Statistical batch process monitoring
Ramaker, H.J.; van Sprang, E.N.M.

Citation for published version (APA):
Statistical Batch Process Monitoring

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. mr. P.F. van der Heijden
ten overstaan van een door het college voor promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit

op dinsdag 6 april 2004, te 10.00 uur
door Hendrik-Jan Ramaker
geboren te Den Burg-Texel

op dinsdag 6 april 2004, te 11.00 uur
door Eric Nicolaas Maurice van Sprang
geboren te Gouda
Promotor: prof. dr. A.K. Smilde

Co-promotor: dr. J.A. Westerhuis

Overige leden: prof. dr. R. Bro
prof. dr. P.J. Gemperline
prof. dr. ir. B. Roffel
prof. dr. R.J.M.M. Does
prof. dr. P.D. Iedema

Faculteit der Natuurwetenschappen, Wiskunde & Informatica

All rights reserved.
Copyright © 2004 by E.N.M. van Sprang and H-J. Ramaker

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, without permission in writing from the authors.

The work described in this thesis was carried out at the Department of Chemical Engineering, Process Analysis and Chemometrics, University of Amsterdam, The Netherlands. This work was financially supported by the council for chemical sciences of the Netherlands Organization for Scientific Research (NWO-CW) and the Netherlands Technology Foundation (STW).
CONTENTS

CHAPTER 1  ♦ BATCH PROCESS MONITORING .................................................. 5
  1.1 General introduction ............................................................................ 6
  1.2 Batch reactors .................................................................................... 6
    1.2.1 Different application areas of batch processing ......................... 7
  1.3 Features of a batch reactor ................................................................. 7
    1.3.1 Advantages of batch processing .................................................. 8
    1.3.2 On-line measurements of batch data ............................................ 8
      1.3.2.1 Engineering variables ......................................................... 9
      1.3.2.2 Spectroscopic data ........................................................... 12
  1.4 Why batch process monitoring? ........................................................... 13
    1.4.1 Safety ......................................................................................... 13
    1.4.2 Process disturbances ................................................................. 13
    1.4.3 Process understanding ............................................................... 14
    1.4.4 Process and product consistency ................................................. 14
    1.4.5 Capability index ......................................................................... 14
  1.5 Statistical process control ................................................................. 15
    1.5.1 Ideas of statistical process control ............................................... 15
    1.5.2 Difference between process monitoring and process control ........ 17
    1.5.3 Methods for batch process monitoring ......................................... 17
      1.5.3.1 Fundamental models .......................................................... 18
      1.5.3.2 Knowledge based approach ............................................... 18
      1.5.3.3 SPC .................................................................................. 19
        1.5.3.3.1 Univariate charts: \( \bar{X} \), CUSUM, EWMA ..................... 19
          1.5.3.3.1.1 \( \bar{X} \) Chart ......................................................... 19
          1.5.3.3.1.2 CUSUM chart ..................................................... 22
          1.5.3.3.1.3 EWMA chart ....................................................... 23
          1.5.3.3.1.4 Average run length ............................................... 24
        1.5.3.3.2 Multivariate control charts ............................................ 25
          1.5.3.3.2.1 Continuous processes without dimension reduction .... 25
          1.5.3.3.3 Megavariate control charts ......................................... 26
            1.5.3.3.3.1 Continuous processes with dimension reduction .... 26
            1.5.3.3.4 Statistical batch process monitoring ......................... 28
        1.5.3.4 Multivariate charts versus univariate charts ....................... 29
          1.5.4.1 Abundance of control charts .......................................... 29
            1.5.4.1.1 False alarm rate .................................................... 29
            1.5.4.1.1.1 Multivariate nature ........................................... 30
    1.6 Nowadays monitoring of batch processes ........................................ 32
      1.6.1 Literature & Software ............................................................... 33
<table>
<thead>
<tr>
<th>CHAPTER 2</th>
<th>I.T.A.-TRAJECTORY ........................................</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>General introduction .......................................</td>
<td>36</td>
</tr>
<tr>
<td>2.2</td>
<td>I: Initial phase ...........................................</td>
<td>37</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Exploratory analysis ......................................</td>
<td>38</td>
</tr>
<tr>
<td>2.3</td>
<td>T: Training phase ..........................................</td>
<td>38</td>
</tr>
<tr>
<td>2.3.1</td>
<td>NOC selection or outlier detection .......................</td>
<td>38</td>
</tr>
<tr>
<td>2.3.1.1</td>
<td>Post-batch analysis .......................................</td>
<td>39</td>
</tr>
<tr>
<td>2.3.1.2</td>
<td>Model building: evolving models ..........................</td>
<td>42</td>
</tr>
<tr>
<td>2.3.1.3</td>
<td>D and SPE-statistic .......................................</td>
<td>43</td>
</tr>
<tr>
<td>2.4</td>
<td>A: Application phase .......................................</td>
<td>44</td>
</tr>
<tr>
<td>2.4.1</td>
<td>On-line monitoring .........................................</td>
<td>45</td>
</tr>
<tr>
<td>2.4.1.1</td>
<td>Fault detection and Fault Diagnosis .....................</td>
<td>45</td>
</tr>
<tr>
<td>2.4.1.1.1</td>
<td>Complementarity of SPE and D-chart .....................</td>
<td>46</td>
</tr>
<tr>
<td>2.4.1.1.2</td>
<td>Contribution plots .......................................</td>
<td>48</td>
</tr>
<tr>
<td>2.4.1.1.3</td>
<td>Example: monitoring a faulty batch with impurity of the feed</td>
<td>49</td>
</tr>
<tr>
<td>2.4.1.2</td>
<td>Process improvement .......................................</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 3</th>
<th>PREPROCESSING .............................................</th>
<th>51</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>General introduction ......................................</td>
<td>52</td>
</tr>
<tr>
<td>3.2</td>
<td>Dynamic time warping of spectroscopic batch data ......</td>
<td>53</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Summary ....................................................</td>
<td>53</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Introduction ...............................................</td>
<td>53</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Theory ......................................................</td>
<td>55</td>
</tr>
<tr>
<td>3.2.3.1</td>
<td>General explanation of DTW ..............................</td>
<td>55</td>
</tr>
<tr>
<td>3.2.3.1.1</td>
<td>The Sample and the Reference ................................</td>
<td>56</td>
</tr>
<tr>
<td>3.2.3.1.2</td>
<td>The grid ....................................................</td>
<td>56</td>
</tr>
<tr>
<td>3.2.3.2</td>
<td>Local constraints ..........................................</td>
<td>57</td>
</tr>
<tr>
<td>3.2.4</td>
<td>DTW algorithm ...............................................</td>
<td>58</td>
</tr>
<tr>
<td>3.2.4.1</td>
<td>Remaining topics ...........................................</td>
<td>64</td>
</tr>
<tr>
<td>3.2.4.1.1</td>
<td>Global constraints ........................................</td>
<td>64</td>
</tr>
<tr>
<td>3.2.4.1.2</td>
<td>Normalization ..............................................</td>
<td>65</td>
</tr>
<tr>
<td>3.2.4.1.3</td>
<td>Normalization for synchronization of batch runs ........</td>
<td>67</td>
</tr>
<tr>
<td>3.2.4.1.4</td>
<td>Direction penalties .......................................</td>
<td>67</td>
</tr>
<tr>
<td>3.2.4.1.5</td>
<td>Multivariate trajectories ..................................</td>
<td>68</td>
</tr>
<tr>
<td>3.2.4.1.6</td>
<td>Calculation of the weight matrix by Kassidas ..........</td>
<td>68</td>
</tr>
<tr>
<td>3.2.4.1.7</td>
<td>New definition of the weights ............................</td>
<td>69</td>
</tr>
<tr>
<td>3.2.5</td>
<td>DTW for batch monitoring ...................................</td>
<td>73</td>
</tr>
<tr>
<td>3.2.6</td>
<td>Industrial example: batch process of urethane resin ....</td>
<td>75</td>
</tr>
<tr>
<td>3.2.7</td>
<td>DTW for spectroscopic data ..................................</td>
<td>77</td>
</tr>
<tr>
<td>3.2.7.1</td>
<td>Weight matrix ...............................................</td>
<td>77</td>
</tr>
<tr>
<td>3.2.7.2</td>
<td>Normalizing ..................................................</td>
<td>81</td>
</tr>
<tr>
<td>3.2.8</td>
<td>Conclusions ..................................................</td>
<td>82</td>
</tr>
</tbody>
</table>
CHAPTER 4  ♦ BLACK MODELS ................................................................. 83
4.1 General introduction ............................................................................. 84
4.2 Critical evaluation of approaches for on-line batch process monitoring 85
   4.2.1 Summary .................................................................................. 85
   4.2.2 Introduction .............................................................................. 85
   4.2.3 Theory ...................................................................................... 87
      4.2.3.1 Modelling of normal operating conditions (NOC) ...................... 87
         4.2.3.1.1 Nomikos & MacGregor (NM) ........................................... 88
      4.2.3.1.2 Wold, Kettaneh, Fridén & Holmberg (WKFH) .................... 89
         4.2.3.1.3 Rännar, MacGregor & Wold (RMW) ............................. 89
   4.2.3.2 On-line monitoring .................................................................. 91
      4.2.3.2.1 Nomikos & MacGregor (NM) approach ............................... 91
      4.2.3.2.2 Zero deviations approach ............................................... 92
      4.2.3.2.3 Current deviations approach ........................................... 93
      4.2.3.2.4 Missing data approach ................................................... 93
      4.2.3.2.5 Rännar, MacGregor & Wold (RMW) approach .................. 94
      4.2.3.2.6 Wold, Kettaneh, Fridén & Holmberg (WKFH) approach ...... 95
      4.2.3.3 Deriving the control limits .................................................... 95
         4.2.3.3.1 D-statistic .................................................................. 95
         4.2.3.3.2 Squared Prediction Error (SPE) ...................................... 96
   4.2.3.4 Summary ............................................................................ 97
   4.2.4 Performance indices ................................................................... 97
      4.2.4.1 Overall type I error (false warning) ...................................... 97
      4.2.4.2 Action Signal Time (AST) and type II error ......................... 98
      4.2.4.3 Summary ........................................................................ 99
   4.2.5 Description of the data ................................................................ 100
      4.2.5.1 Data set 1 ....................................................................... 100
      4.2.5.2 Data set 2 ....................................................................... 101
      4.2.5.3 Data set 3 ....................................................................... 101
      4.2.5.4 Data set 4 ....................................................................... 101
      4.2.5.5 Data set 5 ....................................................................... 102
   4.2.6 Results and discussion ................................................................ 102
      4.2.6.1 AST ............................................................................. 102
      4.2.6.2 Overall type I error ......................................................... 104
   4.2.7 Conclusions and recommendations ............................................ 107
4.3 Fault detection properties of global, local and time evolving models for
batch process monitoring ....................................................................... 108
   4.3.1 Summary ............................................................................. 108
   4.3.2 Introduction ........................................................................... 108
   4.3.3 Statistical batch process monitoring ......................................... 109
      4.3.3.1 Modelling .................................................................... 110
         4.3.3.1.1 Global model .............................................................. 110
         4.3.3.1.2 Time evolving model ............................................... 113
5.2.8.1 Grey model example................................................................. 152
5.2.9 Conclusions.................................................................................. 153

5.3 NIR Spectroscopic monitoring of a series of industrial batch processes using a bilinear grey model........................................ 154
5.3.1 Summary....................................................................................... 154
5.3.2 Introduction................................................................................... 154
5.3.3 Theory.......................................................................................... 156
  5.3.3.1 Structure of the in-line spectroscopic batch wise NIR data..... 156
  5.3.3.2 Range charts and change charts ......................................... 157
  5.3.4 Black model: Principal component analysis (PCA) ............... 158
  5.3.5 Grey model............................................................................... 158
  5.3.6 Experimental............................................................................. 159
    5.3.6.1 Process description............................................................. 159
    5.3.6.2 Instrumentation................................................................. 160
    5.3.6.3 Band assignment and band selection................................. 160
    5.3.6.4 Sampling........................................................................... 161
  5.3.7 Results and discussion.............................................................. 162
    5.3.7.1 Construction of the grey model............................................ 162
    5.3.7.2 Incorporation of external white information into the model... 163
    5.3.7.3 Modelling a single batch run.............................................. 164
    5.3.7.4 Mathematical model robustness.......................................... 165
    5.3.7.5 Modelling multiple batch runs............................................ 167
    5.3.7.6 Estimating the unknown alcohol spectrum.......................... 168
  5.3.8 Qualitative monitoring performance ...................................... 169
  5.3.9 Conclusion................................................................................. 175

5.4 Statistical batch process monitoring using grey models............. 176
5.4.1 Summary....................................................................................... 176
5.4.2 Introduction................................................................................... 176
5.4.3 Theory.......................................................................................... 179
  5.4.3.1 Notation................................................................................. 179
  5.4.3.2 Structure of the data............................................................. 179
  5.4.3.3 Black models....................................................................... 179
  5.4.3.4 Grey models........................................................................ 182
    5.4.3.4.1 Creating grey models..................................................... 182
  5.4.3.5 Batch process monitoring.................................................... 184
    5.4.3.5.1 On-line monitoring of a new batch using the grey Tucker1 model ........................................................................... 184
    5.4.3.5.2 On-line monitoring of a new batch using the grey Tucker3 model ........................................................................... 185
  5.4.3.6 Statistics.................................................................................. 185
    5.4.3.6.1 D-statistic and SPE-statistic............................................. 185
    5.4.3.6.2 Univariate statistics....................................................... 187
  5.4.3.7 Monitoring levels................................................................. 187
5.4.3.8 Post-batch process monitoring ........................................ 188
5.4.3.9 Batch consistency number ........................................... 189
5.4.4 Experimental ......................................................... 190
  5.4.4.1 Process description ............................................... 190
  5.4.4.2 Instrumentation .................................................. 191
  5.4.4.3 Data pre-treatment .............................................. 191
  5.4.4.4 Band assignment and band selection ................................ 191
  5.4.4.5 Mean Centering ................................................ 192
5.4.5 Results and discussion ............................................... 193
  5.4.5.1 Construction of the models ..................................... 193
5.4.5.2 On-line monitoring .............................................. 194
  5.4.5.2.1 First monitoring level with the grey Tucker1 model .......... 195
  5.4.5.2.2 First monitoring level with the grey Tucker3 model .......... 196
  5.4.5.2.3 Second monitoring level grey Tucker1 ................................ 197
  5.4.5.2.4 Second monitoring level grey Tucker3 ................................ 199
5.4.5.3 Results for the BCN ............................................. 201
5.4.6 Conclusions ............................................................ 203

CHAPTER 6   ◆ STATISTICS AND MECHANICS ........................................ 205
6.1 General introduction ....................................................... 206
6.2 Performance assessment and improvement of control charts for statistical batch process monitoring ........................................ 208
  6.2.1 Summary .................................................................. 208
  6.2.2 Introduction ........................................................... 208
  6.2.3 Multivariate statistical process control: the concepts .............. 210
  6.2.4 Multivariate statistical process control: the control charts ........ 210
    6.2.4.1 Continuous processes without dimension reduction .......... 211
    6.2.4.2 Continuous processes with dimension reduction ................ 212
  6.2.5 Mechanics of the $D$ and $SPE$-chart ................................ 214
    6.2.5.1 Post-batch analysis .............................................. 221
  6.2.6 Statistical performance of control charts for batch process monitoring ........................................ 225
    6.2.6.1 On-line batch monitoring ....................................... 225
    6.2.6.2 Further comments on-line statistical batch monitoring ........ 229
  6.2.7 Description of the data ............................................. 229
6.2.8 Results ....................................................................... 232
  6.2.8.1 Performance indices ............................................... 232
    6.2.8.1.1 The overall type I error ........................................ 232
    6.2.8.1.2 Action Signal Time ............................................ 233
  6.2.8.2 Accounting for the difference between NOC and new batches .... 233
  6.2.8.3 Hypothesis testing ................................................ 234
  6.2.8.4 The Bonferroni adjustment ....................................... 235
  6.2.8.5 Further comments on the $D$-statistic ............................ 236
    6.2.8.5.1 Time dependent covariance matrix .......................... 237
6.2.8.5.2 Non-normality ......................................................... 238
6.2.8.5.3 Violation of the $\Sigma_f = \Sigma$ assumption .................. 239
6.2.8.5.4 Degrees of freedom in process monitoring .................. 239
6.2.8.6 Further comments on the squared prediction error (SPE) .... 240
6.2.8.6.1 Non-normality ....................................................... 240
6.2.8.7 Results adjustments to the control charts ..................... 241
6.2.9 Conclusions ............................................................. 243

6.3 The effect of the size of the training set and number of components on the false alarm rate in statistical process monitoring ...................... 245
6.3.1 Summary ............................................................... 245
6.3.2 Introduction ........................................................... 245
6.3.3 Simulation of a continuous process ................................ 246
6.3.4 Multivariate statistical process control .......................... 247
6.3.4.1 Size of the NOC data ............................................. 249
6.3.4.2 Number of principal components ............................... 249
6.3.5 False alarm rate results ............................................. 249
6.3.5.1 Discussion about the false alarm rate results ................ 251
6.3.6 5. Leave-one-out method ........................................... 254
6.3.6.1 Results for the leave-one-out method .......................... 254
6.3.7 Conclusions ............................................................ 256

6.4 Fault manifestation in the $D$ and SPE-chart for different batch monitoring strategies .................................................. 258
6.4.1 Summary ............................................................... 258
6.4.2 Introduction ........................................................... 258
6.4.3 A qualitative description of the $D$ and SPE-statistic ............ 260
6.4.3.1 Mechanics of the SPE and D-statistic .......................... 260
6.4.3.2 Relationship between real process disturbances and the SPE and D-statistic ............................................. 262
6.4.4 Construction of pure $D$-faults .................................... 264
6.4.4.1 General construction of a pure $D$-fault ....................... 267
6.4.4.2 Construction of a pure $D$-fault for the local model ........ 268
6.4.4.3 Construction of a pure $D$-fault for the evolving model .... 270
6.4.4.4 Construction of a pure $D$-fault for the global model ...... 272
6.4.5 Construction of pure SPE-faults ................................... 274
6.4.5.1 General construction of a pure SPE-fault ..................... 276
6.4.5.2 Construction of a pure SPE-fault for the local model ....... 277
6.4.5.3 Construction of a pure SPE-fault for the evolving model ... 277
6.4.5.4 Construction of a pure SPE-fault for the global model ...... 279
6.4.6 Conclusions from SPE and $D$-mechanics ........................ 281
6.4.7 The problem of the embedded error ............................... 282
6.4.8 Case study: Temperature control of a PVC batch reactor ...... 284
6.4.9 Process faults for batch processes ................................ 286
6.4.9.1 Breakage of the correlation: sensor faults .................. 287
6.4.9.1.1 Construction .................................................................................................................. 288
6.4.9.1.2 Results .......................................................................................................................... 290
6.4.9.2 Instantaneous intensification of the correlation .............................................................. 290
6.4.9.2.1 Construction ............................................................................................................... 290
6.4.9.2.2 Results ....................................................................................................................... 292
6.4.9.3 Realistic intensification of the instantaneous correlation between the process variables .......................................................... 292
6.4.9.3.1 Construction ............................................................................................................... 293
6.4.9.3.2 Results ....................................................................................................................... 295
6.4.9.4 Small change breaking the correlation over time ........................................................... 296
6.4.9.4.1 Construction ............................................................................................................... 296
6.4.9.4.2 Results ....................................................................................................................... 297
6.4.10 Conclusions ...................................................................................................................... 297
6.4.10.1.1 On-line monitoring ................................................................................................. 297
6.4.10.1.2 Post batch analysis ................................................................................................. 297

6.5 Correlations in batch process data ...................................................................................... 299
6.5.1 Summary ............................................................................................................................ 299
6.5.2 Introduction ....................................................................................................................... 299
6.5.3 Theory ................................................................................................................................ 300
6.5.3.1 Structure of the data ..................................................................................................... 300
6.5.4 Correlation in batch process data .................................................................................... 301
6.5.5 Matricizing and modelling the three-way array ............................................................... 302
6.5.6 Case study .......................................................................................................................... 306
6.5.6.1 Description of the data ................................................................................................. 306
6.5.6.2 Preprocessing ............................................................................................................... 306
6.5.6.3 Results ......................................................................................................................... 306
6.5.7 Conclusions ....................................................................................................................... 311

♦ FUTURE DIRECTIONS .................................................................................................................. 312
♦ SYMBOLS AND NOTATION .................................................................................................... 316
♦ APPENDICES ............................................................................................................................ 319
♦ SAMENVATTING VOOR NIET INGEWIJDEN ......................................................................... 328
♦ REFERENCES ............................................................................................................................. 332
♦ PUBLICATIONS ......................................................................................................................... 343
♦ RESEARCH ACTIVITIES ............................................................................................................. 345
♦ DANKWOORD ........................................................................................................................... 346
◊ PREFACE

The content of this thesis exists of an introduction to statistical batch process monitoring and a collection of modified articles, which are partly published or submitted to various scientific journals. Each chapter of this thesis can be regarded as a self-containing story and the reader has the choice to omit chapters without a loss of understanding. Due to the structure of this thesis there might be some overlap in content.

The initials HJR (Henk-Jan Ramaker) or EVS (Eric van Sprang) indicate the responsible author(s) of a section or chapter. If the initials (HJR/EVS) are given, both authors are equally responsible. The initials can be found as a footnote at the beginning of a section or chapter.
SUMMARY

This thesis describes the work of four years research carried out at the Process Analysis and Chemometrics group at the University of Amsterdam. The presented work contributes to the field of statistical batch process monitoring (SBPM) and provides the reader with a variety of topics such as an overview of existing methods and SBPM models, introduction of new models, improvements and applications.

First, an introduction is given about batch processes and their importance in industries such as the pharmaceutical or (bio)-chemical industry. The performance of a batch process can be expressed in terms of safety, end-product quality, process consistency, efficiency, and process disturbances. The fluctuations in these batch process parameters are a result of the variation present in a batch process. Multivariate statistical methods are well suited to supervise this variation. SBPM handles the concepts of SPC where statistical thinking holds a central place. The aim is to reduce and monitor the variation of the batch process.

SBPM can be seen as an integral part of a quality management program. In order to successfully implement SBPM in such a program, SBPM is put in a framework called the I.T.A.- trajectory: Initial phase, Training phase and Application phase. The I.T.A.- trajectory provides the user a stepwise approach to the most important parts of SBPM. The initial phase is an explorative analysis of the raw process data. In this phase, it is examined whether the process conditions can be recognized as 'normal operated' and if the process is statistically in-control. In the training phase, a model is built to capture the normal process variation and control charts are constructed. Then, in the application phase, new batches are monitored on-line using the control charts derived in the training phase. If abnormal behaviour is detected, the origin of this unknown variation is located. Finally, the source of variation that caused the process upset is removed and the process is improved.

The first step in the initial phase is to understand batch process data. That is, to study its origin, nature and behaviour. It is often required to preprocess the data before a model can be developed. Variable scaling is often necessary to account for differences in measurement units. Furthermore, synchronization of the batch data is necessary to align the "local time" of each batch. This forces specific events in the batches to occur at the same time in each batch. In this thesis a procedure based on dynamic time warping (DTW) for synchronization of batches is described. The DTW
algorithm is thoroughly explained and it is shown that certain ‘inheritages’ from speech recognition can be omitted.

In the training phase, an SBPM model is constructed. Two distinct classes of SBPM models can be distinguished: black models and grey models. Black models are purely data driven and commonly used in SBPM. There exist a wide variety of black models. The most well known models are explained and evaluated. This evaluation is expressed in terms of the false alarm rate and fault detection performance of the models. Two new black models are introduced: the time evolving and the local model. The results of a comparison study show that it is not always clear which model to use. Grey models are a combination of purely data driven models and prior information such as kinetics or mass/energy balances. The concept of grey models is explained and applied to an industrial polymerization process, which is monitored on-line by Near Infrared spectroscopy. Grey models are found not to give faster fault detection compared to black models but a significant improvement with respect to process understanding.

Based on a SBPM model, control charts are derived. The charts are examined by considering the statistics used for the chart and its mechanics. The mechanics are used to study the manifestation of process faults in the control charts. The initially proposed statistics for process monitoring are discussed. This discussion is partly based on the validity of the statistical assumptions made during the construction of the multivariate control charts. Some suggestions to improve the statistical properties of the charts are proposed. The most important improvements are the new formulation of the null hypothesis as well as the leave-one-out procedure.

The mechanics of the control charts are discussed in two parts. First, the way of how process faults manifest themselves in the control charts. This behavior is examined for various models. By means of a simulation it is shown that in dynamic processes, faults are distributed in an unclear way in both control charts. This is due to the embedded error, but it depends highly on the choice of the model. Second, the ability of each model to capture the auto and cross-correlations present in batch process data is studied. It is shown that most models are well able to capture these correlations. In addition, for each model, the auto-and cross-correlations are distributed over the model parameters differently. Finally, general conclusions and future directions are given.
CHAPTER 1  ♦ BATCH PROCESS MONITORING*

*HJR/EVS
1.1 General introduction

Batch processes are present in many industries and therefore, play an important role. This chapter serves as an introduction to batch processes and statistical batch process monitoring. The advantages of batch monitoring are explained as well as the concepts of statistical process control (SPC). The current state of batch monitoring is outlined with respect to the application areas and research.

The first section gives an introduction of batch processes and for what products batch processes are deployed. In the second section, the principle ideas of statistical process control are briefly explained since statistical batch process monitoring originates from this. Also, some alternatives of statistical batch process monitoring are discussed. Furthermore, the advantages of multivariate control charts versus univariate charts are highlighted. The section ends with an overview of the current state of batch monitoring.

1.2 Batch reactors

Industrial batch reactors are in many ways the simplest reactors. Reactants are instantaneously administrated in a large stainless vessel, which is usually equipped with a stirrer to keep the reactants well mixed. Next, by adding heat or initiators the reaction is initiated and after a certain amount of time the reaction is finished and the reactants have been converted to the desired end-products. After that, the reactor is ready to produce a new batch. There are no reliable numbers concerning the market share of batch processes compared to continuous processes, but this market share is estimated to be around 50%.
1.2.1 Different Application Areas of Batch Processing

Batch processing is done in many industries because of several advantages (these will be discussed later). To show the diversity of products that are produced batch-wise, a list is given of different application areas. (This list is not intended to be complete).

- Chemical industry: polymer reactions (Latex, polystyrene, methylmetacrylate, PVC etc.), resins, coatings.
- Food industry: fat hardening, flavor products, beer.
- Pharmaceutical industry: antibiotics, powder mixing.
- Bio-chemical industry: waste water plant, fermentation, agricultural products.

Clearly, batch processes are common in a variety of industries. Every industry itself has its own characteristics. This is also reflected in the way batch reactors are operated. As an example, in the food industry hygiene is very important. In the pharmaceutical industry regulations for the production of e.g. medicines is very strictly controlled by the FDA (Food and Drug Administration). This makes the introduction of new measurements or recipes more laborious.

1.3 Features of a Batch Reactor

Various features characterize a batch reactor. The most important features are highlighted in the following.

- Batch processes are recipe driven. If during the development phase a completed batch run is known to have produced an on-spec product, the most important process variables from that batch are stored in a recipe. This recipe is then used for future batch runs aiming at consistent production of the on-spec product. This means that initial concentration; temperatures, pressures or dosing profiles of the reactants are fixed. The recipe itself is maintained using precise sequencing and automation of all required operations. During the lifetime of a batch reaction, the recipe itself is continuously adjusted and optimised.
- The reaction time of a batch is finite. The reaction time is often referred to as batch time.
The batch duration can change a lot between different types of batch reactions. For example, a polymerisation reaction may only take half an hour while a fermentation reaction can take several days.

- The conversion of the reactants is usually high (~90%).
- The behaviour of a batch process is non-linear and non-steady-state.
- Batches processes can be operated in stages.
- Batches can also be operated in semi-batch mode. This means that reactants are being added into the reactor slowly instead of instantaneously. Semi-batch reactions have the advantage of good temperature control and the capability of minimizing unwanted side-reactions through the maintenance of a low concentration of one of the reactants.

### 1.3.1 ADVANTAGES OF BATCH PROCESSING

For large volume production, it is not advantageous to use batch processing. In such cases continuous production is preferred. However, batch processing is flexible in the sense that different products can be made using the same equipment. This enables the supplier to quickly respond to the customer needs. In principle, only the recipe needs to be changed in order to produce a different product. Some industries like the food and pharmaceutical industry are highly regulated.

The batch products can be tested and certified batch-by-batch, and from a regulatory point of view, it is easier to deal with. Furthermore, their low-volume, high-value products characterize these industries.

### 1.3.2 ON-LINE MEASUREMENTS OF BATCH DATA

Batch processes are often monitored using real-time measurements. The ideal on-line measurement is simple, disturbance insensitive, cheap and accurate. Unfortunately, the lack of instrument robustness or sample preparation time causes that most plant variables are difficult or even impossible to measure on-line. Process variables in a plant can be categorized according to M. Soroush [1]:

---

-8-
### Table 1

<table>
<thead>
<tr>
<th>Name of group</th>
<th>Brief description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic process variables or engineering variables</td>
<td>Measured readily on-line, to ensure proper operation of plant.</td>
<td>Temperature, pressure, liquid levels, flow rates, feed composition, stirring speed, compressor duty, power supply.</td>
</tr>
<tr>
<td>Plant-product quality indices</td>
<td>Monitored by process engineer to ensure proper operation of plant, rarely available on-line, obtained by laboratory sample analysis</td>
<td>Viscosity, melt viscosity, density, composition, pour point, flash point, octane number, molecular weight distribution.</td>
</tr>
<tr>
<td>End-product quality indices</td>
<td>Quantify product quality, referred as customer specifications, complex and not well understood, until final product is ready, abstract variable (not quantifiable like colour or taste)</td>
<td>Impact/tear/adhesive strength, brittleness, elastic modulus.</td>
</tr>
</tbody>
</table>

**Definition of process variables.**

Data from batch processes often concern the first category of Table 1. This set of data is often referred to as X-block or X-data. End-product quality indices (Y-data) are very interesting from a monitoring point of view. Unfortunately, X-data is often not completed with the Y-data. In the following, two types of engineering variables are briefly discussed. Some figures will be shown to give the reader a taste of what batch data looks like.

**1.3.2.1 Engineering variables**

It was already explained that batch data show non-linear dynamics. By looking at batch data this becomes obvious. The conditions over time are continuously changing, for example deactivation of catalyst in a reactor, fouling of a heat exchanger or fluctuation of concentration and/or flow rate of the feed, etc.
An average number of on-line measurements for a batch process lies between 7 to 15 process variables. In general most datasets are dominated by temperature and pressure measurements. However, a priori, it is not always known which variables are useful for statistical process monitoring.

The first example of typically engineering variables from a batch process concerns the production of paper pulp (TEMBEC-Canada).

For this process, twenty-one process variables are measured on-line. Because of company policy reasons, the time units for the x-axes are omitted as well as the measurement units on the y-axes. Some process variables show a dynamic behaviour with a good signal to noise ratio such as the digester pressure and bottom temperature. Other process variables show dynamic behaviour but suffer from a poor signal to noise ratio (e.g. steam header temperature, steam header pressure). It can be seen that some process variables represent the state of a valve (open/close).
Sometimes the state of a valve is given as a percentage between 0% (close) and 100% (open). Also, the batch-to-batch repeatability for one process variable can be better than another process variable. Another example concerns the production of polyvinyl chloride (PVC) (A.A. Tates et al. [2]). The on-line measurements of fifteen process variables are presented in the following figure:

![Figure 2](image)

*On-line measurements of a batch process for the production of PVC.*

Again, the x-labels and y-labels are omitted. In contrast to the paper pulp data, the heat management of this process is monitored (flow cooling water, reactor temperature, jacket temperature). Since the production of PVC is highly exothermic, it is important to monitor the heat management of the process for safety reasons. Sometimes it is possible to calculate new process variables from the measured ones. These (new) calculated process variables could be added to the data set.

By studying the trajectories of the process variables over time, valuable information can be obtained. This is the first step in analysing batch process data.
1.3.2.2 Spectroscopic data

Another type of batch data comprises the measurement of spectroscopic data (NIR, UV-VIS, Raman). Although commonly this type of data is also referred to as engineering variables, their nature is quite different from e.g. temperature or pressure measurements.

Spectroscopic measurements give mainly information about the chemistry of the process and to some extend physical information whereas temperatures and pressures describe mainly the physics. Also, temperature and pressure profiles are often examined separately, while spectroscopic data is examined as a whole. Furthermore, the number of measured process variables is much larger compared to the traditional engineering datasets. It is not uncommon to measure more than 300 wavelengths. The treatment of spectroscopic data for multivariate analysis is also different (e.g. no scaling because all channels are measured in similar units). In the following figure, spectroscopic X-data from a urethane resin polymerisation process is presented. The data represents the evolution of a single batch run and will be discussed in more detail in Section 5.3.

Figure 3
On-line NIR spectroscopic measurements of a batch process for the production of a urethane resin.
Chapter 1 • Batch process monitoring

It is quite common to present spectroscopic batch data in these 3D pictures. It is obvious that visual inspection of spectral data for non-experts is cumbersome. The analysis of batch data from spectroscopic sensors requires some additional knowledge not commonly known by chemical engineers.

Besides engineering variables and spectroscopic measurements there are more process analytical techniques to monitor the evolution of batch processes. Examples of such techniques are:

- Microcalorimetric monitoring (A. Menert et al. [3]).
- Ultrasound spectroscopy (C. Turner et al. [4]).
- Electronic tongue (C. Turner et al. [4])

1.4 Why batch process monitoring?
Process monitoring is advantageous for many reasons. This is, of course, not only true for batch processes but for industrial processes in general. However, the focus here is on batch processes.

1.4.1 SAFETY
Some batch reactions, like the polymerisation of polyvinyl chloride, are highly exothermic. These exothermic reactions produce a considerable amount of heat that needs to be withdrawn from the reactor to avoid dangerous situations. In such case, it is important to monitor the temperature management. Therefore, process monitoring is used for safety reasons.

1.4.2 PROCESS DISTURBANCES
In the previous example, a process disturbance can lead to undesired side reactions due to temperature upsets. Sometimes the filling of the reactor completely fails because the wrong valve is opened and a different reactant is added to the reactor. As a result, undesired products are formed or unsafe situations can occur. Therefore, process monitoring is done to prevent that process disturbances lead to environmental stress and economic losses.
The batch time of biochemical processes such as the production of antibiotics can take up more than days. Early detection of process disturbances for these types of batch processes is highly wanted. First, detection of disturbances makes it possible to correct the process if possible. Secondly, if the process cannot be corrected anymore, the decision is taken to kill the current process in order to start a new one. If such a decision can be made supported by process monitoring in an early stage, valuable batch time is gained. Therefore, process monitoring leads to time and cost reduction.

1.4.3 PROCESS UNDERSTANDING
One way to optimise a process is to study its behaviour. Knowing exactly how the process reacts on changing process conditions can help process engineers to make adjustments to the recipe of a batch process. Process monitoring is a useful tool for these purposes since the interaction between the various process variables is monitored. The effect of process adjustments can immediately been monitored in multivariate control charts. Process monitoring therefore is helpful for improving the process understanding.

1.4.4 PROCESS AND PRODUCT CONSISTENCY
The consistency of how the batch recipe is realized is reflected in the measure process variables. There will always be a certain amount of variation present in the process parameters. This variation can influence the quality of the end-product. Take as an example the fact that the personnel work in shifts. Every shift will work with slightly different process tuning parameters. This will lead to variation in the batch time of the process and/or product quality. It is important for the producer to reduce this variation as much as possible so that his concurrent position improves. Products that do not fulfil the quality constrains are often mixed with the product from a good batch. In this way, the product still meets the product quality requirements. Process monitoring therefore helps to make the process under consideration more consistent.

1.4.5 CAPABILITY INDEX
In industrial statistics it is common to express the performance of the process consistency in relation with the product consistency in terms of capability indices. The process capability is the repeatability and consistency of the process with respect to
specifications limits of a product parameter. Next, the capability indices are used to analyse whether a process, operating under normal conditions, is capable of meeting the specifications limits of the end-product.

As can be seen from the previous examples it is advantageous to monitor batch processes. These advantages include topics such as environment, health and safety, cost reduction, process consistency and process understanding.

1.5 Statistical process control

The theory of statistical batch process monitoring originated from statistical process control. For that reason, it is interesting to take a look at the concepts of statistical process control (SPC) and the development of statistical batch process monitoring through the years.

1.5.1 Ideas of statistical process control

The basic idea of the statistical approach of process control is that every industrial process is liable to variations. This process variation will affect the quality of the end-product. When the fluctuations in the quality of the end-product are unacceptable, it is worthwhile to study the process itself to see where the process variation enters and how large it is. This thought was developed in the early 20's of the past century (W.A. Shewhart [5]). Initially SPC was developed for the manufacturing industry where the end-product quality is directly measured (e.g. the diameter and length of a nail). Later on, the end-product quality measurements where substituted by measuring the process conditions. Ideally, the observed variance within the process conditions is related to end-product quality.

'Statistical thinking' has two starting points. First, it is assumed that improvement of the end-product is reached by reduction of process variation. To improve the quality of the end-product it is useful to study the variation in the process. Eventually, process monitoring leads to adjustments of the process.

Secondly, it is assumed that this process variation can be well monitored with the use of control charts. Shewhart distinguishes two different kinds of variation:
- Process variation that is inherent to the process itself (Common cause variation).
- Process variation that has an external cause (Special cause variation).

In short, it is aspired to operate a process that is stable, has a minimum of variation and produces an end-product that is on spec. In a process, the following situations can occur:

1) The process is statistically in-control and the end-product fulfils the predetermined specifications. This is the desired situation. Process faults are detected immediately. While detecting the process fault, the cause can be located.

2) The process is statistically in-control but the product does not fulfil the predetermined specifications. This is an unwanted situation. The process is stable, but adjustments are needed. There are two possibilities. First, the process variation is too large and needs to be decreased by e.g. revising the process design. Secondly, it is also possible that not all the required measurements are performed. In other words, the process measurements indicate that the process is stable and reliable while somewhere in the process some process variation is apparent that influences the end-product quality. Further analysis by e.g. experimental design and additional measurements is desired.

3) The process is statistically not in-control and the product fulfils the predetermined specifications. This is also an unwanted situation. During the running time of the process it cannot be predicted whether the product will meet the product specifications. It is also possible that the product will not fulfil these specifications at all. In such cases, localising the cause will be quite difficult.

4) The process is statistically out of control and the product does not fulfil the predetermined specifications. In these highly unwanted situations the whole process needs to be thoroughly examined.
1.5.2 DIFFERENCE BETWEEN PROCESS MONITORING AND PROCESS CONTROL

There is often confusion between process monitoring and process control. Sometimes process monitoring is actually done while it is named process control. This is for instance the case for (multivariate) statistical process control (MSPC). (M)SPC is actually a technique for monitoring processes but the designation suggest that process control is performed.

Process control from a chemical engineering point of view boils down to the automated monitoring with well-defined control actions of a process, in which a computer system is used to regulate the usually continuous operations or processes. Equipment that measures the variables of an industrial process, directs the process according to control signals from the process computer system, and provides appropriate signal transformation. SCADA (supervisory control and data acquisition) is a software program for process control where data is collected in real time from remote locations in order to control equipment and conditions. In the field of process control terms like P, PI and PID controller are commonly used. It is important to note that process control for batch processes is limited because of the non-linear behaviour.

However, the definition of statistical process control is different. That is, SPC is a well-established set of on-line, off-line and at-line mathematical tools, supported by a wealth of software and manual systems, to determine and monitor the level of statistical control demonstrated by a production process. In other words, the level of statistical control is monitored. This is where the term monitoring came in. Therefore, the terms SPC and process monitoring have the same meaning when used in the ongoing to avoid misunderstandings. As of now, multivariate statistical process control of batch processes is referred to as statistical batch process monitoring.

1.5.3 METHODS FOR BATCH PROCESS MONITORING

This chapter gives a summary of possible methods for batch process monitoring. These methods are for the majority based on SPC methods since SPC for batch processes originates from SPC. In order to give a more complete overview also some alternative methods for batch process monitoring are shortly mentioned.
1.5.3.1 Fundamental models
One choice for batch process monitoring is to use fundamental models (P. Nomikos & J.F. MacGregor [6]). These models are based on fundamental mass or energy balances that are commonly found in the field of chemical engineering. With the help of such models estimates of the underlying theoretical states of the process are provided. Often these fundamental models are based on state estimation methods. Such methods combine a fundamental model of the process with on-line process measurements to provide on-line, recursive estimates of the underlying theoretical states of the batch processes (e.g. using a Kalman filter) (P. Nomikos & J.F. MacGregor [6], K.A. Kosanovich et al. [7]).

However, using theoretical fundamental models have some serious drawbacks. Only for a very limited number of processes the models are accurate enough. The chemistry of e.g. a biochemical reaction is far too complicated to be captured in a good working fundamental model. Furthermore, constructing such models is often very time consuming. This makes the use of e.g. state estimators impractical for batch process monitoring purposes.

1.5.3.2 Knowledge based approach
The most pure form of a knowledge-based approach to monitor batch processes is the recipe driven approach that is commonly used. The recipe is composed and adjusted conform the experiences of process operators and researches. In case of process faults, e.g. a sensor failure, the knowledge of the process operator is utilized to take action.

It is obvious that such knowledge is transient since the loss of the process operator also includes the loss of process knowledge. However, each behavioral description can be associated with causal assumptions in a hierarchical structure. For each node in the hierarchy, diagnostic rules can be generated from the behavioral descriptions. Thus the process model is represented by a set of qualitative and quantitative descriptions based on the knowledge of process operators and engineers. These are called rule-based expert systems. This approach is advantageous since it requires no model. However, formulating the behavioral and causal descriptions is just as difficult and time consuming as constructing a fundamental model.

Another approach to the problem of fundamental modeling is the use of neural nets (P. Nomikos & J.F. MacGregor [6]). These neural nets can be used as
pattern identifiers in order to diagnose fault conditions in a batch process. However, such a model needs to be trained with data that contains numerous faulty batches. This is practically never the case for batch processes. Interrogation of neural nets to assign process faults is difficult. Therefore, neural nets are no good candidates for batch process monitoring in terms of fault diagnosis.

1.5.3.3 SPC
Statistical batch process monitoring originated from the field of SPC. The fundamentals of statistical process control were described by W.A. Shewhart [5]. H. Hotelling [8] was one of the first who introduced a multivariate approach for statistical process control to monitor a process of a multivariate nature. J.E. Jackson [9] then applied principal component analysis (PCA) to reduce the dimensionality of the multivariate data. J.V. Kresta et al. [10] proposed a multivariate control technique for continuous processes that uses PCA or projection to latent structures (PLS). Finally, P. Nomikos & J.F. MacGregor [6] extended the concept of multivariate statistical process control for batch processes.

It is important to understand the development of statistical batch process monitoring since terminology and concepts from different steps in this development are often mixed. This leads to confusion and unnecessary mistakes. Therefore, a short overview is given of the development of statistical batch process monitoring starting with classical SPC.

1.5.3.3.1 Univariate charts: $\bar{X}$, CUSUM, EWMA
Univariate charts are based on the measurement of only one property or individual observation such as the mean of e.g. the diameter of a nail. The univariate charts are widely used in e.g. manufacturing industries and make up an important part of quality improvements programs. In the following, the most common univariate charts will be discussed briefly (D.C. Montgomery [11], R.E. Walpole & R.H. Myers [12]).

1.5.3.3.1.1 $\bar{X}$ Chart
Walther S. Shewhart first proposed the general theory of statistical control charts in the early 30's of the last century. A typical control chart is a graphical display of a quality characteristic that has been measured from a sample versus the sample number or time (Figure 4).
In this figure, three lines can be distinguished. The centerline represents the average value (μ) of the quality characteristic. This value corresponds to the in-control state. The other two horizontal lines are the upper control limit (UCL = 3σ) and lower control limit (LCL = -3σ). As long as the points fall between the UCL and LCL, the process is assumed to be statistically in-control. The opposite is true if a point plots outside these control limits.

The process is assumed to be out of control. Action is required to locate the cause of this abnormal deviation from the in-control situation. It should be noted that there are different rules for appointing a signal to be out of control (D.C. Montgomery [11]).

