X_0_k q] \)

\[
Y_{\text{resid}} = Y - \hat{Y}
\]

\[
= \begin{bmatrix} MV_j^D X_0_k q \end{bmatrix} - \begin{bmatrix} MV_j^D X_0_k q \end{bmatrix}
\]

contain the remaining variation which is caused by the disturbances. In other words, the disturbance information in the data is now separated from the open-loop information caused by \( MV_i \) changes, and is captured within the \( Y \)-residuals.

It is not possible (but also not necessary) to transform these residuals into the actual physical disturbance variables, as all we need to retrieve is a disturbance space. This disturbance space is of lower dimension than the number of measured variables. We can build a PCA-model \( T_D = Y_{\text{resid}} \cdot P \) on the \( Y \)-residuals. The dominant principal components will define the disturbance space of the process.

The next similarity optimization step then is to compare the disturbance situation of a current steady-state operation to the situations from the past. For that purpose the PLS- and PCA-model that have been constructed from the historical database are applied to the current process measurements to obtain first the \( Y \)-residuals, and then the current disturbance characteristics \( (t_{D_1}, t_{D_2}, \ldots) \). The latter can now be compared to the disturbance characteristics in the database. One has to calculate the Euclidean distances between the current disturbance characteristics and those from the past, and choose the past points with a distance smaller than a
threshold as nearest neighbors. These nearest neighbors are based on the disturbances as they affect the whole process. If one wants to weigh the disturbance characteristics according to their influence on the quality $q$, the score values from a PLS-model between $T_D$ and $q$ could be used directly in the distance measure.

After a set of nearest neighbors has been selected, one has to identify those that resulted in a more desirable quality than the quality achieved at the current state. Since the disturbance effects of those nearest neighbors with better quality must be similar to the disturbances currently affecting the process, the improved quality must depend on the variables $MV_i$. These operating variables can now be reset either to the values of a particular promising nearest neighbor, or to a weighted average of the $MV_i$ of all the nearest neighbors.

The described concept of Similarity Optimization can be summarized as follows. Consider a model for the quality as:

$$ q = f_1(\text{changes in } MV_i) + f_2(\text{disturbances}) + \text{error} \quad (5.2) $$

where $f_1$ and $f_2$ correspond to the two sources of variation in the data. Since we cannot affect the disturbances ($f_2$), the only handle we have on $q$ are the manipulated variables $MV_i$. By changing them only according to points with similar disturbance characteristics, we effectively use the first part of the model $q = f_1$ only in the region where $f_2$ is constant. In other words, we use the relationship between $MV_i$ and $q$ only in that region of the database where the disturbances are constant and similar to the currently estimated disturbances. This approach ensures that any changes in the $MV_i$ that may be implemented to improve process operation are consistent with the past.

Of course there is always the possibility that in all similar past situations the achieved quality was even worse than the current one. This implies that the process is currently running at its best under the circumstances - as far as the database is concerned. Nothing can be learned in this case from the historical data to improve
quality. However, the database and the models should be updated with the current point so that this knowledge is available for reference in the future.

Obviously, there might exist settings for the manipulated variables $MV_i$ that are very different in structure from the settings recorded in the database, and that could even further improve the current quality. However, no databased approach is able to derive conditions that lie outside the correlation structure or range of the historical data.

5.2.2 Discussion

The Similarity Optimization described in the previous section is based on the necessity to reconstruct information on unmeasured disturbances in order to compare the current process situation to past ones, and learn from similar past situations how the process might be operated in a more successful way than currently done. Three strong assumptions underlie this concept:

1. The database consists of measurements that were taken at times when the process had reached a steady-state, and the mode of operating the process has remained consistent over time - i.e. the covariance structure of the data is not time varying.

2. The disturbances stay at a constant level over extended periods of time. It is therefore possible to implement new settings for the operating variables $MV_i$, and arrive at a new steady-state while the disturbances remain the same.

3. Apart from the manipulated variables $MV_j$ which are correlated with disturbances, there exists a known set of manipulatable variables, $MV_i$, which have been moved independently from disturbances.

It is the third assumption that leads to the main criticism on the approach: if such “independent” $MV_i$ exist - even if they are correlated among themselves - it is possible
to directly extract at least partially causal information from the database. E.g. one can build a PLS-model between these manipulated variables $MV_i$ and the quality $q$. Such a model predicts that part $\hat{q}$ of the quality $q$ that comes from $f_1$ in equation (5.2). If the current $\hat{q}$ value is lower than the highest $\hat{q}$ value of the database, then those combinations of $MV_i$ that result in the highest $\hat{q} = f_1$ should improve the current quality value. However, since the disturbances also affect $q$ ($q = \hat{q} + f_2(\text{disturbances})$), the improved quality may not be as good as the quality achieved in the past, but this improvement is all we can do in the current disturbance situation and within the scope of the historical database. Reconstruction of disturbances may in this approach not even be necessary.

On the other hand, we may not trust the accuracy of this regression model between the $MV_i$-variables and the quality $q$, that is built on all the data regardless of their disturbance situation. It may still make sense to reconstruct disturbances as described above, use the nearest neighbors to build a local model, and hence search for better $MV_i$ settings only in the neighborhood of similar points in the database.

A major problem with any databased approach occurs when it is not known which variables are manipulated by the operators in feedback fashion, and which ones are moved in a manner that is totally uncorrelated with disturbances. Even if there exists a set of independent $MV_i$, nothing can be done unless they can be identified. Any manipulated variables that are known to have been moved in response to a disturbance must be separated from the independent $MV_i$ and assigned to the group of $MV_j^D$-variables that contain disturbance information, and are collected in $Y$. Since it is important that any new process operating settings still follow the same correlation structure that is present in the historical data, these $MV_j^D$ have to remain correlated with the disturbances in the same way as before, and can therefore not be moved to optimize the process.
This leads to the problem that the approach described in section 5.2.1 cannot be applied when all the available manipulated variables have been moved in correlation with disturbances or when it is not known which manipulated variables are correlated with the disturbances. The following simple example illustrates the problem that occurs when the $MV_i$ are partially correlated with disturbances. The variables considered are:

- one manipulated variable: $MV$
- the quality variable: $q$
- a disturbance which affects $q$ as well as a certain process variable $x$ that is observed by the operator for feedback purposes: $D$
- a second disturbance which is independent of $D$, and which affects $q$ but not $x$: $D^*$

We assume that the operator is not perfectly consistent in carrying out the feedback action. The manipulated variable $MV$ contains therefore

- an independent open-loop part: $MV^*$
- in addition to this a feedback part: $-k \cdot D$.

$MV$ and $q$ can be described as follows:

$$MV = MV^* - k \cdot D$$

$$q = b \cdot MV + d \cdot D + D^*$$

$$= b \cdot MV^* + (d - b \cdot k) \cdot D + D^* \quad , \quad (5.3)$$

where $b$, $d$, and $k$ are coefficients. The variables in the above equations can be interpreted as $(n \times 1)$ column vectors containing the $n$ measurements for each variable. $MV^*$, $D$, and $D^*$ can be assumed independent, and are therefore considered
orthogonal (i.e. their scalar product is zero). Least Squares calculates the regression coefficient between \( q \) and \( MV \) as

\[
\hat{b} = \frac{MV^T \cdot q}{MV^T \cdot MV}
\]

\[
= \frac{b \cdot MV^*^T \cdot MV^* - k \cdot (d - b \cdot k) \cdot D^T \cdot D}{MV^*^T \cdot MV^* + k^2 \cdot D^T \cdot D}
\]

\[
= b - \frac{k \cdot d \cdot D^T \cdot D}{MV^*^T \cdot MV^* + k^2 \cdot D^T \cdot D}.
\]

The residual corresponding to the \( Y \)-residuals in section 5.2.1 results here in

\[
q_{\text{resid}} = q - \hat{b} \cdot MV^* \\
= (b - \hat{b}) \cdot MV^* + (d - k \cdot (b - \hat{b})) \cdot D + D^*.
\] (5.4)

\( q_{\text{resid}} \) contains all of the disturbance \( D^* \) and part of the disturbance \( D \), but it also still contains effects due to \( MV^* \). However, in the suggested approach for disturbance reconstruction we rely heavily on the residuals to not contain any open-loop information \( MV^* \). It is therefore worthwhile to investigate if there are cases in which the undesirable term \((b - \hat{b}) \cdot MV^*\) in equation (5.4) becomes negligibly small or zero.