The setup for the \( \bar{X} \) chart will be discussed according to a small example (D.C. Montgomery [11]). The quality characteristic for a manufacture process of piston rings is given by the diameter of the piston ring. It is known that the average value for the diameter (μ = process mean) is 50 mm and the process standard deviation is \( \sigma = 0.015 \) mm. Ideally, the outcome of the average diameter \( \bar{x} \) measured for \( n \) samples of pistons is 50 mm. In practice, this \( \bar{x} \) average will always deviate from \( \mu \). For sample sizes of e.g. \( n = 10 \), the standard deviation of the sample average \( \bar{x} \) is calculated according to equation 1:

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.015}{\sqrt{10}} = 0.0047
\]

This is all the information needed to compute the UCL and LCL. In fact, the following hypothesis test is performed:
\[ H_0 : \mu = 50\text{mm} \]
\[ H_1 : \mu \neq 50\text{mm} \]

With a probability \( \alpha \) of erroneously determining that the process is out of control, the null hypothesis is accepted if \( \bar{x} \) falls between \( 50 - Z_{\alpha/2}(0.0047) \) and \( 50 + Z_{\alpha/2}(0.0047) \). Thus, the UCL = \( 50 + Z_{\alpha/2}(0.0047) \) and LCL = \( 50 - Z_{\alpha/2}(0.0047) \). Sometimes, the value for \( Z \) is taken to be three. Typically, this choice is referred to as the three-sigma limits. The UCL and LCL then are plotted in the control chart. The centre line CL of the control chart reflects the ideal value of the average size of the piston diameter. An example of such a chart is given in Figure 5.

It is important to note that it is assumed that \( \bar{x} \) is normally distributed due to the central limit theorem. Also, it is assumed that the process mean and standard deviation are known.

Closely related to the \( \bar{X} \) chart are the \( R \) chart (center line is the range for the data) and S-chart (monitoring the sample standard deviation). The disadvantage of the \( \bar{X} \) chart is that small changes of the mean are not detected. Only shifts in the mean of magnitude \( 1.5\sigma \) to \( 2\sigma \) or larger are effectively detected.
1.5.3.3.1.2 CUSUM chart

The cumulative sum, or CUSUM chart, is a good alternative to detect small shifts of the mean and was first proposed by E.S. Page [13]. Consider a control chart for the mean with a target for the process mean $\mu_0$, with the following observations $x_1, x_2, x_3, \ldots, x_i$. The cusums are calculated according to:

$$S_1 = x_1 - \mu_0$$
$$S_2 = S_1 - (x_2 - \mu_0)$$
$$S_3 = S_2 - (x_3 - \mu_0)$$
$$\vdots$$
$$S_j = S_{j-1} - (x_j - \mu_0)$$

Often, the starting value for the CUSUM chart is set to zero. Small changes in the mean will result in a relatively large increase in the slope (positive or negative) of the CUSUM chart. For this reason, tabular cusums and the V-mask form of the cusum are used. In the following figure, a CUSUM chart is presented using a V-mask see D.C. Montgomery [11] for a more detailed discussion.

![Example of a CUSUM chart with V-mask](http://www.ub.rug.nl/eldoc/dis/eco/j.e.wieringa/c5.pdf)
1.5.3.3.1.3 EWMA chart

Another alternative to detect small shifts of the mean is the exponentially weighted moving average (EWMA) chart. The performance of the EWMA chart is equivalent to the CUSUM chart, although the EWMA is easier to set up and to operate (D.C. Montgomery [11]). The following chart statistic is plotted:

\[ z_t = \lambda x_t + (1-\lambda)z_{t-1} \]  

(3)

Here, \( \lambda \) is a constant value between zero and one. The starting value of \( z_0 = \mu_0 \), where \( \mu_0 \) is the starting value of the average value \( \mu \) of the quality characteristic. If all the previous values \( z_{t-1}, z_{t-2}, \ldots, z_0 \) are substituted in equation 3, it follows that

\[ z_t = \lambda \sum_{i=0}^{t} (1-\lambda)^i x_{t-i} + (1-\lambda)^t z_0 \]  

(4)

For the calculation of the chart statistic \( z_t \), the means of the past and current are considered. However, the 'older' the mean, the less weight it receives. The centre line, UCL and LCL for the EWMA chart can be calculated as follows:

\[
\begin{align*}
UCL &= \mu_0 + L\sigma \sqrt{\frac{\lambda}{2-\lambda} \left[ 1 - (1-\lambda)^{2t} \right]} \\
CL &= \mu_0 \\
LCL &= \mu_0 - L\sigma \sqrt{\frac{\lambda}{2-\lambda} \left[ 1 - (1-\lambda)^{2t} \right]} 
\end{align*}
\]

An example of an EWMA chart is given in Figure 7:
In the upper graph of Figure 7 a $\bar{X}$ (Shewhart) chart is given of an arbitrary process. In the lower graph of Figure 7 an EWMA chart is given of the same process. As can be seen, the EWMA chart gives a better performance in detecting a small shift from the mean.

1.5.3.3.1.4 Average run length

For univariate charts, decisions regarding sample size and sampling frequency are evaluated through the average run length (ARL) of the control charts (D.C. Montgomery [11]). The average run length is defined as the average number of observations up to and including the first-out-of-control observation. The ARL is a function of $\delta$ where $\delta$ represents a shift of the mean. The ARL ($0$) for any Shewhart chart can be calculated as follows:

$$ ARL = \frac{1}{P(0)} $$

(5)

Where $p$ is the probability that any point exceeds the limits in the control chart.

For a $\bar{X}$ chart with three sigma limits ($P(0) = 0.0027$) the ARL ($0$) when the process is in-control is 370. In other words, every 370 observations an out-of-control
signal is generated while the process is in-control. This is also called a false alarm. The average time to signal (ATS) is calculated as follows:

\[ ATS = ARL \cdot h \]  \hspace{1cm} (6)

Where \( h \) is the time (in hours or seconds) that samples are apart.

1.5.3.3.2 Multivariate control charts

The univariate charts typically deal with single observations. However, when the variables are correlated (which is often the case for an industrial process), superimposing univariate charts is not a very accurate method of monitoring processes because relationships between the variables are not taken into account. In this section, the focus will be on multivariate data originating from continuous processes. The reason for this is that the concepts of multivariate process control for batch processes are similar to multivariate process control for continuous processes. However, in transferring the methodology from continuous to batch processes, there are some issues that are not straightforward. In the following, the concepts of multivariate statistical process control for continuous processes are highlighted. This will be helpful when the theory on statistical batch process monitoring is extensively discussed.

Process measurements of \( J \) process variables are obtained at \( K \) time intervals on a continuous process and collected in the data matrix \( X \) (\( K \times J \)). Preprocessing the process data \( X \) is a first step as will be discussed later. For the following discussion, it is assumed that the data is properly pre-processed. Furthermore, \( X \) is believed to represent in-control observations.

1.5.3.3.2.1 Continuous processes without dimension reduction

In 1947 Hotelling introduced the \( T^2 \) statistic as a method to apply statistical quality control to correlated data with a multivariate nature (H. Hotelling [8]). This summary statistic enables easy control charting and allows for detecting faults of highly correlated measurements. In the case of continuous processes for chemical applications, \( K \) measurements of \( J \) process variables are measured every time interval. These measurements can be arranged in a matrix \( X \) (\( K \times J \)). For a new measurement \( x_{\text{new}} \) (\( J \times 1 \)) the test statistic \( T^2 \) is calculated according to:
\[ T^2 = (x_{\text{avg}} - x_\ast) S^{-1} (x_{\text{avg}} - x_\ast) \]  

where \( S (J \times J) \) is the estimated covariance matrix of \( X \) and \( x_\ast \) is the target value or grand mean. Here, the sample size \( n \) of one single measurement equals 1. Therefore, if the goal is to obtain an in-control set of observations as described in the training phase, the UCL is given by (N.D. Tracy et al. [14]):

\[ \text{UCL} = \frac{(K-1)^2}{K} B(\alpha, J/2; (K - J - 1)/2) \]  

where \( K \) is the number of samples, \( J \) is the number of process variables and \( B \) is the Beta distribution with a confidence level \( \alpha \) and \( (J/2, 1-J-1/2) \) degrees of freedom. The UCL_{avg} to use for detecting changes from the in-control process for new independent observations are given by:

\[ \text{UCL}_{\text{avg}} = \frac{J(K^2 - 1)}{K(K - J)} F(x; J, K - J) \]  

where \( K \) is the number of samples, \( J \) is the number of process variables and \( F \) is the F-distribution with a confidence level \( \alpha \) and \( (J, K-J) \) degrees of freedom.

An important note here is that all measured process variables are used and no dimension reduction takes place. With a large number of correlated process variables this approach breaks down, e.g., \( S \) might become (nearly) singular.

1.5.3.3.3 Megavariate control charts

This section discusses the situation where dimension reduction is required for multivariate statistical process monitoring. This has some important consequences for e.g. the type of control charts that is used. Furthermore, the concepts of this approach are very similar to those of statistical batch process monitoring.

1.5.3.3.1 Continuous processes with dimension reduction

In a chemical process, during process operation many process variables are collected and the dimensions of \( X \) may become large in the second mode \( J \). Since this large number may cause singularity problems, it is advantageous to reduce the dimensionality of the data. PCA is a multivariate statistical method well suited for this
purpose. J.E. Jackson [15] was one of the first who applied a principal component analysis in this context.

A PCA decomposes the matrix $X$ into the sum of $R$ outer products of scores $t$ and loadings $p$ (systematic part) plus a residual part $E$.

$$X = \sum_{r=1}^{R} t_r p_r^T + E = \text{model explained} + \text{residuals}$$

where $T (K \times R)$ is the score matrix, $P (J \times R)$ contains the loadings and $E (K \times J)$ is the residual matrix. A geometrical representation of equation 10 is given in Figure 8.

![Figure 8](image)

*Graphical representation of projecting a sample in a dimension reduced space.*

Once the model plane, spanned by the columns of $P$, is defined, the scores and residuals for a new-scaled measurement $x_{\text{new}} (J \times 1)$ are found by projection on the model plane. This is a regression problem:

$$x_{\text{new}} = \hat{P} t_{\text{new}} + e_{\text{new}}$$  

(11)
The $D$-statistic describes the Mahalanobis distance from the projection on the model plane to the centre of the model plane, and describes the systematic variation in the data. The centre of the model plane corresponds to the average process behaviour. The $SPE$-statistic represents the Euclidian distance between the measurement vector and its projection on the model plane. This distance describes the variation in the data that is inconsistent with the model.

The test statistics are calculated according to

$$
\hat{t}_{\text{new}} = (P'P)^{-1}P'x_{\text{new}} = P'x_{\text{new}} \quad \rightarrow \quad D = t_{\text{new}}'S^{-1}t_{\text{new}} $$

$$
e_{\text{new}} = x_{\text{new}} - \hat{P}t_{\text{new}} \quad \rightarrow \quad SPE = e_{\text{new}}'e_{\text{new}}
$$

Here, $S$ ($R \times R$) is the variance-covariance matrix of the scores $T$ ($I \times R$) and is diagonal since the scores are orthogonal. Subsequently, the test statistics are plotted in the control charts and $x_{\text{new}}$ is assigned to be statistically out of control if one of the control charts signals. The UCL for the $D$-statistic is the same as given in equation 9 except that $I$ is replaced by $R$. The UCL for the $SPE$-statistic is the critical value $SPE_\alpha$. The derivation of this critical value is given by J.E. Jackson [15].

1.5.3.3.4 Statistical batch process monitoring

The principles of statistical batch process monitoring are very similar to the concepts for monitoring continuous processes in a dimension reduced space. First, a sufficient number of batches is collected that represent normal operating conditions (NOC batches). This data set is then modelled in a reduced space using PCA. The latent variables (scores and residuals) form the reference distributions of these batches. The scores form the basis of the $D$-chart whereas the residuals form the base for the $SPE$-chart. The control limits for these control charts are calculated from their reference distributions. Then, the multivariate measurement (at time interval $k$) from a completely new batch is projected on the same dimension reduced space as the NOC batches. The $D$ and $SPE$-statistic for this new batch are calculated from the scores and residuals. These statistics are plotted in the control chart. In this manner, a batch can be monitored in an on-line fashion. It can also be desirable to perform a so-called post-batch analysis. As soon as a new batch is finished, the collection of all the measurements over time are to be projected in the dimension reduced space at once.
The foregoing is a description of statistical batch process monitoring in a nutshell. In the next chapters, every aspect of this approach will be discussed in a step-wise manner. This is called the I.T.A.-trajectory.

1.5.4 Multivariate charts versus univariate charts

It was shown in the previous chapters that univariate charts could be used for process monitoring. For batch processes, a $\bar{X}$ chart can be constructed for each separate process variable. Then, the average value for each process variable is plotted at each time interval $k$. However, these univariate charts have some disadvantages that can be dealt with by using multivariate charts for statistical batch process monitoring. Some of these disadvantages are briefly discussed in the following.

1.5.4.1 Abundance of control charts

The average batch reactor is equipped with ten to thirty sensors that measure typical engineering variables such as pressures and temperatures. If univariate charts are being used, one control chart is constructed for every single process variable. This leads to an overabundance of univariate control charts. This can be considered as a drawback since it is inconvenient to monitor 30 charts at the same time.

This problem becomes even worse when a spectroscopic instrument is mounted to the reactor. Such an instrument records more than 500 -1000 wavelengths every time interval. This can be overcome by using multivariate charts: the $D$ and SPE-chart. The performance of the process can be followed over time in only two control charts. When the chart statistic is out-of-control in one of the charts, the cause of the process fault can be localized using contribution plots (P. Miller et al. [16], J.A. Westerhuis et al. [17]). The ideas about contributions plots are discussed in more detail in the ongoing.

1.5.4.1.1 False alarm rate

If a chart statistic is plotted outside the control limits while the process is actually in-control, a false alarm occurred. The probability of a false alarm using $n$ multiple univariate charts simultaneously plotting in-control is not $1 - \alpha^*$. If a process is in-control, the probability of plotting $n$ means in-control is $(1 - \alpha)^n$ since each chart has a probability of $1 - \alpha$. Thus, the joint probability of a type I error is much larger: $1 - (1 - \alpha)^n$. This gives no trust in the control charts by the process operator. Often, the
process operator picks out only one chart that is of value to the operator. The other charts are being ignored because of the high false alarm rate.

1.5.4.1.1 Multivariate nature
Statistical batch process monitoring models capture the auto-correlation and cross-correlation among the process variables. This information is translated in the \( D \) and \( SPE \)-chart. This gives the operator an extra tool and valuable information about the batch process. If the correlation changes due to a process upset, a signal is given in one of the control charts. The detection of a fault in multivariate control charts is illustrated in the following figure:

![Figure 9](image)

Detection of a fault in multivariate control charts.

Clearly, the fault that occurred around time interval 6 -10 is detected in the \( SPE \)-chart. Univariate control charts for the two faulty process variables are given in the following figure:
It can be seen from Figure 10 that the process fault goes unnoticed using univariate control charts. This happens because the multivariate nature of the process variables is not taken into consideration.

New events that are not captured by the PCA model are detected in the SPE-chart (e.g. a sensor failure). Sometimes the modelled correlation among process variables is still intact but the joint variation is larger than normal. This will result in an out of control signal in the $D$-chart. In this way, faults can be categorized or diagnosed according to the detection in one of the control charts.
1.6 Nowadays monitoring of batch processes

It is interesting to know how monitoring of a batch process is actually performed in practice. First, the theory of statistical batch process monitoring is only ten years old (T. Kourti [18]). It is common that academic research is implemented in industry with a time delay. Clearly, there has been a strong trend in the ascent of data acquisition systems (SCADA) and distributed control systems (DCS) in process plants. It seems that systems like SCADA is well suited for data driven approaches such as statistical batch process monitoring. However, SCADA software originally is designed in a process control environment. The demand for process control software is mainly reliability and a long lifetime. Therefore, developments in this type of software are slow, not to mention of incorporating monitoring tools. Also, the management of the large amounts of data that are being collected is cumbersome. This makes it e.g. almost impossible to withdraw historical data from a database system in a simply manner. These are serious problems that need to be overcome first. However, there are rules within the pharmaceutical industry, dedicated by the FDA, how to automate batch processes. This forces the producer more or less to build modular batch control software. This makes accessing the data a lot more efficient and therefore more attractive to statistical batch process monitoring.

The most common way that industrial batch processes are operated is according to a recipe. Customized software enables the producer to operate these batches fully automated. This is probably the most widely used method for batch process monitoring. Many applications exist where the implementation of these software modules lead to improved product quality and reduced loss of operating time. These software packages do not contain sophisticated monitoring tools like multivariate control charts and the success of these packages can be ascribed to the increased level of knowledge about the batch process itself then due to the software.

However, also commercially software packages are available that offer the opportunity for on-line statistical batch process monitoring. The major drawback of these software packages is that quite often they are not compatible with the platforms used in the process plant. It is foreseen that the next generation software packages will
overcome this obstacle. It also requires a mentality change of the personnel that often sticks to the classical way of monitoring processes.

Considering this positive development together with the increasingly becoming important data-flood problem, it seems the share of statistical batch process monitoring tools will grow compared to conventional monitoring systems.

1.6.1 LITERATURE & SOFTWARE

The pioneering work in this field is presented in the work of Nomikos and MacGregor in the early nineties. Based on their work, several extensions and modifications have been proposed in the last 10 years. In the following, an overview is given of the most relevant publications in this field. The publications are listed under the headings: pioneering articles, statistics, extensions/- methodologies and comparison and evaluations.

| Table 2 |
|-----------------|----------------------------------|
| **Articles**   | **Pioneering**                   |
| **Statistics** | P. Nomikos & J.F. MacGregor [6], P. Nomikos & J.F. MacGregor [19], P. Nomikos & J.F. MacGregor [20], P. Nomikos [21], J.F. MacGregor & P. Nomikos [22] |
|                 | G.E.P. Box [23], J.E. Jackson [9], J.F. Jackson & G.S. Mudholkar [24], J.E. Jackson [15], J.V. Kresta et al. [10], H.-J. Ramaker et al. [25], S. Wold [26], J.E. Jackson & G.S. Mudholkar [24], N.D. Tracy et al. [14], P. Nomikos & J.F. MacGregor [19], E.B. Martin et al. [27] |
| **Extensions & Evaluation** | S. Wold et al. [28], A.A. Tate et al. [2], J.A. Westerhuis et al. [29], E.N.M. Sprang et al. [30], T. Kourti [18], S. Rannar et al. [31], R. Boqué & A.K. Smilde [32], K.S. Dahl et al. [33], P.J. Gemperline et al. [34], A. Kassidas et al. [35], A. Kassidas et al. [36], B. Lennox et al. [37], D.J. Louwerse & A.K. Smilde [38], S. Rannar et al. [31], A.K. Smilde & H.A.L. Kiers [39], R. Boqué & A.K. Smilde [40], E.N.M. Van Sprang et al. [41], J.A. Westerhuis et al. [42] |
| **Applications** | S. Albert & R.D. Kinley [43], B. Lennox et al. [37], E.B. Martin et al. [27], D. Neogi & C.E. Schlags [44], P. Nomikos [45], A.A. Tate et al. [2], B. Lennox et al. [37], K.A. Kosanovich et al. [7], J.A. Westerhuis et al. [46] |

*Literature overview statistical batch process monitoring.*
There are many commercial statistical software packages available. However, commercial software focussing on batch processes is limited. The two leading software packages are SIMCA-P+ developed by Umetrics in Sweden and MSPC+ developed by MDC technology in the U.K. Both packages can be used for off-line and on-line process monitoring.
CHAPTER 2 ♦ I.T.A.-TRAJECTORY*

* HJR/EVS
2.1 General introduction

The basic ideas about SPC were briefly explained in the previous chapter. Within quality management programs such as the 6σ program or TQM (Total Quality Management), SPC is used as an essential tool. A similar program concentrating on the Dutch market is ZVP (Zakelijk Verbeter Programma) by (R.J.M.M. Does et al. [47]). The ZVP is based upon a collection of various quantitative quality programs and consists of a scheme to come to a faultless process, service or product. These steps read as follows:

1) Measure
2) Analyse
3) Improve
4) Guarantee

The same philosophy as ZVP is employed for chemical processes, in particular batch processes. The first initiative for a batch improvement program (Batch Verbeter Programma) is to assign phases according to the concepts of multivariate statistical process control. W.H. Woodall [48] stated that it is very important to improve the communication between practitioners and researchers concerning statistical process control methods. One way to establish an improved communication is to assign phases to the concepts of multivariate statistical process control. Multivariate statistical process control is carried out in three phases: The Initial, Training and Application phase, or the I.T.A.-trajectory. Here, the training and application phase refer to respectively phase 1 and phase 2 from standard SPC terminology. An illustration of the I.T.A.-trajectory is given in the following figure:
Each of the phases from the I.T.A.-trajectory will be discussed in the following chapters.

2.2.1: Initial phase

In the initial phase historical process data is collected and analysed to determine if the data is suited for process monitoring. Because of control actions and safety, it is common that various process variables are measured on-line. This is usually done by means of DCS (distributed control systems) and SCADA (supervisory control and data acquisition system). These systems control the underlying PLC’s (programmable logic controller) and properly store the collected data in a database. Quality measurements of the end-product are usually stored in a LIMS (Laboratory Information Management System) system. Therefore, data are spread in various systems and are not always accessible in an easy manner. Furthermore, it is common that sensors are logged continuously without any markers. The analyst will have to use batch reports to search for the actual time where the batch started. It is not difficult to imagine that this phase is a very time consuming step and often forms a stumbling block in batch monitoring. Another problem often encountered in the initial phase is
missing data because of sensor failure or other malfunction. There are several suggestions in the literature to deal with this problem.

2.2.1 EXPLORATIVE ANALYSIS
Once the data is extracted from the database an explorative analysis is performed. In this analysis the relation and correlation between process variables and end-product quality is determined. Moreover, it is analysed if measured process variables consist of process relevant information. The first step is to look at the trajectories of the measured process variables over time. Often, abrupt faults are easily detected by examining the plots of the individual trajectories. This already tells a lot about the process. Furthermore, in this stage, it is determined if the process is statistically in-control e.g. does the process operate around a steady working point with a certain level of variation. This will be will be defined as normal operating conditions (NOC). In batch process monitoring it is important to specify the normal operating conditions. A first selection is based on the operator knowledge, batch reports and end-product quality results. In the training phase, also multivariate statistical techniques are used to determine the NOC batches in the obtained batch data. This is necessary to detect more incipient faults that are not easily detectable by examining individual trajectories of the process variables. Such faults are typical in industry.

2.3 T: Training phase
The goal of the training phase is to build a well functioning model based on a set of NOC batches. From this model, the SPE and D-control chart are constructed that are used in the application phase for on-line monitoring of completely new batches. The training phase is divided in a few steps. These steps will be described step by step.

2.3.1 NOC SELECTION OR OUTLIER DETECTION
The batches that are collected in the initial phase preferably meet the following requirements. First, the number of batches should be sufficient. The statistical properties of the model are more precise when large numbers of batches are available. Secondly, the batches represent normal operating conditions (NOC). This is very important, since the process variation in the NOC batches serves as a reference distribution. The more this process variation represents NOC conditions, the better
future faulty batches are detected. If e.g. a faulty batch is included in the NOC data, the total amount of process variation increases. As a result, the reference distribution now consists of wrong batches. Therefore, the model is less capable of detecting differences in process variation between the NOC data and a new wrong batch.

2.3.1.1 Post-batch analysis
A single batch can be represented by a matrix $X (K \times J)$, where $K$ is the number of observations and $J$ the number of process variables. $J$ can be either engineering variables such as temperature and pressures, but also channels from spectroscopic measurements. By repeating batch runs, the batches can be stored in a three-way array $X^{I \times J \times K}$, where $I$ is the number of batches. This is illustrated in Figure 12:

Before the batch process monitoring model is built, a selection of NOC batches from the three-way array $X$ is required. This selection is based on multivariate statistical tools and will be referred to as post-batch analysis. An example is given how to select batches from a dataset $X^{53 \times 9 \times 200}$. The dataset concerns a polymerisation of styrene butadiene. From this process, 53 batches are obtained and 9 process variables are measured for a period of 200 time intervals for every batch. The question is: select NOC data by using post batch analysis. The trajectories of the process variables are plotted in the following figure:
It is obvious from Figure 13 that distinguishing between normal and faulty batches is laborious and difficult. The multivariate approach is performed as follows:

**Scaling**

Since PCA is a linear technique, pre-processing of $X$ is required because of the highly non-linear behaviour of batch processes. In this example, the process data $X$ is column-centred and tube-scaled (H.A.L. Kiers [49]), removing the average trajectories of all the process variables and scaling the process variables to unit variance. The pre-processed data in $X$ now represents deviations from the average trajectory and is therefore approximately linear (P. Nomikos & J.F. MacGregor [6], A.K. Smilde [50]).

**Unfolding the data**

The data is arranged in a three-way data array $X$. For post batch analysis, $X$ is matricized in order to perform a PCA analysis. The matricizing is depicted in Figure 14:
It can be seen that matricizing results in the matrix $X^{I \times JK} = X^{33 \times 1800}$.

**PCA**

The next step is to build a PCA model of $X$.

$$X = TP' + E$$  \hspace{1cm} (14) \hspace{1cm}$$

Where $T$ are the scores, $P$ the loadings and $E$ the residuals. In this example, three principal components ($R$), as determined with cross validation, are sufficient to describe $\pm 30\%$ of the variation in $X$.

**$T$ and $Q$-plot**

The scores and residuals that result from the PCA analysis are used to construct two simple graphs, a scatter plot of $T$ and a $Q$-plot. First, a scatter plot of $T$ is simply constructed by plotting the scores for the first column of $T$ against the scores of the second column of $T$. The score plot for the SBR data is given in Figure 15.

![Figure 15](image)

*Figure 15*

*Scatter plot of the scores.*

![Figure 16](image)

*Figure 16*

*$Q$-statistic of the SBR process with $99\%$ confidence limits.*

It can be seen from the score plot that batch 52 and 53 can be regarded as outliers. An individual inspection of these batches will learn what the cause is of this deviating behaviour. In this example the deviations for both batches are caused by impurities in
the feed (P. Nomikos & J.F. MacGregor [6]). From the residuals, the sum of squares of these residuals, denoted as \( Q \), is calculated as follows:

\[
Q_i = \sum_{k=1}^{K} \sum_{j=1}^{J} e_{i,k}
\]

(15)

Note that the sum of squares of the residuals for post batch analysis is referred to as \( Q \). The \( Q \)-statistic can be approximated by a weighted chi-square distribution \( g \cdot \chi^2(b) \) where \( g \) is the weight and \( b \) the degrees of freedom (J.E. Jackson & G.S. Mudholkar [24]). The UCL for the \( Q \)-statistic can be found by using this approximation. The \( Q \)-statistic for all 53 batches are given in Figure 16. From this figure, there are no peculiarities observed in the \( Q \)-statistic. Based on the results of the \( T \) and \( Q \)-plot, batch 52 and 53 are removed from the dataset. The data that remains is said to be statistically in-control. This can be verified by looking at the \( T \)-plot and \( Q \)-plot for the dataset where batch 52 and 53 are removed. Figure 17 and Figure 18 represent the scatter plot and the \( Q \)-statistic respectively. Note that now the scatter plot is a random shot implying that all batches have some variation around a mean trajectory. Also, the \( Q \) statistics do not show any abnormalities.

![Figure 17](image1.png)  
*Scatter plot after removing wrong batches.*

![Figure 18](image2.png)  
*Q-statistic after removing wrong batches.*

### 2.3.1.2 Model building: evolving models

Now that a set of NOC batches is obtained, a model can be constructed for on-line monitoring. The model is used to construct the control charts (SPE and D-chart) for monitoring completely new batches. Different models can be chosen as will be
described in Chapter 4.3. For illustration purposes, a time evolving model (Figure 19) is chosen (see Chapter 4.3).

This model is constructed as follows: at every time interval \( k \), a model is built. For the time evolving model, a matrix \( X_k \) (\( I \times Jk \)) is constructed from the \( k \) frontal slabs of \( X \). This is illustrated in Figure 19.

It can be seen that \( X_k \) is the matricized part of \( X \) until time interval \( k \). Obviously, the size of \( X_k \) increases with time. The evolving model gives \( K \) different model loadings \( P_k \) (\( Jk \times R \)) as can be seen from Figure 19. The computational power of modern computers is no impeding anymore to build these models. The reference distributions for the test statistics are formed by \( T_k \) (\( I \times R \)) and \( E_k \) (\( I \times Jk \)).

2.3.1.3 D and SPE-statistic

The D-statistic is based on the scores \( T_k \) that are calculated at every time interval. The D-statistic for batch \( i \) calculated at time interval \( k \) is given by:

\[
D_{i,k} = t_i^T S_k^{-1} t_i
\]  

(16)

Where \( S_k \) is the variance-covariance matrix of the scores \( T_k \). The D-statistic represents the Mahalanobis distance from the center of the model plane towards the projected sample in this plane.

The calculation of the SPE-statistic is not much different from the \( Q \)-statistic. The term \( SPE \) stands for sum of squared prediction error and is also based on the residuals. The \( Q \)-statistic was based on the residuals calculated for a finished batch.
This $SPE$-statistic is based on the sum of squared residuals of the current part (cp) of the calculated residuals. Suppose a batch evolved until 10 time intervals and the residuals for this batch are calculated using the evolving model. The residual part that is concerned with the measurement at $k = 10$ is taken out. This part is referred to as the current part (cp). From the current part, the sum of squared residuals is calculated and forms the $SPE$-statistic. Then, at time point eleven, the residuals are calculated again and the current part is removed and the $SPE$ is calculated again. Thus, the $SPE$ for batch $I$ at time interval $k$ is calculated as follows:

$$SPE_{I,k} = e_i^e e_i$$  \hspace{1cm} (17)$$

The $SPE$-statistic represents a part of the Euclidian distance from the perpendicular projection of the sample to the model plane. The reason for this is that in this way instantaneous deviations in the residual space are detected. If the entire residual matrix would be used, deviations in the residuals might not be detected or much later because of the contributions from previous observations (averaging out effect).

Once the test statistics are calculated from the NOC data, the control charts can be constructed. These control charts are being used for monitoring completely new batches in the application phase. The control limits for the $D$-chart are found according to equation 9 with $f$ is replaced by $R$. Like the $Q$-statistic, the $SPE$ follows a weighted chi-square distribution $g \cdot \chi^2(h)$. The residuals are fitted to this chi-square distribution where the weight $g$ and degrees of freedom $h$ are optimised. From this distribution, the UCL can be obtained. The control limits of the $SPE$-chart always vary over time, since at every time interval the parameters of the chi-square distribution are estimated using different residuals.

### 2.4 A: Application phase

In the last phase of the I.T.A.-trajectory, new batches are monitored denoted as $x_{new}$. These batches are independent from the NOC data. Preferably, these new batches are operated under the same conditions as the NOC data. If that is the case, the variation in the new batches is comparable to the NOC data. However, if the variation suddenly changes because of a process fault, fast detection of this process upset is
desired. In the following, it is explained how new batches are monitored in the $D$ and $SPE$-chart that were constructed in the training phase.

### 2.4.1 On-line Monitoring

It is explained in the training phase how NOC batches can be described by latent variables (scores and residuals) in a dimension-reduced space. This dimension-reduced space is defined by the model loadings $P$. Since a time evolving model is used, this dimension-reduced space differs at every time interval ($P_k$). From the residuals and scores of the NOC batches, the $SPE$ and $D$-control charts are constructed. In order to compare a new batch $x_{\text{new}}$ with the NOC batches in the dimension reduced space, the measurements up until time interval $k$ ($x_{k,\text{new}}$) are projected onto this space. Once the residuals and scores for the new batch are calculated, the $SPE$ and $D$-statistic can be calculated. The calculation of the $SPE$ and $D$-statistic using a time evolving model reads as follows:

$$t_{k,\text{new}} = (P_k^t P_k)^{-1} P_k^t x_{k,\text{new}} = P_k^t x_{k,\text{new}} \quad \rightarrow \quad D = t_{\text{new}}^t S_k^t t_{\text{new}}$$

(because $P_k^t P_k = I$ by construction)

$$e_{k,\text{new}} = x_{k,\text{new}} - P_k t_{k,\text{new}} \quad \rightarrow \quad SPE = e_{k,\text{new}}^t e_{k,\text{new}}$$

These statistics then are plotted in the corresponding control charts. If the test statistics do not exceed the control limit, the process is said to be in-control. If this is not the case, further action is needed to investigate what caused the test statistic to exceed the control limit.

#### 2.4.1.1 Fault detection and Fault Diagnosis

In process monitoring, an important question is: how fast is a process fault detected in the control charts? The speed of detection is dependent of the quality of the model and the historical data. If fault detection is the most important objective of batch monitoring, it does not really matter if the fault is detected in the $SPE$ or $D$-chart. The detection power of a model is often tested by projection of a faulty batch. For such a tracer batch, the time of the process fault is exactly known. Then, this batch is
monitored and the time it takes to detect the fault is examined. The Action Signal Time (AST) is used to quantify the detection power of the model to pinpoint fault batches. The AST is defined as the time between the introduction of an error and the out-of-control signal.

Next, if a fault is detected, the cause is searched for. This is referred to as fault diagnosis. In statistical batch process monitoring, fault detection plays an important role. The test statistic itself only provides information whether the new batch is statistically in or out of control. However, several techniques exist to investigate the physical nature of the fault that causes the control chart to signal. First, the $SPE$-chart signals faults of a different nature compared to the $D$-chart. Secondly, the contribution of the individual process variables to the $D$ and $SPE$-statistic can be traced back. This will be discussed in the following.

2.4.1.1 Complementarity of SPE and D-chart
The result of the projection of a sample is expressed in terms of the $SPE$ and $D$-statistic. The question therefore is: what kind of relation is there between a real process fault and the result of this fault in the $SPE$ and $D$-statistic? To answer this question, the process measurements are categorised in two groups: i) faults that break the correlation structure or ii) faults that obey the correlation structure but have a more than normal variation. As an example, think of the following situation. The temperature ($T_{\text{pm}}$) and pressure ($P_{\text{pm}}$) are measured in the inner tube of a car tyre, as well as the ambient temperature ($T_a$). Now, as $T_a$ increases because of the sunny weather, so will $T_{\text{pm}}$ and $P_{\text{pm}}$. Thus, the process variables are correlated and behave according to simple physical laws. This is depicted in Figure 20.
Chapter 2 - I.T.A.-trajectory

The white bars represent the point in time where the process behaves under normal conditions. The black bars show the reaction of $T$ and $P$ when $T$ is changing. This is according to the correlation between the process variables. Besides, the variation of the increment of the process variables is considered as acceptable.

The first group of faults is the following. In extreme situations where the surrounding temperature may become very high (e.g. the car has been parked in the burning sun), the pressure and temperature in the inner tube might become dangerously high. Although the process variables behave according to the law of physics, the variation of the process variables is more intense than under normal circumstances. This is illustrated in Figure 20 as the patterned black bar. Such a fault where the variation of the process variables is abnormal high but the correlation between the process variables remains intact will be denoted as intensified correlation for the remaining of this thesis. The other group of faults is of the following. Suppose the temperature in the inner tube $T_{in}$ increases as a result of $T_{ex}$, but because the valve of the car tube is leaking, the pressure $P_{in}$ remains constant. This is not according to the correlation between the process variables, and therefore the correlation is broken. This can be seen from Figure 21. Such disturbances will be denoted as breakage of the correlation.

Often heard remarks in the literature (R. Dunia et al. [51], S. Albert & R.D. Kinley [43], J.V. Kresta et al. [10], P. Nomikos & J.F. MacGregor [6]) are that abnormal variation that still obeys the correlation structure of the process variables is described by the scores (D-statistic) while new events not present in the NOC data will represent itself in the residuals (SPE-statistic).
It is believed that an intensified correlation (e.g. process shifts) is detected in the D-statistic while the breakage of the correlation (e.g. sensor failure) is detected in the SPE-statistic and that both charts are therefore complementary. However, this is a simplification of how a process fault manifests itself in the control charts as will be discussed later (Section 0).

2.4.1.1.2 Contribution plots

An alarm signal in the control charts will tell the user that the process is no longer operating under the specified operating conditions. However, it is not only important to detect that the process is deviating, it is also important to search for the process variables responsible for the alarm and to determine the cause. One of the tools to search for the responsible process variables are contribution plots (P. Miller et al. [16], J.A. Westerhuis et al. [17]). Contribution plots compute the contribution of a single process variable to the monitoring statistics. The process variables with the largest contribution are responsible for the out of control signal and should be analysed further to determine the cause.

The contribution is computed for both the SPE-chart and D-chart. The contribution of process variable $j$ to the SPE is computed by

$$c_{j,k}^{SPE} = c_{jk,new}^2$$

which is simply the squared residual of process variable $j$ at observation $k$. The contribution $c$ of the process variable $j$ to the D-statistic is given by

$$c_{j,k}^{D} = \sum_{j=1}^{IK} t^T S_k^{-1} \left[ x_{jk}^T P_{jk} \left( P_{k}^T P_{k} \right)^{-1} \right]^T$$

(19)

(20)
where \( p_{jk} \) is the \( jk \)-th row of model vector \( P_k \) and \( x_{jk} \) is the \( j \)-th element of the observation vector \( x_k \).

Once the contributions are computed, the responsible process variables are examined and the cause of disturbance is explored and if possible removed.

2.4.1.1.3 Example: monitoring a faulty batch with impurity of the feed
In this example, a new batch is monitored for the SBR process of which it is known that a feed impurity occurred halfway the batch run. The control charts for this batch are given in the following figure:

![SPE and D-chart of a faulty batch](image)

The first deviations are observed in the \( SPE \)-chart around \( k = 15 \) and \( k = 35 \). There, a few observations cross the 95 % confidence limits. However, a clear out of control signal is given starting around time interval \( k = 100 \). The fault is well detected in the \( SPE \)-chart. The AST (95% confidence limit) for the \( SPE \)-chart is 6 observation units.
That is, after six observations after the introduction of the impurity an out of control signal is given.

A contribution plot with an approximate confidence limit for the SPE-chart at time interval \( k = 106 \) (marked with the circle) is given. The contributions reveal that process variable four, five and six and nine have a high contribution to the SPE-statistic (see Figure 23).

![Figure 23](image.png)

Contribution plot: SPE-chart.

The process variables four, five and six represent the temperature measurement inside the reactor vessel and process variable nine the energy release. In fact, around time interval hundred an impurity of 50% above the normal level enters the reactor. As a result, unwanted side reactions add extra heat to the reaction. As a result, the instantaneous rate of energy release (process variable nine) is much higher as expected. This example shows how process upsets can be detected and how contribution plots can help to find the cause of the process fault.

### 2.4.1.2 Process improvement

Once a process upset is detected and diagnosed, the process operator can take further action. It was already explained that process control of batch processes is difficult. However, if it turned out that a sensor failure occurred somewhere along the batch, it is not necessary to adjust the batch since only the equipment is malfunctioning. Sometimes the feed for semi batch processes can be adjusted, if possible, to get the process within specifications again. When no control actions are possible, it can be decided to terminate the batch and start a new one.
CHAPTER 3  ♦ PREPROCESSING*

* This chapter is based on the following publication(s):

3.1 General introduction*

An important step in process modelling is the step prior to model building which is the preprocessing of the data. Preprocessing of the data is often a necessity in order to construct sensible models. The two most applied actions are centering and scaling.

Centering is in principle done if there is a common offset present in the data. In batch process monitoring the three-way array is column or tube centered which removes the average batch trajectory from the data. By removing the average batch trajectory also the main non-linear behaviour is removed from the data, which allows the use of linear models such as PCA.

Scaling is performed to make the scale between different process variables the same. There are different ways to scale process data. For a more detailed discussion on centering and scaling see R. Bro & A.K. Smilde [52]. Another issue in batch process monitoring is the batch run length. It is assumed that the batches are operated in an identical manner and therefore have the same run length. However, in practice this is not always the true. This is a problem since some statistical batch process monitoring models are not applicable for batches with unequal length. Furthermore, even if batches have equal length, the local time may differ. That is, not all the events happen exactly at the same time for all the batches. For these reasons, synchronization of batch trajectories is required. There are some suggestions to deal with this problem (e.g. maturity variables, interpolation). An alternative is the use of dynamic time warping which originates from the speech recognition. In the following, the concept of dynamic time warping is explained and improvements of the algorithm are given.

This chapter introduces the concept of dynamic time warping. This is applied to a spectra batch data set that requires a somewhat different approach. Improvements to the existing algorithm are proposed.

*HJR/EVS

-52-
3.2 Dynamic time warping of spectroscopic batch data*

3.2.1 SUMMARY
This section discusses a method for warping spectral batch data. This method is a modification of a procedure proposed by A. Kassidas et al. [35]. This iterative procedure is based on the dynamic time warping (DTW) algorithm. The symmetric DTW algorithm is discussed in this section. Kassidas defined a certain weight that is received by every process variable in the DTW algorithm. However, high weights are received by process variables that contain no warping information. Therefore, a new definition of these weights is presented. These new weights take into account the amount of warping information of every process variable. The DTW algorithm using the new weights is compared to the procedure suggested by Kassidas. Furthermore, some aspects of this algorithm are optimized for speech recognition, but seem to be not necessary for warping batches. This concerns the normalization of the distance function. This step can therefore be omitted for warping batch data.

3.2.2 INTRODUCTION
Batch processes are very common in chemical, pharmaceutical, food industry and biochemistry. Monitoring these batch processes is wanted for several reasons such as safety, waste-stream reduction, consistency, quality improvement or improved process knowledge. One of the methods for batch process monitoring is based upon multivariate techniques and was introduced by P. Nomikos & J.F. MacGregor [6]. This technique is referred to as statistical batch process monitoring. Many applications and extensions of statistical batch process monitoring have been introduced since.

Most common batch processes are equipped with sensors that measure engineering variables like e.g. temperatures, pressures and flow rates. However, modern process analyzers, like spectroscopic measurement devices, find their way into

Statistical batch process monitoring

batch monitoring. The main advantage of these apparatus is that the measurements contain chemically richer information compared to engineering variables. For statistical batch process monitoring a database is needed with completed batch runs that produced on-spec products. The variation within these data serves as a reference distribution. The performance of independent new batches is compared with this reference distribution. This is achieved using multivariate control charts. These control charts are normally based on PCA or PLS models. For some models it is required that the reference batches have equal duration. However, in practice this is almost never true. Even if batches have an equal runlength, the local time may be different. That is, similar events happen at different time points for e.g. two batches. In such case, the local time of the batches is different and synchronization of similar events is necessary.

The problem of batches with unequal lengths can be overcome by the choice of model (S. Wold et al. [53]). However, the models capable of dealing with batches of unequal length suffer from poor statistics. Another way to deal with this issue is to make use of dynamic time warping (DTW). DTW originates from the world of speech recognition and A. Kassidas et al. [36] proposed a method based on the DTW algorithm that is focused on statistical batch process monitoring. This method was based on batch data that contained engineering variables. Other chemically related warping problems can be found in the papers written by V. Pravdova et al. [54], K. Gollmer & C. Posten [55] and A. Kassidas et al. [35]. Another existing warping technique that can be used for e.g. chromatographic profiles is correlation optimised warping (N.-P.V. Nielsen et al. [56], D. Bylund et al. [57]).

In this work, data from an industrial batch reactor in which a resin is produced, is used. This batch reaction is spectroscopically monitored with NIR. Such NIR datasets are well suited for statistical batch process monitoring. In this section, prior to monitoring, the batches need to be warped because they have unequal length. The warping is based on a method proposed by A. Kassidas et al. [35]. In this section, an explanation of the DTW algorithm is given. Also a strategy is proposed how to warp spectroscopic data since this is not yet discussed in the research area of statistical batch process monitoring. Furthermore, an improved method is presented for selecting the weights matrix needed for warping is presented.

Section 3.2.3 discusses the DTW algorithm. Section 3.2.4 discusses some important added features to the DTW algorithm and explains an alternative method
to calculate weights that are given to each process variable in the DTW algorithm for warping batches. Section 3.2.5 describes the procedure how to warp batches as proposed by Kassidas. Section 3.2.6 describes the industrial example, which is studied. Then, Section 3.2.7 discusses the warping results of the new weights compared to the weights of Kassidas. Also, the effect of normalizing the DTW algorithm is illustrated with an example. Finally, conclusions are given in the Section 3.2.8.