Obviously, it does not make sense to set \( MV^* = 0 \), since this defines \( MV^* \) as a variable moved merely in feedback fashion without any independent variation. The only other situation in which the residual \( q_{\text{resid}} \) would be free of \( MV^* \) is where \((b - \hat{b})\) would be zero. Since

\[
b - \hat{b} = \frac{k \cdot d \cdot D^T \cdot D}{MV^*^T \cdot MV^* + k^2 \cdot D^T \cdot D}
\]

this term becomes zero when the numerator is set to zero. However, this is never the case unless \( k = 0, d = 0, \) or \( D = 0 \), all of which would be contrary to what we assume in this example.

The above excursion shows that there is no reasonable way in which an \( MV \) that is correlated with a disturbance \( D \) that affects the process could be used in
the $X$-matrix of the Similarity Optimization regression model (5.1), and yet lead to separation of open-loop and disturbance information, or in other words disturbance reconstruction. In order to use historical operating data for process optimization we need to separate $MV^*$ and $D$, $D^*$, but this appears to be impossible in cases where $MV$ or all the $MV_i$ are correlated with disturbances $D$.

The insights gained from this example can be extended to the multivariate case where additional process variables $X_{0k}$ and several manipulated variables $MV_i$ are present. If all the manipulated variables are correlated with any of the disturbances affecting the process, then the problem outlined above extends to the multivariate case. The residuals of any regression model will contain not only disturbance information but also part of the open-loop variation. As a consequence the approach of similarity optimization will fail, as would any approach that tries to estimate causal coefficients between these $MV_i$ and the quality $q$.

In Similarity Optimization the extraction of a disturbance space was possible because the disturbances and the other "independent" events could be separated due to the fact that the $MV_i$ are assumed to be uncorrelated with disturbances ($MV_i = MV_i^*$). It is the particular grouping of variables into the $X$- and $Y$-matrices that leads to this result: $X$ must not contain any variables that are correlated with disturbances, or the regression of $Y$ onto $X$ will give residuals that contain a mixture of open-loop and disturbance variation.

### 5.3 Conclusions

This chapter is an attempt to gain a better understanding of the problems that are involved in utilizing routine operating data for process or quality improvement. The main difficulty arises from the non-causal character of such data. While unmeasured disturbances and correlated changes in the manipulated variables distort the causal
relationship between manipulated and process variables as it appears in the database, it is exactly this cause and effect relationship that people often try to extract from such databases. The approach of Similarity Optimization suggested by De Smet (1993) and investigated in this chapter chooses another path: A method is proposed which leads to extraction of the space of the unmeasured disturbances that affect the process. A current operating state that needs improvement can now be compared on the basis of these recovered disturbance characteristics to those points in the historical database which have encountered similar disturbances. Rather than making use of a cause and effect relationship, Similarity Optimization suggests changing the independently manipulated variables according to the settings found in the similar points in the database. By using only those historical data points with similar disturbances, and by changing the manipulated variables in accordance with the manipulated variables of those historical points, we ensure that the new process state still maintains the same covariance structure that has been present in the past.

The discussion of this approach has shown that the success of any databased approach to the optimization problem is closely connected to the nature of the manipulatable variables \((MV_i)\) that one plans to employ in the optimization. Three different cases can be identified:

**Case 1: The MV\(_i\) are independent of disturbances and uncorrelated among themselves.** This situation is similar to the one when the database has been generated through experimental design. By regressing \(Y = [X0_k q]\) onto \(X = [MV_i]\) the causal gains between manipulated and process and quality variables can be obtained. When such regression models are calculated locally in different regions of the database, one can use response surface methods and move the \(MV_i\) in the directions of steepest ascent. In order to identify the proper region of the database that the current situation corresponds to, one may want to reconstruct the disturbances as discussed in section 5.2.1 and choose the
nearest neighbors as a basis for the local model.

**Case 2:** The $MV_i$ are independent of disturbances but correlated among themselves. Due to the correlation among the $MV_i$ it is difficult to obtain good estimates of the individual causal gains between $MV_i$ and process and quality variables. However, one can extract a causal gain between the subspace spanned by the $MV_i$, and the process and quality variables. This can be done by calculating a PCR or PLS regression model where the latent variables define this $MV_i$ subspace. One can either identify directions within this subspace in which to move the $MV_i$ and use the relationship between latent variables and $MV_i$ to obtain the settings for the manipulated variables (similar to what has been described in chapter 2), or one can again search for nearest neighbors on the basis of reconstructed disturbance characteristics, and implement new $MV_i$ settings based on those of the nearest neighbors.