3.2.3 THEORY
There are two problems associated with modeling data profiles of unequal lengths. The first problem is simply the unequal length, meaning that two profiles of unequal length cannot be arranged as two rows in a matrix that serves as a starting point for model building. At first sight this problem can be solved by cutting of the data part of the ‘longest duration’ profile to match the size of the ‘shortest duration’ profile providing that both profiles can now be placed in two rows of a matrix. This points, however, to the second problem: the local time of the profiles is different, e.g., the profiles may have reached a different maturity (such as conversion) in the same time. Clearly, modeling such profiles, e.g. batch trajectories, by placing them in two rows is not appropriate. Synchronization of such profiles can be accomplished by using the dynamic time warping (DTW) algorithm. DTW originates from the field of speech recognition. The DTW algorithm needs to meet some important demands when applied to synchronization of such profiles. First, after warping the profiles need to have the same length. Secondly, certain events that happened in the profiles should occur at the same time. In the following, the DTW algorithm will be explained.

3.2.3.1 General explanation of DTW
In the following a general method for synchronizing all kinds of trajectories is presented. This method is called DTW and is based on a technique named dynamic programming. First, some terminology for DTW is explained. Then, the problem of finding an optimal path through a grid will be introduced. The solution to this problem by using DTW is given and illustrated with a numerical example.
3.2.3.1.1 The Sample and the Reference

The trajectory of the reference vector will be denoted by \( r(J \times 1) \). Ideally, the reference vector is a sample from the dataset that contains all the important characteristics of the data. Therefore, the choice for the reference vector is important. The characteristics in a new sample, denoted by \( s(I \times 1) \), will be synchronized with the reference vector. An example of two trajectories of unequal length is given in Figure 24.

![Two temperature profiles of unequal length.](image)

These trajectories represent the measurements of a temperature recorded during a batch run. It can be seen that these trajectories are different in length \( (I \neq J) \). The objective is to synchronize these trajectories to equal length \( K \).

3.2.3.1.2 The grid

The sample and the reference form the base of a grid. This grid consists of an x and y-axis. The length of \( r \) and \( s \) gives the dimensions of the grid. The length of the x-axis is given by the sample \( s \) and the length of the y-axis is given by the reference \( r \). This has been illustrated in Figure 25.

![Construction of the grid.](image)

The cell represents a grid point (see Figure 25). A grid point is denoted as \( w(k) \) where \( k \) is a running index for a new common axis of the sample and reference.

A sequence of grid points will be referred to as the trajectory or path \( w \). The path \( w \) is denoted as:
Chapter 3 ♦ Preprocessing

\[ w = \{ w(1), w(2), \ldots w(k), w(k+1), \ldots w(K) \}, \quad \max(I, J) \leq K \leq I + J \]  

where the number of grid points in the path is given by \( K \). The path \( w \) is also referred to as the warping function. The starting and ending point of \( w \) are defined: the starting point \( w(k=1) \) of this trajectory is given by \( (i=1, j=1) \) where the ending point \( w(k=K) \) of this trajectory is given by \( (i=I, j=J) \). For any point along the path, \( w(k) \) can be written as \( w(k) = (i(k), j(k)) \). An example of such a path is given in Figure 26.

![Figure 26](image)

*Construction of a path through the grid.*

Notice that many paths can be constructed that connect the begin point and end point of the grid. Each path may have a different pathlength \( K \). Furthermore, the size of \( K \) cannot be determined a priori.

3.2.3.2 Local constraints

The construction of a path \( w \) is subjected to local constraints. Consider the grid point \( w(k) \) illustrated as the black dot in Figure 27:

![Figure 27](image)

*Representation of the local constraints.*

This grid point \( (i, j) \) is a part of a trajectory \( w \). The predecessor \( w(k-1) \) of this grid point \( (i, j) \) can be any of the gray dots. That is:

\[ w(k-1) = (i-1, j), (i-1, j-1) \text{ or } (i, j-1) \]
Statistical batch process monitoring

Going from grid point \((i, j)\) to \((i+1, j+1)\) on some path, either \(i+1 > i\) or \(j+1 > j\). This guarantees that the path cannot go 'back in time' and therefore the slope between to consecutive grid points lying on a path is always \(\geq 0\). This is a requirement of the DTW algorithm. Equation 22 shows that the predecessor of grid point \(w(k)\) can be chosen from a set of three.

In a symmetric DTW algorithm, the outcome of a path will not be different if the sample and reference are interchanged because the reference and sample are treated equally. This is not true for an asymmetric algorithm. In an asymmetric algorithm vertical steps are not allowed so \((i, j-1)\) is not a possible predecessor of \((i, j)\). In this way, data points from the reference could be omitted allowing time scale compression of the pattern. This can be favorable for some applications where it is not necessary to assume that the best path includes all the data points of both the sample and reference (e.g. slow changes of the pattern compared to the time interval). However, for e.g. batch processes all the data should be taken into consideration by the algorithm, thus the symmetric DTW algorithm is used.

### 3.2.4 DTW ALGORITHM

As explained before, many paths can be constructed. Each path is characterized by a distance. The cumulative distance \(D(i, j)\) from the beginning to point \((i, j)\) along a path is given by:

\[
D(i(l), j(l)) = \sum_{l=1}^{L} d(i(l), j(l)) = \sum_{l=1}^{L} d(w(l))
\]

(23)

where \(d(w(l))\) is some distance measure between two points of the sample and reference. Here, the running index \(l\) denotes that the path does not connect the beginning point of the grid with the end-point. If this is true, the running index of the path is denoted with the index \(k\). The distance measure between the sample and reference at grid point \((i, j)\) is given by:

\[
d(w(l)) = \left( s(j(l)) - r(i(l)) \right)^2
\]

(24)
Chapter 3 • Preprocessing

The goal of DTW is to find the optimal path \( w \) (that connects the beginning and ending point of the grid) with length \( K \) that minimizes the function \( D(i, j) \) in Equation 23:

\[
D(I, J) = \min_w \left[ \sum_{k=1}^{K} d(w(k)) \right]
\] (25)

One way of finding the path that solves equation 25 is as follows. Construct all possible paths through the grid, calculate the cumulated distance for each path and select the path that gives the smallest outcome. However, already for relative small grids the number of possible paths is very large. The computational load to solve this problem on an average computer is far too high. A very elegant solution to this problem is dynamic programming. Dynamic programming is a mathematical technique, which guarantees to find the optimum path without having to calculate the distance along all possible paths. It is based on the Markov Chain property. In the following, the construction of the path that gives the minimum accumulative distance is explained. A numerical example is used to illustrate dynamic programming.

**Step 1: Initializing the grid**

Every grid point \((i, j)\) connects the measurements of the sample \(s(j)\) and reference \(r(i)\). For every grid point the distance between \(s(j)\) and \(r(i)\) is calculated according to equation 24. This is illustrated in Figure 28:
Statistical batch process monitoring

The values chosen for s and r are arbitrary and serve only as a numerical example. It can be seen from Figure 28 that the grid has been filled with the local distance measures.

**Step 2: Calculate the cumulated distances**

The next step is to fill the complete grid with the cumulated distances. These distances are calculated in a column or row-wise manner but this is an arbitrary choice. Here, it is chosen to fill the grid in a column-wise manner starting at the point \((i=1, j=1)\). The next grid point therefore is \((1,2)\). For this grid point the predecessors must be determined according to the local constraints illustrated in Figure 27. It can be seen that from the three predecessors, two steps are not valid (the diagonal step and the horizontal step) since they fall outside the grid. Thus the only predecessor can be the grid point \((1,1)\). Therefore the best cumulated distance along the path from the starting point \((1,1)\) to \((1,2)\) is given by:

\[
D(1,2) = D(1,1) + d(1,2) = 49 + 25 = 74
\]

\[
(26)
\]
The next grid point is given by $(1,3)$. For the same reasons as grid point $(1,2)$ two preceding steps are cancelled. The best cumulated distance along the path from $(1,1)$ to $(1,3)$ is now given by:

$$D(1,3) = D(1,2) + d(1,3) = 74 + 16 = 90$$

This is repeated until the last grid point of the first column is reached. The same exercise can be done for the first row of the grid. In this case, the predecessor of the grid point in the first row can only originate from a horizontal step.

Once the best cumulated distances are calculated for the first column and first row, the next grid point to consider is $(2,2)$. In this example it is chosen to fill the remaining of the grid in a column wise matter. It is up to the user to do this either column or row wise. This situation is sketched in Figure 29.

---

**Reference**

*Calculating the accumulated distance for every grid point.*

The bold numbers in the first row and column represent the best cumulated distances that are already calculated. The grid point $(2,2)$ is black circled and the possible
predecessor grid points are dotted circled. The best cumulated distance along the path from the starting point to \((2,2)\) is given by:

\[
D(2,2) = \min \left\{ \begin{array}{l}
D(1,2) + d(2,2) \\
D(1,1) + d(2,2) \\
D(2,1) + d(2,2)
\end{array} \right. = \min \left[ \begin{array}{c}
74 + 100 \\
49 + 100 \\
193 + 100
\end{array} \right] = 49 + 100 = 149
\] (28)

This exercise is repeated for every grid point and in this manner the complete grid can be filled with the best cumulated distances.

An important property of the DTW algorithm is the Markov chain property. This property reads as follows. All the information until grid point \(w(k)\) is enclosed in \(w(k)\). Therefore, to go the next step, in this case \(w(k+1)\), it is only necessary to consider \(w(k)\).

**Step 3: Finding the optimal path with backtracking**

Once the grid is filled with cumulated distances, the optimal path from the begin point to the end point can be derived. This is done in a back wise manner starting from the point \(D(1,J)\). From this point the preceding points are considered. The point that gives the lowest cumulated distance is selected. From this selected point, the next three preceding points are considered and the one that gives the smallest accumulative distance is selected etc. This is repeated until the grid point \(D(1,1)\) is reached. It might happen in backtracking that preceding grid points have the same value for the cumulated distance. It is necessary to consider the predecessors of those grid points in order to find the solution. If diagonal steps are preferred, the choice can also be based on this criterion.

The selected points form the optimal path \(w\) through the grid that minimizes the distance function from Equation 25. The optimal path is illustrated in Figure 30 as the marked numbers.
This path forms the joint time axis for the reference and sample. The accompanying values for both the sample and reference can be easily found. Forward tracking of the optimal path does not give the same results as backtracking. For example, the first 5 grid points that result from forward tracking are given by the sequence of \{ (1,1),(1,2),(1,3),(1,4),(1,5) \} (see Figure 30). These grid points are not the same as the last five grid points from the path that resulted from backward tracking that are given by \{ (1,1),(1,2),(1,3),(1,4),(2,5) \}. Also, by forward tracking to find the optimal path in Figure 30, the grid point \((I,J)\) is not reached.
The warping result of the two temperature trajectories is given in Figure 31.

![Warping result of two temperature profiles.](image)

It can be seen from this figure that the joint time axes $K \neq I \neq J$. Each grid point from the optimal path $w$ represents a pair of data points from the reference and sample. These data points are taken from the sample and reference one by one. In this way the warped sample and reference can be constructed and plotted as in Figure 31. Also, the warping result for these temperature profiles would have been the same if the sample and reference were interchanged. This is a property of the symmetric DTW algorithm. Also, the joint time $K$ axis cannot be expressed in seconds anymore. The joint time axis now represents a dimensionless non-linear time scale.

### 3.2.4.1 Remaining topics

This Section describes additional features that complete the DTW algorithm. This complete DTW algorithm is used in the ongoing of this section when applied to statistical batch process monitoring.

#### 3.2.4.1.1 Global constraints

Global constraints or adjustment window conditions exclude certain regions of the grid in which the optimal path can lie. This may save computation time that is important for online applications. Also unrealistic warps can be omitted. Global constraints can also follow from certain choices of the local constraints. This is not the case for the choice of local constraints in this section. An illustration of the global constraint is given in Figure 32:
The grid points marked as NaN are no part of the search in the grid anymore. This global constraint is chosen in such a way that $|i(k) - j(k)| \leq r$ where $r$ can be chosen by the user.

**3.2.4.1.2 Normalization**

This Section discusses the possibility to normalize the minimization problem as formulated in Equation 25. This minimization problem minimizes two aspects: the cumulative distance and the number of steps. A path that follows many horizontal and vertical steps contains more steps compared to a path with many diagonal steps. As a result, the number of contributions of $d(i,j)$ to $D(I,J)$ is dependent of the path length. Hence, diagonal paths will be favored. Although this does not necessarily have to be a problem, there are situations where paths that deviate from the diagonal are also important to consider, e.g. patterns with the same length but different time scaling. Making the selection of the path independent of the number of steps can be done in two ways.
The first way is described by J.N. Holmes [58] and is the easiest to understand. When a diagonal step is taken, twice the value of $d(i, j)$ is added. This can be understood from Figure 27. If the grid point $(i, j)$ is reached by a vertical step from grid point $(i - 1, j - 1)$ followed by a horizontal step (or vice versa), two times a local distance $(d(i, j) + d(i - 1, j))$ is added since two steps are taken. Going to $(i, j)$ starting from $(i - 1, j - 1)$ by a diagonal step requires only one step. To compensate for this, twice the value of $d(i, j)$ is added for the diagonal step. The calculation of the cumulative distance from Equation 28 now becomes:

$$D(2, 2) = \min \begin{cases} D(1, 2) + 1 \cdot d(2, 2) \\ D(1, 1) + 2 \cdot d(2, 2) \\ D(2, 1) + 1 \cdot d(2, 2) \end{cases} \quad (29)$$

Concluding, the cumulative distance function can be made independent from the number of steps by the following equation:

$$D(i(l), j(l)) = \min \begin{cases} D(i - 1, j) + d(i, j) \\ D(i - 1, j - 1) + 2 \cdot d(i, j) \\ D(i, j - 1) + d(i, j) \end{cases} \quad (30)$$

The second way to make the selection of the path independent of the number of steps is described by H. Sakoe & S. Chiba [59]. The cumulative distance along a path from the beginning of trajectory to point $(i, j)$ can be defined by:

$$D(i(l), j(l)) = \sum_{i=1}^{L} d(i(l), j(l)) \cdot \nu(l) = \sum_{i=1}^{L} d(w(l)) \cdot \nu(l) \quad (31)$$

where $\nu(l)$ is a so-called nonnegative weighting coefficient or normalization weight. The time normalized distance between a sample and reference is given by:

$$D(I, J) = \min_w \left[ \sum_{k=1}^{K} d(i(k), j(k)) \cdot \nu(k) \over \sum_{k=1}^{K} \nu(k) \right] \quad (32)$$

-66-
where the denominator is employed to compensate for the effect of steps $K$. The characteristics of the fundamental definition in Equation 32 of the time-normalized distance depend on the choice of the normalization weights $v(k)$. The minimization of Equation 32 is a difficult problem. If the denominator from Equation 32 is independent of the warping function, it can be put outside of the brackets:

\[
D(I, J) = \frac{1}{K} \min_v\left[ \sum_{k=1}^{K} d(i(k), j(k)) \cdot v(k) \right]
\]  \hspace{1cm} (33)

One way of obtaining Equation 33 is by choosing $v(k) = 1$ for a horizontal or vertical step and $v(k) = 2$ for a diagonal step. In this case, the term outside the brackets becomes $I + J$ and is indeed independent of the number of steps. Although both methods normalize the distance function, it is done in different ways.

### 3.2.4.1.3 Normalization for synchronization of batch runs

With speech recognition, a spoken word is matched with a library of words. This library of words has certain characteristics. If the pronunciation of the word by the speaker is somewhat slow in the beginning but fast in the end, the characteristics of the library words are different from the spoken word. In such cases, normalization makes the matching independent of the pronunciation of a person. However, the objective of warping batch processes is different. A pair of batches is fixed and there is no library to match. Therefore, the objective for normalizing the DTW algorithm is missing in case of batch monitoring and can therefore be omitted.

### 3.2.4.1.4 Direction penalties

By using direction penalties (J.N. Holmes [58]), the path can be forced to follow the diagonal, or not. That is, vertical or horizontal steps are given some kind of weight. These direction penalties are not to be confused with the normalization weights. A vertical direction penalty is denoted as $vp$ and a horizontal direction penalty is denoted as $hp$. The larger these penalties are chosen, the more favored the diagonal path becomes. The cumulative distance is calculated according to:
\[
\begin{aligned}
D(i(l), j(l)) &= \min \left[ D(i-1, j) + d(i, j) + hdp \\
& \quad D(i-1, j-1) + d(i, j) \\
& \quad D(i, j-1) + d(i, j) + vdp \right]
\end{aligned}
\]

The values of the penalties should be chosen by experiment and their magnitude should be in the same range as the distance measure.

### 3.2.4.1.5 Multivariate trajectories

DTW has been explained for single trajectories so far. The extension to multivariate trajectories is straightforward and needs only minor changes in the algorithm. For multivariate trajectories, the sample \( S(I \times N) \) represents a matrix as well as \( R(J \times N) \) where \( N \) is the number of variables. In the DTW algorithm only the calculation of the distance measure is different for multivariate trajectories and is calculated as:

\[
d(i(k), j(k)) = \left[ s_{i(k)} - r_{j(k)} \right]^T Z \left[ s_{i(k)} - r_{j(k)} \right]
\]

where \( Z \) is a positive definite matrix. If \( Z \) is chosen to be identity, the Euclidean distance is taken. The weight matrix \( Z \) can be used in the DTW algorithm to give more weight to certain variables.

### 3.2.4.1.6 Calculation of the weight matrix by Kassidas

Kassidas suggested to use \( Z \) to give more weight to variables that are ‘consistent’ from batch to batch, e.g., a steadily increasing value of the measured conversion is more consistent than a noisy measurement of the level of the reactor. The warping of batches should rely more on these consistent variables. One choice is to fix the weights for every process variable. However, this requires extensive knowledge about the process that is not always available. Kassidas developed a method to increase or decrease the weights for the variables in an iterative manner. That is, the weight matrix is initialized and all the batches are warped and synchronized. From the synchronized trajectories and the reference, the average trajectory is calculated. For the first batch and first process variable, the difference with the average trajectory is calculated. These differences are squared and summed over time resulting in a number. This number is calculated for every batch. The reciprocal sum of all these
numbers represents the weight that is given to a particular process variable. This weight is calculated for every process variable in the same manner. Then, the weight matrix is updated and the whole procedure (warping and synchronizing) is repeated a number of times. The details of this iterative procedure can be found in A. Kassidas et al. [35].

If the trajectory of a process variable deviates not much from the average trajectory, a high weight is received. Because of this, the dynamics of the signal over time are ignored. It does not matter if the trajectory of the process variable over time is curved-like or nearly horizontal, as long as the trajectories are close to the average trajectory.

Kassidas already suggested using a convergence criterion for \( Z \) to stop the iterative procedure although no suggestion of what type of criterion was given. In this section, a convergence criterion of \( Z \) is used to decide whether a next iteration is needed using the definition of the weights by Kassidas. The convergence criterion reads as:

\[
\frac{|\text{diag}(Z)|_u - |\text{diag}(Z)|_{u-1}}{|\text{diag}(Z)|_{u-1}} < 0.002
\]

where \( u \) is the number of iterations. Iteration of \( Z \) more than the arbitrary value 0.002 does not significantly improve the result further. The converged matrix \( Z \) gives an indication of which process variables are consistent and useful in the DTW algorithm.

### 3.2.4.1.7 New definition of the weights

It is explained in the foregoing what characteristics are needed for the trajectory of a process variable to receive a high weight. An important question is how this affects the DTW algorithm. Consider the warping of two horizontal lines that have the same value but different lengths. The optimal path can lie everywhere in the grid and the lines will always be perfectly warped. In other words, such a process variable contains no warping information. In the definition of the weights by Kassidas such process variables are given a high weight. As a result, the outcome of the optimal path is dominated by these variables. It was explained that such a path could be more or less ‘random’. This is disadvantageous if the multivariate data also consists of e.g. curved-like process variables, e.g. kinetic profiles. Such profiles are sensitive to incorrect
warping. In other words, these process variables contain important warping information.

In this chapter, an alternative definition for $Z$ is presented. These weights account for the ‘warping information’ of each process variable. It can be seen from Equation 35 that every process variable has its own contribution to the local distance. That is, for every process variable a single grid can be constructed with local distances. The contour plot of such a single grid for a process variable that contains important warping information is expected to show a clear valley that runs from the start to the end point of the grid. The valley refers to grid points representing small distances. It is preferred that these process variables with sharp valleys will be given a high weight. The shape of the optimal path should follow these valleys as much as possible. A similar surface plot for the horizontal profiles is expected to be rather flat since the local distances are similar. Process variables corresponding to such a surface plot should be given a low weight. The calculation of $Z$ is performed as follows:

- The initial values for the diagonal elements of $W$ are set to one.
- The first batch $X_{i=1}$ is warped on $X_{n_f}$. The warping results in a path $w_{i=1}$ with length $K$. The coordinates of the unsynchronized path are required to update $Z_{c=1}$ according to the new definition. The path is depicted in Fig 33a:

![Fig 33a](image)

*Fig 33a
Optimal path.*

- A grid is filled with the local distances calculated for the first process variable. The first variable from $X_{r=1}$ and $X_{n_f}$ is illustrated in Fig 33b. The surface plot of this grid is given in Fig 33c. Such a surface plot is constructed for all...
the process variables. Therefore, $N$ different surface plots can be constructed where each plot represents the local distances between the sample and reference for the $n$-th process variable.

For the first process variable ($n = 1$), the optimal path is mapped to the surface plot. This is illustrated in Fig 33c. Those grid points are selected that correspond to the coordinates from the optimal path $w_1$ that was found for batch $X_1$ and $X_{\text{ref.}}$. Let the mean of these elements be denoted as $MLD_{n=1}^{\text{opt path}}$ where $MLD$ refers to mean local distance. The mean of local distances for those grid points that are not on these path coordinates are denoted as $MLD_{n=1}^{\text{off path}}$. The first diagonal element of the weight matrix $Z_i$ is given as:

$$\zeta_{n=1,n=1} = \frac{MLD_{n=1}^{\text{off path}}}{MLD_{n=1}^{\text{opt path}}} \quad (37)$$

If the path follows the valley of the grid, the $MLD_{n=1}^{\text{opt path}}$ will be relatively small compared to $MLD_{n=1}^{\text{off path}}$. This results in a high weight.
The second diagonal element of $Z_{ii=1}$ is calculated using the surface plot that represents the local distance for the second process variable between the sample and reference etc.

![Path mapped on surface plot](image)

**Fig 33c**
*Optimal path mapped on surface plot of local distances.*

- This whole procedure is repeated for each batch $X_f$. The weight matrix $Z$ is taken as the average over $Z_0$. $Z$ is normalized in such a way that the sum of the diagonal elements equals the number of process variables.
- After the first iteration ($ii = 1$), the weight matrix is updated. The whole procedure to calculate the update of $Z$ is repeated and stopped when a convergence criterion is met. The convergence criterion of equation 36 is also used. The criterion of 0.002 is chosen since more iterations have negligible effects on the warping results.

Variables with a high weight reflect variables that contain important warping information. Variables with a low weight are not very sensitive to incorrect warping, such as horizontal lines.
3.2.5 DTW FOR BATCH MONITORING

The only DTW related work in the area of statistical batch process monitoring was done by A. Kassidas et al. [35]. Kassidas suggested an iterative procedure for warping unequal batch trajectories. This procedure was applied to an industrial polymerization dataset. The details of this algorithm can be found in A. Kassidas et al. [35].

In this section, the iterative procedure from Kassidas is also used for warping and synchronizing spectroscopic data including the usage of global constraints. However, some aspects of this procedure are changed. These issues are related to scaling, the weighting matrix \( Z \) and the number of iterations. Furthermore, the DTW algorithm is exploited without normalization or direction penalties. This is different from Kassidas where the distance function is normalized according to Equation 33.

The iterative procedure can be split into three steps: Step A, Step B and Step C. These steps will be discussed in the following.

Step A: Scaling
Before applying DTW, the batches need to be scaled properly. Let the sample \( S \) be represented as \( X_{ij} \) where \( i \) is the number of time intervals, \( N \) the number of process variables and \( \epsilon = 1, ..., C \) a running index for the number of batches. Kassidas suggested to calculate the average range of each process variable for one batch. This is repeated for every batch. From these average ranges for all batches the total average range is calculated. Then, each variable is divided by this total average range. Since spectroscopic data usually has the same measurement units of similar magnitude, the scaling step in the iterative procedure for DTW using spectroscopic data can be omitted.

Step B: Warping
In this step the DTW algorithm is applied. Two important issues of this step are highlighted. The first issue concerns the calculation of the weight matrix \( Z \). This is already explained in the previous section. The second issue is the choice of the reference batch \( R = X_{nm} \) \( (J \times N) \). Before the start of step B, one out of all batches is selected as the reference batch. This reference batch is fixed and all the other batches are warped one by one. Every warped batch is connected to the reference via an optimal path \( w_r \). It is important to realize that the length of \( w_r \) is different for all the warping results. Thus, every batch will still have a different length after warping.
Kassida used the batch, which was the closest to the average duration of all batches as the reference batch. It is recommended to check whether this batch is also representing reasonable normal operating conditions. An alternative way to select a reference batch is by selecting that batch which gives the best result in terms of the final monitoring. This can be done by considering all batches as reference batch and select the one which gives the fastest detection of faulty batches, e.g. using the AST (E.N.M. Van Sprang et al. [41]). This requires, however, a sufficient number of faulty batches to make this procedure reliable. In this case study, there was only one faulty batch available. Hence, the procedure of Kassida was used to select the reference batch.

Step C: Asymmetric synchronization

The batches have been warped with the reference batch in step B. This results in \((C-1)\) pairs of \(X_{nf}\) and \(X_r\). The length of each pair of batches will be different from each other. This will be taken care of in step C. Let Figure 34 represent the warping result of some pair \(X_r\) and \(X_{nf}\) after step B is finished:

It can be seen from this figure that the optimal path contains some consecutive horizontal steps marked with the black dots. This implies that several time points of batch \(X_r\) are aligned with only one time point from \(X_{nf}\). This might happen in situations where the process variables of \(X_r\) remain almost constant for a while and are aligned with \(X_{nf}\). The proposed method takes the average value for \(X_r\) over the consecutive horizontal steps. In this way, the length of the optimal path will equal...
the size of $X_{nf}$. If the black dots from Figure 34 are being replaced with a white dot representing the average value over that range, the length of the path is 10. This is exactly the length of $X_{nf}$. This is done for all the pairs of batches, and as a result every batch $X_i$ will become of the same length of the reference batch $X_{nf}$. This is called an asymmetric operation that synchronizes all $X_i$ in a way that all have the same duration as $X_{nf}$ (A. Kassidas et al. [35]).

For batches that have a shorter duration than $X_{nf}$, the procedure works the same. For this asymmetric operation, the average values must always be calculated in the time direction of $X_i$. This is important if the axes of the grid are being interchanged with $X_{nf}$ and $X_i$. Because of the asymmetric operation, the intensity signals of $X_i$ are averaged over a number of time points. This is an important disadvantage of the procedure since variation in the data is averaged. This variation serves as a reference distribution for statistical batch process monitoring.

### 3.2.6 Industrial Example: Batch Process of Urethane Resin

The data used in this section comprises an industrial multi-stage fed-batch process of making a urethane resin. A more detailed description can be found in the paper written by E.N.M. Van Sprang et al. [41]. Only the first stage of this process is used in this study. The reaction that takes place in the first stage reads as:

$$OCN - R_1 - NCO + HO - R_2 \rightarrow OCN - R_1 - NH - CO - O - R_2$$

The urethane resin is formed from the reaction of a di-isocyanate with an alcohol. The first stage of this process can be split into three steps:

1) the tank is filled with a di-isocyanate.
2) an alcohol (reactant) is added drop wise to start the reaction.
3) all the di-isocyanate has reacted and the process is kept at a constant temperature.

This reaction is carried out using a temperature profile while the reactants are added in precisely known stoichiometric ratios. A NIR spectrometer with a 2 nm resolution is used to record the spectra in the range of 1100 nm to 2500 nm. The regions of the
active compounds are as follows: di-isocyanate \( \sim 1910 - 1940 \, \text{nm} \), alcohol \( \sim 1398 - 1421 \, \text{nm} \) and product \( \sim 1485 - 1503 \, \text{nm} \). A total of 17 batches are measured. The duration \( t \) of each batch varies between 69 and 140 time intervals. These differences are mainly caused by change of the shift during the night. The beginning \( (1100 \, \text{nm} - 1300 \, \text{nm}) \) and ending \( (2100 \, \text{nm} - 2500 \, \text{nm}) \) of the spectrum are mainly instrumental noise and are for this reasons discarded. This results in 400 wavelengths recorded at every time interval. The raw spectra as well as the second derivate spectra for one of the batches has been plotted in Figure 35a and Figure 35b:

![Figure 35](image)

*a: Raw spectra of a batch.
b: Second derivative spectra of a batch showing absorbing regions.
c: Outcome of the weight defined by Kassidas.
d: Outcome of the weight according to new definition.

The second derivative spectra (using a Savitsky Golay filter) are computed because of offset problems with the spectra.
3.2.7 DTW FOR SPECTROSCOPIC DATA

The warped batches in the dataset can be used to build monitoring models. First, the results of applying the Kassidas procedure will be discussed. The average calculation time to warp two batches with where the length of \( J \) and \( I \) is approximately 100 and \( N = 400 \) takes about 2 minutes on a Athlon 1.25 GHz desktop. This calculation includes the computation of the new weights.

3.2.7.1 Weight matrix

As explained, the spectroscopic instrument records absorption at 400 wavelengths every point in time. It is interesting to examine the weights that the algorithm calculated. A suitable batch is chosen as the reference batch. The results of \( Z \) as defined by Kassidas after 10 iterations are presented in Figure 35c. In this figure, the weight given to every single wavelength can be read from the y-axis. The wavelengths where mostly the pure compounds dominate are bold. Clearly, it is not the active compounds that are given the highest weight. Especially the region of 1300 nm – 1350 nm is given the highest weight. This region corresponds with the baseline of the spectrum as can be seen from Figure 35a. It will be shown in the following that these regions contain no warping information and therefore should be given a low weight instead. Besides this issue, the weight matrix was not converging. This indicates that the weight matrix \( Z \) is troublesome and the values obtained after 10 iterations are not reliable.

To illustrate the problems of the Kassidas weights more in detail, the trajectory over time for the wavelengths 1316 and 2058 nm have been plotted for all the batches after warping and synchronization (Figure 36). The crosses represent the average trajectory for these two wavelengths.
Figure 36

*Warped time profiles of two wavelengths using Kassidas method.*

Figure 35c shows that the wavelength at 1316 nm is given the highest weight and the wavelength of 2058 nm a very small weight (marked as circles in Figure 35c). It is apparent from Figure 36 that the differences of the batches with the average trajectory are small for wavelength 1316. This can be understood since this wavelength represents the baseline. There is not much variation in this baseline. This is different from the wavelength at 2058 nm. Figure 36 illustrates that the deviation from the average trajectory is larger. Therefore, less weight is given to this wavelength.

The results of the new weight matrix are presented in Figure 35d. The new weights converged properly according to the convergence criterion. The time profile of two different wavelengths have been illustrated in Figure 37.
The wavelength at 1360 nm that corresponds to the baseline is given a small weight (see Figure 35d). This is not true for the wavelength at 1738 nm. This wavelength has a high signal to noise ratio. This feature is given a high weight as can be seen from Figure 35d. Clearly, the regions that correspond to the baseline are given a smaller weight as compared to weights defined by Kassidas. This is true for all the flat regions in the spectra that are now given a low weight. These regions contain no warping information. Figure 35d shows that the peaks of the weight-profile correspond to the tops and valleys of the second derivative spectra. The new weights therefore reflect the underlying nature of the data much better as compared to Figure 35c. Therefore, only the important parts of the spectra are considered by the weights.

Furthermore, Figure 35d indicates that not only the regions corresponding to the active compounds are given a high weight. Also other regions of spectra contain valuable warping information like the region of 2000 nm – 2100 nm. In this case it is not sensible to perform a wavelength selection based upon prior knowledge about the absorbance regions of the active compounds since important warping variables are discarded. This is in contradiction for calibration models, where temperature effects are unwanted and wavelength selection is preferable.

The warping results of the method by Kassidas and the method presented in this section is shown in Figure 38:

![Figure 37](image)

*Warped time profiles of two wavelengths using new method.*
The spectra over time for two batches have been warped. Not all the spectra over time are plotted (only $\lambda = 1920\text{nm} - 1923\text{nm}$) because this results in an unclear picture. The iterative procedure was repeated ten times for both approaches. It can be seen from this Figure that the definition of the new weights without the normalization step (Figure 38c) gives better warping result as compared to Kassidas (Figure 38b). Also, the new weights converged according to the criteria whereas the Kassidas weights did not.

The weight matrix is also tested for typical engineering datasets where temperatures, pressures etc. are measured. The outcome for the weight matrix gave good results. Those variables that are considered to be important are given the highest weight by the algorithm. Such variables are e.g. temperature set points or conversion measurements.
3.2.7.2 Normalizing

It was explained in the theory section that the normalization step originates from speech recognition applications and can be omitted for warping batch data. The effect of normalizing the data is illustrated in the following figure:

![Normalized vs. Not Normalized](image)

**Figure 39**

*Results of warping concentration profiles with and without normalizing.*

From the spectral dataset, as described in Section 6, concentration profiles are estimated for the reactant, intermediate and product. A detailed description is given in E.N.M. Van Sprang et al. [60]. These profiles for seventeen batches are warped using batch five as the reference batch. The left frame of Figure 39 shows the results of warping the concentration profiles of the product using normalization in the DTW algorithm. The right frame of Figure 39 shows the results of warping the concentration profiles of the product without using normalization. Clearly, normalizing the DTW algorithm gives worse results. However, this dataset comprises only three process variables (concentration profiles) where the spectral dataset contains 400 process variables (wavelengths). In this case, the problem of normalization is more difficult to investigate by visually inspecting the warped profiles. It seems that the fewer process variables the dataset contains, the more...
pronounced the problem of normalization becomes. The effect of normalization is also investigated for typically engineering dataset that contain relatively few process variables compared to spectral datasets. These results also showed that normalizing the DTW algorithm gives worse results compared to not normalizing.

3.2.8 CONCLUSIONS
In this section a strategy is proposed to warp spectral batch process data. The dynamic time warping algorithm is discussed and explained. It is shown that the normalization step can be omitted in this algorithm since applications from speech recognition are different from warping batches. Furthermore, it is shown using the definition of Kassidas that high weights are given to process variables that contain no warping information. For this reason a new definition of the weights is presented in this section. It can be concluded that the performance of the DTW algorithm, without normalization and new definition of the weights, improves the results.
CHAPTER 4 • BLACK MODELS*

*This chapter is based on the following publication(s):


4.1 General introduction*

Batch process monitoring as proposed by P. Nomikos & J.F. MacGregor [7] works with data driven models such as PCA. An advantage of this approach is that there is no fundamental knowledge of the process needed. The proposed approach is therefore general applicable, which is an advantage especially with (bio) chemical processes, which can often be very complex.

An important step in batch process monitoring is the selection of historical batch runs. This choice implicitly defines the normal or acceptable process variations. After that, the selected batches are modelled to capture normal process variation. However, the model choice is not straightforward since there are many models to choose from as will become clear later.

The content of this chapter consists of three sections. The first section discusses the evaluation of existing methods and approaches for batch process monitoring. The performance of the different models is validated by means of analysing six batch processes and performance indices are defined accordingly. The second section introduces two new models for batch process monitoring. The models are discussed and compared against a benchmark model. For this purpose six batch processes are studied and the performance of the models is discussed using predefined performance indices. The third section is an application in which a regression model is used to monitor a batch polymerisation process. The process of interest is monitored using both multivariate control charts as well as with traditional univariate control charts.

* HJR/El'S

-84-
4.2 Critical evaluation of approaches for on-line batch process monitoring *

4.2.1 SUMMARY
Since the introduction of batch process monitoring using component models in 1992, different approaches for statistical batch process monitoring have been suggested in the literature. This is the first evaluation of five proposed approaches so far. The differences and similarities between the approaches are highlighted. The derivation of control charts for these approaches is discussed. A control chart should give a fast and reliable detection of disturbances in the process. These features are evaluated for each approach by means of two performance indices. First, the action signal time for various disturbed batches is tested. Secondly, the probability of a false warning in a control chart is computed. In order to evaluate the five approaches, five different data sets are studied: one simulation of a batch process, three batch processes obtained from industry and one laboratory spectral data set. The obtained results for the performance indices are summarised and discussed. Recommendations helpful for practical use are given.

4.2.2 INTRODUCTION
Batch processes are widespread in the chemical, pharmaceutical, food and biotechnical industries. Batch processes are characterised by finite duration, non-steady-state behaviour, high conversions and, most importantly, the recipe-driven approach. If during the development phase a completed batch run is known to have produced an on-specification product, the most important process variables from that batch are stored in a recipe. This recipe is then used for future batch runs, with an aim to consistent production of the on-specification product. However, due to the complexity of most industrial batch reactions the recipe-driven approach is not always sufficient. Monitoring of batch processes is advantageous for several reasons: safety, cost reduction, quality control, fault diagnosis, fault detection, process improvement or a better process understanding.

Since the pioneering work of multivariate statistical process control (MSPC) for batch processes by J.F. MacGregor & P. Nomikos [22] and P. Nomikos & J.F. MacGregor [6], batch process monitoring (BPM) has been applied to several different processes. Furthermore, commercial companies have developed BPM software and incorporated some of the proposed methods. There is a clear trend in the ascent of supervisory control and data acquisition systems (SCADA) and distributed control systems (DCS) in process plants. The combination of these systems and BPM seems to be very promising.

The main concept of BPM is to model the common-cause variation present in batch runs obtained under normal operating conditions (NOC). This model is subsequently used to determine whether a new batch corresponds to normal operating behaviour or not. Therefore, the monitoring performance depends heavily upon this NOC data.

Since the introduction of BPM in 1992, various researchers have proposed several variants and extensions to the original methodology proposed. S. Wold et al. [53] introduced a variant of multivariate batch modelling for monitoring and diagnosis purposes. An alternative way of arranging the data is the most important feature of this work. S. Rannar et al. [31] used an approach based on a recursive multi-block (hierarchical) PCA/PLS method. D.J. Louwerse & A.K. Smilde [38] extended BPM by the use of PARAFAC (PARAllel FACtor analysis) and Tucker three-way models. Additionally, improved control charts were developed using a modified PCA method. R. Boqué & A.K. Smilde [40] used multivariate statistical procedures based on multiway covariates regression models. F.B. Martin & A.J. Morris [61] introduced a control chart based on a nonparametric method. B.M. Wise et al. [62], J.A. Westerhuis et al. [42] and K.S. Dahl et al. [63] applied and compared several alternatives for multivariate statistical analysis of batch process data.

The aim of this chapter is to summarise and critically evaluate the different approaches proposed to statistical batch process monitoring in the literature so far. From the proposed approaches a representative selection is made. Only approaches, which build component models of process data, are considered, that is, no regression models between process conditions and product quality data are compared. The focus is on component models since the subject of regression models is a field of its own. The particular properties, advantages and disadvantages of each proposed approach are investigated. This is the first systematic treatment of type I error and type II error.
for a variety of methods, which is very useful for practice. The evaluation is performed using five data sets covering a range of industrial and laboratory processes. The different approaches are compared by the use of two performance indices: the overall type I error and the Action Signal Time (AST).

A summary of each approach is given in the Section 4.2.3 using a uniform notation. The similarities and differences in terms of how the data is modelled and used for on-line monitoring are highlighted. Furthermore, this section describes the derivation of the control charts used for the $D$- and $SPE$-statistics. Section 4.2.4 introduces and defines the two performance indices, the overall type I error and AST, used to evaluate and compare the approaches. Section 4.2.5 gives a short description of the data sets studied. The results of the evaluation are given in Section 4.2.6. Finally, conclusions are drawn and recommendations are given Section 4.2.7.

4.2.3 THEORY

4.2.3.1 Modelling of normal operating conditions (NOC)
The $J$ process variables measured over $K$ time points from a single NOC batch run are stored in matrix $X_{J \times K}$. A common way of storing several NOC batch runs is to stack each run in a three-way data array $X_{I \times J \times K}$. The different batches are denoted by $i = 1, \ldots, I$ and form the first mode of $X$. The process variables are denoted by $j = 1, \ldots, J$ and form the second mode of $X$. The time points at which measurements are taken are denoted by $k = 1, \ldots, K$ and form the third mode of $X$. Note that it is assumed here that each batch is of the same length although, despite the recipe-driven approach, this is not always the case. One solution to the problem of unequal batch length is to use a 'maturity' variable as an alternative to time for the third mode axis, such as percentage of conversion (P. Nomikos & J.F. MacGregor [19], A.A. Tates et al. [2]). Other possibilities are the use of dynamic time warping (A. Kassidas et al. [35]) or the PARAFAC2 model (H.A.L. Kiers [64]).

Various kinds of process variables are monitored during a batch run, for example temperatures, pressures, flow-rates or spectral data such as infrared absorbances. The automation of modern batch processes facilitates the storing of process variables within a database. The process variables are usually highly correlated and, therefore, a multivariate approach is required. There are often only a few underlying phenomena driving the process. Latent variable methods, such as principal
component analysis (PCA), aim to extract these underlying factors from the data to form a reduced-dimension subspace, thus making the large amounts of highly correlated data more comprehensible. The number of components, R, to use can be selected using different criteria such as cross-validation (S. Wold [26]).

Prior to building a latent variable model, preprocessing of \( X \) is performed, usually in the form of centering and scaling. Centering is performed so that the actual model describes deviations from a mean. The process variables are typically measured in different units and scaling is required to give each process variable equal influence prior to modelling.

4.2.3.1.1 Nomikos & MacGregor (NM)

In order to derive a component model, \( X \) is first matricized (H.A.L. Kiers [49]). P. Nomikos & J.F. MacGregor [6] matricize \( X \) in the batch direction resulting in a matrix \( X \) \((I \times JK)\) where the second mode is nested within the third mode. Each frontal slice \( X_k \) \((I \times J)\) is placed next to \( X_{k+1} \) \((I \times J)\) as shown in Figure 14.

The purpose of monitoring using the NM approach, is to detect deviations from the desired level of operation, which is the mean trajectory of the batches in time. Therefore, in order to model the deviations from the mean trajectory, the mean trajectory is removed by centering the data across the columns of \( \mathbf{X}^{[JK]} \). To account for the difference in measurement units, slab-scaling (H.A.L. Kiers [49]) is used to give each process variable equal variance. That is, all measurements for each process variable, \( X_i \) \((I \times K)\), are scaled to unit sum of squares (R.A. Harshman & M.E. Lundy [65]). An alternative to this form of scaling is tube-scaling better known as autoscaling, whereby each column of \( \mathbf{X}^{[JK]} \) is divided by its standard deviation. Although this way of scaling is not preferred because it can disturb the multilinear nature of the data (R.A. Harshman & M.E. Lundy [65]), there are indications that the difference between slab-scaling and tube-scaling is not very crucial for batch process data (J.A. Westerhuis et al. [42]).

The preprocessed matrix \( \mathbf{X}^{[JK]} \) is used to build a PCA model. For the NM approach the model is given by:

\[
\mathbf{X}^{[JK]} = \mathbf{T} \mathbf{P}' + \mathbf{E} \quad (38)
\]

where \( \mathbf{T} \) \((I \times R)\) are the scores, \( \mathbf{P} \) \((JK \times R)\) the loadings and \( \mathbf{E} \) \((I \times JK)\) the residuals.
4.2.3.1.2 Wold, Kettaneh, Fridén & Holmberg (WKFH)

S. Wold et al. [53] matricize \( \mathbf{X} \) in the process variable direction giving \( \mathbf{X}^{K \times J} \) as shown in Figure 40.

Each horizontal slice \( \mathbf{X}_{i+1} (J \times K) \) is placed beneath the horizontal slice \( \mathbf{X}_i (J \times K) \). Here the third mode is nested within the first mode. Each column (process variable) of \( \mathbf{X}^{K \times J} \) is centered by subtracting its column mean and scaled by its standard deviation. Note that this centering operation does not remove the mean trajectory of the batches in time. This approach models the correlation between the process variables and uses the score trajectories to monitor the process. The PCA model for WKFH becomes:

\[
\mathbf{X}^{K \times J} = \mathbf{T} \mathbf{P}' + \mathbf{E}
\]

where \( \mathbf{T} (KI \times R) \) are the scores, \( \mathbf{P} (J \times R) \) the loadings and \( \mathbf{E} (KI \times J) \) the residuals.