**Case 3:** The $MV_i$ are correlated with disturbances: $MV_i = MV_i^* - k \cdot D$.

It is easy to see that for $k \to 0$, $MV_i \to MV_i^*$ which is the situation in cases 1 and 2. On the other hand, if $MV_i^* \to 0$, the manipulated variables are moved in pure feedback fashion. Then, there is no open-loop information in the data, and the $MV_i$ should not be adjusted in any other way than to compensate for the disturbances as before ($MV_i = -k \cdot D$). The most likely case is the intermediate one where $k \neq 0$ and $MV_i^* \neq 0$. As discussed in section 5.2.2, it is not possible in this case to separate the disturbances from the open-loop information. We believe that in these cases no process optimization can be done relying only on the available data.

On this account, the only useful databases for process improvement or optimization appear to be the trivial cases 1 and 2. However, one has to be very lucky to encounter such a situation. We expect real databases typically to be of type 3.
In most cases one is just not sure which manipulated variables - if any - contain open-loop information, and whether the operator has moved them in relation to disturbances or not. Therefore, one cannot draw much information from the database. At this point, any attempt at using such typical process data for process optimization seems futile.

Although this result is somewhat disappointing, it might have been expected from Fisher's statement (see section 5.1), that not much can be done with such "happenstance data". However, one should limit the extent of this statement to the area of process optimization and control that basically require cause and effect information. In other areas such as process monitoring and diagnosis, methods have been developed which make powerful use of process operating databases. An overview over multivariate statistical methods in these areas is given in (Kourti and MacGregor 1995). The key there is that these applications do not need causal information. The interest is merely in the common cause variation, which is indeed present in process operating data.
Chapter 6

Summary and Conclusions

Large amounts of process data are readily available for many industrial processes. Some potential uses for these data for problems in process engineering have been investigated in this thesis. Although databased approaches are restricted by the range and the structure within the available data, they can still be very helpful in extracting useful information and providing insights before more involved studies are applied or experiments are designed. The advantages of databased approaches lie in their ability to produce first results relatively fast and at low cost.

The underlying theme of this thesis has been the use of historical process data in connection with latent variable regression tools to address problems of product quality design and process operating improvement.

In chapter 2 a methodology has been developed based on latent variable techniques that uses historical process data on available product grades to determine process conditions which achieve a desired new grade quality. A requirement is that this new grade quality $y_{des}^T$ lies within the range of the grades that already exist. A method that tests the feasibility of the production of this new grade has been proposed. In finding new process conditions that achieve $y_{des}^T$, it is of great importance that these conditions be not only feasible but also consistent with the conditions
used in the past, and therefore retain the same covariance structure as present in
the data set. This is the key point which makes latent variable methods so useful in
this context. They model not only the relationship between operating conditions and
quality specifications but also the covariance structure within the process conditions
($X$-space). An inversion of such a model has been proposed which leads to conditions
that maintain the covariance structure of the past and are in some way consistent
with the operating strategies of other grades. It has been shown that standard re-
gression models and their inversion do not respect the covariance structure of the
data and cannot provide this consistency. The different approaches discussed have
been summarized in a unified framework that highlights the underlying mathematical
concepts in the proposed techniques.

Since the operating space $X$ is often of higher dimension than the space
spanned by quality variables $Y$, the inversion of the latent variable models in these
cases results in not just a single set of process conditions but in a whole window of
conditions that offers a choice to the operator and plant engineer. Implementation of
these conditions may not lead to the desired quality in one step. Iterations between
experiments and redesigning the conditions may be necessary.

An extension of the methodology to nonlinear model inversion has been pre-
sented. However, the risks involved in employing these nonlinear techniques in this
context have been discussed in some detail at the end of chapter 2.

The method developed in chapter 2 has been applied to three different indus-
trial processes in chapter 3. Since it was not possible in any of the applications to
implement new process conditions in the plant to verify that the predicted conditions
would indeed lead to the desired product quality, the approach had to be tested on the
available grades. In each of the examples one or more of the given grades were taken
out of the databases and treated as a new desired grade for which the corresponding
process conditions had to be found. The difficulties encountered in these examples
were mainly in regard to quality and quantity of the data involved. However, in spite of these problems the method exhibited promising results.