4.2.3.1.3 Rännar, MacGregor & Wold (RMW)

S. Rännar et al. [31], describe an algorithm for adaptive batch process monitoring using hierarchical PCA.
The concept of adaptive process monitoring is summarised in Figure 41. Here, \( X_k \) represents a frontal slice of \( X \). First, a block scores vector, \( b_k \), which represents the local variation at time point \( k \), is computed. Then this block scores vector is placed in the consensus matrix, \( B_k \), along with a super scores vector, \( t_{k-1} \), which summarises previous process variation up to time point \( k-1 \). The block scores can be weighted by the adaptive parameter, \( d \). Using the consensus matrix, \( B_k \), a new super scores vector, \( t_k \), is computed which now represents the total variation up to time point \( k \). The same procedure is repeated for time point \( k+1 \) until the end of the batch run.

The adaptive nature of this approach is determined by the adaptive parameter \( d \). If \( d \) is high, much weight is given to the current process measurement relative to all previous variation. For \( d = \infty \) this algorithm is equivalent to performing a PCA on each separate block, \( X_k \). In general, increasing \( d \) will lead to more variance being explained by the model. If \( d \) is chosen to be zero, all super scores, \( t_k \), will be the same as the first super score, \( t_1 \), and the algorithm is not adaptive at all. An advantage of tuning the adaptive parameter for each block would be that the model can then adapt to different stages in the process. Note that, as yet, there exists no theory on how to tune \( d \). Therefore, in this evaluation two values of \( d \) are chosen, 0.3 and 1.0, which are the same as those used in the example by S. Rannar et al. [31].
4.2.3.2 On-line monitoring

This section describes how to monitor the evolution of a new batch using the different approaches. All the approaches studied use the concept of projecting an observation vector $x_k$ onto a space defined by $P$. This projection results in scores $t$ and residuals $e$ which are computed by:

\[
\begin{align*}
    t_k &= x_k P (P'P)^{-1} \\
    \hat{x}_k &= t_k P' \\
    e_k &= x_k - \hat{x}_k
\end{align*}
\]

(40)  
(41)  
(42)

The content of the observation vector $x_k$ and the size of the matrix $P$ depend on the approach chosen as will be explained. This can be seen as an important distinctive feature between the different approaches. Each observation vector is preprocessed before projection using the means and scaling parameters obtained from the NOC data. The features of the different methods will be discussed below.

4.2.3.2.1 Nomikos & MacGregor (NM) approach

As stated previously, the matrix $P$ of the NM approach has dimensions $(JK \times R)$. To compute the scores $t_k$ and the residuals $e_k$ according to equation (3)-(5), the observation vector $x_k$ must have the dimensions of $(JK \times 1)$. This is depicted in Figure 42.
The observation vector consists of three parts: the past measurements, a current measurement and the unknown future measurements. Take for example a measurement at time interval $k = 10$. The vector $x_k$ exists of the previous measurements $k = 1,...,9$, the current measurement $(k = 10)$ with size ($J \times 1$) and the future measurements $k = 11,...,K$. The latter is the basic problem of this approach, that is, the future behaviour of the new batch must be inferred somehow because the batch run is not completed.

P. Nomikos & J.F. MacGregor [6] suggested two approaches to fill in the unknown part of $x_k$: the zero deviations approach and the current deviations approach.

4.2.3.2.2 Zero deviations approach

The unknown future measurements are assumed to behave at the desired level of operation as defined from the NOC data. The remainder of the batch is assumed to have zero deviations from the mean trajectory.
For a process disturbance, this can be seen as an optimistic scenario. If, for example, a temperature is currently too high, its future evolution is assumed to return directly to the desired level of operation.

### 4.2.3.2.3 Current deviations approach

The current deviations approach assumes that the future measurements will continue to deviate from the desired level of operation at the same level as present at time interval $k$.

For a process disturbance, this can be seen as a pessimistic scenario, because it is assumed that a process variable which is too high/low will remain so for the rest of the batch run.

### 4.2.3.2.4 Missing data approach

P. Nomikos & J.F. MacGregor [6] suggested using the missing data approach to overcome the need for assuming the future evolution of a batch.

Using the missing data approach, the future process measurements are regarded as missing values as shown in Figure 43. Only the part of the loadings $\mathbf{P}$ up until time point $k$ are used to compute the scores and residuals. This part is denoted as $\mathbf{P}_k$. Thus
the loadings $\mathbf{P}_k$ are now 'growing' in time and have dimensions $(k\times R)$. The vector $\mathbf{x}_k$ necessarily has to be of the size $(k\times 1)$. Thus a current measurement is simply added to the past measurements in $\mathbf{x}_k$. The least squares solution is used to calculate the scores $t_k$.

\[ t_k = (\mathbf{P}_k' \mathbf{P}_k)^{-1} \mathbf{P}_k' \mathbf{x}_k \]  

(43)

Note that only when $k = K$ the term $(\mathbf{P}_k' \mathbf{P}_k)^{-1}$ is equal to identity (because $\mathbf{P}_k$ is then the same as $\mathbf{P}$ which is an orthogonal matrix). A poor monitoring performance can be expected in the early stage of the batch run, because the reduced space on which $\mathbf{x}_k$ is projected is not well defined.

### 4.2.3.2.5 Rännar, MacGregor & Wold (RMW) approach

The RMW approach builds local models on each frontal slice $\mathbf{X}_k$, giving $K$ local models. As can be seen from Fig. 2, this results in local loadings $\mathbf{P}_k$ ($J\times R$). The observation $\mathbf{x}_k$ ($J\times 1$) projected on the local reduced space spanned by $\mathbf{P}_k$ corresponds to the measurement at time $k$. The super scores are computed in a slightly different way to that described in the paper by S. Rännar et al. [31].

\[ b_{rk} = \mathbf{x}_k^{(r)} \mathbf{P}_k (\mathbf{P}_k' \mathbf{P}_k)^{-1} \]  

(44)

\[ t_{rk} = (t_{r,k-1})' d \cdot b_{rk} \mathbf{W}_k (\mathbf{W}_k' \mathbf{W}_k)^{-1} \]  

(45)

The difference between both approaches is that in this study the scores are rescaled so as to carry the variance captured by the model, this being necessary for monitoring purposes. Note that the scores for each component must be computed separately. Therefore the observation vector is deflated to compute the score for the $n^{th}$ component by:

\[ \mathbf{x}_k^{(r+1)} = \mathbf{x}_k^{(r)} - t_{rk} \mathbf{p}_{rk} \]  

(46)

A clear advantage of this approach is the fact that no filling in procedure is needed because only local models are used. However, the use of many different models does make this adaptive monitoring algorithm computationally more complex.
4.2.3.2.6 Wold, Kettaneh, Fridén & Holmberg (WKFH) approach

With this approach, PCA is performed on \( X^{KxJ} \) and to compute the scores \( t \) and residuals \( e \) according to equations 40-42, the observation vector \( x_k \) has dimensions \((J \times 1)\). Thus \( x_k \) is the measurement at time interval \( k \), which can be projected on \( P \) \((J \times R)\). The advantage of this approach is that no assumptions about future measurements are necessary.

4.2.3.3 Deriving the control limits

In general, on-line monitoring of batch processes is done using two types of control-charts: the \( D \)-chart to monitor deviation within the model relative to the centerpoint and the squared prediction error (SPE) chart to monitor deviations from the model. Note that it is also possible to construct control charts for individual scores. In this section attention is paid only to the \( D \)- and SPE-charts to avoid an abundance of results.

Statistical limits for the \( D \) and SPE statistics are derived from the NOC data. The \( I \) batches from the NOC data are projected on the model resulting in \( I \) values of the \( D \) and SPE at each time interval. Statistical distributions for these test statistics are derived and used to compute control limits using a certain \( \alpha \) value.

4.2.3.3.1 D-statistic

The Hotelling \( T^2 \) statistic is called the \( D \)-statistic when a reduced space with \( R \) components is used instead of \( x_k \) with \( J \) or \( JK \) variables. In general, the \( D \)-statistic is given by

\[
D_k = (\mathbf{t}_{\text{new},k} - \overline{t}_k)^T \mathbf{S}_k^{-1} (\mathbf{t}_{\text{new},k} - \overline{t}_k) \frac{I(I - R)}{R(I^2 - 1)} \sim F(R, I - R)
\]

where \( \mathbf{t}_{\text{new},k} \) \((R \times 1)\) are the scores of the new batch at time interval \( k \) and \( \overline{t}_k \) \((R \times 1)\) contains the means of the columns of the score matrix \( \mathbf{T}_k \) \((I \times R)\). \( \mathbf{S}_k \) \((R \times R)\) is the variance-covariance matrix of \( \mathbf{T}_k \) and is an estimation of the variation across the batches. Since the scores are linear combinations of a large number of variables, according to the central limit theorem the scores are independently normally distributed across the batch direction. The test statistic \( D_k \) follows an \( F \)-distribution with \( R \) and \( I-R \) degrees of freedom (N.D. Tracy et al. [14]).
Notice that for the NM approach, \( \bar{t}_k = 0 \) since the data is column mean centered. Furthermore, it is assumed that \( \mathbf{S}_k \) is constant in time and is equal to the variance-covariance matrix of the scores calculated when all data is available, \( \mathbf{T}_{k=K} \).

For the RMW approach, \( \bar{t}_k = 0 \) also, but the variance covariance matrix \( \mathbf{S}_k \) is now calculated for each time interval \( k \) using the super scores \( \mathbf{T}_k (IxR) \) of the NOC batch runs.

For the WKFH approach, the matrix \( \mathbf{T} (KxR) \) from equation 39 is rearranged to obtain a matrix \( \mathbf{T}_k (IxR) \) for each time interval \( k \), i.e. the \( I \) rows corresponding to time interval \( k \) are selected from \( \mathbf{T} \) and used to form \( \mathbf{T}_k \). Here \( \bar{t}_k \) is unequal to zero and \( \mathbf{S}_k \) is the variance-covariance matrix of the score matrix \( \mathbf{T}_k \).

### 4.2.3.3.2 Squared Prediction Error (SPE)

For all approaches the \( SPE_k \) statistic is computed by:

\[
SPE_k = \sum_{j=1}^{I} \varepsilon_{j,k}^2 \sim g \chi^2
\]

where \( \varepsilon_{j,k} \) is the prediction error of the process variable \( j \) at time interval \( k \). The control limits for the \( SPE \) are obtained by fitting a weighted \( \chi^2 \)-distribution to the reference distribution obtained from the NOC data at each time point. In equation 48, parameter \( h \) represents the degrees of freedom and parameter \( g \) the weight to account for the magnitude of \( SPE_k \). These two parameters can be estimated in different ways. In this study the estimation is done according to J.E. Jackson & G.S. Mudholkar [24], whereby \( g \) and \( h \) are functions of the eigenvalues of the residual variance covariance matrix at each time interval \( k \). In order to increase the number of observations available for each estimation, P. Nomikos & J.F. MacGregor [19] used a time window from \( k-2 \) to \( k+2 \). This approach is also used here.

Note that the use of a \( \chi^2 \)-distribution implicitly assumes normality of the errors, which may not always be true in practice. However, as the parameters of the \( \chi^2 \)-distribution used to calculate the \( SPE \) limits are obtained directly from the moments of the sampling distribution of the actual NOC data, this approximating distribution is found to work well even in cases where the errors are not normal.
4.2.3.4 Summary
To evaluate the different approaches, it is important to realize that the approaches are conceptually different. In the NM approach, a model is constructed which captures the correlation between the process variables and the correlation of the process variables in time. In this approach, each batch run is regarded as an object and has by construction the disadvantage that the future behaviour of the batch run is not known during on-line monitoring.

The RMW approach captures the correlation of the process variables at a certain time interval. Since the local models are constructed in a recursive way, the time behaviour is included implicitly, the extent to which can be tuned using the parameter $d$.

The WKFH approach models the correlation patterns between the process variables that occur during the NOC batch runs. By rearranging the scores, the dynamic behaviour of the process is implicitly captured by the average score trajectories.

4.2.4 Performance Indices
This evaluation uses two performance indices to evaluate how well each approach performs in terms of fault detection. The two indices are the overall type I error and the action signal time (AST).

4.2.4.1 Overall type I error (false warning)
The first performance index used is the type I error. The type I error is related to the null hypothesis ($H_0$) that the process is in-control. If $H_0$ is rejected when it is true, a type I error has occurred. Thus,

$$\alpha_{\text{imposed}} = P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ is true})$$

(49)

To calculate the actual value of $\alpha$, P. Nomikos & J.F. MacGregor [19] used equation 49 to calculate the probability of an overall type I error.

$$\alpha_{\text{actual}} = P(\text{overall type I error}) = \frac{\sum \text{false warnings}}{IK}$$

(50)
In the present section the following procedure is used to calculate $a_{\text{actual}}$. One batch is removed from the NOC data set and a model is built on the remaining $I-1$ batches. Control limits for the $D$-chart and $SPE$-chart are derived from this model. Then the left out NOC batch is monitored. As this batch is from the NOC data, it is assumed to be in-control over the entire trajectory and, therefore, a crossing of a control limit is considered to be a false warning. The procedure is repeated for every NOC batch and the total number of false warnings for the NOC data is calculated. This leave-one-out approach is used so as to maintain independence between the monitored batch and the NOC model.

4.2.4.2 Action Signal Time (AST) and type II error

To assess the type II error, it is common in statistical process control to use average run length (ARL) curves to study the efficiency of a control chart. However, such a rigorous analysis requires a large amount of data or a theoretical treatment. Both alternatives are not feasible in batch monitoring approaches. Therefore, in this study the Action Signal Time (AST) is used to evaluate the detection power of a method and serves as an indication of the type II error. The AST is defined as the time between the introduction of an error and the out-of-control signal, as can be seen in Figure 44. The signal is said to be out-of-control only if at least three consecutive points are outside the control limit. The choice of the number of observations that define an out-of-control signal is according to practice, when action is only undertaken after some consecutive points have been out of control. In this study, different batch processes are studied. Each process has a different batch run length. To compare the AST of the different processes, the relative AST is used. The relative AST (RAST) is the ratio between the AST and the total erroneous time trajectory (ET) of the batch run. The ET is the time period starting from the introduction of a disturbance until the end of the batch run (see Figure 44). Note that there are several possibilities of defining the RAST with each definition having pros and cons. In this evaluation a choice is made to use the given definition of the RAST.
Note that there is distinction between an out-of-control signal, defined as three points above the control limit, and a warning, defined as one point above the control limit.

In ordinary hypothesis testing there is a clear relationship between the type II error and the type I error: by choosing a lower $\alpha$, the type I error decreases and the type II error increases. Due to the distinction between an out-of-control signal and a warning, such a direct relationship is not present anymore between the type I error and the AST. There will remain a tendency, however, for the AST to become larger when choosing a lower $\alpha$, which also decreases the type I error.

4.2.4.3 Summary

The emphasis of the comparison lies on the detection of erroneous batches. Therefore, the AST is the most relevant index of the two. The overall type I error is mainly a tool for confirmation whether the control limits are at an acceptable level. If the type I error is not satisfactory, the user can raise or lower the limits until an acceptable level of $\alpha$ is reached.
4.2.5 DESCRIPTION OF THE DATA

In order to evaluate the different approaches for on-line monitoring, the approaches are applied to five data sets. Each data set has different specific features as will be explained in the following. The NOC data sets are used to compute the overall type I error and disturbed batches are used to compute the AST. The time of disturbance occurrence and the dimensions of the three-way data array for the studied data sets are given in Table 3.

| Table 3 |

<table>
<thead>
<tr>
<th>Data set 1</th>
<th>Data set 2</th>
<th>Data set 3</th>
<th>Data set 4</th>
<th>Data set 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>50</td>
<td>36</td>
<td>46</td>
<td>67</td>
</tr>
<tr>
<td>J</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>K</td>
<td>200</td>
<td>100</td>
<td>116</td>
<td>58</td>
</tr>
<tr>
<td>Disturbed batches</td>
<td>Batch 106'</td>
<td>Batch 50</td>
<td>Batch 41</td>
<td>Batch 14</td>
</tr>
<tr>
<td>Time of disturbance</td>
<td>1</td>
<td>56</td>
<td>58</td>
<td>1</td>
</tr>
</tbody>
</table>

*exact time of disturbance is known.

The times of introduction of the errors for the different processes are given in Table 3. Out-of-control signals before the fault occurred are ignored. For data sets 1 and 5 prior knowledge of the exact time of disturbance is available. For the remaining data sets the time of introduction of the error is obtained after a thorough data analysis, using process knowledge, and discussion with the process engineers.

4.2.5.1 Data set 1

Data set 1 consists of a simulated emulsion co-polymerisation of styrene-butadiene. This process is well described in P. Nomikos & J.F. MacGregor [6]. This data set consists of 50 NOC batches for which 9 process variables are measured during 200 time intervals. The following process variables are measured: two flow-rates, four temperatures, density, conversion and rate of energy release. Two disturbed batches are available to calculate the AST. The first has an initial organic impurity
contamination in the butadiene feed. The second has a similar, but larger, contamination halfway the batch run.

4.2.5.2 Data set 2
Data set 2 is an industrial (Du Pont) two-stage polymerisation process with 36 selected NOC batches for which 10 process variables are measured during 100 time intervals. One disturbed batch is available to calculate the performance indices, which was known to have a poor end-product quality. This data set is described in P. Nomikos & J.F. MacGregor [19].

4.2.5.3 Data set 3
Data set 3 consists of an industrial (DuPont) polymerisation in an autoclave with 47 NOC batch runs. This data set is also referred to by K.A. Kosanovich et al. [7]. The polymer is produced through five stages during these stages the process is monitored by 8 process variables (six temperatures and two pressures) during 116 time intervals. Prior to model building, the raw data is linearly interpolated to align the NOC batches, this is described in R. Boqué & A.K. Smilde [40]. Three disturbed batches are available to calculate the AST. A univariate analysis of batch 41 clearly shows disturbances for two pressure and four temperature measurements halfway the batch run. This is probably due to an upset in the first stage of the process. For batch 46 and batch 47, univariate analysis shows disturbances at the end of the batch run for one pressure and five temperature measurements.

4.2.5.4 Data set 4
This data set is from an industrial (Shell) suspension polymerisation of polyvinylchloride (PVC) in a batch reactor. The data set consists of 67 batch runs obtained under NOC and 15 process variables were measured for a period of 58 time intervals. The NOC batches are aligned using the conversion as a maturity variable. A more detailed description is given by A.A. Tates et al. [2]. The process variables consist of: eight temperatures, condenser duty, agitator speed and power supply, two mass streams of cooling water, batch reactor level and pressure. Three disturbed batches are available to calculate the AST. Analysis of univariate plots of the process variables is used to study the disturbed batches. Analysis of batch 17 shows that the temperature and the amount of refrigerant water to the jacket are disturbed at the end
of the batch run. Batch 24 shows disturbances for the batch level and the power supply to the agitator. Batch 49 reveals disturbances for the temperature and amount of cooling water, and a temperature disturbance within the reactor.

4.2.5.5 Data set 5

Data set 5 consists of a laboratory spectroscopic batch process of a two-step biochemical conversion reaction described in S. Bijlsma et al. [66] and can be obtained from the Process Analysis & Chemometrics website of the University of Amsterdam. The data set consists of 27 NOC batch runs measured at wavelengths 300-500 nm during 271 time intervals. One batch run with a pH disturbance at time interval 138 is available to compute the AST.

4.2.6 RESULTS AND DISCUSSION

Prior to analysis, the number of principal components for each method is determined using cross validation (S. Wold [67]). It is found that a three-component model is enough to explain common-cause variation for all approaches and also for all data sets. This is in agreement with the number chosen by the authors for data sets 1 and 5.

4.2.6.1 AST

The results for the average AST for the control charts with a 95% confidence limit are given in Fig. 6. These results were obtained as follows. First, each of the disturbed batches (see Table 1) is monitored using the SPE-chart and the D-chart for each approach. This results in 2 AST values for each disturbed batch and 20 AST values for each approach. Although the fault could appear first either in the SPE-chart or in the D-chart, from a practical point of view the control chart, which gives the shortest AST, is of interest. Therefore using the shortest AST of the two charts will result in 10 AST values (one for each disturbed batch) for each approach. As stated earlier, the relative AST is used for comparing the AST of different processes. To obtain an overall picture of the RAST, the RAST is averaged out over the data sets for each monitoring approach. A similar procedure is followed for the control charts with 99% confidence limits.
Figure 45
The average RAST found for 95% confidence limits.

In Figure 45, the results are divided into three groups. The first group is formed by the missing data approach and the RMW approach ($d = 0.3$), for which the shortest average AST is seen. For data sets 1 and 2, P. Nomikos & J.F. MacGregor [19] also reported a superior performance for the missing data approach in terms of AST. As described earlier, however, for the first 10% of the batch run, the D-chart for the missing data approach is not found to be useful as not enough data is available to obtain a reliable score $t$ and, hence, a reliable $D$. This leads to many false warnings in the beginning of the batch run. There are three batches with a disturbance immediately from the beginning. As will become apparent in the next section, both the missing data and the RMW ($d = 0.3$) approaches give high overall type I errors and so although a fault seems to be detected quickly by the $D$-chart, this could be due to the shortcomings of these approaches.

The second group is formed by the NM current deviations and NM zero approach. The third group is formed by the WKFH and RMW ($d = 1$). Both groups are generally found to perform well with a small difference in the AST. From Figure 45 can be seen that the third group has a slightly larger AST.
A similar pattern to that seen in Figure 45 is found for the control charts with 99% confidence limits and the results are not given here. The differences between the average AST for the 95% and 99% limits are small. This implies that the test statistic crosses both the 95% and 99% confidence limit at the moment of detection. Therefore, the detection of a fault is not very sensitive to the exact confidence limit as long as reasonable values for these limits are chosen.

Note that using the NM approach, faults are generally detected in the $SPE$-chart. Only if the disturbance occurred at the beginning of a batch, then the $D$-chart sometimes signals faster than the $SPE$-chart. Contrary, in the WKFH approach most of the faults are detected in the $D$-chart. This is due to the different construction of the models.

### 4.2.6.2 Overall type I error

In general, it is expected that $a_{\text{actual}}$ is slightly higher than $a_{\text{imposed}}$ for the $SPE$. This is because the $SPE$ limits are approximated from the NOC residuals and it is assumed that the residuals from a new batch come from the same population. In fact, the residuals from a new batch will generally be slightly higher than the NOC residuals since the NOC data is modelled in a least-squares way in which the residuals are minimised. This is reflected in the overall type I error. The $D$-statistic is approximated by a $F$-distribution, which takes into account the fact that the new batch is independent from the NOC data (N.D. Tracy et al. [14]).

The $a_{\text{actual}}$ values for the overall type I error for the $SPE$ using 95% confidence limits are given in Figure 46.
Results for the overall type 1 error for the SPE95.

The solid line represents the value for $a_{\text{imposed}}$. Comments on Figure 46 are as follows:

- It is clear that the value for $a_{\text{actual}}$ is indeed slightly higher than $a_{\text{imposed}}$. This is also in agreement with the results found by P. Nomikos & J.F. MacGregor [19]. It seems to be a general conclusion for these approaches, the data sets used being dissimilar in nature.
- The RMW approach ($d = 0.3$ and $d = 1$) has the highest value for $a_{\text{actual}}$ for all data sets. For this approach, where local models at each time interval $k$ are used, the NOC data is modelled relatively well and the NOC residuals are correspondingly low. Thus, the difference between the NOC residuals and those for a new batch are higher than for other approaches.
- The patterns obtained using 99% confidence limits for the SPE lead to the same conclusions as when 95% limits were used.
The results for the overall type I error for the $D$-statistic using 95% confidence limits are given in Figure 47. Comments on Figure 47 are as follows:

- The spectral data set (data set 5) gives different results from the engineering data sets.
- The $a_{\text{actual}}$ values of the RMW approach with $d = 0.3$ are far too high. The low value for the adaptive parameter $d$ implies that a relatively high importance is given to the process history for that batch. This may mean that a false warning, such as a spike in the data, can continue to influence the $D$-chart over a length of time and, therefore, high values for the overall type I error are seen. The weighting parameter $d$ can be tuned in such a way that $a_{\text{imposed}}$ equals $a_{\text{actual}}$, which may provide a method for determining the correct value to use for $d$. 

![Figure 47](image-url)
The WKFH approach and the RMW approach with \( d = 1 \) give reasonable values for \( \alpha_{\text{actual}} \).

In general, the filling-in approaches, NM current deviations and NM zero approach have values for \( \alpha_{\text{actual}} \), which are too low.

The results found for the missing data approach must be interpreted with care. It has already been mentioned that the \( D \)-chart at the beginning of the batch is unreliable, resulting in too many false warnings and, therefore, a high overall type I error.

The patterns obtained using 99% confidence limits for the \( D \)-statistic lead to the same conclusions as when 95% limits were used.

### 4.2.7 Conclusions and Recommendations

In this section an evaluation is made of five different approaches for monitoring batch processes. The theory behind each approach is explained together with the specific features. The most important distinction between the approaches is the way in which the correlation structures in the data, both between process variables and in time, are modelled, sometimes leading to a need to assume the future behaviour of a new batch during on-line monitoring. To evaluate the different approaches, performance indices were used and applied to five different data sets. Some recommendations on using the methods are given as follows:

- The results of the evaluation presented here suggested that the NM current deviations, NM zero- and the WKF approach on average give the lowest reliable AST.

- The control charts for the NM missing data approach are found to be particularly unreliable at the start of the batch and should, perhaps, be ignored until data from enough time points is available.

- The method presented in this study for estimating the overall type I error could be used in practice to correctly set the confidence limits depending upon the monitoring approach being used. This also provides a method for determining the correct value to use for the adaptive parameter \( d \) in the RMW approach, i.e. \( d \) can be tuned in such a way that \( \alpha_{\text{imposed}} \) equals \( \alpha_{\text{actual}} \).

- The detection of a fault in the control charts is not very sensitive to the exact confidence limit as long as reasonable values for these limits are chosen.
4.3 Fault detection properties of global, local and time evolving models for batch process monitoring*

4.3.1 SUMMARY
This chapter treats the fault detection capabilities of alternative methods for batch process monitoring. Two new methods are investigated and compared to an existing one (the benchmark). A description of the models is given and the performance is discussed by means of fault detection performance indices. The performance indices used are the overall type I error and the action signal time. In order to evaluate the performance of the models in terms of the overall type I error and action signal time, six different batch process data sets have been used. The data set comprises four industrial data sets, one simulated data set and a laboratory spectral data set.

4.3.2 INTRODUCTION
Batch processes are very important in chemical, food and pharmaceutical industry. The flexibility of batch plants makes them interesting for the manufacturing of a large variety of products. In order to accomplish a consistent product quality, batch processes are recipe driven. This requires a high demand on the information technology systems used in batch plants. These technology systems are coupled to databases containing e.g. historical data about previous batch runs like temperatures, pressures etc. This historical data can be evaluated to give better insight process information.

The batch reactor itself is usually equipped with various sensors to measure process variables like temperatures, pressures or agitator speed. If this batch is known to have produced an on-spec product, the variation of these process variables during the batch run might be considered as allowable. Multivariate statistical methods can be used to capture the variation of this reference distribution for all successful batches.

Then the variation of the process variables for completely new batches is compared to the reference distribution to see if the batch is in-control.

The introduction of multivariate statistical process control applied to batch processes was done by J.F. MacGregor & P. Nomikos [22]. Their concepts have been evaluated (E.N.M. Van Sprang et al. [41]) and extended using different models (S. Wold et al. [53], S. Rannar et al. [31], D.J. Louwerse & A.K. Smilde [38], R. Boqué & A.K. Smilde [40]). Different commercial software packages based on these methods are developed. A drawback of the approach described by Nomikos and MacGregor is that the future behaviour of a new batch must be estimated using a filling-in approach.

In this chapter, two new types of models for batch process monitoring are introduced: local models and time evolving models. P. Nomikos & J.F. MacGregor [6] already suggested to use the time evolving model as 'the most valid method' to overcome the filling-in requirement. There is no need to estimate the future behaviour of a new batch for both these models. The performance of the evolving and local model is tested and compared to one of the most commonly used global models. Six different datasets are used for this purpose. Section 4.3.3 explains the concepts of statistical batch monitoring using global, evolving and local models. The explanation of the performance indices used in this study is given in Section 4.3.4. A brief overview of the datasets used to test the different models is presented in the Section 4.3.5. The outcome of the performance indices after applying the models to the different datasets is presented in Section 4.3.6. Section 4.3.6.3 shows an example of a wrong batch that is monitored using all three models. Then, in Section 4.3.7 conclusions are drawn.

4.3.3 STATISTICAL BATCH PROCESS MONITORING
The following notation is used: the number of batches is denoted by $i = 1, \ldots, I$. The number of measured process variables is given by $j = 1, \ldots, J$ and the number of time intervals is denoted by $k = 1, \ldots, K$. For a single batch run, the measurements of $J$ process variables over $K$ time intervals are stored in the matrix $X_i (J \times K)$. Matrices are denoted as uppercase bold characters. Several batch runs can be stacked in a three-way data array $\mathbf{X} (I \times J \times K)$ denoted as an uppercase underlined bold character. The batches $I$ form the first mode of $\mathbf{X}$, the process variables $J$ the second mode and the
time intervals $K$ the third mode. The matrix denoted as $\mathbf{X}_k (I\times J)$ represents a frontal slab of $\mathbf{X}$.

Prior to model building, $\mathbf{X}$ should contain batches that are known to have produced an on-spec product under normal operating conditions (NOC). The choice whether a batch is NOC is not straightforward but should be supported by a combination of expert knowledge from process operators and statistical founding by the model builder. It has important consequences once the data is assigned as NOC. If so, $\mathbf{X}$ represent batches that are by definition statistically in-control and statistical parameters are estimated based on these NOC batches. Besides NOC selection, the data in $\mathbf{X}$ needs to be properly pre-processed. Centring across the batch mode is applied to remove the non-linear behaviour of the trajectories in the data (P. Nomikos & J.F. MacGregor [6]). As a result, centring leads to modes describing deviations from a nominal trajectory. Process variables are often measured in different units. To give each process variable an equal importance, the data is scaled to e.g. unit standard deviation. Here, the data-array $\mathbf{X}$ is slab-scaled (H.A.L. Kiers [49]), but other types of scaling exist. In the following, it is assumed that $\mathbf{X}$ is properly scaled, centred and represents NOC data.

4.3.3.1 Modelling
The following section describes the three different methods for on-line statistical batch process monitoring. The first method is based on global models with filling-in of current deviations. This is a commonly used method and has been well described in literature (P. Nomikos & J.F. MacGregor [6], E.N.M. Van Sprang et al. [41]). The other types of methods, based on local and time evolving models are new and will be introduced in this chapter.

4.3.3.1.1 Global model
First, the properly scaled data-array $\mathbf{X}$ is matricized in the batch direction resulting in the matrix $\mathbf{X} (I\times J\times K)$ where the second mode of $\mathbf{X}$ is nested within the third mode. The two dots on top of $\mathbf{X}$ refer to the current deviations approach. A PCA is performed on $\mathbf{X}$ to calculate the loadings $\mathbf{P} (J\times K\times R)$, scores $\mathbf{T} (I\times R)$ and residuals $\mathbf{E} (I\times J\times K)$. This is illustrated in Figure 48.
Chapter 4 • Black models

The next step is to construct a matrix $\bar{X}_k$ ($I \times JK$) at every time interval $k$. This is illustrated in Figure 49:

$\bar{X} = [x_1 \ x_2 \ x_3 \ x_4]$

Figure 48

Construction of $\bar{X}$ for the global model.

Figure 49

Construction of $\bar{X}_k$ for the global model to calculate the scores and residuals.
It can be seen that the frontal slabs of \( \mathbf{X} \) form the matrix \( \mathbf{X}_k \). The part of \( \mathbf{X}_k \) until time interval \( k \) \((\mathbf{X}_k^{1:k})\) consists of \( \mathbf{X}_1, \ldots, \mathbf{X}_k \), while the remaining part \((\mathbf{X}_k^{k:k})\) is filled with the frontal slab \( \mathbf{X}_k \). Then, the matrix \( \mathbf{X}_k \) is projected on the loadings \( \mathbf{P} \) to find the scores \( \mathbf{T}_k \) \((I \times R)\) and residuals \( \mathbf{E}_k \) \((I \times J)\). These scores and residuals form the reference distributions for the test statistics \( D \) and \( SPE \).

To derive these test statistics for a new independent batch \( \mathbf{x} \), at every time interval \( k \) a vector \( \mathbf{x}_k \) \((J \times 1)\) is constructed in the same manner as \( \mathbf{X}_k \). This vector is projected on the loadings \( \mathbf{P} \) to find the scores \( \mathbf{t}_k \) \((R \times 1)\) and residuals \( \mathbf{e}_k \) \((J \times 1)\) as can be seen from Figure 49. The calculation of the \( SPE \) and \( D \)-statistic is found in the first column of

Table 4 as well as an overview of the notation used for the current deviations approach.

<table>
<thead>
<tr>
<th>NOC data</th>
<th>Global</th>
<th>Evolving</th>
<th>Local</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data (( \mathbf{X} ))</td>
<td>( \mathbf{X}_k ) ((I \times J \times K))</td>
<td>( \mathbf{X}_k ) ((I \times J))</td>
<td>( \mathbf{X}_k ) ((I \times J))</td>
</tr>
<tr>
<td>Loadings (( \mathbf{P} ))</td>
<td>( \mathbf{P} ) ((J \times K \times R))</td>
<td>( \mathbf{P}_k ) ((J \times R))</td>
<td>( \mathbf{P}_k ) ((J \times R))</td>
</tr>
<tr>
<td>Scores (( \mathbf{T} ))</td>
<td>( \mathbf{T}_k ) ((I \times R))</td>
<td>( \mathbf{T}_k ) ((I \times R))</td>
<td>( \mathbf{T}_k ) ((I \times R))</td>
</tr>
<tr>
<td>Residuals (( \mathbf{E} ))</td>
<td>( \mathbf{E}_k ) ((I \times J \times K))</td>
<td>( \mathbf{E}_k ) ((I \times J))</td>
<td>( \mathbf{E}_k ) ((I \times J))</td>
</tr>
<tr>
<td>Covariance matrix (( \mathbf{S} ))</td>
<td>( \mathbf{S}_k = \mathbf{T}_k \mathbf{T}_k^\top ) ((R \times R))</td>
<td>( \mathbf{S}_k = \mathbf{T}_k \mathbf{T}_k^\top ) ((R \times R))</td>
<td>( \mathbf{S}_k = \mathbf{T}_k \mathbf{T}_k^\top ) ((R \times R))</td>
</tr>
<tr>
<td>New batch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data (( \mathbf{x} ))</td>
<td>( \mathbf{x}_k ) ((J \times 1))</td>
<td>( \mathbf{x}_k ) ((J \times 1))</td>
<td>( \mathbf{x}_k ) ((J \times 1))</td>
</tr>
<tr>
<td>Scores (( \mathbf{t} ))</td>
<td>( \mathbf{t}_k ) ((R \times 1))</td>
<td>( \mathbf{t}_k ) ((R \times 1))</td>
<td>( \mathbf{t}_k ) ((R \times 1))</td>
</tr>
<tr>
<td>Residuals (( \mathbf{e} ))</td>
<td>( \mathbf{e}_k ) ((J \times 1))</td>
<td>( \mathbf{e}_k ) ((J \times 1))</td>
<td>( \mathbf{e}_k ) ((J \times 1))</td>
</tr>
<tr>
<td>( SPE \sim g \chi^2 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D \sim F(( R, I - R ))</td>
<td>( \frac{I(I-R)}{R(I^2-1)} \cdot \mathbf{t}_k \mathbf{S}_k^{-1} \mathbf{t}_k )</td>
<td>( \frac{I(I-R)}{R(I^2-1)} \cdot \mathbf{t}_k \mathbf{S}_k^{-1} \mathbf{t}_k )</td>
<td>( \frac{I(I-R)}{R(I^2-1)} \cdot \mathbf{t}_k \mathbf{S}_k^{-1} \mathbf{t}_k )</td>
</tr>
</tbody>
</table>

*Overview of the used notation for the different models.*
The abbreviation ‘cp’ refers to the current part of a vector (or matrix) indicated by \( R_i(J_k - J + 1): J_k \). That is, only the part corresponding to the \( J \) process variables measured at time interval \( k \) are of interest. This notation is read as follows. Suppose a matricized matrix has 5 rows and 200 columns. The process variables (\( J = 10 \)) are nested within the time mode (\( K = 20 \)) as a result of matricizing. Suppose only the measurements at time \( k = 3 \) are of interest. In other words only the columns 21 until 30 are of interest. This is denoted as 5, 21:30.

The SPE-statistic follows approximately a weighted chi-square distribution (G.E.P. Box [23], J.E. Jackson & G.S. Mudholkar [24]) while the D-statistic follows an F-distribution (N.D. Tracy et al. [14]).

### 4.3.3.1.2 Time evolving model

It was shown for the current deviations approach that the ‘future’ part of the data is assumed to behave according to the current measurement of the process variables. This may be considered as a drawback and can be avoided by using a time evolving model. For a time evolving model, a matrix \( \bar{X}_{k} \) (\( K \times Jk \)) is constructed from the \( k \) frontal slabs of \( \bar{X} \). This is illustrated in Figure 50:

\[
\begin{align*}
\bar{X}_1 & = X_1 \\
\bar{X}_2 & = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \\
\bar{X}_3 & = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix} \\
\bar{X}_4 & = \begin{bmatrix} X_1 & X_2 & X_3 & X_4 \end{bmatrix}
\end{align*}
\]

Figure 50

*Construction of \( \bar{X}_k \) for the evolving model.*
The reversed arc on top of $\mathbf{X}_k$ refers to the time evolving modelling approach. It can be seen that $\mathbf{X}_k$ is the matricized part of $\mathbf{X}$ until time interval $k$. Obviously, the size of $\mathbf{X}_k$ increases with time. Another difference with the current deviations approach is that for every $\mathbf{X}_k$ a PCA model is built. This gives $K$ different model loadings $\mathbf{P}_k$ ($J \times R$) as can be seen from Figure 50. The computational power of modern computers is no impeding anymore to build these models. The reference distributions for the test statistics are formed by $\mathbf{T}_k$ ($I \times R$) and $\mathbf{E}_k$ ($I \times J$).

The scores $\mathbf{t}_k$ ($R \times 1$) and residuals $\mathbf{e}_k$ ($J \times 1$) for a new batch are found by projecting $\mathbf{x}_k$ ($J \times 1$) on the corresponding loadings. $\mathbf{x}_k$ is a vector that contains the measured process variables until time interval $k$. The second column of Table 4 gives an overview of the notation used and calculation of the test statistics for the time evolving modelling approach.

4.3.3.1.3 Local model

Local models are built for each frontal slab of $\mathbf{X}$. Therefore, the data matrix $\mathbf{X}_k$ ($I \times J$) equals the $k$-th frontal slab of $\mathbf{X}$. The arc on top of $\mathbf{X}_k$ refers to the local modelling approach. The construction of $\mathbf{X}_k$ is depicted in Figure 51:

![Figure 51](image)

$\mathbf{X} = \mathbf{X}_1 \quad \mathbf{X}_2 = \mathbf{X}_2 \quad \mathbf{X}_3 = \mathbf{X}_3 \quad \mathbf{X}_4 = \mathbf{X}_4$

$\mathbf{X}_1 = \mathbf{X}_1 \quad \mathbf{X}_2 = \mathbf{X}_2 \quad \mathbf{X}_3 = \mathbf{X}_3 \quad \mathbf{X}_4 = \mathbf{X}_4$

$\mathbf{X}_1 = \mathbf{T}_1 \mathbf{P}_1 + \mathbf{E}_1 \quad \mathbf{X}_2 = \mathbf{T}_2 \mathbf{P}_2 + \mathbf{E}_2 \quad \mathbf{X}_3 = \mathbf{T}_3 \mathbf{P}_3 + \mathbf{E}_3 \quad \mathbf{X}_4 = \mathbf{T}_4 \mathbf{P}_4 + \mathbf{E}_4$

$\mathbf{t}_1 = \mathbf{P}_1 \mathbf{x}_1 \quad \mathbf{t}_2 = \mathbf{P}_2 \mathbf{x}_2 \quad \mathbf{t}_3 = \mathbf{P}_3 \mathbf{x}_3 \quad \mathbf{t}_4 = \mathbf{P}_4 \mathbf{x}_4$

$\mathbf{e}_1 = \mathbf{x}_1 - \mathbf{P}_1 \mathbf{t}_1 \quad \mathbf{e}_2 = \mathbf{x}_2 - \mathbf{P}_2 \mathbf{t}_2 \quad \mathbf{e}_3 = \mathbf{x}_3 - \mathbf{P}_3 \mathbf{t}_3 \quad \mathbf{e}_4 = \mathbf{x}_4 - \mathbf{P}_4 \mathbf{t}_4$

Figure 51

Construction of $\mathbf{X}_k$ for the local model.
For each data matrix $\tilde{X}_k$ a PCA model is built that gives the scores $\bar{T}_k$ ($I \times R$) and residuals $\tilde{E}_k$ ($I \times f$). This way of building models may be considered as straightforward and simple while no filling in of the future part of a new batch is required.

The measurement of the process variables at a certain time interval forms the vector $\bar{x}_k$ ($f \times 1$). This vector is projected onto the model loadings $\tilde{P}_i$ to find the scores $\tilde{t}_i$ ($R \times 1$) and residuals $\tilde{e}_i$ ($f \times 1$). The third column

Table 4 of gives an overview of the used notation for the local modelling approach.

### 4.3.4 PERFORMANCE INDICES

In this chapter two new models are compared to a well established method. In order to qualify each method, two performance indices are used: the overall type I error and the action signal time (AST). The following gives a description of these two performance indices.

#### 4.3.4.1 Overall type I error

In this Chapter (4.3) the control charts for monitoring the SPE and $D$-statistic are set with an imposed confidence level of 5%, that is $\alpha_{impord} = 0.05$. As stated before, an NOC batch is by definition statistically in-control. Therefore, the overall null hypothesis $H_0$ is formulated as follows: $H_0$ is true if an NOC batch is in-control during the entire batch run. If this is not true, the null hypothesis is rejected and an overall type I error has occurred. Such a batch is said to generate a false alarm.

It is taken into account that approximately 5% of all the batches will generate a type I error. In order to set the limits for the control charts, approximations have to be made. That is, the scores are assumed to be normally distributed. This is approximately true because of the central limit theorem. Also, the squared residuals are approximately $\chi^2$ distributed with an estimated degree of freedoms. The quality of these approximations might differ between the models. Therefore, it is useful to check the validity of these approximations by checking whether the value of $\alpha_{impord}$ is close to the actual value of $\alpha$ ($\alpha_{actaul}$).

This procedure is accomplished by separately monitoring the NOC batches and checking if indeed 5% of the NOC batches generate a false alarm. Let the
indicator variable $\tau$, be one if a NOC batch has generated a false alarm and zero otherwise. Then the value of $\alpha_{\text{actual}}$ is calculated as follows:

$$\alpha_{\text{actual}} = \frac{\sum_{i=1}^{l} \tau_i}{l}$$  \hfill (51)

If the value of $\alpha_{\text{actual}} = \alpha_{\text{imposed}}$, it can be concluded the limits are quite good. This conclusion can be drawn sharper if the number of batches is considerably large. Therefore, as many batches as possible should be tested in order to obtain reliable results. In this six different datasets are used with a total of 273 batches. The outcome of the overall type I error per dataset varies significantly. For small data sets the standard deviation of the overall type I error can be considerable (0.05±0.04). A global average of all datasets with a smaller standard deviation can be obtained by performing a meta-analysis.

The validity of the overall $H_0$ is sequentially tested at every time interval $k$. This makes it the problem of a composite $H_0$. That is, because there is a possibility of 5% that an NOC batch generates a false alarm at every time interval, the probability of rejecting the overall $H_0$ is high. This is a well known problem and can be overcome by adjusting the value of $\alpha_{\text{imposed}}$ according to the Bonferroni adjustment (G.W. Snedecor & G.C. Cochran [68]):

$$\alpha_{\text{actual}} = \frac{\alpha_{\text{actual}}}{K}$$  \hfill (52)

The price that must be paid for this adjustment of the control limit is that the limit becomes more conservative.

It should be mentioned that the overall type I error is a useful tool to see whether caution is needed in a particular case. Often the actual overall type I error is larger than the imposed one. In practice, limits are often simply adjusted to obtain an acceptable value in cases where the overall type I error is too high.
4.3.4.2 Action signal time (AST)
The second performance index used in this is the action signal time (AST). The AST is related to the detection of a process fault. Preferably the AST is as short as possible and this can be related to the type of model used for batch process monitoring. The AST is defined as the time elapsed between the introduction of a process fault and the action signal above limits in the control chart. Here, the action signal is taken as three consecutive points lying outside the confidence limit. The actual choice of the action signal is taken from a practical point of view. In a real process it is unlikely that the process operator will take immediate action when a control chart gives one signal.

When comparing AST results for different datasets, there is the problem that each dataset has a different number of time points. Hence comparing absolute AST values over datasets is not meaningful. In order to make a meta-analysis possible a relative AST (RAST) is calculated as the percentage of the total batch time $K$. For the different datasets used in this study, a total number of 13 faulty batches are available to test the detection ability of the various modelling types. For the same reasons as explained for the OT1 error, a global average of the RAST over all thirteen batches is taken. This is necessary for a sensible comparison between the RAST for the different models.

4.3.5 Model validation data
Six different types of datasets including NOC and faulty batches have been used to calculate the OT1 error and AST. These datasets have been described in earlier work (E.N.M. Van Sprang et al. [41]). Other references for a further detailed description of these datasets are given in the description of the datasets. This section briefly describes the most characteristic features of the datasets. An overview of the size of $X$ for the datasets is given in Table 5 as well as the occurrence of a disturbance in case of the available faulty batches.
Table 5

<table>
<thead>
<tr>
<th>Data set $(I \times J \times K)$</th>
<th>Disturbed batches</th>
<th>Fault start</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set 1 $(67 \times 15 \times 58)$</td>
<td>Batch 14</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Batch 17</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Batch 24</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Batch 49</td>
<td>46</td>
</tr>
<tr>
<td>Data set 2 $(27 \times 67 \times 271)$</td>
<td>Batch 2</td>
<td>138</td>
</tr>
<tr>
<td>Data set 3 $(47 \times 3 \times 100)$</td>
<td>Wrong batch 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Wrong batch 2</td>
<td>1</td>
</tr>
<tr>
<td>Data set 4 $(50 \times 9 \times 200)$</td>
<td>Batch 106</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Batch 99</td>
<td>100</td>
</tr>
<tr>
<td>Data set 5 $(47 \times 8 \times 116)$</td>
<td>Batch 41</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Batch 46</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Batch 47</td>
<td>94</td>
</tr>
<tr>
<td>Data set 6 $(36 \times 10 \times 100)$</td>
<td>Batch 50</td>
<td>56</td>
</tr>
</tbody>
</table>

Description of the data sets and disturbed batches.

4.3.5.1 Dataset 1
This dataset comprises 67 batches from a former industrial PVC process of Shell. Fifteen process variables were measured being mainly temperatures and pressures within the reactor and cooling system of the reactor. Three faulty batches are used for the AST. A more detailed description of this process can be found in A.A. Tates et al. [2].

4.3.5.2 Dataset 2
This data set is characterised by the spectroscopic measurements at 271 wavelengths of a two step laboratory bio-reaction. One faulty batch with a pH-disturbance is used for the AST. A detailed description of this reaction is given in S. Bijlsma et al. [66].

4.3.5.3 Dataset 3
This process describes the fat hardening process in a industrial batch reactor of Loders Croklaan (the Netherlands). Only three process variables are measured: the H₂ flow and temperature and pressure in the reactor. Two faulty batches are used with initial problems: the end-product quality of these batches was unsatisfying. More information can be found in A.K. Smilde & H.A.L. Kiers [39].

-118-
4.3.5.4 Dataset 4
This dataset is used in the pioneering work of P. Nomikos & J.F. MacGregor [6] and is a simulated styrene-butadiene polymerization process. Two faulty batches were simulated both with impurities of the feed.

4.3.5.5 Dataset 5
This dataset is an industrial multi-staged polymerization of a polymer product by DuPont (K.A. Kosanovich et al. [7]). Three batches with a process upset are available to calculate the AST.

4.3.5.6 Dataset 6
Dataset 6 is an industrial two-staged process from DuPont (P. Nomikos & J.F. MacGregor [19]). One faulty batch that is known to have poor end-product quality is available to calculate the AST.

4.3.6 RESULTS
This section discusses the outcomes of the performance indices for the three models described before. Also, a short evaluation is given about the chosen number of principal components as well of the amount of explained variation for each model.

4.3.6.1 Modelling
For every model, the chosen number of principal components (PC) for each dataset is given in Table 6:
### Table 6

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Explained variance</th>
<th>Evolving</th>
<th>Local</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(x100%)</td>
<td>R</td>
<td>(x100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>61%</td>
<td>3</td>
<td>67%</td>
</tr>
<tr>
<td>3</td>
<td>75%</td>
<td>2</td>
<td>86%</td>
</tr>
<tr>
<td>4</td>
<td>26%</td>
<td>3</td>
<td>30%</td>
</tr>
<tr>
<td>5</td>
<td>41%</td>
<td>3</td>
<td>53%</td>
</tr>
<tr>
<td>6</td>
<td>56%</td>
<td>3</td>
<td>64%</td>
</tr>
<tr>
<td>Average</td>
<td>52%</td>
<td>3</td>
<td>60%</td>
</tr>
<tr>
<td>2</td>
<td>99%</td>
<td>3</td>
<td>99%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>98%</td>
</tr>
</tbody>
</table>

Overview of the amount of explained variance and number of principal components for each model.

It can be seen from this table that for the global model a number of 3 components is chosen for all datasets except for dataset 3 where 2 components are chosen. This number of components described a sufficient part of the variation and this number of components resulted from cross validation methods (S. Wold [67]). The amount of explained variation at each time interval for dataset 1 is represented in Figure 52:

![Amount of explained variance for dataset 1](image-url)
The calculation of the amount of explained variation at every time interval is given by the equation in the first column of Table 6. It can be seen from Figure 52 that the explained variation is varying over time. The changes in the curve are caused by the start or ending of a different phase of the process. Apparently some phases in the process can be better described than others. The average amount of explained variation over time for dataset 1 can be found in the first column of Table 6 and is 61%. Such numbers are also calculated for the other datasets 3,4,5 and 6 resulting in an average of 52% as denoted in Table 6. The variation within some datasets is better explained than in others due to the differences in background of the datasets. The average amount of explained variation for dataset 2 is considerably higher compared to the other datasets. This is typical for spectroscopic data. This dataset was not considered in the calculation of the average over time and all datasets.

For the evolving model two principal components were chosen to describe the data. It requires an extensive analysis to decide on the number of principal components for each time interval $k$ and this may be considered as a serious drawback of the evolving model. Because it is not the goal to perform this analysis, two principal components are chosen for all datasets. Figure 52 gives an impression of the amount of explained variation over time for data set 1 using an evolving model. The equation to calculate this number is given in the fourth column of Table 6. Progressing over time, the amount of explained variation steadily decreases. This can be understood from the following. At every time interval new variation is added to the process. In the beginning two components are still enough to capture the variation, but by introducing more and more new events, the two components become less capable of fitting the data. Yet, the average explained variation over time and datasets 1,3,4,5 and 6 is 60%, which is sufficiently enough. This number is higher than for the global model (52%). For the same reasons as described before, the amount of explained variation is almost 100% for dataset 2.

For the local model one principal component was chosen to describe the datasets. As for the evolving models an extensive analysis is required to decide on the number of principal components for each time interval. Without going into detail, there is a clear indication that taking more than one component results in over fitting the data. This will be the subject of a follow-up. Again, the amount of explained variation over time for dataset 1 is given in Figure 25. As for the global model, the amount of explained variation fluctuates over time. The shape of both curves for the
local and global model is quite similar. The average amount of explained variation over time and dataset 1, 3, 4, 5, and 6 is 55%. This is remarkably high for just one principal component. Apparently the underlying phenomena present locally at each time interval can be well described by only one component. This makes the local model highly adaptive to changes within the process. Also for a local model the explained variation for dataset 2 is almost 100%. Due to the set-up of the simulated dataset 4, the amount of explained variation is low compared to the other datasets. This is not only true for the local model but also for the evolving and global model. The representativeness of this simulated dataset can, therefore, be questioned.

4.3.6.2 Overall type I error and RAST

The results found for the OT1 error and RAST for both the SPE and D-chart are presented in Table 7 and Table 8.

Table 7

<table>
<thead>
<tr>
<th></th>
<th>SPE</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evolving (2 pc’s)</td>
<td>5%</td>
<td>11%</td>
</tr>
<tr>
<td>Global (3 pc’s)</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>Local (1 pc)</td>
<td>7%</td>
<td>9%</td>
</tr>
</tbody>
</table>

*Overall type I error for the various models (average over all datasets).*

Table 8

<table>
<thead>
<tr>
<th></th>
<th>SPE</th>
<th>D</th>
<th>SPE or D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evolving (2 pc’s)</td>
<td>39%</td>
<td>93%</td>
<td>36%</td>
</tr>
<tr>
<td>Global (3 pc’s)</td>
<td>49%</td>
<td>69%</td>
<td>48%</td>
</tr>
<tr>
<td>Local (1 pc)</td>
<td>58%</td>
<td>82%</td>
<td>53%</td>
</tr>
</tbody>
</table>

*RAST of the various models.*

The actual OT1 error for the global model is reasonable for the SPE-chart (6%) and D-chart (3%). The same is true for the actual OT1 error of the SPE-chart (7%) for the local model, although the D-chart (9%) gives a slightly too high OT1 error. This is mainly due to the contribution of dataset 4, which is the simulated dataset. For the evolving model, the SPE-chart gives a satisfying outcome for the OT1 error (5%) whereas the OT1 error for the D-chart is slightly too high (11%) although not
disturbingly. Because the OT1 error for the models is more or less in the same range, it is possible to compare the outcomes of the RAST.

For a fixed alpha, the RAST not only depends on the type of model used but also on the degrees of freedom associated with the hypothesis test performed. That is, if $\alpha_{\text{repeal}}$ is fixed then the RAST depends on the degrees of freedom that are estimated at every time interval for the chi-square distribution for the $SPE$. The residuals calculated for a global model are based on more data than for e.g. the local model, therefore the degrees of freedom might be different. A study showed that the differences in the estimated degrees of freedom for the different models for the chi-square distribution were very small by comparing the $t$-values (more variation over time than over methods). This result supports the idea that the RAST really shows the differences between the models.

The evolving model gives the fastest RAST for the $SPE$-chart. On the average it takes 39% of the batch time to detect a fault using this model. Together with the good results for the OT1 error, the $SPE$-chart works fine for the evolving model. The RAST for the $D$-chart is high for all models. For the evolving model it even takes 93% of the batch time to detect a fault. It can be concluded from the results for the RAST error that most faults are detected in the $SPE$-chart. This result will be thoroughly discussed in Part II of this series.

The column ‘$SPE$’ and ‘$D$’ in

Table 8 indicate how fast a fault is detected in the ‘$SPE$’ or ‘$D$’ chart. Another aspect is ‘how fast is a faulty batch signalled’ in general? This is irrespective of whether it is detected in the $SPE$ or $D$-chart, since only the signalling itself is important for detection, not the particular chart in which this occurs. The last column of

Table 8 gives the answer to this question: irrespective of the type of chart, what is the RAST of the model, averaged over the faulty batches. Then it is clear that the evolving model detects the fastest (36%), followed by the global (48%) and local model (53%). Another aspect is, of course, fault diagnosis that might be easier in the $SPE$-chart than the $D$-chart. These aspects are covered in Chapter 0.

4.3.6.3 Example of a faulty batch

In this section an example is shown of a batch that is monitored using all three models. For this purpose, batch 24 (see Table 5) is chosen from the first dataset. This
batch contains a process fault at time interval six where the temperature in the top of the condensor is too high. A univariate plot of this process variable is given in Figure 53:

In this figure the solid lines represent the $3\sigma$ control limits and the dotted line the average value of the condensor temperature for all batches. The limits reveal that there is a lot of variation in this temperature signal around time interval 9 – 15, but for this particular batch this variation starts earlier than normal. Around time interval 9, the temperature is corrected and behaves normal again. The control charts for the global, evolving and local model are given in respectively Figure 54, Figure 55 and Figure 56.
Figure 54
*Monitoring a wrong batch in the SPE and D-chart for the global model.*

Figure 55
*Monitoring a wrong batch in the SPE and D-chart for the evolving model.*
Statistical batch process monitoring

The upper graph of these figures represents the SPE-chart and the lower graph the D-chart. The dotted lines represent the 95% confidence limits and the solid lines the 99% confidence limits. The value of the D-limits is different for the three models. This is because these limits are based on the number of principal components that are different for each model. The action signal for this fault is generated in the SPE-chart with the same RAST for all three models. Notice that false warnings are generated by the D-chart for the global model at time interval four and five.

The SPE-limits for the different models reflect more or less the amount of explained variation. If this amount of explained variation is high, the limits for the SPE-chart are low. Considering the value of the limits of middle part of the SPE-charts for all models, the data can be well described. The end of the batch is more difficult to describe considering the increasing limits for all three limits. This is also reflected in the amount of explained variation. The beginning of the batch is well described by the local and evolving model where the global model seem to have trouble describing the first part of the data. All three models have difficulties describing the data around time interval 10.

Figure 56
Monitoring a wrong batch in the SPE and D-chart for the local model.
4.3.7 CONCLUSIONS

The performance of three different models for batch process monitoring has been evaluated using two performance indices: the overall type I error and the relative action signal time. The results were obtained using six different datasets. The evolving and local model needs no filling in procedure where the global model does. The following conclusions can be drawn:

- The OT1 error and the RAST for the evolving model and local model give satisfying results. For this reason these models offer good and simple alternatives to the well established global model using current deviations.
- Most faults are first detected in the SPE-chart by all three models. The reason for this is discussed in Part II of this.
- The evolving model gives the fastest detection of a fault and this is an advantage compared to the other models. A fine-tuning can be made by carefully selecting the number of components for each time interval. This can also be considered as a drawback. However, with a relatively straightforward way of selecting the number of components, the evolving model performs well.
- The use of a local model can be considered as easy and straightforward. For the same reasons as for the evolving model, a fine-tuning can be made to select on the number of components. This might further improve the results found for the local model.
4.4 Batch Process Monitoring Using Online MIR Spectroscopy*

4.4.1 SUMMAR Y
Many high quality products are produced in a batch wise manner. One of the characteristics of a batch process is the recipe driven nature. By repeating the recipe in an identical manner a desired end product is obtained. However, in spite of repeating the recipe in an identical manner, process differences occur. These differences can be caused by a change of feedstock supplier or impurities in the process. Because of this, quality differences might occur in the end product quality or unsafe process situations arise. Therefore, the need to monitor an industrial batch process exists. An industrial process is usually monitored by process measurements such as pressures and temperatures. Nowadays, due to technical developments, spectroscopy is more and more used for process monitoring. Spectroscopic measurements have the advantage of giving a direct chemical insight in the process.

Multivariate statistical process control (MSPC) is a statistical way of monitoring the behaviour of a process. Combining spectroscopic measurements with MSPC will notice process perturbations or process deviations from normal operating conditions in a very simple manner. In the following an application is given of batch process monitoring. It is shown how a calibration model is developed and used with the principles of MSPC. Statistical control charts are developed and used to detect batches with a process upset.

4.4.2 INTRODUCTION
The batch wise production of high quality products such as pharmaceuticals, polymers or bio-chemicals plays an important role in process industry. A batch process is a recipe driven approach, that is, if a new process is operated in a similar manner as the previous one, according to the recipe, it is expected that a similar end product will be obtained. Since batch processes are often highly complex, it is not straightforward to maintain identical operating conditions and (undesired) variations between the batches.

may occur. There are many sources in a process that can cause these variations. As an example, a different initial feedstock or a change in heat exchange due to impurities. This uncontrolled variation may lead to poor quality end products or unsafe process situations. It is therefore desirable to keep the batch-to-batch variation in-control.

The batch-to-batch variation can be monitored using multivariate statistical batch process control (MSPC). MSPC for batch processes or batch process monitoring was initially developed by J.F. MacGregor & P. Nomikos [22] and, since then many extensions and applications have been presented in the literature (S. Rannar et al. [31], S. Wold et al. [53], D. Neogi & C.E. Schlags [44], D.J. Louwerse & A.K. Smilde [38], J.A. Westerhuis et al. [42], A.K. Smilde & H.A.L. Kiers [39]). A very convenient tool is the use of on-line spectroscopic measurements for batch process monitoring. It is a fast and, non-destructive measurement giving direct chemical insight of the spectroscopic active compounds present in the system. This is in contrast to engineering variables such as temperatures and pressures, which are indirectly related to the chemistry in the process. A list of examples of on-line spectroscopic measurements can be found in the review of J. Workman in J. Workman, Jr. et al. [69].

In the present study, batch process monitoring is performed. A batch polymerisation process of poly(methylmethacrylate) (PMMA) has been monitored using mid infrared (MIR) spectroscopy and a calibration model, combined with the principles of batch MSPC. The calibration model and the statistical theory are explained in the following section before considering the experimental set-up and instrumentation used.

4.4.3 Theory

4.4.3.1 Calibration
Projection to latent structures (or partial least squares, PLS) is a commonly used technique in spectroscopy to model a block of predictors $X$ (spectra) and a block of response variables $y$ (the concentrations), (P. Geladi [70]). PLS is a decomposition which maximises the covariance between the $X$ ($I \times J$) and $y$ ($I \times 1$) block. The matrices $X$ and $y$ are decomposed into a number of $R$ score vectors $T$ ($I \times R$), loading
vectors $\mathbf{P} \ (I \times J)$ and $\mathbf{q} \ (R \times 1)$, and residual matrices $\mathbf{E} \ (I \times J)$ and $\mathbf{f} \ (I \times 1)$. The PLS decomposition is given in matrix notation in equation 53 to 55.

$$\mathbf{X} = \mathbf{T}\mathbf{P} + \mathbf{E} \quad (53)$$

$$\mathbf{y} = \mathbf{T}\mathbf{q} + \mathbf{f} \quad (54)$$

$$\mathbf{T} = \mathbf{XW}(\mathbf{P}'\mathbf{W})^{-1} \quad (55)$$

where the matrix $\mathbf{W} \ (J \times R)$ is an orthonormal weight matrix.

Using the model parameters $\mathbf{W}$, $\mathbf{P}$ and $\mathbf{q}$, the concentration $\hat{y}$ is predicted on-line each time a new observation $\mathbf{x}_{\text{new}} \ (J \times 1)$ is measured according to equation 58 and equation 59.

$$\hat{\mathbf{t}}_{\text{new}} = \mathbf{x}_{\text{new}}\mathbf{W}(\mathbf{P}'\mathbf{W})^{-1} \quad (56)$$

$$\hat{y} = \hat{\mathbf{t}}_{\text{new}}\mathbf{q} \quad (57)$$
4.4.4 ON-LINE MONITORING

The principle of multivariate statistical batch process monitoring (BPM) is used to monitor the batch polymerisation reaction. Statistical batch process monitoring is divided in three phases according to the I.T.A.-trajectory; the Initial phase, the Training phase and the Application phase (H.-J. Ramaker et al. [71]).

The first phase is the initial phase. This phase consists of measuring and collecting historical process data from previous batch runs. In this study, the process measurements are MIR spectra. An example of the measured process spectra is given in Figure 57.

The second phase is the training phase. In this phase it is determined which of the historical batch runs are assigned as batches that are operating under normal conditions (NOC). After the selection, a process model based on a calibration set is developed to predict the concentration during the evolution of the process. Since the model is based on a calibration set which is representative for the batch process under NOC, it can also be used for statistical batch process monitoring. The third phase, is monitoring a new batch using the developed control charts in phase two. The model combined with a historical data set is used to determine the statistical control limits. The data set consists of batches that are operated under normal operating conditions.

A new batch is monitored by using equation 56. With every new observation $\mathbf{x}_{\text{new}}$, a new score vector $\mathbf{t}_{\text{new}}$ ($R \times 1$) is computed according to equation 57. The model parameter $\mathbf{P}$ is used to compute model residuals $\mathbf{e}_{\text{new}}$ ($f \times 1$) for each observation.
The predicted concentration \( \hat{y} \), the new score vector \( \hat{t}_{nw} \) and the residuals \( e_{nw} \) are used to compute the monitoring statistics.

In order to develop control charts with proper control limits, the historical NOC batches must follow the same procedure, as a new batch would do. This results in the vector \( y_i \) (\( N \times 1 \)), consisting of the predicted concentrations at each interval \( i \). The matrices \( \hat{T}_i \) (\( N \times R \)), consisting of the score vectors from the \( N \) batch runs at interval \( i \) and the residual matrix \( E_i \) (\( N \times J \)), which consists of the model residuals computed corresponding to interval \( i \). The NOC matrices are used to develop three types of control charts; the concentration chart, the \( D \)-chart and the \( SPE \)-chart.

### 4.4.4.1 Concentration chart

The control limits of the concentration chart are determined using the vector \( y_i \). The mean \( (\bar{y}_j) \) and standard deviation \( (\bar{s}) \) are computed from \( y_i \). The upper and lower control limits for each time interval \( i \) are given by

\[
\begin{align*}
UCL_i &= \bar{y}_j + t_{(N-1,\alpha/2)}\bar{s}, \\
LCL_i &= \bar{y}_j - t_{(N-1,\alpha/2)}\bar{s},
\end{align*}
\]

where \( t \) is the Student \( t \) value with \( N-1 \) degrees of freedom with a \( \alpha/2 \) significance level. The concentration chart monitors deviations in the predicted concentration compared to the average monomer decrease.

### 4.4.4.2 D-chart

The \( D \)-statistic is the Hotelling statistic \( T^2 \) in a reduced space \( R \). This statistic is a Mahalanobis distance between the center of that space and the new obtained scores \( \hat{t}_{nw} \). The \( D_i \) statistic for each interval \( i \) is computed by

\[
D_i = (\hat{t}_{nw} - \overline{t}) S^{-1} (\hat{t}_{nw} - \overline{t}) - \frac{R(I^2 - 1)}{I(I - R)} F(R, I - R)
\]

where \( \overline{t} \) are the mean scores obtained from the historical score matrix \( \hat{T}_i \). The matrix \( S_i \) (\( R \times R \)) is the variance-covariance matrix computed with the historical score matrix.
\(T\) and \(F(R,J-R)\) represents the \(F\)-distribution with \(R\) and \(J-R\) degrees of freedom. The control limits are obtained from the \(F\) distribution with 99\% significance level.

The concentration chart and the \(D\)-chart are directly correlated. That is, the \(D\)-chart defines a statistic based on the score vector and monitors the batch process in the reduced space and the concentration chart monitors the concentration that is predicted using the score vector. Therefore, the concentration chart and \(D\)-chart will give more or less similar results.

### 4.4.4.3 The SPE-chart

The SPE statistic is a residual statistic and equals the sum of squared residuals summed over all wavelengths \(J\) at interval \(i\).

\[
SPE_i = \sum_{j=1}^{J} e_i^2 \sim \chi_i^2
\]  

(62)

The SPE statistic follows a weighted \(\chi^2\) distribution with \(b\) degrees of freedom and a weight \(g\). The parameters \(b\) and \(g\) are estimated from the NOC residuals \(E_i\). If the SPE statistic is large, it means that the spectrum is different from the spectra of the calibration samples. This could mean that a new compound is present in the system, or that the probe is fouling.

### 4.4.5 EXPERIMENTAL

The reaction of interest is a batch polymerisation of methylmethacrylate (MMA) to polymethylmethacrylate (PMMA) via a free radical polymerisation process. Azobisisobutyronitrile (AIBN) has been used as the free radical initiator. The reaction scheme is shown in Figure 58.

![Figure 58](AIBN initiated polymerisation of methylmethacrylate.)

The polymerisation was performed in a 500 ml reactor with heating/cooling jacket under a nitrogen atmosphere to prevent termination or inhibition by impurities. The reaction was constantly stirred.
using a magnetic stirrer and kept at a temperature of 80°C using a temperature controlled water bath (Thermo NESLAB). An initial solution of 0.36 mole methylmethacrylate in 0.89 mole toluene was heated at 80°C prior to the start of the reaction. The reaction was initiated using a solution of 0.9 mmole azobisisobutyronitrile in 0.04 mole toluene.

The reaction was monitored by collecting MIR spectra using the SpectraProbe Linx MID spectrometer from SpectraProbe Ltd. (UK). The instrument utilised a two bounce silicon attenuated total reflection (ATR) crystal yielding spectra in the range 1020 cm\(^{-1}\) to 1936 cm\(^{-1}\). The detector is a patented MID IR detector, which operates at room temperature and requires no cooling elements. The detector is a fixed 128 pixel pyroelectric array, which SpectraProbe has developed for use in the Mid IR range and could collect radiation from 2.5 to 10 microns. Furthermore, the detector is currently fixed to see the 5 to 10 micron range. The detector uses a grating rather than an FTIR system to address the detector array. The spectral resolution of the detector is in the range of 4cm\(^{-1}\) @1020cm\(^{-1}\) and 16cm\(^{-1}\)@1923cm\(^{-1}\). Spectra were collected over a four minute period and compared against a ten minute background spectra previously acquired in toluene as an auto-zero of the instrument.

4.4.6 RESULTS AND DISCUSSION
An example of the measured process spectra is given in Figure 59. The spectra are taken from a single batch run prior to injecting the initiator and the bands of the spectroscopic active compounds have been assigned according to A. Olinga et al. [72].

![Figure 59](image_url)

*Figure 59*  
*Process spectra with band assignment according to A. Olinga et al. [72].*
A PLS calibration model has been constructed to predict the monomer concentration during the polymerisation. The calibration model was based on 12 standard solutions with known concentrations of methylmethacrylate (MMA) / polymethylmethacrylate (PMMA). Each calibration sample was measured five times at 80 °C.

Prior to modelling, wavenumber selection and data pre-processing were performed, resulting in the use of the wavenumber range of 1450 cm\(^{-1}\)-1140 cm\(^{-1}\). This range was chosen to follow the decrease of MMA. Data preprocessing was applied by taking the first derivative spectra using a Savitsky-Golay filter (window size 3) and the number of PLS components was determined by cross validation. Three components were used in the PLS model and the concentration was predicted with a 5 % precision. The calibration concentrations versus the estimated concentrations are given in Figure 60.

The results of monitoring the monomer concentration of batch 1 are given in Figure 60 where the decrease of the monomer during the process is shown. After the initial five measurements, the initiator was added to the system and the reaction started. This is reflected by the decrease of the monomer concentration in time from this point.

A total of 18 batch runs obtained under normal operating conditions (NOC) were available for developing statistical control charts. One batch is used to validate the model and 2 batch runs with known process perturbations were available for testing the control charts. A description of the faulty batches is given in Table 9.
Table 9

<table>
<thead>
<tr>
<th>Batch no.</th>
<th>process upsets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>normal batch, no upsets</td>
</tr>
<tr>
<td>2</td>
<td>different operating temperature</td>
</tr>
<tr>
<td>3</td>
<td>small addition of styrene in the system</td>
</tr>
</tbody>
</table>

*Batch runs with known process upsets.*

The behaviour of the batch was analysed and checked under normal operating conditions and the developed control charts used to study the behaviour. The batch 1 measurements, given in Figure 61, were monitored and are presented in Figure 62 A, B and C. In Figure 62 A, the concentration chart is given with 99 % control limits. The first measurement prior to injecting the initiator is set as time $t=1$. It can be seen that the batch behaves within the specified limits and therefore is statistically in-control. The $D$-chart in Figure 62 B, which is related with the Concentration chart, gives a similar picture. Figure 62 C shows the normalised $SPE$-chart with 99 % control limits. From this figure it can be seen that the process has no unknown phenomena or unmodelled variation. The batch behaves under normal conditions.

The control charts have been used to monitor the perturbed batches given in Table 9. Batch 2 has been run at a lower temperature ($70^\circ$C). The three control charts are given in Figure 63 A, B and C. The Concentration chart and $D$-chart indicate that the first measurement is on the border of acceptable variation, the second measurement is clearly out of control. The concentration chart shows that the monomer is decreasing very slowly indicating that there is a slow conversion. The $D$-chart supports the concentration chart and shows that the process is deviating from normal conditions as time evolves. The $SPE$-chart shows that no unmodelled or non-linear variation is detected. This is expected since the absorbing compounds do not change and therefore the spectra fit the model.
Figure 62
Monitoring batch 1 using statistical control charts.

Figure 63
Monitoring batch 2 using statistical control charts.
Batch 3 had small fractions of polystyrene added to the system. The results are given in Figure 64 A, B and C. The fraction polystyrene is added after 10 minutes. It seems that the addition of polystyrene is affecting the process kinetics. This is shown in the concentration chart and the D-chart. The conversion of MMA is slowing down due to the polystyrene addition. Polystyrene is a new compound in the system. A new or unknown compound should be observed in the SPE-chart. However, in this specific situation, this is not the case. The SPE-chart gives no out of control signal since polystyrene absorbs around 699 cm⁻¹. This wavenumber is not used in the current calibration model and is therefore not detected in the SPE-chart.

![Graphs A, B, C](image)

Figure 64

*Monitoring batch 3 using statistical control charts.*

### 4.4.7 CONCLUSION

A batch polymerisation process was monitored on-line using mid infrared spectroscopy with a Spectraprobe Linx spectrometer. The instrument is compact and easy to hand over to an industrial process. A calibration model has been constructed and used to monitor the decrease of the monomer concentration. The concentration is predicted with a 5 % precision. Furthermore, the calibration model is used together with the principles of statistical batch process monitoring and different batches are monitored using control charts. It has been shown that the control charts easily detect unusual behaviour and that statistical process control is therefore a useful method for monitoring batch-to-batch variations.
CHAPTER 5  ♦ GREY MODELS*

*This chapter is based on the following publication(s):


5.1 General introduction*

The building of process models is done for various reasons. In the situation of batch process monitoring the aim of the model is to describe the process under normal operating conditions. Next, new batches are fitted to this model and the model fit is used as a measure how well the new batches behaves according to normal operating conditions.

The model choice is not univocal; there are many options to describe a process. With respect to the model choice there are two extremes possible. The first is a mechanistic model based on first principles such as process kinetics or energy balances. These models are sometimes called hard or white models. The second is a data driven approach such as neural nets or PCA. These models are sometimes called soft or black models. Each of these models has advantages and disadvantages. A short overview is given in Figure 65.

Grey Models: Idea

<table>
<thead>
<tr>
<th>Hard models</th>
<th>Grey models</th>
<th>Soft models</th>
</tr>
</thead>
<tbody>
<tr>
<td>First principles</td>
<td>External information</td>
<td>Empirical models</td>
</tr>
<tr>
<td>PDE, ODE</td>
<td></td>
<td>Latent Variables</td>
</tr>
</tbody>
</table>

- Good extrapolation
- Adaptive
- Fundamental insight
- Poor fit to data
- Poor monitoring

- Good extrapolation
- Adaptive
- Fundamental insight
- Good fit to data
- Good monitoring

- Good fit to data
- Good monitoring
- Poor extrapolation
- Hardly adaptive
- Poor fund. insight

Figure 65
Overview of white, black, and grey models.

* HJR/FES.
White models are characterised by their fundamental insight and adaptive nature. Furthermore, the models can be extrapolated. A disadvantage is that the models do not always have a good fit and do not model batch-to-batch differences. The black models are characterised by their high fit and ability to model batch-to-batch differences. A disadvantage of the black models is that the models are hardly adaptive and cannot easily be extrapolated. Furthermore, the models give a poor fundamental insight.

The idea of grey (hybrid) models is to combine both black and white models to have the advantages of both models. That is, the ability to model the batch-to-batch differences with a model that has a good data fit and a good model interpretation.

This chapter consists of three sections. The first section introduces the concepts of incorporating external process information in order to improve the performance of batch process models for statistical batch process monitoring. It is explained how two types of external process information can be distinguished and how these can be incorporated into the batch process models. Finally, a small example is given of how a grey model is used to monitor a biochemical reaction.

The second section describes the construction of a grey model of an industrial batch process, which is monitored on-line with NIR spectroscopy. It is shown how external process information is collected and how this is used to construct a grey model. Next, the grey model is used to follow the behaviour of the batch process and compared with other approaches to monitor the batch behaviour.

The third section gives a full strategy for statistical batch process monitoring using grey models. Two types of grey models are introduced and the derivation of these models is explained in detail. The two models are used for statistical batch process monitoring and their performance is validated. The section concludes with a summary of the differences and similarities between the models and recommendations for use in practice are given.
5.2 Improved monitoring of batch processes by incorporating external information*

5.2.1 SUMMARY
In this Section an overview is given of statistical process monitoring with the emphasis on batch processes and the possible steps to take for improving this by incorporating external information. First, the general concept of statistical process monitoring of batches is explained. This concept has already been shown to be successful according to the number of references to industrial applications. Incorporating external information can enhance the performance of statistical process monitoring of batch processes. Two types of external information can be distinguished: batch-run specific and process specific information. Various examples of both types of external information are given. Several ideas of how to incorporate the external information in model development are discussed. The concept of incorporating process specific information is highlighted by an example of a grey model. This model is applied to a biochemical batch process that is spectroscopically monitored.

5.2.2 INTRODUCTION
In the field of process analysis, process monitoring and process control, a common aim is the building of models that describe the nature or sources of variation in time. The choice of model to use depends not only upon the particular purpose of the model (e.g. data exploration, optimization, monitoring and process control), but also upon the type of process itself. One option is to apply first-principle engineering knowledge such as mass/energy balances, reaction kinetics and stoichiometries etc. However, a full first-principle model is usually not available, either because the complexity of most industrial processes, or the time and expense required, prevent the

development of such a model. For example, a biochemical process such as the production of antibiotics is far too complex to be described by a physicochemical model.

An alternative approach is to apply empirical models, based on subspace methods such as principal component analysis (PCA), to the measured data. These methods are well understood and widely used in the field of process analysis. Automation of modern process plants increases the potential of these data driven approaches. Chemical reactors are typically equipped with various process analyzers such as temperature, pressure and mass-flow sensors. Data from these sensors are continuously stored in databases by fully automated systems. Therefore, enormous amounts of historical data are available. Subspace methods are able to make the data more comprehensible by describing it in terms of factors (principal components) representing the main multivariate trends underlying the data. The models are purely data driven, and the factors are identified solely in terms of describing maximum variance in the data. They are not necessarily easily interpretable as to their underlying physicochemical cause. For example, principal components are usually constrained to be orthogonal, although there is no reason to believe this is true of the actual sources of variation in the process. Empirical models can, thus, be limited in their use in further understanding of the process through model interpretation. Furthermore, they have rather poor extrapolation properties, being data driven.

A third way is to combine the two approaches, i.e. incorporate additional information into empirical models of the process data. Although only partial information about the physicochemical nature of the process may be known, this should not be ignored, but can very well be used in combination with an empirical model so as to provide as full a model of the process as possible. Some advantages of this approach are discussed later, but it can be said here that in a recent discussion of statistical process control. W.H. Woodall [48] noted that the incorporation of external information may help to improve process 'understanding, modeling and reducing variability over time', as well as narrowing the gap between the industrial engineers (who usually prefer first-principle models) and academic statisticians (who tend towards the use of empirical models which describe the data well).

The processes considered in this thesis are batch processes, although the methodology is equally applicable to continuous processes and, indeed, chemical
process modeling in general. Furthermore, the main focus is on modeling with the aim of statistical process monitoring, although again the methodology discussed is also suitable for other goals.

Many chemical, pharmaceutical, and food products are made in batch processes. Batch processes are characterized by finite duration, non-steady-state behavior, high conversions and, most importantly, the recipe-driven approach. Monitoring the behavior of such batch processes is important for process optimization, meeting specification limits and safety. Pioneering work in the area, called batch process monitoring has been done by Nomikos and MacGregor (P. Nomikos & J.F. MacGregor [6]) has been successfully applied to real processes on several occasions (K.A. Kosanovich et al. [7], D. Neogi & C.F. Schlags [44], A.A. Tates et al. [2]).

The goal of this section is to classify different types of external information, and to describe how this information is used to improve the monitoring of batch processes.

5.2.3 ADVANTAGES OF INCORPORATING EXTERNAL INFORMATION

The goals of batch process monitoring are threefold:

i) Detecting abnormal variation
ii) Locating in which group of variables of the process the abnormal variation occurred.
iii) Finding and solving the cause(s) of the abnormal variation.

The idea of using external information is that with respect to at least one of the three goals of batch process monitoring, better results are obtained. Hence, either the detection capability is improved and/or the diagnostic capabilities (locating and finding the causes for the fault) are improved. To some extent this idea has already been proven to be valid. This is discussed later.

5.2.4 TYPES OF EXTERNAL INFORMATION

There are different types of external information. Two categories of external information can be distinguished: batch-run specific information and process specific information. These two categories will be discussed in the following sections.
5.2.4.1 Batch-run specific information

Batch-run specific information can be knowledge or additional measurements associated with individual batch-runs. A clear example of such information is the product quality. Each batch-run leads to a product with a (unique) specific quality. Preferably the quality of all batches would be exactly equal, but in practice this is not the case. So the quality measurement of the product gives additional information on how the batch was run. Another example of batch-run specific information is the quality of the feed that is used at the start of the batch-run. Consider for example a fat hardening process where the feed consists of soybeans. Due to changes in the origin of the beans and the time of year, the product quality may change significantly. Additional batch-run specific information can range from a simple list of the different countries from which the beans were obtained to a GC analysis of the triglycerides in the beans. Other examples of batch-run specific information are:

- **Reactor conditions**: The aging of a catalyst or the initial charging of the reactor with the feed.
- **Cleaning of the reactor**: After completion of a batch the reactor is often cleaned. If impurities remain, cross-contamination can cause problems. The cleaning history can be of particular interest in such a case.
- **Reactor number**: A recipe can be used for multiple reactors. Each separate reactor may have its own characteristics due to (un)known causes.
- **Environmental conditions**: Seasonal influences may affect important reaction conditions such as ambient-temperature, humidity etc.
- **Upstream conditions**: In some cases, the feedstocks used are the result of a previous chemical process with its own variability. Knowledge of this upstream process can be incorporated. Conditions of prior storage can influence the properties of reactants or additives. For example, the additives needed to perform the reaction are stored in a tank or a cold store.
- **Batch duration**: The duration of a batch should be the same for each batch-run due to the recipe-driven approach. Sometimes this is not the case.
5.2.4.2 Process specific information

Process specific information relates to the nature and sources of information within the process itself. This means that although each batch-run may be slightly different from the nominal average batch-run, the underlying physicochemical causes influencing each batch are the same for a given process. An example of process specific information may be a set of ordinary differential equations describing the relationship between temperature and pressure. These equations should be applicable to each batch-run for a given process in a given reactor. Another example would be where the kinetics and/or the stoichiometries for one or more of the reactions occurring within the reactor are known. Again, this information will apply to each batch-run, although the exact value of kinetic rate constants may vary slightly from batch to batch. More examples could be:

- **Dosing profiles:** For a fed batch reaction, additives are supplied according to a recipe. The exact time of addition can be of interest in such batches.
- **The ideal trajectories of process variables:** Since batch processes are recipe driven, the ideal trajectories should be the same for each batch-run. Such ideal trajectories might be known from theory and process data can be expressed as deviations from these ideal trajectories.
- **Process stages:** Some batch processes are carried out in multiple stages e.g. after a specific time the reactor is depressurized.
- **Grouping of process variables:** Highly correlated variables such as a number of temperatures in one reactor or wavenumbers in spectra can be grouped together and scaled appropriately.
- **Instrumental characteristics:** Sometimes characteristics such as accuracy and reliability are available.
- **Absorbance spectra of pure components:** In case of on-line spectroscopic monitoring, the absorbance spectra of pure components are often known and can be incorporated as a source of external information.

It is important to distinguish between these two types of external information because the methods to incorporate such information in batch process monitoring methods differ between the two categories.
5.2.5 Methods to Incorporate Batch-run Specific Information

The way of using the additional batch-run specific information depends on the type of information. If the information available is influenced by the batch-run, such as the product quality, then the arrangement of data depicted in Figure 66 can be used. In Figure 66, the process measurements of the first batch-run are collected in the first horizontal slab of $\mathbf{X}$. The corresponding product quality measurements are collected in the first row of $\mathbf{Y}$. There are different ways to deal with this data structure. One of the ways is to rearrange the three-way array $\mathbf{X}$ to a two-way array $\mathbf{X}$ and use a PLS regression model to build a relationship between $\mathbf{X}$ and $\mathbf{Y}$ (P. Nomikos & J.F. MacGregor [73], T. Kourt et al. [74]). An alternative route is to leave the three-way array $\mathbf{X}$ intact and build directly a regression model between $\mathbf{X}$ and $\mathbf{Y}$ using multiway regression (R. Boqué & A.K. Smilde [32], R. Bro [75]). Both mentioned alternatives have the property that the empirical model of $\mathbf{X}$ is steered towards describing directions of variation important for product quality.

Apart from being able to predict product quality for an ongoing batch-run, this sometimes amounts in better detection and diagnosis of abnormal behavior (R. Boqué & A.K. Smilde [32]).

There is also a disadvantage in this approach. Process variables which might not be of direct importance for the product quality tend to get less weight in the model of $\mathbf{X}$. Therefore, abnormal variation in such process variables that may cause a process disturbance, might not be detected. In order to overcome this problem, a different type of PLS model could be built that also describes the variation in $\mathbf{X}$ that is not relevant for the product quality. This could be arranged by using an orthogonal signal correction approach. Orthogonal signal correction (OSC), (S. Wold et al. [76]) removes information from $\mathbf{X}$ that is uncorrelated with the product quality $\mathbf{Y}$. After the removal, a PLS model is built between the remainder of $\mathbf{X}$ and $\mathbf{Y}$. In a
monitoring scheme, the OSC components as well as the PLS components can be used. In this way, there is a split between the variation in $X$ that is relevant for $Y$ and a part that is not relevant for $Y$. If both parts are monitored separately, this reveals additional information in case of an out-of-control situation. It can be determined whether the process upset affects the product quality or whether it does not. Such information might be valuable for deciding whether the process should be stopped. Furthermore it might be helpful in fault diagnosis since some types of disturbances will affect the product quality and others will not.

If additional batch-run specific information is available before the batch-run is started such as the feed quality, up-stream information or information on the history of the autoclave, this may be used to set specific parameters at the start of the batch-run. Also this information can be used to give an early prediction of the product quality that might be produced. It is possible to build monitoring charts based only on this external information. When such a chart alarms, it is better not to start the batch-run.

For on-line process monitoring of the batch-run, information on the feed can be used in the following arrangement according to Figure 67:

In this figure, measurements of the feed quality of the first batch are collected in the first row of $Z$, and these measurements correspond to the measurements of the first batch obtained during the batch-run. A multiblock component model is built in this case. The final components $T$ are weighted averages of the principal components of the additional information ($Z$) and of the principal components of the batch process measurements ($X$). In the monitoring phase both types of information are used to decide whether the batch is in-control. In case of an out-of-control situation, the model can be interrogated and may reveal information as to whether the process upset was mainly
caused by a feed of inferior quality or because during the batch-run a disturbance was present or a combination of both.

Again there exist different alternatives to deal with the data structure as presented above. One alternative is to rearrange $X$ into a two-way array and apply two-way multiblock models such as consensus PCA (S. Wold et al. [77]). Another route is to leave the multiway array $X$ intact and apply a multiway multiblock component model directly to the data (A.K. Smilde et al. [78]).

If both types of batch-run specific information are available, e.g. product quality and up-stream information, than the arrangement can be extended to a multiblock regression model according to Figure 68:

In this way the final components $T$ are weighted averages of the components of the additional information and of the components of the batch process measurements, but they are steered towards describing directions of variation important for product quality. This gives the same advantages as described above when only the product quality $Y$ was available as additional information. Again there are different methods to deal with such an arrangement (J.F. MacGregor et al. [79], A.K. Smilde et al. [78], J.A. Westerhuis et al. [80]).

New methods are in the process of being developed that use the batch-run specific information that is known before the batch-run is started to set process parameters to specific values. E.g. for a certain type of feed it might be necessary to use higher temperatures during the run or extend the duration of the run. Furthermore, it can be predicted whether low quality feed can still be processed to inspect products with the extra effort during the process. This will lead to different limits on the feed specifications, and maybe feed of lower quality can be used, which is of course economically beneficial.
Summarizing, some methods already exist to incorporate certain types of additional batch-run specific information. These methods may produce models that are better in detecting special variation in the process or that are better in diagnosing the cause or specific process disturbances. Other methods are still being developed and more research and experience is necessary.