Using some of the concepts developed in chapter 2 the related problem of transferring production of a grade from a plant A to another plant B was treated in chapter 4. The discussion of the data structure in this problem addresses how process conditions and quality data of the two plants relate to each other. In particular it has been shown that it is important to compare the quality spaces of the two plants. A test has been suggested which - based on the available data - indicates which of the plant A grades seem feasible for production in plant B. Only grades that are consistent with the covariance structure of the already existing plant B grades should be included in the approach.

A method has been outlined which uses all the data on plant B and the quality data of those plant A grades that fall in the same space as plant B grades. An extended PCR-model is built and then inverted to predict process conditions in plant B for the grade so far only produced in plant A.

As in the chapter on product design, the approach relies again on a model that describes not only the relationship between process conditions and quality measurements but also the structure of the process conditions. This allows for construction of new process conditions that are physically feasible and follow the operating strategies of the past grades. If the space spanned by the process conditions is higher dimensional than the space spanned by the quality measurements the solution again comprises a whole region of operating conditions.

Future work in this area might investigate how such a methodology could be of assistance in scaling up laboratory experiments from a pilot plant to a full scale process. Another issue related to the concept presented in chapter 4 is the problem where product quality produced by two or more plants has to be matched.

The last chapter differs slightly from the contents of the rest of the thesis.
CHAPTER 6. SUMMARY AND CONCLUSIONS

It has been motivated by a number of published attempts trying to exploit the potentials of large historical operating databases to gain causal information for process performance optimization. Since the underlying difficulties and issues involved in this problem are typically not recognized, it seemed appropriate to take a closer look at this problem. However, the scope of this investigation was more of an exploratory nature aimed at shedding some light onto the issues and possibly providing some guidance for any future research in this area.

The major difficulties with this problem of databased process optimization are the extremely correlated nature of typical process operating data, the presence of unmeasured disturbances that affect most variables, and the effect of feedback control arising from automatic controllers or from operator actions. An approach referred to as "Similarity Optimization" has been suggested. It uses latent variable models (PCA and PLS) to extract information on the disturbances, and nearest neighbor methods to find similar conditions that can potentially lead to improved performance. Initially it was hoped that by only looking at past situations with similar disturbance occurrences instead of trying to estimate causal coefficients between variables, the problems mentioned above could be overcome. However, it was recognized that the approach is prone to fail under the same conditions as the investigated published approaches would. The main conclusion drawn from this study states that a databased approach to process improvement or optimization is feasible only in the rare case when there exist manipulated variables that are known to be independent of the disturbances. In section 5.2.2 it was discussed, how correlation between manipulated variables and disturbances compounds the causal relationships among the variables with disturbance variation. This seems to render any attempts at extracting useful information for process optimization futile. Future work in this area will probably have to look more deeply into this issue about correlation between manipulated variables and disturbances. For now we have to agree with Fisher that not much can be done with
real operating data in the area of process optimization.

Throughout the course of this thesis latent variable methods have proven to be very useful for problems dealing with historical operating data. This is due to the highly correlated nature of such data and the ability of latent variable methods to use this correlation to reduce the dimensionality of the problem. By working with the few underlying independent latent variables instead of the many collinear process variables it becomes possible to respect and maintain the covariance structure that is present in the historical data. Preservation of this structure when predicting new process conditions - for product design, quality improvement or any other purpose - is extremely important in order to obtain reliable results. It ensures that the new conditions are still within the region of validity of the model obtained from the data. The methods suggested in this thesis are limited - as are all databased approaches - to the range and the structure of the available data. However, this work shows that within these limitations much can be gained and learned from historical process data - depending of course on the quality and the information content of the data. Even if the databased methods do not lead to perfect results in one step, they can be used to lessen the necessary efforts and costs involved in designed experiments.

There is still a lot to be done in this field. In particular the area of process optimization or improvement using databases needs to be explored further. Extensions to the use of dynamic data should be investigated. Applications to batch processes would probably benefit from more complex methodologies that deal with the full trajectories rather than breaking the trajectories up into a few descriptive variables as has been done in this thesis. It is hoped that the ideas presented here on these topics will lead to further work in the area.
Appendix A

Construction of the “Null Component” $x_{null}^T$ for PCR-Inversion

The calculation of the null component is illustrated here for the inversion of a PCR model. However, the concepts are the same for the null component of any other latent variable method that models both the relationship between $X$ and $Y$, as well as the $X$-space. For a different model, in particular equations (A.2) and (A.3) would have to be adapted with the specific parameters used in that model.