5.2.6 METHODS TO INCORPORATE PROCESS SPECIFIC INFORMATION
Incorporating process specific information in empirical models can be done in several ways. Some examples of how to incorporate process specific external information follow.

5.2.7 DATA PREPROCESSING AND WEIGHTING
Available first-principles engineering knowledge can be used to derive new variables based on those already measured. Just as the use of logarithmic transforms can sometimes help remove non-linearity inherent in the data, these new variables can be used to provide data that are better conditioned for empirical linear models. Another possibility lies in the use of weighted least-squares algorithms as an alternative to ‘blind’ scaling whereby each variable is simply divided by its standard deviation. To estimate parameters in an empirical model of $X$, a least-squares optimization is often used. Different weights can be given to different process variables based on the signal-to-noise ratio of each measurement, with noisy measurements being given a low weight. In a similar way, the use of appropriate data preprocessing can also be used to incorporate process specific information about the noise levels of the process measurements into the model, by scaling down noisy variables. Finally, if it is known that the process variables fall into known groups, for example a set of 1001 spectral absorbances at different wavelengths and two temperature measurements, it is sensible to block the variables in two groups and to scale each group separately so as to give each measurement block appropriate significance.

5.2.8 GREY MODELS
Hybrid or semi-mechanistic models, as described in the engineering literature (D.C. Psichogios & L.H. Ungar [81], M.L. Thompson & M.A. Kramer [82], R. Simutis et al. [83], H.A.B.T. Braake et al. [84], P.M. Bentler & S.Y. Lee [85]), combine two different
types of models, typically by fitting a first-principles model to the data and then using an empirical model to fit any remaining systematic variation not already fitted. An example would be the use of a first-principles model to describe the molecular weight distribution during polymer production and then applying a neural network to the residuals to account for further systematic, but non-ideal, behaviour. The advantage of this approach is that it is possible to combine models with totally different characteristics, for example the hard model may be linear and the empirical model non-linear. Furthermore, it is possible to discover to what extent the measured data corresponds to the proposed hard model of the process. A disadvantage of the hybrid modelling approach is that the sequential nature of the fitting (i.e. hard model followed by soft model) means that the overall fit is not necessary optimal in a least-squares sense, although an iterative fitting procedure could be envisaged.

Another way of applying first-principles models is known as data reconciliation. This technique adjusts the measured data, which contains random and systematic errors, so as to conform to known mechanistic models such as mass and energy balances (C.M. Crowe [86], H.W. Tong & C.M. Crowe [87]).

In the psychometrics and chemometrics literature, some attention has been paid to constrained least-squares models, whereby the model parameters are restricted to reflect the underlying physicochemistry of a process. For example, if it is known that a concentration profile to be estimated should be non-negative and/or unimodal, then the model parameters can be constrained appropriately (A. de Juan et al. [88], R. Bro & S. De Jong [89]). An extension of this approach to batch process data has been called grey modelling.

Grey models are hybrid models in that a fundamental part is combined with an empirical part. The white part of the model is constrained to fit known external process information ('Known causes'). The black part of the model is unconstrained and fits further systematic variation, which is not explained by the white part ('Unknown causes'). Together, the white and black parts form the grey model, as shown in Figure 69.
Unsystematic variation in the data is not modelled and forms the residuals. Unlike the hybrid model described above, grey models can be estimated as a single least squares optimisation function for which multiple constraints of a various nature are applied according the external process specific information available. The advantage here is that the model parameters are calculated simultaneously, which can be shown to be beneficial in isolating the white and black sources of variation (S.P. Gurden et al. [90]). For batch process data, the combination of grey and three-way models is found to be particularly useful because of the ease with which external information about each of the modes of the three-way array (e.g. batch number, process variable or wavelength, time) can be incorporated.

5.2.8.1 Grey model example
As an example, consider the two-step consecutive reaction,

\[ A + B \stackrel{k_1}{\rightarrow} C \stackrel{k_2}{\rightarrow} D + E \]

where \( k_1 \) and \( k_2 \) are pH dependent. This reaction is monitored using UV-Vis spectroscopy, with only A, C and D being spectroscopically active (S. Bijlsma et al. [91]). If a series of batch runs are measured then a three-way array of spectroscopic data is produced: \( X \) (batch number × wavelength × time). In building a model of the measured data, three pieces of external information are used to form the white part of the model: (i) the pure spectra of A and D are known; (ii) the kinetic scheme (although not the rate constants) are known; (iii) the Beer-Lambert law of linear additivity is applicable. These pieces of information were incorporated into the model to form the white part, which now reflects to what extent each individual batch-run obeys the kinetics and spectra for this process. Using this approach, it was also possible to estimate the pure spectrum of the intermediate compound, C, and the actual values of the rate constants \( k_1 \) and \( k_2 \) for the two reactions. Two black model components were necessary to describe other sources of variation, in this case
changes in $k_1$ due to temperature or pH fluctuation and a ‘spike’ which was an artifact of the spectrometer. A more detailed treatment of this example is given elsewhere (S.P. Gurden et al. [90]).

The grey model can be used for post-batch analysis as well as for on-line monitoring of a running batch (J.A. Westerhuis et al. [46]). An example of this is shown in Figure 70, where it is clear that a fault occurs after 26 minutes in a new batch. This agrees with reality because in this batch the pH was deliberately changed slightly after 26 minutes and the reaction is pH sensitive.

![Figure 70](image)

**Figure 70**
*On-line use of a grey model. After 26 minutes the SPE breaks the 95% (dotted line) and the 99% (solid line) control limits.*

Summarizing, there are some methods available in data analysis to deal with incorporating process specific information. The potential of these methods is not fully explored yet.

### 5.2.9 CONCLUSIONS

It is shown that external information related to batch processes is readily available in practice. This external information can be divided into two groups: batch-run specific and process specific information. The examples shown for both types of external information are just a grab from all possible sources of external information. Different methods are in the process of development to improve batch process monitoring by incorporating external information. The potential of these methods to incorporate external information is not yet fully explored. Therefore more applications are needed in order to develop the methods to their full capability.
5.3 NIR Spectroscopic monitoring of a series of industrial batch processes using a bilinear grey model*

5.3.1 SUMMARY
A good process understanding is the foundation for process optimisation, process monitoring, end-point detection and estimation of the end product quality. Performing good process measurements and the construction of process models will contribute to a better process understanding. To improve the process knowledge it is common to build process models. These models are often based on first principles such as kinetic rates or mass balances. This type of models is also known as hard or white models. White models are characterised by being generally applicable but often having only a reasonable fit to real process data. Other commonly used types of models are empirical or black-box models such as regression and neural nets. Black-box models are characterised by having a good data fit but lacking a chemical meaningful model interpretation.

Alternative models are grey models, which are combinations of white models and black models. The aim of a grey model is to combine the advantages of both black-box models and white models. In a qualitative case study of monitoring industrial batches using NIR spectroscopy, it is shown that grey models are a good tool for detecting batch-to-batch variations and an excellent tool for process diagnosis compared to common spectroscopic monitoring tools.

5.3.2 INTRODUCTION
To achieve consistent end product quality in industrial batch processes, it is required to have a consistent batch process operation. Therefore, batch processes are carried out according to a predefined recipe in which reactants are added with a specific flow rate or in certain volumes at specific times. Moreover, process parameters such as dosing rates, temperatures, chilling water, flow rates etc. give the input parameters for

---

industrial process control and process performance monitoring. The traditionally measured variables give an insight into the thermo- and hydrodynamic properties of the system, but not directly into specific chemical compounds. By applying spectroscopy directly to the full-scale process as another measurement, chemical information is added to the process operation. These spectra may contain information about (i) the nature of the spectroscopically active compounds present in the reactor, (ii) the changes in concentration of these compounds during the batch run and (iii) the physical changes (temperature, particle size distribution) of the reaction mixture. This provides a new opportunity for process understanding, monitoring and control at a level previously not possible with standard temperatures and pressures: process analysis at a chemical level. Common examples of successful applications of on-line spectroscopy in the process industries are the on-line analysis of alkane gas mixtures (H.F.M. Boelens et al. [92]), the on-line analyses of the blend homogenity (S.S. Sekulic et al. [93]), end-point detection (H.W. Ward, II et al. [94]), the characterisation of batch reactions (P.J. Gemperline et al. [34]) and the statistical monitoring of batch processes (J.A. Westerhuis et al. [46], R. de Maesschalck et al. [95]).

In the field of spectroscopy, the use of empirical modelling techniques such as Principal Component Analysis (PCA), Multivariate Linear Regression (MLR) or Projection to Latent Structures (PLS) are common tools to achieve the desired purposes such as process monitoring or end product quality prediction. These tools generate empirical or black box models. Such models fit the data properly, but cannot be generalised to different situations and do not always generate process insight. White models (models based on first principles), on the other hand, do provide insight and are more general. However, white models do not fit the process data that well, and are not able to explain batch-to-batch differences.

Grey (hybrid) models try to combine the advantages of black and white models. A special advantage of grey models is the ability to include known external information, e.g. fundamental knowledge of the process at laboratory scale. The grey model fits the process data using the external information as constraints on the model parameters. In this way fundamental knowledge and empirical models are combined. The combination of external process information with empirical models has shown many successful applications. Examples are: estimating the kinetic constants from
spectroscopic measurements (S. Bijlsma et al. [66]), exploring for unknown kinetic mechanisms (E. Furusjö et al. [96]), obtaining an improved process understanding (S.P. Gurden et al. [90]) and curve resolution (E.A. Sylvestre et al. [97], R. Tauler & D. Barcelo [98]). For a more detailed discussion of grey models, see S.P. Gurden et al. [90].

The aim of this study is twofold: First, to show how a grey model is constructed using in-line industrial NIR spectroscopic measurements from production batches. Second, to give a qualitative performance of the grey model compared to commonly used spectroscopic-tools in testing of the batch-to-batch consistency.

5.3.3 Theory
5.3.3.1 Structure of the in-line spectroscopic batch wise NIR data
It is assumed that the Lambert-Beer law holds, that is, there is a linear additive contribution of the R pure chemical compounds to the response spectra. For a single batch run, this is given by

\[ X = \sum_{r=1}^{R} c_r s_r^T + E = CS^T + E \]  (63)

where \( X \) (\( I \times J \)) are the measured spectra of \( I \) subsequent time intervals at \( J \) wavelengths, \( c_r \) (\( I \times 1 \)) is a vector consisting of the concentration profile for compound \( r \) and \( s_r \) (\( J \times 1 \)) is the corresponding normalized pure compound spectrum. The matrix \( E \) consists of unsystematic variation such as noise.

Visualisation of process spectra
Measured process spectra contain a large amount of chemical information about the process. However, extracting this information is sometimes difficult since an overload of process spectra is easily obtained due to the nature of the technique. There are several possibilities for extracting the relevant chemical information from the spectra. Some of these are discussed in the following.
Chapter 5 • Grey models

Wavelength tracking
A simple way of looking at time resolved process NIR spectra is to use prior process information such as the knowledge of the major compounds present in the reaction, along with knowledge of the pure spectra of these compounds which may be found in the literature. Specific compound information relating to concentration changes is obtained by monitoring the response of a selective band corresponding to a chemical compound as a function of batch process time. However, with many types of spectroscopy it is not always possible to monitor the compounds of interest individually, since there is a lack of selectivity. NIR spectroscopy, for example, often consists of broad and overlapping peaks, which means that each compound will probably not have a selective peak. Another disadvantage of this strategy, is that only a small fraction of the available information present in the spectra is used, e.g. an unknown interferent showing up in a particular batch run absorbing at a wavelength not selected goes unnoticed.

5.3.3.2 Range charts and change charts
To obtain a general idea about the evolution of an individual batch run given the measured spectra, process-monitoring charts such as range- or change charts can be used (F.A. DeThomas et al. [99]). The range chart \( RC \) shows the difference between two successive spectra in time, pair wise taken at each individual wavelength \( j \) and summarised over all \( J \) wavelengths.

\[
RC_j = \sum_{i=1}^{J} (x_y^i - x_y^{i-1})^2
\]

where \( x_y \) is absorbance measured at the \( i \)-th step in time at wavelength \( j \). Instantaneous changes are captured by this control chart.

The average change \( (AC) \) chart can be thought of as a moving average with a window of 2. That is, the average contribution of a window of two subsequent time intervals at wavelength \( j \) is summed over \( J \) wavelengths. The average change chart is more likely to capture trends in the process since fast deviations are averaged out. The \( AC_j \) is given as a function of time by:
Both charts present the evolution of the batch process by a single value in time, however, the information presented is rather general since it does not provide a direct insight into specific chemical compounds in the process.

5.3.4 **Black model: Principal component analysis (PCA)**

A popular tool for analysing spectroscopic process data is principal component analysis (PCA). PCA is an empirical method where the measured spectra $X$ are decomposed into three matrices: The score matrix $T (I \times R)$, the loading matrix $P (J \times R)$ and the residual matrix $E (I \times J)$, where $R$ is the number of components. PCA solves the following problem:

$$\min_{T,P} \|X - TP^T\|^2$$

with the restrictions that $TT^T$ is diagonal and $PTP$ equals $I$ (identity). The score matrix $T$ captures the systematic variation in time. The loading matrix $P$ captures systematic information in the spectral direction and residual matrix $E$ consists of unmodelled variation. By plotting the scores as a function of time, the evolution of the process can be followed.

5.3.5 **Grey model**

The aim of a grey model is to have a process model from which the model parameters have a direct chemical or physical meaning. This is achieved by incorporating known chemical or other process knowledge into an empirical model. In the present study the empirical model is given by

$$X = CB^T + E$$

where $X (I \times J)$ are the measured spectra with $J$ wavelengths during $I$ observations, $C (I \times R)$ consists of the temporal profiles of the $R$ chemical compounds and $B (J \times R)$ consists of the spectral profiles. The unmodelled variation is given in matrix $E (I \times J)$. The model given in equation 5 is a bilinear model. External process information may
be available which has chemical or physical significance; e.g. non-negativity of concentration profiles. Sometimes, there is prior knowledge about the process kinetics or pure compound spectra.

The model of equation 67 also underlies curve resolution methods (W.H. Lawton & E.A. Sylvestre [100], R. Tauler et al. [101]). There are different ways to estimate the parameters of equation 67 under restrictions. A versatile way of estimating is alternating least squares (R. Tauler et al. [101], R. Bro & S. De Jong [89]). Therefore, the parameters of equation 67 are estimated using alternating least squares with restrictions.

In the present modelling problem, which is the modelling of a full-scale batch process, the spectral data matrix \( \mathbf{X} \) is decomposed into three matrices: one which summarises the time invariant factors in the data (e.g. pure chemical compound spectra), one which summarises the time variant factors in the data (e.g. concentration profiles of the chemical compounds) and one which represents the unmodelled variation (C.E. Miller [102]). For the current modelling problem another common factor is introduced: the scale invariant factor. The scale invariant factor has identical profiles for both the laboratory-scale process and the full-scale process. With this scale invariant factor, it is possible to incorporate laboratory-scale knowledge into the full-scale model. In the current modelling problem it is assumed that the time invariant and the scale invariant factors are identical: they are the pure compound spectra.

As stated before, a grey model consists of a white (known) part and a black (unknown) part. The white part of the model consists of the time invariant and scale invariant factors. Both factors are related to the pure compound spectra. The time variant factors (the time-resolved profiles) are the black part of the model.

5.3.6 EXPERIMENTAL

5.3.6.1 Process description

The process of making a urethane resin is a multi-stage fed-batch process. In this study, only the first stage is of interest. The main reaction in the first stage is given by:

\[
OCN - R_1 - NCO + HO - R_2 \xrightarrow{\Delta} OCN - R_1 - NH - CO - O - R_2
\]
where a di-isocyanate reacts with an alcohol to a urethane resin with an extra NCO-group for further OH addition in subsequent steps.

The operation of this first stage is divided into three steps. The first step, after cleaning and purging the tank with N\textsubscript{2}, the tank is filled with a di-isocyanate. In the second step a second reactant (alcohol) is added drop-wise and the reaction starts. The third step is the part of the process where all the di-isocyanate has reacted and reaction mixture is kept at a constant temperature. This process is a non-isothermal process, it is operated according a predefined change in temperature. All reactants are added in precisely known stoichiometric ratios. Furthermore, it is known that a side reaction occurs. The alcohol reacts with the urethane to an undesired side product coupling to the second NCO group.

5.3.6.2 Instrumentation
The full-scale process spectra are obtained using an in-line FOSS NIR SYSTEMS spectrometer with a 2 nm resolution and a transflection probe with a 2 x 2 mm path length in the range of 1100 nm to 2500 nm. The laboratory-scale compound spectra are obtained using a Perkin-Elmer FT-NIR spectrometer with a 2 nm resolution and a 2 mm path length in the range of 1200 nm to 2500 nm.

5.3.6.3 Band assignment and band selection
From the literature, E.W. Crandall & N.J. Ashvin [107], it is known that the di-isocyanate (NCO ~ 1920 nm), the alcohol (OH ~ 1398 nm - 1421 nm) and the product (NH ~1485 nm – 1503 nm) are spectroscopically active in the near infrared range (1100 nm – 2500 nm). Furthermore, it is known that the monitored reaction is exothermal and that a predefined temperature gradient exists. The functional groups present in the compounds are shape sensitive for temperature changes. This is not desired since the pure compound spectra are considered as time invariant factors. Therefore a selection of spectral bands is made in such a way that the bands are not too much affected by temperature changes and consists of sufficient chemical process information.

The selected range is validated by laboratory experiments. That is, pure compound spectra of di-isocyanate and alcohol were measured at different temperatures. Based on the observed changes in the measured spectra, a wavelength
selection was made compromising between temperature effects and chemical compound information. Figure 71 shows the second derivative spectra and the wavelength range selected for a single batch run. The circles highlight the wavelength ranges corresponding to the functional groups (NCO, OH and NH) in the system.

The 2nd derivative spectra of the selected wavelength range used for process modelling.

5.3.6.4 Sampling
A total of 18 batches is available \( (N = 18) \). For every batch run \( n \) \( (n = 1.. N) \), the selected \( J \) wavelengths (see Figure 71) obtained at observation \( i \) are stacked in a matrix \( X_s \) \( (I_s \times J) \). Note that each batch run \( n \) has a different number of total sampled spectra \( I_s \). The collected spectra have a baseline offset, which is removed by the use of second derivative spectra estimated using a Savitzky-Golay filter (A. Savitzky & M.J.E. Golay [103]).

The window size is determined by computing the second derivative using windows with a different number of channels. Visual inspection showed that the best results were obtained using a window of 21 channels and a second-order polynomial.
5.3.7 RESULTS AND DISCUSSION
This section is divided into two subsections. In the first subsection, the construction of the grey model is discussed in detail. The second subsection presents and discusses the results of the qualitative monitoring performance of batch-to-batch consistency of the approaches suggested previously.

5.3.7.1 Construction of the grey model
The first step in constructing a grey model is collecting the available (white) process knowledge. With respect to the process knowledge, the following is known: in the first step, only di-isocyanate is present in the reactor. In the second step, alcohol is added and the reaction starts. The third step is a transition from the second step where all di-isocyanate and alcohol has reacted and only a mixture of the product and a particular amount of the undesired side-product are present in the reactor.

With the present knowledge, it is possible to obtain estimates of the pure di-isocyanate spectrum and the end mixture spectra under process conditions. The estimations are done according to a selective window approach (M. Maeder [104]). For di-isocyanate, the spectra corresponding to the first step for all batch runs are collected. These spectra correspond to di-isocyanate under process conditions. The collected spectra are stacked in a matrix. To assure that only di-isocyanate is present in the data, a rank analysis is performed. A rank one matrix is found. This suggests that no interferents are present and the average spectrum of the collected spectra can be used as the pure compound spectrum of di-isocyanate. To obtain the end mixture spectrum, a similar procedure is followed. That is, a rank analysis is applied to data corresponding to the third step of the process. A rank one matrix is found implying that only one component is present in the data. Since end mixture spectra are used from different batch runs, it suggests that the ratio between the end product and undesired product is constant for the different batches.

At this point in the analysis, two estimated spectra are available: the pure spectrum of di-isocyanate and the product mixture spectrum. Both estimates are based on process knowledge. The pure compound spectrum of the alcohol is unknown from the process. However, this information can be obtained from laboratory experiments. In order to use laboratory-scale information, it is interesting to know whether laboratory measurements correspond to measurements obtained from the full-scale process. Therefore, also spectral information of di-isocyanate is
obtained from laboratory measurements. This is shown in Figure 72. The spectrum of di-isocyanate obtained under process conditions and the spectrum obtained under laboratory conditions are given in Figure 72a. The difference spectrum is given in Figure 72b. There are only minor differences between the full-scale measurements and laboratory measurements implying the spectra to be reasonably scale invariant.

In spite of the different measurement conditions such as a different probe, spectrometer and path length, the spectra obtained are similar in shape. They differ in intensity due to path length differences. These differences, however, are captured in the concentration profiles.

![Figure 72](image)

**Figure 72**

*a) The normalised laboratory absorbance spectrum and the normalised full-scale absorbance spectrum.*

*b) The difference spectrum between the normalised laboratory absorbance spectrum and the normalised full-scale absorbance spectrum.*

From the obtained results it seems reasonable to use the laboratory-scale spectrum of alcohol in the grey model.

### 5.3.7.2 Incorporation of external white information into the model

The collected white knowledge discussed in the previous section is incorporated as constraints into the bilinear model given in equation 68. This gives the following problem to be solved:

$$\min_{C,B} \left\| X - CB^T \right\|^2$$  \hspace{1cm} (68)
where $B$ ($J \times R$) represents the time and scale invariant factors (pure compound spectra) and $C$ ($I \times R$) the time variant factors (concentration profiles). Equation 68 is a restricted least squares problem and is solved with $B$ fully restricted to equal the pure compound spectra and non-negativity imposed on the $C$ profiles (R. Bro & S. De Jong [89]). Moreover, having matrix $B$ fully specified in the least squares algorithm will result in a unique solution for the model parameters as shown in Appendix A.

5.3.7.3 Modelling a single batch run
The analysis starts with a single batch run under normal operating conditions (NOC), which is modelled using the grey model in equation 68. The total variation present in the data $X$ is represented as 100% of the total sum-of-squares (SS). The performance of the model is judged by the amount of explained variation of the total variation, and is computed by:

$$SS \% = \frac{\|CB^T\|^2}{\|X\|^2} \times 100\% \quad (69)$$

The model captures 99.65% of the total sum-of-squares of the measured process spectra $X$. The compound profiles found for the relative concentrations are given in Figure 73. The circles mark the time of the start of alcohol addition (Time = 10) and the end of alcohol addition (Time = 59), which corresponds with the process operator log.

![Figure 73](image)

*The concentration profiles obtained by the Grey model vs. time.*

*The circle marks the start and end of alcohol addition corresponding to the operator log.*

-164-
The estimated profiles give a clear view of the chemistry in the first stage of the process. During the first stage only di-isocynate is present in the reactor. During addition of the alcohol (Time = 10 - 58), a decrease is observed for di-isocyanate and the formation of the product is observed. The product is formed until all di-isocyanate has reacted (Time = 66). Furthermore, the middle graph of Figure 73 shows the behaviour of the alcohol during the reaction. That is, the introduction of the alcohol at Time = 10 and the stop of alcohol addition after Time = 58. The dosing of the alcohol is an important procedure in the process. It is therefore satisfactory that the obtained alcohol profiles correspond to the process operator log. Furthermore, it is expected that the observed profiles are influenced by addition of the alcohol; they reflect not only pure chemistry but also the technological operation of the process. Hence, the concentration profiles are affected by: i) the process kinetics, ii) dilution phenomena due to the addition of the alcohol, iii) volume expansion due to increasing temperature, iv) diffusion and mixing phenomena, since the probe is located at the bottom of the reactor, a small delay is expected in observing new phenomena in the process.

5.3.7.4 Mathematical model robustness

In this subsection an impression is obtained about the influence of the spectral variation on the resolved concentration profiles. It is important to realize that the robustness applies to the statistical variability of the model parameters. Rotational ambiguity is not a problem, since the model parameters can be estimated uniquely (see above). The estimations of the pure compound spectra are based on spectra collected from the 18 historical batch runs. From these, two spectral data sets were generated to estimate the pure compound spectra. The first data set with di-isocyanate spectra and the second data set with the product mixture spectra. By assuming that each individual spectrum is an estimation of the pure compound spectrum, and using each spectrum as a restriction in the grey model, a series of models is computed and an impression is obtained of the concentration profile variation caused by the estimation error in the di-isocyanate spectrum and product mixture spectrum. For this purpose, one batch is selected and is modelled repeatedly according to the following resampling procedure:
Statistical batch process monitoring

a) The first individual di-isocyanate spectrum of the data set, together with the average end mixture spectrum and laboratory measured alcohol spectrum is used to estimate the di-isocyanate profile in time.

b) Repeat step a) for the second individual spectrum of di-isocyanate.
Step b) is repeated until all individual estimated pure di-isocynate spectra (30 in total) are used.

Note that this is a worst-case scenario because the profiles in the grey model are obtained using the average spectrum, which is always more stable than an individual spectrum.

A similar procedure is used for the product mixture spectrum. The obtained concentration profiles are given in Figure 74a and Figure 74b.

Figure 74
a) The concentration profiles obtained by the model by resampling the di-isocyanate spectrum vs. time.
b) The concentration profiles obtained by the model by resampling the end mixture spectrum vs. time.
c) The modified second derivative spectra used for resampling the alcohol.
d) The concentration profiles obtained by the model by resampling the end mixture spectrum vs time.
Figure 74a shows that the variation present in the various di-isocyanate spectra influences the di-isocyanate concentration profiles in the start of the process. Furthermore, it is shown that the time profiles of the alcohol and the product are not much influenced by deviations in the di-isocyanate spectrum. Figure 74b shows the time profiles obtained for different end mixture spectra. From this it can be seen that variation in the profiles caused by estimation variation in the end-mixture spectrum do not influence the shape of the profile, only the intensity.

Since the alcohol spectrum is obtained from laboratory measurements it is not possible to do a variation analysis based on the sampled process spectra. Therefore, a shape robustness test is performed. The influence of the change of shape of the spectrum on the time resolved profiles are studied. This test is performed by adding peaks and by changing the shape of peaks in a realistic way thereby modifying the shape of the alcohol spectrum. A total of ten alcohol spectra are constructed. The modified second derivative spectra, used for modelling are given in Figure 74c. The main shape changes are made in the range of 1450 nm to 1550 nm, which represents the (small) OH contribution in the spectrum (Figure 71). The time profiles of all compounds found for the various modified alcohol spectra in the grey model are given in Figure 74d. Slightly modifying the alcohol spectrum gives a variation in intensity of all profiles but not in the shape. Since the alcohol profile is similar for the different spectral disturbances, this suggests that the alcohol profile is mostly dominated by the large contribution in the range of 1560 nm to 1650 nm.

It can be concluded that measurement variation present in the spectra does influence the intensity of the resolved time profiles but the general shape of the profiles is consistent.

5.3.7.5 Modelling multiple batch runs
The grey model can also be used to model multiple batch runs at the same time. This is done in the same manner as suggested by R. Tauler et al. [105]. The 18 batch runs are stacked on top of each other with the selected spectral mode $f$ as a common factor for all batch runs. This results in a new matrix $X_{\text{total}}(I_{\text{total}} \times f)$ where $I_{\text{total}} = \sum I_n$ and $I_n$ is the number of measured spectra corresponding to batch run $n$. The matrix $X_{\text{total}}$ is modelled according to:
\[
\min_{C_{\text{total}}, B} \| X_{\text{total}} - C_{\text{total}} B^T \|^2
\]  

(70)

where \( C_{\text{total}} \) consists of the concentration profiles of the 18 batch runs. \( C_{\text{total}} \) is estimated with \( B \) fixed and non-negativity constraints on \( C_{\text{total}} \). The matrix \( C_{\text{total}} \) is rearranged into \( C \) for each batch run \( n \). In this manner the differences and equalities of the batch runs become apparent. By specifying the model as above, equal concentration profiles are found as with the modelling of a single batch run.

5.3.7.6 Estimating the unknown alcohol spectrum

The grey model of equation 70 can also be used to estimate unknown compound spectra. This is interesting since the alcohol spectrum under process conditions is unknown. Obtaining a good estimating of the alcohol spectrum with the grey model is not straightforward since the concentration of alcohol present in the reactor is relatively low compared to di-isocyanate and the product mixture. The low alcohol concentration makes it difficult to estimate a reliable spectrum using the grey model, which is the reason why a laboratory-measured alcohol spectrum was used in the grey model. Still it is interesting to study the possibility of estimating the alcohol spectrum with the grey model in equation 70. This is done for two reasons: First, to see whether the proposed grey model has curve resolution capabilities for a component at low concentration. Second, it is an extra check on the validity of imposing a laboratory scale spectrum on the model parameters.

The model is applied to \( X_{\text{total}} \) with non-negativity constraints on \( C_{\text{total}} \) and two columns of \( B \) restricted. The first column of \( B \) is restricted to equal the di-isocyanate spectrum and the third column of \( B \) is restricted to equal the end mixture spectrum. The second column of \( B \) corresponds to the alcohol spectrum, which will be estimated by the model.

The alcohol spectrum obtained with curve resolution and the measured alcohol spectrum is given in Figure 75.
From the sensitivity analysis it is known that a peak in the range of 1560 nm – 1650 nm, dominates the measured alcohol profile. Figure 75, shows that the dominant range of the spectrum is slightly shifted compared to the measured alcohol spectrum. Furthermore, spectral differences are observed in the range of 1480 nm – 1540 nm: this range has a large contribution of NH. Other spectral differences are observed in the range of 1840 nm – 1930 nm and this range corresponds to NCO. Since the spectral differences are mainly observed in the temperature sensitive ranges, it suggests that the shift and differences are caused by the temperature gradient present in the system. The obtained concentration profiles, however, are reasonable which was shown in the robustness analysis.

5.3.8 QUALITATIVE MONITORING PERFORMANCE

In the following it is discussed how the grey model of equation 70 performs in detecting and studying the batch-to-batch consistency. A grey model is constructed on $X_{total}$ with the laboratory- scale spectrum of the alcohol imposed, because this would resemble the case in practice when scale invariant constraints are imposed. The relative concentration profiles (C) found by the grey model are compared with the techniques as discussed previously. The comparison focuses on two issues. First, the performance of detecting an inconsistent batch and secondly, the diagnostic possibilities of the techniques used are discussed. For illustration, three batch runs are selected to show the performance of the different techniques. The first batch, Batch 1, is a normal batch run obtained under normal operation conditions. For the second batch, Batch 2, it is prior from DCS data known that a manual intervention occurred
in the dosing system of the alcohol. The third batch, Batch 3, has a large spike in the spectral response, which was caused by an instrumental error (cooling failure).

![Wavelength tracking](image)

**Figure 76**

(a) *Wavelength tracking, the NCO channel at wavelength 1920 nm vs. time.*

(b) *The OH channel at wavelength 1428 nm vs. time.*

(c) *The NH channel at wavelength 1498 nm vs. time.*

The first technique is wavelength tracking. Three wavelengths are selected to monitor the three batches. The di-isocyanate is monitored at a wavelength of 1920 nm, the alcohol at 1428 nm and the product at 1498 nm. The results of wavelength tracking are given in Figure 76. The top graph represents the di-isocyanate, the middle graph represents the alcohol and the bottom graph represents the product. The profiles of Batch 1 give the trend, which the profiles of Batch 2 and Batch 3 should follow. Batch 2, which has a manual intervention in the dosing of the alcohol, starts to show a distinct profile from Batch 1 at Time = 37 in the middle graph. A similar conclusion can be drawn for the bottom graph, although not so clear as for the middle graph. This is different for Batch 3 where a clear disturbance is observed in the product profile at Time = 79.
The graphs in Figure 76 present a general trend of the process. There might be expected that prior to the start of the reaction a contribution of the di-isocyanate is present, since this is the only compound present in the system. From Figure 76, it is clear that this is not the case. This is due to the nature of the used spectroscopy. There is not a selective region for the di-isocynate, alcohol or the product. Moreover, intuitively wrong signals are observed for the product. The signal is decreasing after the start of reaction where an increasing signal is expected. Furthermore, the tracking signal is not very stable since it is hard to find a selective region in the NIR spectra. Moreover, the compounds monitored are sensitive to temperature changes, which can also affect the tracking signal.

The second technique is the Change chart. The three batches are given in Figure 77. In Figure 77 can be seen that there is no clear distinction between the three batches except for Batch 3, where the magnitude of the upset in the spectral response is large. The change around Time = 37 goes unnoticed in this graph. This implies that the Change chart is only suited for monitoring if large disturbances occur. Furthermore, this chart has a lack of generating pure chemical compound information during process monitoring.

The third technique is the Range chart. The results are given in Figure 78. Due to the nature of the technique, emphasis is placed on differences between spectra.
Figure 78 it becomes clear that Batch 2 starts to show a small distinction in the tracking signal compared to Batch 1 and Batch 3 around Time = 19. In case of doubt, it is clear that Batch 2 is deviating from Batch 1 and Batch 3 at Time = 37. With respect to Batch 3, it is easy to see that the batch behaves similar to Batch 1 until the upset around Time = 79. Although this chart performs well for detecting deviations, it has, similar to the Change chart a lack of generating pure chemical compound information. Because, from this chart it is not clear what the cause of deviation in Batch 2 is.

Technique four is the black box model. A three component model is built to model the process. The PCA scores are monitored as a function of time and given in Figure 79. With respect to Batch 2, in the bottom graph of Figure 79, there is a clear deviation in the score profile compared to Batch 1 and Batch 3. The deviation starts around Time = 18, representing the manual intervention during the batch run. For Batch 3, the manual intervention at Time = 79 is well observed in the top graph of Figure 9. Furthermore, small deviations are observed in the middle and bottom graph at Time = 79. The process can well be monitored using the model scores. However, the scores are a linear combination of the model loadings $P$. This means that it is not possible to make a distinction between the score profiles and assign the score profiles to a specific chemical compound. Hence, again diagnostics are severely hampered.
The scores vs. time. The top graph represents the scores of the first PC vs. time. The middle graph represents the scores of the second PC vs. time. The bottom graph represents the third PC vs. time.

The last technique used is the grey model. The results for the grey model are given in Figure 80. The top graph in Figure 80 represents the behaviour of the di-isocyanate. The middle graph represents the alcohol behaviour and the product formation is given in the bottom graph.
From Figure 80, monitoring Batch 2, it is shown that the dosing of the alcohol is different compared to Batch 1 and Batch 3. In the middle graph, the alcohol profile shows an excess of alcohol combined with a slower formation of the product. This implies that the reaction is performing slower compared to Batch 1 and Batch 3. Since the reaction did not proceed as fast, a manual intervention took place and the alcohol flow was shutdown. Around Time = 28, the process seems to speed up which is reflected in the formation of the product and the di-isocyanate profile (Time = 32). At Time = 38 all alcohol has reacted and the process stagnates. The remainder of the alcohol is added at Time = 46 and the process continues until all di-isocyante has reacted.

With respect to Batch 3, it is shown that the batch is operated similar to Batch 1 until an upset occurs at Time = 79. Moreover, it is shown that the observed upset has no influence on the product concentration (Time > 80) implying that the process is operated according to specifications.

The qualitative performance of the five techniques studied is summarised in Table 10.

Table 10

<table>
<thead>
<tr>
<th></th>
<th>detection</th>
<th>diagnosis</th>
<th>model simplicity</th>
<th>wavelength tracking</th>
<th>Change chart</th>
<th>Range chart</th>
<th>PCA</th>
<th>grey model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td>--</td>
<td>--</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>--</td>
</tr>
</tbody>
</table>

*Qualitative performance of monitoring the batch-to-batch consistency.*

Considering the criteria of Table 10, the Change chart is not recommended for monitoring and detecting an inconsistent batch. The Change chart is found to have a poor performance in detection and a bad performance in diagnosis. Table 10 shows that it is possible to detect a process upset using wavelength tracking, furthermore, the simplicity of the approach could be seen as an advantage. In this study, it is not recommended to use wavelength tracking since there is not a clear selective region for the individual compounds. Moreover, temperature effects have an influence on the tracking signal.

The Range chart, PCA and the grey model are found to be excellent in detection and are suited for monitoring and detecting inconsistent batches. However, the Range chart and PCA do not perform well for fault diagnosis.
The grey model gets an excellent mark for diagnosis. The advantage of the grey model of being able to have a direct physical-chemical interpretation of the process, makes it favourable to use the grey model for detection and diagnosis of batch-to-batch consistency. A disadvantage of the grey model could be the complexity of the model.

5.3.9 CONCLUSION

A detailed description is given of how a grey model is constructed to model an industrial process of which NIR spectroscopic measurements are available. The stability of the obtained profiles are studied by a robustness analysis were it is shown that the obtained results are stable. Furthermore, it is shown how the grey model can be used for extracting unknown compound spectra. Then the grey model is compared with four different techniques (wavelength tracking, Range chart, Change chart and PCA) to monitor the batch-to-batch consistency using NIR spectroscopy in a case study. The grey model gives a clear and direct insight in physical chemical phenomena of the process, which is an advantage, compared to other techniques. Therefore, it is recommended to use the grey model to monitor and study batch-to-batch consistency. The next step in subsequent research is a more quantitative comparison of the 18 batches with the focus on statistical limits useful for process monitoring.
5.4 Statistical batch process monitoring using grey models*

5.4.1 SUMMARY
A complete strategy for monitoring industrial batches processes using grey models is presented including fault detection and fault diagnosis tools. Grey models is a novel concept in batch process modelling and monitoring. A grey model is a hybrid model in-between hard (white) process models and soft (black) models combining the advantages of both approaches. The principles of grey models are explained and it is shown how these models can be constructed. For this purpose an industrial batch process is available which is spectroscopically monitored. It is explained how the spectroscopic measurements are combined with prior process knowledge. To show the versatility of the strategy, two types of grey models are constructed and used for statistical batch process monitoring. The two models are compared and validated for both on-line monitoring and post-batch analysis. For the latter, the batch consistency number (BCN) is introduced in order to have a fast and simple post-batch analysis. The results show how these models help to detect and diagnose process upsets. The use of grey models for batch process monitoring results in a fast detection of process upsets and a good fault diagnosis.

5.4.2 INTRODUCTION
Batch processes are a significant part of today’s production in the food and (bio-) chemical industries. Batch processes are recipe driven and produce low volume and high value products. Because of control actions and for safety reasons, it is common to carry out process measurements such as temperature, pressure or flow measurements. A disadvantage of these process measurements is that they do not give direct insight in the process chemistry. To have chemical insight during the batch run, spectroscopic measurements can be performed (J. Workman, Jr. et al. [69], S.P. Gurden et al. [90], E.N.M. Van Sprang et al. [106], J.A. Westerhuis et al. [46]). For a quick detection of process upsets or unusual process behaviour it is common to use

---

the process measurements for monitoring purposes. In the nineties P. Nomikos & J.F. MacGregor [6] introduced the concept of multivariate statistical monitoring of batch processes. Since the introduction of the method, many extensions and applications have been reported in the literature (S. Rannar et al. [31], S. Wold et al. [53], R. Boqué & A.K. Smilde [40], E.B. Martin et al. [27], D.J. Louwerse & A.K. Smilde [38]).

In principle, batch process monitoring is divided into three phases according to the I.T.A.- trajectory The Initial phase, the Training phase and the Application phase. The Initial phase consists of collecting a set of historical batch runs, which are monitored on-line and operated under normal operating conditions (NOC).

In the Training phase, suspicious batch-runs or batch-runs which are not considered to be obtained under normal operating conditions (NOC) are removed from the historical data set. When a representative historical data set of NOC batches is obtained, a model is built to model the common cause variation. With the obtained process model, test statistics are computed and statistical confidence limits are derived to construct control charts. In the Application phase a new batch is monitored by means of control charts derived in the Training phase.

A batch process can either be monitored on-line or post-batch. With on-line monitoring, the aim is to have a fast detection of process upsets. Then, if an alarm is given in the control charts, the model is used to locate and diagnose the cause of the alarm. In post-batch analysis, the aim is to study the general behaviour of batch runs and obtain an insight in the differences and similarities between batches.

There are a variety of possible models to capture the process behaviour e.g. Principal Component Analysis (PCA), Multivariate Linear Regression (MLR) or Projection to Latent structures (PLS). Such models are known as empirical or black box models. These models fit the data properly, but cannot be generalised to different situations and do not always generate a good process insight. Another type of models are white models, which are models based on first principles and do provide process insight. However, white models do not fit the process data that well, and are not able to explain batch-to-batch differences.

A combination of the previous black box and white models are grey (hybrid) models. Grey models try to combine the advantages of the black box and white models. That is, a good model interpretation and a good data fit. A grey model is
constructed by including known external information as white knowledge, e.g. fundamental knowledge of the process at laboratory scale or trajectories of process variables, into a black model. The grey model fits the process data using the external information as constraints on the model parameters. In this manner fundamental knowledge and empirical models are combined. A more detailed discussion of grey models can be found in S.P. Gurden et al. [90]. The concept of combining white information with black models has been reported in several publications (E.N.M. Van Sprang et al. [60], S. Bijlsma et al. [66], A. de Juan et al. [107]). The grey concept implicitly divides the data into a part consisting of variation due to known causes (white part) and into a part consisting of systematic variation due to unknown causes (black part). This makes the model very suited to search for unknown phenomena in the data (H.-J. Ramaker et al. [71]). Other examples are the estimation of kinetic constants using spectroscopic process data and the knowledge of reaction kinetics or improving curve resolution models by combining hard kinetic models with soft models (S. Bijlsma et al. [66], A. de Juan et al. [107]).

In the following, two types of grey models are presented and validated for batch process monitoring. The models are validated for both on-line monitoring and post-batch analysis. The first type is a grey Tucker1 model and the second type is a grey Tucker3 model. The grey Tucker1 model is a relative simple model and can directly be used for on-line batch monitoring. If the grey Tucker1 model is used in post-batch analysis a rearrangement of the data is needed. The second type is a grey Tucker3 model. This model is more complex compared to the grey Tucker1 model and cannot be used directly for on-line batch process monitoring. However, this model is directly applicable for post-batch analysis.

Moreover, the models are used to capture the process behaviour and to derive control charts for process monitoring. The performance of the models is validated for on-line detection and diagnosis and also for post-batch analysis. For this purpose, the batch consistency number (BCN) is introduced as a simple analysis to study the differences and similarities between batches.

The outline of this article is as follows: After the Introduction, the concepts and construction of grey models is discussed together with the derivation of the monitoring statistics in the Theory section. Next, in the Experimental section, the
structure of the process data is discussed and the construction of grey models is
explained. Furthermore, it is shown how the models are used for statistical batch
process monitoring. After that, the results are presented and discussed in the section
Results and Discussion. Finally, conclusions will be given.

5.4.3 Theory

5.4.3.1 Notation
Scalar quantities are written as lowercase italics (x) and vectors as lowercase bold characters (x).
Uppercase bold characters (X) represent matrices and three-way arrays are given as
underlined uppercase bold characters (X). The symbol T refers to the transpose.

5.4.3.2 Structure of the data
A single batch run which is monitored on-line at K intervals is represented by a matrix
X (K x J) with J the number of process variables. In case of a series of multiple batch
runs, the individual batch runs can be stacked on top of each other and form the
three-way array X (I x J x K) with I the number of batch runs.

5.4.3.3 Black models
There are two sensible directions to matricize a three-way (H.A.L. Kiers [49]). The first
direction is in the batch direction, that is, the three-way array will be matricized in
a matrix X_{t x K}. The second direction is in the process variable direction, this will result
in a matrix X_{K x J}. The matricizing direction will have an influence on the
interpretation of the model parameters as will become clear in the following. Both
matrices can be modelled using a Tucker1 model. Assume the three-way array is
matricized in the process variable direction a single element in X_{K x J} is modelled by

\[ x_{k,j} = \sum_{r=1}^{R} e_{k,r} b_{jr} + e_{k,j} \]

or in matrix notation,

\[ X_{K x J} = C B^T + E \]
where matrix $C$ ($KI \times R$) consists of time related profiles, matrix $B$ ($J \times R$) consists of information regarding the correlation between the process variables and matrix $E$ is the residual matrix. If the three-way array is matricized in the batch direction a single element in $X_{i,j,k}$ is modelled by

$$X_{i,j,k} = \sum_{r=1}^{R} a_{i,r} p_{r,j} + e_{i,j,k} \tag{73}$$

or in matrix notation,

$$X_{i,j,k} = AP + E \tag{74}$$

where matrix $A$ ($I \times R$) describes the differences between the batches matrix $P$ ($JK \times R$) consists of information regarding the systematic variation and correlation between the process variables and matrix $E$ is the residual matrix. As shown, the matricizing direction determines the content of the model parameters and therefore also its interpretation.