In order for $x_{pred}^T = \hat{x}_{new}^T + x_{null}^T$ to maintain the covariance structure of the historical $X$-matrix, one has to resort again to the scaled latent variables ($u^T$) of the $X$-space:

$$x_{pred}^T = (\hat{u}_{new}^T + u_{null}^T) \cdot \Sigma_A \cdot V_A^T .$$  \hspace{1cm} (A.1)

It is the component $u_{null}^T$ we are looking for. As mentioned previously in section 2.2.3, $x_{null}^T$ - and therefore $u_{null}^T$ - has to come from the $(A - k)$ dimensional subspace in $X$ that is not spanned by $M_{PCR}^T$. At the same time, adding this component $x_{null}^T = u_{null}^T \cdot \Sigma_A \cdot V_A^T$ to $x_{new}^T = \hat{u}_{new}^T \cdot \Sigma_A \cdot V_A^T$ should not alter the values of $y_{des}^T$.
which are determined by $\hat{x}_{\text{new}}^T$. Since $u_{\text{new}}^T \cdot \Sigma_A \cdot B = \gamma_{\text{dest}}^T$, this requires that

$$u_{\text{null}}^T \cdot \Sigma_A \cdot B = 0$$

which with $B = \Sigma_A^{-1} \cdot U_A^T \cdot Y$ is equivalent to

$$u_{\text{null}}^T \cdot U_A^T \cdot Y = 0$$

$$(1 \times A) \cdot (A \times n) \cdot (n \times k)$$

(A.2)

Any vector $u_{\text{null}}^T$ lying in the left nullspace of $U_A^T \cdot Y$ is a solution to this equation (A.2). Singular value decomposition of $U_A^T \cdot Y$ reveals this nullspace:

$$U_A^T \cdot Y = [G_1 : G_2] \cdot \Sigma_{U_A Y} \cdot V_{U_A Y}^T$$

$$(A \times k) \cdot (A \times (A-k)) \cdot (A \times k) \cdot (k \times k)$$

(A.3)

$G_2^T$ defines the space within which $u_{\text{null}}^T$ can lie:

$$u_{\text{null}}^T = \lambda^T \cdot G_2^T$$

$$(1 \times A) \cdot ((A-k) \times A)$$

Mathematically $\lambda^T$ is arbitrary in magnitude and direction. However,

$$x_{\text{pred}}^T = (u_{\text{new}}^T + \lambda^T \cdot G_2^T) \cdot \Sigma_A \cdot V_A^T$$

must lie within the range of historical $X$ data. As mentioned in section 2.2.3, different $x_{\text{pred}}^T$ have to be calculated for different $\lambda^T$-vectors, and only those $x_{\text{pred}}^T$ should be retained that have $t_{\text{pred}}^T$-scores within the score cluster of historical process conditions.
Appendix B