The second option is to model the three-way array with a Tucker3 model. The Tucker3 model decomposes the three-way array $X$ into the matrices $A$, $B$, $C$ and $G$. A single element in the three-way array is given by

$$X_{i,j,k} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{i,p} b_{q,j} c_{r,k} g_{pqr} + e_{i,j,k} \tag{75}$$

or in matrix notation,

$$X_{i,j,k} = AG(C \otimes B)^T + E \tag{76}$$

where matrix $A$ ($I \times P$) captures the systematic differences between the batches. Matrix $G$ ($P \times QR$) consists of the interaction parameters between the three reduced components. Matrix $C$ ($K \times R$) consists of time related profiles and matrix $B$ ($J \times Q$) consist of information concerning the correlation between the process variables.

The relation between the Tucker3 and the Tucker1 model is explained in the following. A Tucker3 model tries to compress the data in each of the three directions
(I \times J \times K) into a reduced number of latent variables \( P, Q, R \). \( P \) is the number of latent variables in the batch direction \( l \), \( Q \) is the number of latent variables in the process variable direction \( f \) and \( R \) is the number of latent variables in the time direction \( K \). It is also possible to compress only two directions and then the model becomes

\[
X_{b/JK} = AG(I \otimes B)^T + E
\]  

(77)

with \( G(P \times QJK) \) and \( I(Q \times Q) \) the identity matrix. This is called a Tucker2 model (P.M. Kroonenberg [115]). Of course, it is also possible to compress only in one direction and the model becomes:

\[
X_{b/JK} = AG + E
\]  

(78)

with \( A(I \times P) \) and \( G(P \times JK) \). Comparing equation 78 to equation 71 it can be seen that both models are identical. Both models model the matricized three-way array with a bilinear model. The model in equation 71 is called a Tucker1 model. In fact, a Tucker1 model with the proper restrictions is equal to a Tucker3 model.

In short, both models are well able to describe the variation in the data. The difference between the two models is that the Tucker3 model is able to reduce the data with a different number of latent variables in each of the three directions (batch, variable, time). Interactions between the three reduced directions are given in the core array \( G \). In contrast, the Tucker1 model only reduces the data in one direction.

Since batch processes are recipe driven, it is assumed that the batches are very similar. However, in practice, batches will have slightly different run lengths, which are reflected in a different number \( (K_i) \) of sampled spectra. There are several ways of synchronizing the batches. Batches can be synchronised using maturity variables (A.A. Tates et al. [2]), interpolation (R. Boqué & A.K. Smilde [40]) or dynamic time warping (DTW) (A. Kassidas et al. [35]). In this study the differences in batch run length are dealt with using a DTW algorithm, which is adapted to handle spectroscopic data (H.-J. Ramaker et al. [108]). For simplicity, the synchronization is done in a post-batch manner although it is also possible to apply this technique on-line. More suggestions can be found in T. Kourt [18].
5.4.3.4 Grey models
Grey models are hybrid models. That is, a part of the model consists of known systematic variation and is combined with an unknown empirical part. The known or "white" part of the model is constrained to reflect available external process information such as process kinetics. The empirical or "black" part of the model is unconstrained and fits further systematic variation that is not explained by the white part. Together, the white and black parts form the grey model.

Unsystematic variation in the data is not modelled and forms the residuals. The model parameters are calculated simultaneously with the aim to maximize the total amount of variation explained in the data. The separation between the white and black part of the model makes it possible to isolate the unknown sources of systematic variation and to study the cause of this unknown variation.

5.4.3.4.1 Creating grey models
The black models are made grey by estimating the models in a single least squares optimisation with multiple constraints applied according to prior external process information ("white" knowledge). First, a general discussion is given of the Tucker1 model in equation 71. In Figure 81, assume Near Infrared (NIR) process spectra in matrix X. Matrix X is decomposed in the matrices C_w, B_w, C_b, B_b and the residual matrix E. Information related to time and concentration profiles are represented in the matrices C_w and C_b and spectral information is captured in the matrices B_w and B_b. The first step in creating a grey model is collecting the available white knowledge. Having NIR process spectra the following white information can be thought of: The spectral response is a linear sum of the contribution of the individual compounds according to the Lambert-Beer law.

Sometimes the reaction kinetics are known and can be imposed on matrix C_w. If the reaction kinetics are not known, it is possible to impose physical sensible restrictions such as non-negativity or unimodality on the profiles in C since concentrations can never be negative. If pure compound spectra are available they can be imposed on matrix B_w. If there are no pure compound spectra available, restrictions can be imposed such as non-negativity.

The model implicitly imposes the Lambert-Beer law if a Tucker1 model is used since this model assumes linear addition of the individual components and no interaction. After imposing all white information on the model, additional unrestricted
components can be added to the model to capture unknown systematic variation in the data.

A similar reasoning holds for the Tucker3 model. The NIR process spectra are stored in the three-way array $X$ as shown in Figure 82. The model decomposes the three-way array $X$ into a white part and a black part. The white part is defined by the matrices $A_w$, $B_w$, $C_w$ and $G_w$, which consist of prior process information. With the Tucker3 model, the Lambert-Beer law is incorporated by restricting elements in the core array $G_w$ to equal zero (S.P. Gurden et al. [90]). If available, pure compound spectra are imposed on matrix $B_w$ and process kinetics on matrix $C_w$. Just as in the grey Tucker1 model, adding new components to the model will define the black part. The black part of the model is defined by the matrices $A_b$, $B_b$, $C_b$, $G_b$. Note that the scores $A_w$ are not white but black since there are no assumptions made about $A$ or restrictions placed on $A$. 

![Diagram](image-url)
5.4.3.5 Batch process monitoring

In spite of the close relationship between the two models, they are substantially different. Both models have a different principle of modelling. The grey Tucker1 model is a model with the aim of modelling a time interval $X^t(I \times K)$ and the grey Tucker3 model is a three-way model with the aim to model the entire three-way array $X(I \times J \times K)$. Therefore, both models generate a different test statistic as will become clear in the following. Furthermore, the test statistic of the grey Tucker1 model has a direct physical interpretation e.g. the concentration behaviour whereas the test statistic of the grey Tucker3 model does not have a direct physical interpretation.

5.4.3.5.1 On-line monitoring of a new batch using the grey Tucker1 model

For on-line monitoring with a grey Tucker1 model, the local observation vector $x^l(I \times 1)$ is used. A new batch is monitored by projecting the local observation vector $x^l$ on $B$ according to equation 79.

$$\hat{c}_k = \left[ x^l'B\left(B^TB\right)^{-1}\right]^T$$

( 79 )

However, the unconstrained solution may result in negative values, which is not wanted since $\hat{c}_k$ are relative concentrations and can never be negative. Therefore, the new batch is monitored by regressing the observation $x^l$ onto $B$ using non-negativity constraints (R. Bro & S. De Jong [89]) just as in modelling the historical batches in the trainings phase according to

$$\min_{\mathbf{c}} \|x^l - B\mathbf{c}\|^2, \text{ with } \mathbf{c} \geq 0$$

( 80 )

The local residuals are computed by

$$\mathbf{e}^l = x^l - B\hat{c}_k$$

( 81 )

The obtained estimates $\hat{c}_k$ and the local residuals $\mathbf{e}_k^l$ will be used as test statistics for process monitoring.
5.4.3.5.2 On-line monitoring of a new batch using the grey Tucker3 model

Monitoring a new batch using the grey Tucker3 model is more complicated than batch process monitoring using a Tucker1 model. Consider the Tucker3 model with

\[ V = \left( G(C \otimes B)^T \right)^T \]  

(82)

A new batch is monitored by projecting the observations \( x \) (\( JK \times 1 \)) onto the model \( V \) (\( JK \times P \)).

\[ a_{\text{av}} = \left[ x^T V (V^T V)^{-1} \right]^T \]  

(83)

As can be seen, the monitoring procedure requires an observation vector \( x \) (\( JK \times 1 \)) of full length \( JK \). This is not available in on-line monitoring since at time \( k \) only the observations are known until \( k \). Therefore, assumptions have to be made about the future unknown observations \((k+1)\) to \( K\). For this purpose, the method of current deviations is used (P. Nomikos & J.F. MacGregor [6]). That is, it is assumed that the batch remains at the same level (current deviation) for the remainder of the batch run. The residuals are computed according to

\[ e_k = x_k - V a_{\text{av}} \]  

(84)

5.4.3.6 Statistics

The statistical confidence limits are derived using a leave one out principle (D.J. Louwerse & A.K. Smilde [38]). In this approach, every batch is left out from the population and treated as if it was a new batch. The remaining batches are used to build a model. The left out batch is projected onto the model and the obtained scores and residuals are stored for deriving the confidence limits of the control charts.

5.4.3.6.1 D-statistic and SPE-statistic

The two multivariate statistics are the D-statistic and the SPE-statistic (P. Nomikos & J.F. MacGregor [20]). The D-statistic monitors deviations within the model relative to the center point and the SPE-statistic monitors deviations from the model. The statistical limits for the D and SPE statistics are derived from the NOC data. The I
batches from the NOC data are monitored resulting in \( I \) values of the \( D \) and \( SPE \) at each time interval. Statistical distributions for these test statistics are derived and used to compute control limits.

**D-statistic**
The Hotelling \( T^2 \) statistic is called the D-statistic when a reduced space with \( R \) components is used instead of \( x_k \) with \( J \) or \( JK \) variables. In case of the grey Tucker1 model, the \( D \)-statistic is given by

\[
D_k = \left( \hat{c}_{av,k} - \hat{c}_k \right) \tilde{S}_k^{-1} \left( \hat{c}_{av,k} - \hat{c}_k \right) \frac{I(I-R)}{R(I^2-1)} \sim F(R, I-R) \tag{85}
\]

where \( \hat{c}_{av,k} (R \times 1) \) are the estimates of the relative concentrations of the new batch at time interval \( k \) and \( \hat{c}_k (R \times 1) \) contains the means of the columns of matrix \( C_k (I \times R) \) which are the estimated relative concentrations of the NOC data. The matrix \( \tilde{S}_k (R \times R) \) is the variance-covariance matrix of \( C_k \). \( D_k \) follows an \( F \) distribution with \( (R, I-R) \) degrees of freedom. In case of the grey Tucker3 model, the \( D \)-statistic is given by

\[
D_k = \left( \hat{a}_{av,k} - \hat{a}_k \right) \tilde{S}_k^{-1} \left( \hat{a}_{av,k} - \hat{a}_k \right) \frac{I(I-P)}{P(I^2-1)} \sim F(P, I-P) \tag{86}
\]

where \( \hat{a}_{av,k} (P \times 1) \) are the scores of the new batch at time interval \( k \) and \( \hat{a}_k (P \times 1) \) contains the means of the columns of the score matrix \( A_k (P \times L) \). \( \tilde{S}_k (P \times P) \) is the variance-covariance matrix of \( A_k \), which consists of the scores obtained from the NOC data.

**Squared Prediction Error (SPE)**
For both models, the \( SPE_k \) statistic is computed by

\[
SPE_k = \sum_{j=1}^{I} \hat{e}_{j,k}^2 \sim \chi_{I}^2 \tag{87}
\]
where $\varepsilon_k$ is the prediction error of the process variable $j$ at time interval $k$. The control limits for the SPE are obtained by fitting a weighted $\chi^2$-distribution to the reference distribution obtained from the NOC data at each time point. In equation 87, parameter $h$ represents the degrees of freedom and parameter $g$ represents the weight to account for the magnitude of $SPE_k$. These two parameters can be estimated in different ways (P. Nomikos & J.F. MacGregor [6]).

In this study the estimation is done according to J.E. Jackson & G.S. Mudholkar [24], whereby $g$ and $h$ are functions of the eigenvalues of the residual variance covariance matrix at each time interval $k$.

### 5.4.3.6.2 Univariate statistics

The individual test statistics $a_{k,r}$ and $c_{k,r}$ can be monitored using univariate control charts. The upper and lower control limits for the Tucker1 model are computed by

$$UCL_{k} = \bar{e}_{k,r} + t_{(1-\alpha/2)} \cdot s_k \left( 1 + \frac{1}{I} \right)^{1/2} \quad \text{and} \quad LCL_{k} = \bar{e}_{k,r} - t_{(1-\alpha/2)} \cdot s_k \left( 1 + \frac{1}{I} \right)^{1/2}$$

where $\bar{e}_{k,r}$ is the mean computed from matrix $C_k$ and $s_k$ is the corresponding standard deviation computed from $C_k$. The $t$ is the Student $t$ value with $I-1$ degrees of freedom with a $\alpha$ significance level and $I$ the number of historical batches. The control limits for the Tucker3 model are derived in a similar way according to

$$UCL_{k} = \bar{a}_{k,r} + t_{(1-\alpha/2)} \cdot s_k \left( 1 + \frac{1}{I} \right)^{1/2} \quad \text{and} \quad LCL_{k} = \bar{a}_{k,r} - t_{(1-\alpha/2)} \cdot s_k \left( 1 + \frac{1}{I} \right)^{1/2}$$

where $\bar{a}_{k,r}$ is the mean computed from matrix $A_k$ and $s_k$ is the corresponding variance computed from $A_k$.

### 5.4.3.7 Monitoring levels

To simplify the operation of statistical process monitoring in terms of detection and diagnostics, different monitoring levels are defined. The first level is the process or batch level. This includes monitoring the process using multivariate statistics such as
the $D$-statistic and the $SPE$-statistic. Then, if an out of control signal is given in the control charts, a second monitoring level is used. That is, the process variables or components are studied individually using contribution plots (P. Miller et al. [16], J.A. Westerhuis et al. [17]) and univariate statistics to identify and locate the cause of the out of control signal.

5.4.3.8 Post-batch process monitoring

The aim of post-batch monitoring is to find similarities and differences between batches. For this purpose, both models can be used. The batches are monitored in a similar way as with on-line monitoring. For the grey Tucker3 model an observation vector $\mathbf{x}$ ($JK \times 1$) is available. This vector is projected onto the model and from the obtained scores and residuals $D$- and $Q$-statistics are computed. The derivation of the $D$-statistic is identical to the on-line situation and is computed according to

$$D = (\mathbf{a} - \bar{\mathbf{a}}) \mathbf{S}^{-1} (\mathbf{a} - \bar{\mathbf{a}}) \frac{I(P)}{P(I^2 - 1)} \sim F(P, I - P) \quad (90)$$

where $\mathbf{a}$ ($P \times 1$) is the score vector found for the observation vector and $\mathbf{S}$ ($P \times P$) is the variance covariance matrix of $\mathbf{A}$ ($I \times P$). $D$ follows an $F$-distribution with $(P, I-P)$ degrees of freedom. The residual statistic is computed by means of the $Q$-statistic (P. Nomikos & J.F. MacGregor [20]) given by

$$Q_i = \sum_{j \leq k} \varepsilon_{jk}^2 \sim \chi^2_b$$

(91)

where $\varepsilon_{jk}$ is the residual at interval $jk$ and $Q_i$ follows a weighted chi-square distribution with $b$ degrees of freedom.

Post-batch monitoring using the grey Tucker1 model is more complex since the data has to be rearranged prior to analysis. The test statistics are based on the local observation vector $\mathbf{X}_k^r$ ($I \times 1$) and reflect the behaviour of the process at time interval $k$. Thus, each batch generates a vector $\mathbf{c}_k$ ($R \times 1$) and a vector $\mathbf{e}_k$ ($J \times 1$) for each time interval. The generated vectors are stored in the matrices $\mathbf{C}_{total}$ ($I \times RK$) and $\mathbf{E}_{total}$ ($I \times JK$) respectively. By constructing a PCA model on the matrix $\mathbf{C}_{total}$, a post-batch
analysis is performed. This is analogue to the "top level" proposed by S. Wold et al. [54]. Once a PCA model is built, the statistics are similar to the statistics as described for the grey Tucker3 model. As a result two residual matrices have to be analysed. That is, the constructed matrix $E_{\text{red}}(I \times JK)$ and the residual matrix from the PCA model.

5.4.3.9 Batch consistency number

As stated previously, performing a post-batch analysis with grey models results in multiple control charts. For the grey Tucker1 model, one $D$-chart for the scores and two $Q$-charts since there are two residual matrices to be studied. The grey Tucker3 model is studied with a $D$-chart and a $Q$-chart. It is much more convenient to have single index summarising the behaviour of a batch since it is not very easy to express the behaviour of a single batch run with the approach described above. For this purpose, the batch consistency number ($BCN$) is introduced. The $BCN$ is a single index defined according to the ideas of A. Raich & A. Cinar [109] which were further elaborated by H.H. Yue & S.J. Qin [110]. The idea of the $BCN$ is to combine both test statistics into a single index summarising the behaviour of a batch. A derivation of the $BCN$ is given in Appendix D.

For the grey Tucker3 model the $BCN$ is computed according to

$$BCN = x^T \theta x \sim \chi^2_h$$

(92)

where $x (JK \times 1)$ is the observation vector used in the grey Tucker3 model and

$$\theta = \frac{V (V^T V)^{-1} S^{-1} (V^T V)^{-1} V^T}{F(I, I - R)} \left( I - \tilde{V} \right)^T \left( I - \tilde{V} \right)$$

(93)

where $\tilde{V} = V (V^T V)^{-1} V^T$ and $I$ is the identity matrix. The $BCN$ follows a weighted chi-square distribution with $h$ degrees of freedom (H.H. Yue & S.J. Qin [110]).

Computing the batch consistency number for the grey Tucker1 model is done in a slightly different manner since this model has a local observation vector. The $BCN$ is found by computing first the local batch consistency numbers ($BCN_k$) for
Statistical batch process monitoring

Each time interval. After that, the local batch consistency numbers are summarised and divided by the number of time intervals. The local $BCN_k$ is computed according to

$$BCN_k = x_k^T \Theta x_k \sim g\chi^2$$

(94)

where $x_k^i (j \times 1)$ is the local observation vector and

$$\Theta = \frac{B(B^TB)^{-1} S_k^{-1} (B^TB)^{-1} B^T I(I-R) B(B^TB)^{-1} B^T I(I-R) + (I - \hat{B})(I - \hat{B})}{F(I,I-R) R(I^2 - 1) g\chi^2}$$

(95)

where $\hat{B} = B(B^TB)^{-1} B^T$. From the expression in equation 95, it follows that the observation vector is regressed on matrix $B$ in an ordinary least squares manner. However, this regression does not take into account the imposed non-negativity restrictions. This is done with an extra regression step as described previously. Therefore, $x_k^i B(B^TB)^{-1}$ is replaced with $\hat{c}_k$. The same reasoning holds for the regression of the observation vector onto $\hat{B}$. This regression step is now replaced by the residuals $e_k$. Therefore, the local $BCN_k$ is computed for every interval $k$ according to

$$BCN_k = \frac{\hat{c}_k S_k^{-1} \hat{c}_k^T I(I-P) + e^T e}{F(I,I-R) P(I^2 - 1) g\chi^2}$$

(96)

after that, the $BCN$ is found by

$$BCN = \frac{1}{K} \sum_{k=1}^{K} BCN_k \sim g\chi^2$$

(97)

5.4.4 EXPERIMENTAL

5.4.4.1 Process description

The process of making a urethane resin is a multi-stage fed-batch process. In this study, only the first stage is of interest. The main reaction in the first stage is given by:
\[
OCN - R_1 - NCO + HO - R_2 \xrightarrow{\text{RC}} OCN - R_1 - NH - CO - O - R_2
\]

where a di-isocyanate reacts with an alcohol to a urethane resin with an extra NCO-group for further OH addition in subsequent steps.

The operation of this first stage is divided into three steps. The first step, after cleaning and purging of the reactor with \( \text{N}_2 \), is filling the reactor with a di-isocyanate. In the second step a second reactant (alcohol) is added drop-wise and the reaction starts. The third step is the part of the process where all the di-isocyanate has reacted and reaction mixture is kept at a constant temperature. This process is operated according a predefined change in temperature. All reactants are added in precisely known stoichiometric ratios. Furthermore, it is known that a side reaction occurs. The alcohol reacts with the urethane to an undesired side product coupling to the second NCO group of the di-isocyanate.

5.4.4.2 Instrumentation
The full-scale process spectra are obtained using an in-line FOSS NIR SYSTEMS 6500 spectrometer with a 2 nm resolution and a transfection probe with a 2 x 2 mm path length in the range of 1100 nm to 2500 nm. The laboratory-scale compound spectra are obtained using a Perkin-Elmer FT-NIR 9x-III spectrometer with a 2 nm resolution and a 2 mm path length in the range of 1200 nm to 2500 nm.

5.4.4.3 Data pre-treatment
The collected spectra have a baseline offset, which is removed by the use of second derivative spectra estimated using a Savitzky-Golay filter (A. Savitzky & M.J.E. Golay [103]). A visual inspection showed that the best results were obtained using a window of 21 channels and a second-order polynomial. There are 18 batches available \((I = 18)\) for this study. It is known that 4 of the available batch runs are not operated under the defined normal operating conditions and are therefore excluded from the model. The remaining 14 batches are used for modelling and validation. For every batch run \(i (i = 1 \ldots I)\), the selected \(J\) wavelengths are stacked in a three-way array \(X (I \times J \times K)\).

5.4.4.4 Band assignment and band selection
From the literature (E.W. Crandall & N.J. Ashvin [111]) it is known that the diisocyanate (NCO \( \sim 1920\) nm), the alcohol (OH \( \sim 1398\) nm - 1421 nm) and the
product \((\text{NH} \sim 1485 \text{ nm} - 1503 \text{ nm})\) are spectroscopically active in the near infrared range \((1100 \text{ nm} - 2500 \text{ nm})\). Furthermore, it is known that the monitored reaction is exothermal and that a predefined temperature gradient exists. The functional groups present in the compounds are shape sensitive for temperature changes. This is not desired since the pure compound spectra are considered as time invariant factors. Therefore, a selection of spectral bands is made in such a way that the bands are not too much affected by temperature changes and consist of sufficient chemical process information. The selected range is validated by laboratory experiments. That is, pure compound spectra of di-isocyanate and alcohol were measured at different temperatures. Based on the observed changes in the measured spectra, a wavelength selection was made compromising between temperature effects and chemical compound information (E.N.M. Van Sprang et al. [60]).

The upper graph in Figure 83 shows the spectra obtained during a single batch run. In the lower graph Figure 83, the second derivative spectra and the selected wavelength range is given. The circles highlight the wavelength ranges corresponding to the functional groups (NCO, OH, CH and NH) present in the system.

5.4.4.5 Mean Centering
To study the differences between batches it is common to column center the process data across the batch mode in the situation where the three-way array is modelled using a Tucker3 model. This operation removes the common variation in the batches and the remainder of the data consists of the differences between the batches around a zero mean.
To understand the effect of the centering operation, the data is assumed to have systematic variation present in all the batches and unsystematic variation specific for each batch. The systematic variation is process specific with linear differences between the batches. The unsystematic variation is caused by non-linear deviations and experimental error and therefore, typically batch specific. If the data consists of only systematic variation such as well behaved spectra according to the Lambert-Beer law, there is no difference between the model loadings $B$ and $C$ obtained from centered process data and model loadings obtained from non-centered process data (Appendix D).

However, process data usually consists of additional unsystematic variation. If after centering the systematic variation is relatively small compared to the unsystematic variation such as non-linear behaviour, this will cause estimation problems of the model loadings. Therefore, in this study, the models are constructed using non-centered data.

### 5.4.5 RESULTS AND DISCUSSION

#### 5.4.5.1 Construction of the models

The model parameters are estimated using restricted alternating least squares based on prior knowledge of the polyurethane process. The following is known of the urethane process: The law of Lambert-Beer holds, that is, linear addition and no interaction between the individual compounds is allowed. Furthermore, there are estimations of the pure compound spectra available from a previous study (E.N.M. Van Sprang et al. [60]). Also, it is known that the concentration can never be negative; therefore, non-negativity constraints are imposed on the concentration profiles. An overview of the prior knowledge is summarised in Table 11. Furthermore, it is pointed out where this knowledge is incorporated in the models.

<table>
<thead>
<tr>
<th>Prior knowledge</th>
<th>Tucker 1</th>
<th>Tucker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambert-Beer law</td>
<td>Implicit</td>
<td>G</td>
</tr>
<tr>
<td>Pure compound spectrum</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Non-negativity</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

*Prior information available of the polyurethane process.*
In the grey Tucker1 model, the model implicitly gives the Lambert-beer law. The pure compound spectra are imposed as restrictions on the columns of B and non-negativity constraints are imposed on the profiles of C. In the grey Tucker3 model, restricting relevant elements of the core array G to equal zero according to imposes the Lambert-beer law \( G = [g_{11} 0 0 0 \ g_{12} 0 0 0 \ g_{13}] \). The pure compound spectra are incorporated in matrix B and non-negativity constraints are imposed on C.

A priori, the number of black components is unknown. Therefore, a series of models was constructed starting with the simplest model consisting of white knowledge only. Each successive model had an additional black component. Next, the magnitude of the root mean square error (RMSE) of experimental variation was used as a reference to make a decision about the number of black components. The RMSE of experimental variation was estimated with spectra obtained from the period prior to the addition of the alcohol because in this period there is only di-isocyanate present in the reactor. From the magnitude of the RMSE of both white models it was concluded that additional black components do not consist of any significant systematic variation and therefore, are not included in the models. The first model in this study is a grey Tucker1 model with three components \( R = 3 \). The second model is a grey Tucker3 model with \( P, Q, R = [1 \ 3 \ 3] \).

5.4.5.2 On-line monitoring

The monitoring performance of both models is validated. From prior analysis, it is known that four batches of the historical data set are operated in a different manner. Only two of the four faulty batches are used for testing the models and control charts since there are only two distinct types of faults. The remaining two batches have a similar type of fault and are therefore not used for testing. Furthermore, one normal batch is selected as a reference and is excluded from the model. This results in three batches for performance testing: batch1, batch 2 and batch 3 and thirteen batches for modelling the NOC behaviour. Batch 1 is a normal batch operated under NOC. Batch 2 has a clear different dosing profile of the alcohol and batch 3 had a manual dosing of the alcohol instead of the normal automatic dosing which results in slightly different compound profiles.
5.4.5.2.1 First monitoring level with the grey Tucker1 model

The three batches are monitored using the D and SPE-chart. Both charts are presented with 99% confidence limits in Figure 84. The upper graph in Figure 84 corresponds to the normalised SPE-chart. That is, the SPE-statistic is normalised for the 99% confidence limit. There, it is shown that not one of the monitored batches crosses the confidence limit. This implies that there is no significant breaking of the correlation between the channels and the batches behave within the model plane. Deviations in the model plane are observed in the D-chart. In the D-chart, the lower graph of Figure 84, the differences between the batches become apparent. Batch 1 behaves as expected and no alarm is observed. For batch 2 there is a clear out of control signal observed starting at Time = 7. This excessive signal in the D-chart indicates that there is a significant change in the model plane. The batch continues to be out of control until Time = 35 where it operates below the confidence limits until the batch crosses the limits again at Time = 44. The operation of batch 3 shows multiple crossings of the confidence limits. The first major crossing at Time = 31 to 33 and the second at Time = 40 to 50 implying there are two events during the process.

Figure 84
Monitoring level 1 for a grey Tucker1 model. Monitoring batch 1, batch 2 and batch 3.
Because of the nature of the grey Tucker1 model, the concentration behaviour of the three chemical compounds in the reactor is captured by the $D$-chart. Deviations and disturbances such as contaminations are likely to be captured in the SPE-chart. This however, will of course only occur if the disturbances are spectroscopically active.

5.4.5.2.2 First monitoring level with the grey Tucker3 model

The same batches are monitored using the grey Tucker3 model. The normalised SPE-chart and $D$-chart with 99 % confidence limits are given in Figure 85. The SPE-chart shows that none of the batches crosses the confidence limit, implying that there are no significant breakings of the correlation. In the $D$-chart is shown that the normal batch behaves as expected. Batch 2 gives a clear out of control signal at Time = 8 and stays above the confidence limits during the entire batch run. Observing the behaviour of batch 3, it can be seen that this batch touches the confidence limit around Time = 13 to 28. This implies that this batch operates on the borders of specified normal conditions. For the remainder of the run the batch behaves within the specified control limit.

**Figure 85**

*Monitoring level 1 for a grey Tucker3 model. Monitoring batch 1, batch 2 and batch 3*
In general, both models are detecting the same deviations although the time of observation may differ. Furthermore, the grey Tucker1 model is more sensitive in detecting deviations. This can be explained by the nature of the models. The grey Tucker3 model is a global model where the behaviour of a batch is captured by modelling the autocorrelation and cross-correlation between the process variables with the matrices $G$, $B$, $C$. Linear differences from the auto and cross correlation are captured by the scores in matrix $A$. Unsystematic variation and non-linear differences are captured by the residuals $E$.

It is shown that if a process is monitored with a global model, the process history is taken into account up to time $k$ and will lead to graduate changes if an event occurs.

With the grey Tucker1 model it is easy to understand that this model will only take into account the correlation between the process variables and not the auto and cross correlation over time. However, the auto and cross correlations of the process are implicitly captured by the control limits. Here, it is shown that this model is well able to detect deviations in the process. In this specific situation, a semi-batch process monitored by means of NIR spectra. The nature of the process data is defined in such a way that the autocorrelation is captured by the concentration profiles and the model captures the correlation between the spectra. In principle there is no need for taking into account the entire process history since this is implicitly captured by the current measurement. Concluding, for on-line monitoring there is no need for the more complex grey Tucker3 model.

5.4.5.2.3 Second monitoring level grey Tucker1

It is clear from the control charts that batch 2 is not operating under NOC conditions and that batch 3 is operating on the border of the specified confidence limits. In the second monitoring level the aim is to diagnose the cause of the observed deviations in the control charts. First, the grey Tucker1 model will be discussed.

With the grey Tucker1 model it is possible to study the behaviour of the individual chemical compounds using univariate control charts. The behaviour of the individual compounds for batch 2 is given in Figure 86 with 95 % and 99 % confidence limits. The upper graph corresponds to the di-isocyanate, the middle graph corresponds to the alcohol and the lower graph corresponds to the product. For the
alcohol profile, it is obvious that its behaviour is different from normal conditions and is assigned as the cause for the out of control signal in the $D$-chart. The observed profiles suggest that there is no alcohol present in the reactor. However, the reaction takes place implying there is a reactant added to the system. Since it is known that there are no out of control signals observed in the SPE-chart, it is assumed that the added alcohol is similar to normal conditions. Furthermore, it is known that the NIR probe is situated in the lower part of the reactor. Since the alcohol is added drop wise in the top of the reactor, it is well possible that the alcohol has reacted before it can reach the NIR probe. This implies that either the reaction kinetics are faster than usual or a lower amount of alcohol is added to the system.

![Graphs of compound profiles for batch 3](image)

Figure 86

*The second monitoring level using a grey Tucker1 model for batch process monitoring.
Monitoring batch 2.*

The compound profiles for batch 3 are given in Figure 87. Looking at the diisocyanate profiles and the product profiles it seems that the reaction behaves according to normal operating conditions. Having a closer look at the profiles, the
differences become apparent. At the start of the reaction the process behaves as expected until Time = 10 where the alcohol concentration is higher than expected. It seems that the reaction is inhibited. After Time = 20, it seems that the reaction is accelerating until there is no alcohol present in the reactor. This is marked as an out of control signal in the $D$-chart at Time = 30. Next, the alcohol is increasing which is expressed as a signal in the $D$-chart at Time = 40.

![Graphs showing di-isocyanate, alcohol, and product concentrations over time](image)

Figure 87
*The second monitoring level using a grey Tucker3 model for batch process monitoring.*

5.4.5.2.4 **Second monitoring level grey Tucker3**

In case of the grey Tucker3 model it is not straightforward to compute the individual component profiles. However, in this study it is possible to compute the individual contribution of the chemical compound to the $D$-statistic because there are no interactions between the chemical compounds allowed. The derivation of the individual contributions is given in Appendix C.
The individual contributions of the three components (di-isocyanate, alcohol and product) for batch 2 are given in Figure 88 a, b, and c respectively. It is easy to see that from Time = 9 the alcohol has the largest contribution to the D-statistic and therefore, the alcohol can be assigned as the main cause for the out of control signal in the D-chart. For batch 3, the individual contribution of the di-isocyanate, alcohol and product are given in Figure 88 a, b and c. It is known that it behaves on the border of the specified confidence limits. For this batch, the individual contributions do not give a clear indication which compound has the largest contribution to the D-statistic. In general, the contributions of the grey Tucker3 model are less clear compared to the compound profiles found with the grey Tucker1 model. An advantage of the grey Tucker1 model is the direct interpretation of chemistry in the reactor whereas the grey Tucker3 model only gives an indication which component is deviating.


5.4.5.3 Results for the BCN

The BCN is computed (as described in Section 5.4.3.9) for all batches and presented with a 95-percentile limit. That is, the value below of which 95 percent of the reference BCN will fall. The BCN is computed in a leave-one-out manner similar to the derivation of the control charts. The results for the BCN obtained with a grey Tucker 1 model are given in Figure 90. Numbers 15 to 17 in Figure 90 are the indices of the three monitored batches. The results found for the post-batch analysis of the three monitored batches correspond with the results found for the on-line situation since it can be expected that if a batch is out of control with on-line monitoring, this batch is also out of control with the post-batch analysis.
The BCN of the normal batch (number 15) is of such a magnitude that it falls within the magnitude of the NOC batches. For batch 2 (number 16) a large BCN is observed, indicating that this batch is not operated in a consistent manner. For batch three (number 17) it is known that it behaves on the border of normal operating conditions and the BCN shows that the batch behaved in a consistent manner compared to the NOC batches. Furthermore, two NOC batches (number 2 and number 4) have a relatively large BCN. Both batches are considered as NOC batches according to the operators log. An individual inspection learned that these batches had a deviating dosing profile of the alcohol compared to the average alcohol profile.

The results for the grey Tucker3 model are given in Figure 91. In this figure, the numbers 15 to 17 are the indices of the three monitored batches. As can be seen, the BCN of the normal batch lies in the same order of magnitude as the NOC batches. Batch 2 (number 16) has a large BCN indicating that batch 2 is not operated under NOC. The magnitude for batch 3 (number 17) lies at the same level as most of the NOC batches indicating that the overall consistency is within NOC variation. It is interesting to observe that compared to the previous figure, the batch at number 4 has relatively large BCN and the batch at number 2 lies within NOC variation. Since both models are different by nature there is not a clear explanation for this behaviour.
5.4.6 CONCLUSIONS
A strategy is presented to monitor batch process with grey models. To this extent, the batch process monitoring performance of a grey Tucker1 model and a grey Tucker3 model are studied using an industrial case study. Both models are used to capture the NOC behaviour of the process of interest and to derive statistical control charts in order to detect deviations from NOC and to diagnose the cause of these deviations. The differences between the two models are summarised in Table 11.

The grey Tucker1 model has unique model parameters because of the enforced restrictions. The grey Tucker1 model has a better fit compared to the grey Tucker3 model, which is also reflected in the sensitivity of detecting deviations from normal operating conditions. The strong part of the grey Tucker1 model is its on-line monitoring capability, especially in assigning causes of abnormal behaviour once an on-line control chart detects a fault. This is due to an easy interpretation of its model parameters in terms of within-batch variation. The grey Tucker1 model is relatively poor in post-batch monitoring and judging batch-to-batch consistency.

The grey Tucker3 model has also unique model parameters because of the enforced restrictions. However, this model is much more rigid compared to the grey Tucker1 model which is reflected in the poorer model fit. The use of a grey Tucker3 model for post-batch analysis is straightforward. A BCN can be defined and interpreted. The model parameters have an easy interpretation in terms of between-batch variation. This makes the grey Tucker3 model very useful for post-batch monitoring and analysis. The grey Tucker3 model has a relatively poor performance in on-line monitoring, especially in terms of diagnosing faults.

Summarizing, a grey Tucker1 model is more suited for on-line monitoring and a grey Tucker3 model for post-batch monitoring. The grey Tucker1 model focuses more on within-batch variation whereas the grey Tucker3 model focuses more on between-batch variation. The models are, therefore, complementary.
CHAPTER 6 ◆ STATISTICS AND MECHANISTICS*

* This chapter is based on the following publication(s):


H.J. Ramaker, E.N.M. van Sprang, J.A. Westerhuis and A.K. Smilde, The effect of the size of the training set and number of principal components on the false alarm rate in statistical process monitoring, Chemometrics and intelligent laboratory systems, accepted.

6.1 General introduction *

One of the pillars of statistical batch process monitoring is multivariate statistics. It is therefore interesting to investigate and understand these statistics. The most important statistical properties evaluated in this chapter, are the false alarm rate and action signal time. These indices are outlined in Chapter 4.2. The size of the false alarm rate is directly related to the definition of the null hypothesis. In this chapter, a new definition of the null hypothesis is presented.

Also, the effect of the number of samples in the dataset as well as the number of components for the PCA model on the false alarm rate is investigated.

Besides the false alarm rate and action signal time, other statistical properties are of interest. These properties, e.g. normality assumption of the scores/residuals, are discussed separately for the $D$ and $SPE$-chart. Especially the $D$-chart is suspicious because of its poor performance in detecting faults compared to the $SPE$-chart. Several suggestions are given to improve the statistical properties of the control charts. The most important improvements are the Bonferroni adjustment and leave-one-out procedure.

The detection characteristics of the control charts can also be related to mechanistic properties of the control charts (e.g. how process faults manifest in the control charts). For this reason, these mechanistic properties are thoroughly discussed. It is shown how process faults with certain characteristics are distributed in both the control charts. These characteristics are coupled to correlation present in batch data. In order to understand these correlations better, a description of what types of correlation can be expected in batch process data is presented. Furthermore, it is shown how well different models are capable of modelling these correlations.

This chapter consists of four sections. The first section describes a performance assessment of control charts for statistical batch process monitoring. The statistical properties of the $SPE$ and $D$-chart are discussed and tested for six different datasets. Improvements for these charts are presented and tested on the same datasets. In the second section, the effect of the number of components and size of the dataset is studied with respect to the false alarm rate. In the third section of

* HJR/EVS

-206-
this chapter, the mechanics of the SPE and D-chart are thoroughly discussed. It is shown how process faults manifest themselves in these control charts for three different models: global, local and time evolving models. The fourth section of this chapter explains what different types of correlations are present in batch process data. It is shown how the different models are able to capture these correlations.
6.2 Performance assessment and improvement of control charts for statistical batch process monitoring*

6.2.1 SUMMARY
This section describes the concepts of statistical batch process monitoring and the associated problems. It starts with an introduction in process monitoring in general, which is then extended to batch process monitoring. The mechanics of the control charts for process monitoring are studied. Furthermore, the performance of the control charts is discussed by means of two performance indices: the overall type I error and the action signal time. The problems of the existing approach are highlighted and illustrated by e.g. the mechanistic properties of the control charts. Improvements are suggested and checked with the performance indices. To evaluate the effect of the proposed improvements as well as to assess the performance of the existing approach, six different data sets have been used: four industrial data sets, a simulated batch reaction and a laboratory spectral data set.

6.2.2 INTRODUCTION
The batch-wise production of chemicals, bio-chemicals, pharmaceutics and foods is common in industry. Monitoring such processes is necessary for several reasons: safety, improving product quality and a better understanding of the process. Besides knowledge based or state estimation approaches, multivariate statistical approaches can be used for monitoring batch processes.

The fundamentals of statistical process control were described by W.A. Shewhart [5]. H. Hotelling [8] was one of the first who introduced a multivariate approach to statistical process control to monitor a process of a multivariate nature. J.F. Jackson [9] then applied principal component analysis (PCA) to reduce the dimensionality of the multivariate data. B.M. Wise & N.L. Ricker [112] and J.V. Kresta

et al. [10] proposed a multivariate control technique that uses PCA or projection to latent structures (PLS). Finally, P. Nomikos & J.F. MacGregor [6] extended the concept of multivariate statistical process control using PCA to batch processes. This is called statistical batch process monitoring. Their approach to batch process monitoring is a purely data driven approach based on linear multivariate analysis methods. The relevant variation of a process is captured in a dimension-reduced model. This is advantageous in terms of interpretability and comprehensibility. Two test statistics are derived from this model and both are monitored using control charts.

In the paper by E.N.M. Van Sprang et al. [41] several problems were encountered with respect to batch process monitoring. In that paper the results of a comparison study between different approaches for on-line batch process monitoring is described. Two performance indices were used to qualify the different approaches. Also, the validity of assuming normality for the underlying distributions of the test statistics was checked. It was concluded that one of the control charts works reasonably well despite violating the normality assumption. The other control-chart has a poor performance with respect to the indices while the normality assumptions are reasonable.

The aim of the current section is to give an overview of the problems encountered with on-line batch process monitoring and propose possible solutions. It will be shown that the concepts from continuous process monitoring are not straightforwardly applicable to batch process monitoring. The mechanistic and statistical properties of the control charts for batch process monitoring are studied. Ideas are given on how to improve the performance of the control charts.

This section is organised in the following way. Section 6.2.3 gives a brief summary of the concepts of multivariate statistical process control considering the different phases in monitoring. In Section 6.2.4, the control charts for multivariate statistical process control are discussed for continuous processes with and without dimension reduction. An understanding of the type of process faults detected by the control charts is given. Also, the extension of multivariate statistical process control to post-batch analysis is discussed. This is followed by explaining on-line batch process monitoring. In Section 6.2.7, a description of the datasets used as examples is given. In Section 6.2.8, problems of statistical batch process monitoring are discussed and
possible solutions are applied to the examples. Finally, conclusions will be given in the last Section 0.

6.2.3 MULTIVARIATE STATISTICAL PROCESS CONTROL: THE CONCEPTS

W.H. Woodall [48] stated that it is very important to improve communication between practitioners and researchers concerning statistical process control methods. One way to establish an improved communication is to assign phases to the concepts of multivariate statistical process control. In this section, multivariate statistical process control is carried out in three phases: The initial, training and application phase. Here, the training and application phase refer to respectively phase 1 and phase 2 from standard SPC terminology.

The initial phase consists of probably the most time consuming part: that is the collection and cleaning up of the process data. The goal of the training phase is to obtain a set of in-control observations, build a suitable model for these observations and derive control limits from this model. For this reason, the data must be checked for outliers that do not represent an in-control process. This step in the training phase resembles phase 1 from SPC terminology. The model can be further extended by incorporating existing knowledge of the process under consideration (A.K. Smilde et al. [113]). This can be helpful in terms of e.g. fault diagnosis or increased process understanding.

The application phase starts with monitoring new independent observations to detect abnormal deviations from the in-control process. This step refers to phase 2 from SPC terminology. Besides detection of process upsets, fault diagnosis has to take place in this phase. The detection and diagnosis of the fault may lead to an improvement of the process.

6.2.4 MULTIVARIATE STATISTICAL PROCESS CONTROL: THE CONTROL CHARTS

The concepts of multivariate process control for batch processes are similar to multivariate process control for continuous processes. However, in transferring the methodology from continuous to batch processes, there are some issues that are not straightforward.
In the following an overview is given how the concepts of multivariate statistical process control for continuous process are extended to on-line batch process monitoring. This is done is a stepwise manner where the related problems for each step are highlighted and discussed.

Process measurements of \( J \) process variables are made at \( I \) time points on a continuous process and collected in the data matrix \( X (I \times J) \). Preprocessing the process data \( X \) is a first step. Generally, mean centering of the variables is a good idea and also scaling if the measured units of the process variables are different. In this section it is assumed that the data is properly pre-processed. Furthermore, \( X \) is believed to represent in-control observations.