Data for the LDPE Example

Table B.1: Manipulated Variable Values (X) for the 15 LDPE Grades

<table>
<thead>
<tr>
<th>Grade</th>
<th>F1</th>
<th>F2</th>
<th>Cs1</th>
<th>Cs2</th>
<th>P</th>
<th>Tin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.74</td>
<td>0.75</td>
<td>0.022</td>
<td>0.021</td>
<td>2378</td>
<td>438</td>
</tr>
<tr>
<td>2</td>
<td>0.41</td>
<td>0.41</td>
<td>0.016</td>
<td>0.015</td>
<td>2290</td>
<td>460</td>
</tr>
<tr>
<td>3</td>
<td>0.44</td>
<td>0.44</td>
<td>0.069</td>
<td>0.071</td>
<td>2923</td>
<td>454</td>
</tr>
<tr>
<td>4</td>
<td>0.32</td>
<td>0.31</td>
<td>0.079</td>
<td>0.079</td>
<td>2861</td>
<td>478</td>
</tr>
<tr>
<td>5</td>
<td>0.23</td>
<td>0.27</td>
<td>0.036</td>
<td>0.037</td>
<td>3501</td>
<td>473</td>
</tr>
<tr>
<td>6</td>
<td>0.74</td>
<td>0.76</td>
<td>0.094</td>
<td>0.092</td>
<td>2201</td>
<td>436</td>
</tr>
<tr>
<td>7</td>
<td>0.90</td>
<td>0.87</td>
<td>0.082</td>
<td>0.083</td>
<td>2068</td>
<td>421</td>
</tr>
<tr>
<td>8</td>
<td>0.41</td>
<td>0.42</td>
<td>0.114</td>
<td>0.113</td>
<td>2905</td>
<td>472</td>
</tr>
<tr>
<td>9</td>
<td>0.17</td>
<td>0.14</td>
<td>0.055</td>
<td>0.055</td>
<td>3516</td>
<td>492</td>
</tr>
<tr>
<td>10</td>
<td>0.33</td>
<td>0.35</td>
<td>0.094</td>
<td>0.094</td>
<td>3059</td>
<td>475</td>
</tr>
<tr>
<td>11</td>
<td>0.59</td>
<td>0.62</td>
<td>0.099</td>
<td>0.100</td>
<td>2506</td>
<td>452</td>
</tr>
<tr>
<td>12</td>
<td>0.48</td>
<td>0.48</td>
<td>0.070</td>
<td>0.071</td>
<td>2528</td>
<td>466</td>
</tr>
<tr>
<td>13</td>
<td>0.49</td>
<td>0.52</td>
<td>0.046</td>
<td>0.047</td>
<td>2710</td>
<td>469</td>
</tr>
<tr>
<td>14</td>
<td>0.25</td>
<td>0.23</td>
<td>0.035</td>
<td>0.034</td>
<td>3346</td>
<td>477</td>
</tr>
<tr>
<td>15</td>
<td>0.31</td>
<td>0.34</td>
<td>0.037</td>
<td>0.036</td>
<td>3220</td>
<td>483</td>
</tr>
</tbody>
</table>
Table B.2: Quality Variable Values (Y) for the 15 LDPE Grades

<table>
<thead>
<tr>
<th>Grade</th>
<th>Conv</th>
<th>MWn</th>
<th>MWw</th>
<th>LCB</th>
<th>SCB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1390</td>
<td>32128</td>
<td>420875</td>
<td>1.077</td>
<td>23.9</td>
</tr>
<tr>
<td>2</td>
<td>0.1069</td>
<td>35838</td>
<td>283415</td>
<td>0.766</td>
<td>23.0</td>
</tr>
<tr>
<td>3</td>
<td>0.1364</td>
<td>27651</td>
<td>197719</td>
<td>0.848</td>
<td>25.5</td>
</tr>
<tr>
<td>4</td>
<td>0.1119</td>
<td>27215</td>
<td>142446</td>
<td>0.671</td>
<td>25.3</td>
</tr>
<tr>
<td>5</td>
<td>0.1396</td>
<td>31348</td>
<td>184145</td>
<td>0.641</td>
<td>26.7</td>
</tr>
<tr>
<td>6</td>
<td>0.1264</td>
<td>24507</td>
<td>222908</td>
<td>1.097</td>
<td>24.1</td>
</tr>
<tr>
<td>7</td>
<td>0.1337</td>
<td>25664</td>
<td>322753</td>
<td>1.229</td>
<td>23.5</td>
</tr>
<tr>
<td>8</td>
<td>0.1220</td>
<td>23180</td>
<td>124943</td>
<td>0.784</td>
<td>26.3</td>
</tr>
<tr>
<td>9</td>
<td>0.1143</td>
<td>30561</td>
<td>134897</td>
<td>0.479</td>
<td>26.3</td>
</tr>
<tr>
<td>10</td>
<td>0.1232</td>
<td>25298</td>
<td>134874</td>
<td>0.717</td>
<td>26.2</td>
</tr>
<tr>
<td>11</td>
<td>0.1257</td>
<td>24032</td>
<td>171741</td>
<td>0.963</td>
<td>25.1</td>
</tr>
<tr>
<td>12</td>
<td>0.1157</td>
<td>27186</td>
<td>175084</td>
<td>0.824</td>
<td>24.7</td>
</tr>
<tr>
<td>13</td>
<td>0.1264</td>
<td>29070</td>
<td>203804</td>
<td>0.843</td>
<td>25.2</td>
</tr>
<tr>
<td>14</td>
<td>0.1289</td>
<td>32094</td>
<td>180039</td>
<td>0.610</td>
<td>26.0</td>
</tr>
<tr>
<td>15</td>
<td>0.1309</td>
<td>30472</td>
<td>177312</td>
<td>0.685</td>
<td>26.3</td>
</tr>
</tbody>
</table>
References


REFERENCES


REFERENCES