### 6.2.4.1 Continuous processes without dimension reduction

In 1947 Hotelling introduced the \( T^2 \) statistic as a method to apply statistical quality control to correlated data with a multivariate nature. This summary statistic enables easy control charting and allows for detecting faults of highly correlated measurements. In the case of continuous processes for chemical applications, at every time interval \( J \) process variables are measured and repeated \( I \) times. These measurements can be arranged in a matrix \( X (I \times J) \). For a new measurement vector \( x_{new} (J \times 1) \) the test statistic \( T^2 \) is calculated according to:

\[
T^2 = (x_{new} - x_{\bar{X}}) S^{-1} (x_{new} - x_{\bar{X}}) \tag{98}
\]

where \( S \) is the estimated covariance matrix of \( X \) and \( x_{\bar{X}} \) is the target value or grand mean. Therefore, if the goal is to obtain an in-control set of observations as described in the training phase, the UCL is given by (N.D. Tracy et al. [14]):

\[
UCL = \frac{(I-1)^2}{I} B(\alpha; J / 2; (I - J - 1)/2) \tag{99}
\]

The \( UCL_{new} \) to use for detecting deviations from the in-control process for new independent observations are given by:

\[
UCL_{new} = \frac{J (T^2 - 1)}{I (I - J)} F(\alpha; J; I - J) \tag{100}
\]
An important note here is that all the process variables are used and no dimension reduction takes place. With a large number of process variables this approach breaks down, e.g., $S$ might become (nearly) singular.

6.2.4.2 Continuous processes with dimension reduction

In a chemical process, the dimensions of $X$ may be become large in the second mode $J$. It is therefore advantageous to reduce the dimensionality of the data. Principal component analysis (PCA) is a multivariate statistical method well suited for this purpose. J.E. Jackson & G.S. Mudholkar [24] was one of the first to apply a principal component analysis for this type of problem. Reducing the dimensionality of the data results in two test-statistics. The derivation of these test statistics will follow. In this section, these test statistics are referred to as $D$-statistic and Squared Prediction Error (SPE) statistic.

A PCA decomposes the matrix $X$ into the sum of $R$ outer products of scores $t$ and loadings $p$ plus a residual part $E$:

$$X = \sum_{i=1}^{R} t_i p_i^T + E = TP + E$$

(101)

where $T$ ($I \times R$) is the score matrix, $P$ ($J \times R$) contains the loadings and $E$ ($I \times J$) is the residual matrix. A geometrical representation of equation 101 is given in Figure 92.
Figure 92

Geometrical interpretation of PCA for a two component model in a three dimensional space. The i-th row of \( X \) represents a sample and can be drawn as a vector \((x_i)\) in the space. The subspace where this sample is orthogonally projected on is spanned by the columns of \( P \). The distance \((e_i)\) from \( x_i \) to the model plane is given is the i-th row of \( E \). The co-ordinates, or scores, of the projection in the plane are given by the i-th row of \( T \).

Once the model plane, spanned by the columns of \( P \), is defined, the scores and residuals for a new scaled measurement \( x_{\text{new}} \) \((j \times 1)\) are found by projection on the model plane. This is a regression problem:

\[
x_{\text{new}} = Pt_{\text{new}} + e_{\text{new}} \quad \text{(102)}
\]

where \( t_{\text{new}} \) is estimated using least squares. Two test statistics are defined. The \( D-\)statistic describes the Mahalanobis distance from the projection on the model plane to the centre of the model plane, and describes the systematic variation in the data. The centre of the model plane corresponds to average process behaviour. Therefore, the closer the projection of a sample is to this centre, the better. The \( SPE-\)statistic (Squared Prediction Error) represents the Euclidian distance between the
measurement vector and its projection on the model plane. This distance describes the variation in the data that is consistent with the model.

For the ease of illustration, the number of PCA components \( R \) is chosen to be two. The test statistics are calculated according to:

\[
\begin{align*}
\mathbf{t}_{mx} &= \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{x}_{mx} = \mathbf{P}'\mathbf{x}_{mx} \\
\mathbf{e}_{mx} &= \mathbf{x}_{mx} - \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}\mathbf{P}'
\end{align*}
\]

\[
\rightarrow D = \mathbf{t}_{mx}'\mathbf{S}^{-1}\mathbf{t}_{mx} \quad \text{(103)}
\]

\[
\rightarrow \text{SPE} = \mathbf{e}_{mx}'\mathbf{e}_{mx} \quad \text{(104)}
\]

Here, \( \mathbf{S} \) is the variance-covariance matrix of the scores \( \mathbf{T} \) and is usually diagonal because the scores \( \mathbf{T} \) are orthogonal. Subsequently, the test statistics are plotted in the control charts and \( \mathbf{x}_{new} \) is assigned to be statistically out of control if one of the control charts signals. The UCL for the \( D \)-statistic is the same as given in equation 100 except that \( J \) is replaced by \( R \). The UCL for the \( \text{SPE} \)-statistic resembles the critical value \( \text{SPE}_\alpha \). The derivation of this critical value can be found in J.F. Jackson & G.S. Mudholkar [24].

6.2.5 MECHANISTICS OF THE \( D \) AND \( \text{SPE} \)-CHART

Detecting a fault at the \( D \) or \( \text{SPE} \)-chart is important, but of equal importance is the diagnosis of the fault found. There is a general believe that new events in a process not present in the NOC data are detected by the \( \text{SPE} \)-chart, e.g. sensor failure. The \( D \)-chart diagnosis abnormal variation in the process that still obeys the correlation structure of the process variables as reflected in the matrix \( \mathbf{P} \). Hence, this is a distinction between faults detected in the \( \text{SPE} \) and \( D \)-chart. The following citations were found in the literature:

R. Dunia & S.J. Qin [114]: "A significant increase in \( \text{SPE} \), which measures the distance from the principal component subspace, indicates a breakdown of the normal correlation". A significant increase in \( T^2 \) can be due to a fault or normal change in the process throughput that conserves the correlation structure ".

S. Albert & R.D. Kinley [43]: "...mainly in the direction of the \( \text{SPE} \) statistic, which indicates that warnings were generated because of new events rather than changes in correlation structure ".

-214-
P. Nomikos & J.F. MacGregor [19]: "If a new batch is still operating in the same way as the batches in the reference database, but still has a larger than normal variation in its measurements, this will show up clearly as large deviations of the $t$ scores from the origin of the reduces space. In the case in which a totally new fault that is not represented in the database occurs, the principal components will then not be able to describe correctly the variation. Thus the new observations will move off the MPCA plane, resulting in large values of the $SPE$.

Similar statements can be found in J.V. Kresta et al. [10], K.A. Kosanovich et al. [7], W. Lin et al. [115], B. Lennox et al. [37].

To investigate the diagnostic properties of the $SPE$ and $D$-chart, a detailed analysis is made of the mechanics of the $SPE$ or $D$. Therefore, three types of faults are distinguished and discussed: i) a fault detected only in the $D$-chart ii) a fault detected only in the $SPE$-chart iii) a fault detected in both charts. It will be shown in the following how these faults can be envisioned and that it is not easy to establish a simple relationship between the type of fault and the detection signal.

i) **Faults detected only in the $D$-chart**: The vector $\mathbf{x}_i$ ($J \times 1$) in Figure 93 represents a sample of $J$ process variables at time $i$ under normal operating conditions, hence the control charts do not signal an error for $\mathbf{x}_i$. The ellipse represents the UCL for the $D$-statistic.
At the next time point \( i+1 \), the vector \( \mathbf{x}_{i+1} \) represents a measurement containing a process fault at time \( i+1 \). This sample has such properties that it only manifests itself in the \( D \)-chart. This implies that the length of the residual vector does not break the limit given by \( SPE^\alpha \), or \( \mathbf{e}_{i+1} = \mathbf{e}_i \). Only the co-ordinates after projection of \( \mathbf{x}_{i+1} \), given by \( \begin{bmatrix} t_1 & t_2 \end{bmatrix}_{i+1} \), are significantly different for \( \mathbf{x}_{i+1} \) as compared to the co-ordinates for \( \mathbf{x}_i \). Thus, this particular fault only changes within the plane. Clearly, the projection of \( \mathbf{x}_{i+1} \) results in a distance further away from the centre of the plane and now falls outside the confidence limits. It is insightful to 'translate' this information from the latent variable space to the raw process variables. In other words, if the scores change and the residuals stay constant due to a process disturbance, how can such an event happen in the original process variables? This is only possible if the new measurement moves in the process variable space parallel to the model plane. In Figure 94, the black bars (\( \mathbf{x}_i \)) represent the scaled values for 15 process variables from a real process.
Let the white bars represent the process event $x_{i+1}$. This process event has been constructed. Considering $x_i$ and $x_{i+1}$, the correlation between the process variables is consistent and not broken, only more intense than under normal operating conditions. This is considered to be a multivariate fault where the correlation pattern is not disturbed and to be detected in the $D$-chart. Only if the process variables behave exactly as depicted in Figure 94, this will result exclusively in a change within the plane spanned by the loadings.

\textit{ii) Faults detected only in the SPE-chart:} The vector $x_i$ ($J \times 1$) in Figure 95 represents a sample of $J$ process variables at time $i$ under normal operating conditions, hence the control charts do not signal an error for $x_i$. 
Again, the vector $x_{i+1}$ represents a measurement containing a process upset at time $i+1$. This sample has such properties that it only manifests itself in the SPE-chart. This implies that the scores have a limited space to vary. In this example $t_{i+1} = t_i$. Only the length of the residuals given by $\|e_{i+1}\|^2$ are significantly different for $x_{i+1}$ as compared to the residuals of $x_i$. Thus, this particular upset moves perpendicular from the plane. The translation of this fault in terms of raw process variables is given in Figure 96. The black ($x_i$) represents the scaled values for 15 process variables and the white bars represent the process upset $x_{i+1}$.
The fault is constructed in such a way that process variable 6 is erroneous. This becomes obvious from Figure 96. To detect this specific fault only in the $SPE$-chart, all the other process variables must change to some degree to be perpendicular to the model (i.e. in the direction $(I - PP^T)$). That is, some process variables are hardly affected, some are breaking the correlation structure and some will follow the correlation structure. To move perpendicular away from the plane in the same direction as $e_y$, all process variables must contribute. Only if the process variables behave as depicted in Figure 96, the fault will be detected exclusively in the $SPE$-chart.

The reason for all these changes in the process variables necessary for a fault only to show up in the $SPE$ is that the model plane is tilted meaning that all the process variables contribute to the model. In other words, none of the process variables are orthogonal to the plane.

It may become obvious that for a fault to show up exclusively in the $SPE$ or $D$-chart the process variables have to change in a very unique and specific way. That
is, to find a fault exclusively to be detected in the SPE or D-chart can be considered as extreme.

iii) Faults detected in both charts

From practice it is seen that most process upsets are detected in both the control charts. The reasoning for this observation will be illustrated in the following using an example.

For a fault to be detected exclusively in the $D$-chart all the process variables must behave as depicted in Figure 94. In reality there are dynamics in the process. This means that not all the process variables will immediately respond to such a process upset. As an example, suppose the cooling of a reactor is insufficient. Obviously, the area near the cooling jacket such as the wall of the reactor is heated faster than the area in the middle of the reactor. If the sensors are located in the middle, it will take some delay for the sensors to respond to the process upset. As a result, first the correlation pattern will be broken by the sensor located at the wall before the others get affected. A breakage of the correlation pattern is more likely to be detected in the $SPE$-chart.

For a fault to be detected exclusively in the $SPE$-chart all the process variables must behave as depicted in Figure 96. Hence, simple faults like a sensor failure will also always affect the $D$-chart, because not all the process variables change according to the pattern of Figure 96. This problem is known in the factor analysis literature as embedded error (E.R. Malinowski [125]).

In mathematical terms embedded error can be explained as follows. Let the vector $\mathbf{x}$ represents the raw data defined as:

$$\mathbf{x} = \mathbf{x}^* + \mathbf{e} \quad (105)$$

where $\mathbf{x}^*$ is the pure data and $\mathbf{e}$ is the true experimental error. The pure data $\mathbf{x}^*$ can be written exactly according to the model, that is:

$$\mathbf{x}^* = \mathbf{P}\mathbf{t}^* \quad (106)$$

Here, it is assumed that $\mathbf{P}$ reflects the true underlying correlation in the process. The scores $\mathbf{t}$ for $\mathbf{x}$ are found by projecting $\mathbf{x}$ according to:
\[ t = P'x = P'(x' + e) = P'x' + Pe \neq t' \] (note that \( t' = P'x' = PPT'= t' \) ) (107)

Obviously, the scores \( t \) found are not the true scores \( t' \) because part of the error mixes within the scores \( t \). Now, suppose the true experimental error looks like:

\[ e = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \]

where there is a sensor failure on the sixth process variable. It can be seen from equation 107, by projecting \( e \) on \( P \) that this projection is mixed within the scores \( t \). The SPE-chart is based upon the extracted error \( e^0 \) and is found by:

\[
e^0 = x - Pt = x - PP'x = x - PP'(x' + e) = x' + e - PP'(x' + e) = (x' - PP'x') + (e - PP'e) = (x' - x') + e - PP'e \] (108)

and thus \( e^0 \neq e \). As a result, the SPE-chart is based upon the part of the experimental error \( e \) that is not projected. Clearly, the fault is distributed in both the monitored scores \( t \) and residuals \( e^0 \) by nature. Therefore, as a consequence of applying PCA for monitoring purposes, the charts can only be complementary if \( e^0 = e \) (Figure 96). It was shown that this situation is extremely rare.

Summarised, a process disturbance very likely will not behave in such a way that it is detected in either the \( D \) or \( SPE \)-chart. The fault describing the insufficiently cooled reactor is only exclusively detected in the \( D \)-chart if the time is given to all the other process variables to be affected. The sensor failure can never cause the other process variables to change in such a way that it will be detected exclusively in the \( SPE \)-chart. As a result, faults will mix up or detected in both charts. This hampers the diagnostic capabilities of the control charts.

6.2.5.1 Post-batch analysis
So far the focus has been on continuous processes. The reason for this is that the arguments related to the \( SPE \) and \( D \)-chart as discussed in the previous chapters also apply to multivariate statistical process control for batch processes. The same
problems associated with the mechanistics of the control charts are encountered in batch process monitoring, and even more severe.

The objective of post-batch analysis is to assign a finished batch to be statistically in or out-of-control by comparing the batch to a reference distribution. In the initial phase of statistical batch monitoring, batch data are collected from the plant and are judged for their normal operating behaviour. Apart from using external information regarding producing on-spec products, batches are also subjected to post-batch analysis for judging whether or not they have been run under normal operating conditions (NOC). The selected batches are referred to as NOC batches.

Commonly batch process data is stored in such a way that it fits a three-way data array $X$. Suppose $J$ process variables are measured over $K$ time points from a single NOC batch run, and stored in a matrix $A (J \times K)$. A simple way of storing several batch runs is to stack each run in the three-way data array $X (I \times J \times K)$ according to Figure 97.

![Figure 97](image)

Arrangement of batch process data prior to model building.

The different batches are denoted by $i = 1, \ldots, I$ and form the first mode of $X$. The process variables are denoted by $j = 1, \ldots, J$ and form the second mode of $X$. The time points at which measurements are taken form the third mode of $X$ and are denoted by $k = 1, \ldots, K$. Here, it is assumed that the NOC batches represent a process that is statistically in-control. Therefore, the variation between batches from the NOC data $X$ is considered to be common cause variation that is to be modelled.

Since PCA is a linear technique, pre-processing of $X$ is required because of the highly non-linear behaviour of batch processes. In this study $X$ is column-centred and slab-scaled within each process variable (H.A.L. Kiers [49]), thereby removing the average trajectories of all the process variables and giving all process variables $a \text{ priori}$ equal weight. The pre-processed data in $X$ now represents deviations from the average trajectory and is therefore approximately linear (P. Nomikos & J.F. MacGregor [6]).
The data-array $X$ is matricized (H.A.L. Kiers [49]) in the batch direction $I$ resulting in $X_{i\times j\times k}$. Because of matricizing $X$, the process variable mode and time mode are nested. The number of columns of this $X$ is much larger as compared to data collected from continuous processes because all process variables at each point are considered as new variables. For an average number of process variables, this number may be as high as 2000. For a batch process monitored with spectroscopic instruments this number can easily be 100,000, thus, dimension reduction is necessary.

The derivation of the model and calculation of the test statistics is not different from the continuous processes with dimension reduction. Therefore, a PCA model on $X_{i\times j\times k}$ is built according to

$$X_{i\times j\times k} = TP' + E$$

where $T$ ($I\times R$) are the scores, $P$ ($J\times K\times R$) the loadings and $E$ ($I\times K\times R$) the residuals of the model. Every row of $X$ represents a completed NOC batch.

The test statistics for a new independent finished batch $x$ are calculated in the same manner as equation 103 and equation 104. For post-batch analysis the $SPE$-statistic is referred to as the $Q$-statistic (P. Nomikos & J.F. MacGregor [19]) to distinguish it from the on-line statistic $SPE$ (see later), that is the sum of squared residuals over all time periods.

The data structure of $x$ has very important consequences for diagnosis and detection. Besides the correlation between the process variables for this type of batch data, there is also correlation in the time direction. That is, if a process fault that changes the correlation structure occurs somewhere halfway the batch run, it can never be detected uniquely in the $D$-chart. This situation has been illustrated in Figure 98.
The data represent the measurement of three temperatures in a batch reactor. The batch duration for this process is 15 time intervals. The x-axis of Figure 98 describes the vector $\mathbf{x}^{15}$ when the 3 process variables are measured at the 15 time points. In the figure each measurement of the process variables is separated by small blank spaces. A process fault, indicated by the white bars, starts at time interval 12: the temperatures are simultaneously affected. The black bars represent the in-control situation where the process would still be normal. Clearly, the correlation pattern for the process variables at time interval 12-15 is the same. However, the test statistics are calculated by projection of the entire batch run, including time interval 1-11. Since for time interval 1-11 the correlation pattern is not changing, the projection of $\mathbf{x}$ can never lead to detection of the fault exclusively in the D-chart. By projecting the entire batch run, the correlation pattern considering all the process variables is broken. Breakage of the correlation patterns is likely to be detected in the $Q$-chart.

Another consequence of the data-structure of batch data can be understood from the following. Figure 99 represents the same temperature measurements in the batch reactor.

![Figure 99](image)

*Construction of batch data including sensor fault.*

Suppose one of the temperature sensors gives an abnormal signal due to a sensor failure at time interval 8. It was shown for the continuous case that these types of faults are only detected uniquely in the SPE-chart under specific conditions: all process variables must be changed in a certain way. To calculate the $Q$-statistic, the entire batch run is projected. Thus not only the two temperature measurements at time interval 8 have to change, but also the 'past' measurements from time interval 1-7 and future measurements of time interval 8-15 need to change. Clearly, this cannot happen. Therefore, this fault can never be detected uniquely in the $Q$-chart.
Apart from the problems mentioned above, the problem of embedded error becomes even more pronounced in the $Q$-statistic. Moreover, since the $Q$-statistic takes a summation over the time direction, large squared residuals at a certain time point tend to be levelled out by small squared residuals at other time points (P. Nomikos & J.F. MacGregor [19]). This also hampers the detection performance of the $Q$-statistic.

Summarising, most process faults during a batch run will break the correlation pattern. This makes diagnosing difficult. The breakage of the correlation will show up mainly in the $Q$-chart and a portion of the fault creeps into the $D$-chart due to the embedded error. Hence, the diagnostic capabilities of the $Q$ and $D$-chart are limited. Moreover, due to the limited detection power of batch charts, using post batch analysis for deciding whether or not a training batch is an NOC batch is difficult.

6.2.6 Statistical performance of control charts for batch process monitoring

In the first part of this chapter, the problems associated with the mechanics of the control charts are discussed. In the following chapters, the statistical properties of these control charts are discussed. Problems related to these statistic properties are highlighted and suggestions for improvements are given. The properties of the charts are tested by means of performance indices. The effect of the improvements is tested using several batch process datasets.

6.2.6.1 On-line batch monitoring

It is also desirable to monitor a batch from the beginning until its completion. That is, at every time interval $k$, the batch $x_k$ is tested whether it is statistically in or out-of-control by comparing the test statistics $SPE$ and $D$ for $x_k$ with their reference distribution. This is called on-line batch monitoring.

A completely new batch independent from the NOC data will be denoted as $x_k$. This batch will be monitored on-line, $x_k$ will be projected on to the space generated by the model the model $P$ although the batch has not finished yet. The loadings $P$ are the loadings as given in equation 109. In order to project this particular batch on $P$ to obtain the scores needed for monitoring, the dimensions of $x_k$ must
match the dimension of the matrix $P$. Therefore $x_k$ is constructed as depicted in Figure 100.

Suppose at time interval $k$ a measurement of $J$ process variables comes in. This will be denoted as the current part of $x_k$. All the measurements until $k$ are denoted as the past part of $x_k$. Therefore, the future part of $x_k$ from $k+1$ to $K$ must be estimated somehow. Thus at each time interval $k$, the vector $x_k$ is constructed.

In this study the current deviations approach (P. Nomikos & J.F. MacGregor [6]) is chosen to fill in the future part of $x_k$. E.N.M. Van Sprang et al. [41] used several performance indices and data-sets to show that this approach works reasonably well compared with other methods for on-line batch process monitoring.

The principle of batch process monitoring is comparing a new batch $x_k$ ($JK\times 1$) against a reference distribution. Therefore, in order to have a sensible comparison, the NOC batches are treated in a similar fashion as $x_k$. Suppose there are 67 NOC batches and the batch duration $K$ equals 58. To monitor the new batch $x_k$, 58 different vectors are constructed (see Figure 100) according to the current deviations approach, each denoted by $x_k$. Therefore, 58 different reference distributions are required, each denoted by $X_k$ ($J\times K$). The matrix $X_k$ consists of three parts: a part with known process measurements, current measurements and a part with unknown process measurements. The latter part is estimated using the current deviations approach. This procedure is repeated for each time interval $k$. 

-226-
Once $X_k$ is known for each time interval, the matrix is projected on the model using the loadings $P$ in order to obtain the necessary scores $T_k$ ($I \times R$) and residuals $E_k$ ($I \times K$). These scores and residuals serve as reference distributions for the test statistics $D_k$ and $SPE_k$.

Thus, at every time interval $k$, the batch $x_k$ is tested whether it is statistically in or out-of-control by comparing the $SPE_k$ and $D_k$-statistic for $x_k$ with their reference distribution. The derivation of the test statistics is explained in the following.

**D-statistic for on-line batch monitoring**

The calculation of the $D$-statistic for a new batch $x_k$ involves the scores. To calculate this test statistic, first the Mahalanobis distance ($MD$) is calculated to take variances and covariances between the scores into account (K.V. Mardia et al. [116]):

$$t_k = (PP)^{-1}P'x_k = P'x_k 
\rightarrow \quad MD_k = t_k'S^{-1}t_k$$

(110)

where $t_k$ ($R \times I$) are the scores calculated at time interval $k$, and $S$ ($R \times R$) is the variance-covariance matrix of $T_k$. The scores are considered to be multivariate normally distributed due to the central limit theorem (P. Nomikos & J.F. MacGregor [19]). The $D$-statistic is the Mahalanobis distance times a correction factor, and is calculated according to (P. Nomikos & J.F. MacGregor [19]):

$$D_k = MD_k \cdot \frac{I(I-R)}{R(I^2-1)}$$

(111)

and follows an $F(R, I - R)$ distribution, assuming that the new scores are drawn from the same distribution as the reference distribution (N.D. Tracy et al. [14]). Another assumption involving the $D$-statistic is that the covariance matrix $S$ of the scores $t_k$ of the new batch equals the covariance matrix $S$ from the reference distribution $T_k$. The limits for the control-chart are not time dependent. This is because the degrees of freedom for the $F$-distribution are determined by $R$ and $I$, which are fixed.

If a fault is not apparent from the beginning of a batch the correlation structure is always broken by definition and will therefore be detected in the $SPE$-chart. Thus after roughly 10% of the batch time the $D$-chart can be discarded as well. Furthermore, even if the fault is apparent from the start, it will break the correlation
structure due to the filling-in procedure. This is not true only for the current deviations approach, but holds for all filling-in methods. Only when the batch is completely finished, and the fault was apparent from the start until the end, a post batch analysis can result in a signal in the D-chart as explained earlier. It might become obvious from this that the D-chart is useless for on-line monitoring purposes using any filling-in procedure.

P. Nomikos & J.F. MacGregor [19] proposed also a third method to deal with the problem of an unfinished new batch. In this method \( x^*_i \) is projected onto the part of \( P \), which is associated to the measurements up to and including time point \( k \). This is called the projection approach and it is recommended to use only after at least 10% of the new batch has finished, otherwise no reliable statistics will be obtained. From the analysis above it becomes clear that the projection approach will not solve the problems of the D-statistic either. There is only one specific situation where the D-chart, using the projection approach, is useful. Suppose there is abnormal variation within the process variables from the start of the batch that behaves according to the correlation structure. Since the projection approach does not use a filling-in procedure, such a fault will indeed show up in the D-statistic.

**SPE-statistic for on-line batch monitoring**

The calculation of the SPE-statistic for a new batch \( x_k \) involves the residuals. Using the scores \( t_k \) from equation 110, the residuals \( e_k \) \((1\times1)\) are calculated according to:

\[
e_k = x_{k,j-1}^{\text{J}+1} - t_k P_{j-1}^{\text{J}+1} \rightarrow \text{SPE}_k = e_k' e_k (112)
\]

Notice that the residuals are calculated using only the current part of \( x_k \) and \( P \) indicated by \( R, (\text{J}k - \text{J}+1): \text{J}k \). That is, only the part corresponding to the \( J \) process variables measured at time interval \( k \) are of interest. This notation is read as follows. Suppose the matrix \( P' \) has 5 rows and 100 columns. The process variables \((J = 10)\) are nested within the time mode \((K = 10)\) as a result of matricizing. Suppose only the measurements at e.g. time \( k=3 \) are of interest. In other words only the columns 21 until 30 are of interest. This is denoted as \( 5, 21:30 \).

The effect of the embedded error is decreased for the \( \text{SPE}_k \) by the fact that only the current part of the calculated residuals is used. Moreover, the \( \text{SPE}_k \) only
considers squared residuals at a certain time point, and does not sum-up over time as does the Q-statistic. This increases the detection power of the $SPE$-chart. The $SPE_k$ follows approximately a weighted $\chi^2$ distribution (P. Nomikos & J.F. MacGregor [19]) with $h$ degrees of freedom and weight $g$ to account for the magnitude of $SPE_k$, assuming that $x_k$ is multivariate normally distributed with expectation zero (G.E.P. Box [23]). These parameters $g$ and $h$ are estimated (J.E. Jackson & G.S. Mudholkar [24]) at each time interval $k$ using the proper reference distribution: the current part of $E^i$ of size $I \times J$. The limits for a $SPE$-chart will vary in time. This is because the degrees of freedom at each time interval are estimated from the residuals.

### 6.2.6.2 Further comments on-line statistical batch monitoring

The batch is statistically out of control when the limits are broken in one of the control charts. If the limits are broken the cause of the deviation from the normal operating conditions can be located using contribution plots (P. Miller et al. [16], J.A. Westerhuis et al. [17]). Since the (on-line) control charts $D$ and especially $SPE$ are much more sensitive than the post-batch versions ($D$ and $Q$), it is recommended to use on-line monitoring to judge whether training batches are NOC batches or not (apart from *a priori* information concerning on-spec products).

### 6.2.7 DESCRIPTION OF THE DATA

In order to evaluate the effect of improvements (see Section 6.2.8) for on-line monitoring, six different types of data sets are used as an example. The different features of these data sets will be explained in the following. The NOC data sets are used to compute the overall type I error and disturbed batches are used to compute the action signal time (AST). The definition of these performance indices is given in Section 6.2.8. The time of disturbance occurrence for the erroneous batches and the dimensions of the three-way data array $X$ for the studied data sets are given in Table 12.
Table 12

<table>
<thead>
<tr>
<th>Data set (IxJxK)</th>
<th>Disturbed batches</th>
<th>Fault start</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set 1 (67 x 15 x 58)</td>
<td>Batch 14</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Batch 17</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Batch 24</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Batch 49</td>
<td>46</td>
</tr>
<tr>
<td>Data set 2 (27 x 67 x 271)</td>
<td>Batch 2*</td>
<td>138</td>
</tr>
<tr>
<td>Data set 3 (47 x 3 x 100)</td>
<td>Wrong batch 1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Wrong batch 2</td>
<td>1</td>
</tr>
<tr>
<td>Data set 4 (50 x 9 x 200)</td>
<td>Batch 106*</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Batch 99*</td>
<td>100</td>
</tr>
<tr>
<td>Data set 5 (47 x 8 x 116)</td>
<td>Batch 41</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Batch 46</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Batch 47</td>
<td>94</td>
</tr>
<tr>
<td>Data set 6 (36 x 10 x 100)</td>
<td>Batch 50</td>
<td>56</td>
</tr>
</tbody>
</table>

Description of the data sets and disturbed batches. * = exact occurrence of fault known

Data set 1
This data set is from an industrial (Shell) suspension polymerisation of polyvinylchloride (PVC) in a batch reactor. The data set consists of 67 batch runs obtained under NOC and 15 process variables were measured for a period of 58 time intervals. The NOC batches are aligned using the conversion as a maturity variable. A more detailed description is given by A.A. Tates et al. [2]. The process variables consist of: temperatures inside reactor and cooling water, condenser duty, agitator speed and power supply, two mass streams of cooling water, batch reactor level and pressure. Four disturbed batches are available to calculate the AST. Analysis of univariate plots of the process variables is used to study the disturbed batches. Batch 14 shows a disturbance in the amount of cooling water through the condensor. Batch 17 indicated a disturbance in the temperature of the top condensor. Batch 24 suffers from disturbances for the batch level and the power supply to the agitator. Batch 24 suffers from disturbances for the batch level and the power supply to the agitator. Analysis of batch 46 revealed that the temperature and the amount of refrigerant water to the jacket are disturbed at the end of the batch run.
Data set 2
Data set 2 consists of a laboratory spectroscopic batch process of a two-step biochemical conversion reaction described in S. Bijlsma et al. [66], and can be obtained from the Process Analysis & Chemometrics website of the University of Amsterdam. The data set consists of 27 NOC batch runs measured at wavelengths 300-500 nm during 271 time intervals. One batch run with a pH disturbance at time interval 138 is available to compute the AST.

Data set 3
This data from Loders Croklaan concerns a food manufacturing batch process during which the catalysed hydrogenation of bean oil is performed. Three process variables (H₂ flow, pressure and temperature) are measured over 100 time points for 50 batches. These batches were synchronised based on the rate of conversion. A more detailed description of the process can be found in A.K. Smilde & H.A.L. Kiers [39]. Two disturbed batches are used in this study: the first batch has a too high H₂ flow as well as a too high rising pressure, the second batch suffers from a deactivated catalyst. Besides, the product quality of the first batch was of low quality.

Data set 4
Data set 4 consists of a simulated emulsion co-polymerisation of styrene-butadiene. This process is well described in the paper by P. Nomikos & J.F. MacGregor [6]. The data set consists of 50 NOC batches for which 9 process variables are measured during 200 time intervals. The following process variables are measured: two flow-rates, four temperatures, density, conversion and rate of energy release. Two disturbed batches are available to test the detection power of the methods. The first (batch 106) has an initial organic impurity contamination in the butadiene feed. The second (batch 99) has a similar, but larger, contamination halfway the batch run.

Data set 5
Data set 5 consists of an industrial (DuPont) polymerisation in an autoclave with 47 NOC batch runs. This data set is also referred to by K.A. Kosanovich et al. [7]. The polymer is produced through five stages during these stages the process is monitored by 8 process variables (six temperatures and two pressures) during 116 time intervals. Prior to model building, the raw data is linearly interpolated to align the NOC
batches, this is described in R. Boqué & A.K. Smilde [40]. Three disturbed batches are available to calculate the AST. A univariate analysis of batch 41 shows clear disturbances for two pressure and four temperature measurements halfway the batch run. This is probably due to an upset in the first stage of the process. For batch 46 and batch 47, univariate analysis shows disturbances at the end of the batch run for one pressure and five temperature measurements.

**Data set 6**

Data set 6 is an industrial (Du Pont) two-stage polymerisation process with 36 selected NOC batches for which 10 process variables are measured during 100 time intervals. One disturbed batch is available to calculate the performance indices, which was known to have a poor end-product quality. This data set is described in P. Nomikos & J.F. MacGregor [19].

For all data sets except the sixth one, three principal components are chosen to describe the data in $\mathbf{X}^{p \times k}$. For data set six, two components were sufficient. The data is cleaned from outliers, and thus represents NOC batches.

**6.2.8 Results**

All the issues and problems discussed in the previous section will be discussed and illustrated using the case study of the several batch processes. The problems and issues, however, hold in general for all batch process monitoring problems.

**6.2.8.1 Performance indices**

The next two indices are used to test the performance of the proposed methods: the overall type I error and the action signal time (AST). The definition of these indices will be discussed in this section.

**6.2.8.1.1 The overall type I error**

On-line monitoring of batches can be regarded as a sequential hypothesis-testing problem. A batch is tested at every time interval $k$ whether it is statistically in-control by comparing the test statistics $SPE$ and $D$-statistic to their reference distribution. There are two hypotheses of importance, that is, the null hypothesis $H_0$ (the batch is in-control) and the alternative hypothesis $H_1$ (the batch is out of control). The
rejection of $H_0$ while it is true is called a type I error. The size of this error, or the probability of a type I error, is given by:

$$\alpha_{\text{imposed}} = P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ is true})$$  \hspace{1cm} (113)

In this study $\alpha_{\text{imposed}}$ is set to 5%. It is useful to check the performance of the control charts by comparing the actual $\alpha$ to $\alpha_{\text{imposed}}$. This is done by sequentially monitoring all NOC batches. Since an NOC batch is by definition statistically in-control, a signal in one of the charts can be regarded as a type I error. The performance of a chart is good if $\alpha_{\text{actual}}$ is close to $\alpha_{\text{imposed}}$.

### 6.2.8.1.2 Action Signal Time

The action signal time (AST) is used to study the efficiency of a control chart to detect faults. A more commonly used criterion in SPC for investigating the type II error is the average run length (ARL) curves. However, theoretical derivations of ARL statistics are difficult if not impossible in the case of batch process monitoring. An alternative might be an extensive and rigorous simulation study but that is not the.

The action signal time is defined as the time between the introduction of an error and the out-of-control signal. Three consecutive points outside the control limit defines the out-of-control signal. This definition of the out-of-control signal is common practice since process operators will unlikely undertake action after just one crossing of the limit.

### 6.2.8.2 Accounting for the difference between NOC and new batches

In on-line batch monitoring, the scores and residuals of new batches are compared with their reference distributions for NOC batches. The NOC model is calculated in such a way as to minimise the residual variance of the NOC batches. This is reflected in the NOC scores and residuals. As the NOC model is based upon a limited number of batches, it is likely that the projection of independent new batches from the same population will give higher residuals (and lower scores) than the NOC batches, because the new batches did not take part in the fitting of the model to the data.

In order to make the NOC scores and residuals more comparable to those of new batches, a leave-batch-out procedure is used, where each NOC batch is treated as a new independent batch. This approach was suggested by D.J. Louwerse & A.K.
Smilde [38]. A model is built on the remaining NOC batches, and the NOC batch left out from the model is projected to calculate the scores and residuals. This is repeated for all NOC batches, and all of the calculated scores and residuals form the new NOC distribution of scores and residuals. This procedure is explained in more detail in Appendix E. Clearly, these scores are not calculated in order to minimise the residuals variance. Therefore, this distribution of scores and residuals is better comparable to those found by projection of a new independent batch.

6.2.8.3 Hypothesis testing

A crucial point in calculating \( \alpha_{\text{actual}} \) is the definition of the null hypothesis. The null hypothesis used by Nomikos and MacGregor is specified for each separate time point \( k \): a batch is in-control at a certain time point \( k \) if it does not exceed the limit for that time point. Such a hypothesis is then tested sequentially for all \( K \) time points during the batch run (P. Nomikos & J.F. MacGregor [19]). Every rejection of the null hypothesis for an NOC batch is a false alarm. When all \( I \) NOC batches are subjected to monitoring then \( \alpha_{\text{actual}} \) can be calculated by:

\[
\alpha_{\text{actual}} = P(\text{overall type I error}) = \frac{\sum \text{false alarm}}{IK}
\]

(114)

E.N.M. Van Sprang et al. [41] used the same method to calculate \( \alpha_{\text{actual}} \). They found an \( \alpha_{\text{actual}} \) which is slightly too high for the SPE-chart and too low for the D-chart for various datasets.

A more realistic null hypothesis is: a batch is in-control if it is in-control during the entire batch run. As soon as an NOC batch breaks the limits of the control chart somewhere along the time trajectory, the null hypothesis that the batch is in-control is rejected and a false warning has occurred. After breaking the limits no further testing is required. This \( H_0 \) is more realistic than the one used by P. Nomikos [21] because it tests whether a whole batch is in-control and not only at a certain point. Moreover, the \( H_0 \) used by Nomikos and MacGregor will result in 5% false alarms in the whole batch trajectory of each new batch (by construction). Stated otherwise, the \( H_0 \) of P. Nomikos [21] will result in too tight control limits, which will be a problem in practice, which has indeed already been observed (P. Nomikos [21]).
The \( \alpha_{\text{actual}} \) for the new null hypothesis can be calculated as follows. Let the indicator variable \( \sigma_i \) be 1 if this event is true, and zero if the limits are not broken somewhere along the time trajectory. Then, the calculation of \( \alpha_{\text{actual}} \) is now given by:

\[
\alpha_{\text{actual}} = P(\text{overall type I error}) = \frac{\sum_{i=1}^{I} \sigma_i}{I}
\]

(115

The new null hypothesis is a composite hypothesis and testing this hypothesis comes down to testing sequentially individual hypotheses: one hypothesis for each point in time. The difference between this new way of testing and the previous approach is that the latter does not consider the overall (composite) hypothesis, which causes the problems as indicated above.

6.2.8.4 The Bonferroni adjustment

Because of the sequential testing problem in both the \( D \) and \( SPE \)-chart, the probability of improperly rejecting the null hypothesis is given by:

\[
P(\text{Overall type I error}) = 1-(1-\alpha)^{K}
\]

(116

where \( \alpha = \alpha_{\text{imposed}} \). For example, if \( \alpha = 0.05 \) and if \( K \) is equal to 58, then the probability that an overall type I error occurs is equal to 0.95. Clearly, the magnitude of 0.95 \( >> \) \( \alpha \). In other words, the probability of rejecting a batch while it produced a good product is 95%. This is a problematic situation for practical purposes. Setting the limits according to the Bonferroni correction (G.W. Snedecor & G.C. Cochran [68]) solves this problem. To assure the probability of an overall type I error reflects the value of \( \alpha \), \( \alpha_{\text{imposed}} \) is calculated according to:

\[
\alpha_{\text{imposed}} = 1-(1-\alpha)^{1/K} = \frac{\alpha}{K}
\]

(117

For the previously example, where \( K = 58 \), the overall type I error is now approximately equal to 0.05. Clearly this is more in agreement with the expectations of choosing the confidence limit of 95%.

-235-
In chemical engineering practice it is common to set $\alpha_{\text{impure}}$ to 99.9\% to reduce the number of false alarms (D.J. Louwerse & A.K. Smilde [38]). Applying Bonferroni and the leave-one-out approach motivates this practical usage from a theoretical background.

The Bonferroni procedure ignores important information about the data, that is the correlation structure of the test statistics. This will result in too conservative limits leading to decreased detection speed of erroneous batches. Therefore the critical level of $\alpha_{\text{impure}}$ can be adjusted according to the Dubey and Armitage-Parmar (D/AP) procedure (A.J. Sankoh et al. [117]). They suggest to use the following adjusted value for $\alpha_{\text{impure}}$:

$$\alpha_{\text{impure}} = 1 - (1 - \alpha)^{1/m}$$

where $m = k^{1-\bar{r}_k}$ \hspace{1cm} (118)

Here, $\bar{r}_k$ is the mean correlation coefficient between the K-test statistics and is calculated as follows. Each NOC batch is monitored in a leave-one-out procedure, and the test statistics $D$ and $SPE$ are stored in the $i$-th row of the matrices $D$ ($I\times K$) and $SPE$ ($I\times K$). Then for both matrices the correlation matrix is calculated. Now, $\bar{r}_k$ is the mean value for the off-diagonal elements and can be calculated for both the $SPE$ and $D$-chart. If the correlation is zero, the D/AP procedure is equal to the Bonferroni adjustment. The average value over all datasets of $\bar{r}_k$ for the $SPE$ fluctuates around 0.26 and for the $D$-statistic around 0.71. These numbers are reasonable since the residuals describe non-systematic variation and therefore behave erratic. At the other hand the $D$-statistic captures the structured variation in the data and will therefore be correlated.

**6.2.8.5 Further comments on the D-statistic**

The following general improvements to the $D$ and $SPE$-chart were suggested in the foregoing: the leave-one-out approach, a new definition of $H_0$, the Bonferroni correction and the D/AP procedure. Although the $SPE$-chart is clearly improved, these suggestions will not improve the performance of the $D$-statistic since the performance of this statistic is severely hampered due to its construction. The problem of the $D$-statistic resulting from filling-in procedures, however, can be solved partly by changing the way to calculate the $D$-statistic. This is described in Section
6.2.8.5.1. The rest of this Section gives some auxiliary comments with respect to the D-statistic.

**6.2.8.5.1 Time dependent covariance matrix**

As can be seen from equation 110, the calculation of the D-statistic is based upon the covariance matrix \( S \) (\( R \times R \)) and \( S \) is assumed to be constant over time. This covariance matrix is calculated according to:

\[
S = \frac{1}{I-1} (T_k)'T_k 
\]  

(119)

P. Nomikos & J.F. MacGregor [7] already mentioned in their paper that it is possible to calculate a covariance matrix for each point in time. This would result in:

\[
S_k = \frac{1}{I-1} (T_k)'T_k 
\]  

(120)

The impact of using a time varying covariance matrix is shown Figure 101.

Here, \( S_k \) is calculated for the first dataset. The dotted lines in Figure 101 represent the \( R \) diagonal elements (variances) of \( S_k \) where the solid line represents the squared sum.
of the off-diagonal elements (covariance) $S_k$. This figure shows that the covariance matrix cannot be assumed to be constant over time. Moreover, it can be seen from Equation 120 that the scores $T_k$ are used to calculate $S_k$. These scores $T_k$ are affected by the filling-in procedure. Using $S_k$ compensates to some extent for the problems of using the filling-in procedure. Therefore, it is strongly recommended to use a time dependent $S_k$.

The fault detection is clearly improved as can be seen from an example in Figure 102. Here, the $D$-statistic using $S_k$, marked by the star, clearly shows a more dynamic behaviour compared to the situation where $S$ is used. Moreover, the faulty batch goes unnoticed in the $D$-chart when $S$ is used!

![Figure 102](image)

Figure 102
$D$-chart using time independent and time varying covariance matrix.

### 6.2.8.5.2 Non-normality

The scores used to calculate the $D$-statistic are assumed to be normally distributed. P. Nomikos & J.F. MacGregor [19] use the central limit theorem to justify the assumption of normality. N.D. Tracy et al. [14] stated in their paper that the results for deriving control limits depend on the validity of this assumption. That is, the rejection of the null hypothesis using a F-test is very sensitive to small departures
from normality. The effect on the limits increases when the distributions are either heavy-tailed or light-tailed (G.M. Miller [118]).

In this study, normality plots of the scores $T_k$ indicated that the scores did not depart severely from normality. It can be concluded that the normality for all data sets is acceptable, and therefore the F-test is allowed.

6.2.8.5.3 Violation of the $\Sigma_f = \Sigma$ assumption

If the variation of the process variables in the application phase can be described by the covariance matrix $\Sigma_f$, then it is be assumed that $\Sigma_f = \Sigma$. An estimate of the true population value $\Sigma$ is given by S. S.J. Wierda [119] investigated the effect of the violation of this assumption, that is, $\Sigma_f \neq \Sigma$. Wierda concluded the following: the true probability that the $D$-chart produces a false alarm can be considerably smaller or larger than $\alpha$ if the above stated assumption is violated.

Reasons for violation of this assumption are twofold. First, the estimate $S$ or $S_k$ is inaccurate because of the limited number of batches. Secondly, a new monitored batch will always differ from the NOC batches and therefore the assumption is violated to a certain extent by definition. S.J. Wierda [119] suggest to use a control chart that tests the hypothesis that the covariance remains stable over time. In fact, this is exactly what the $SPE$-statistic is testing. In other words, as soon as the $SPE$-chart detects a breakage of the correlation pattern, the $D$-statistic becomes unreliable. The argumentation does not only apply to on-line monitoring, but also for post-batch analysis.

6.2.8.5.4 Degrees of freedom in process monitoring

In the current methodology for deriving control limits for the $D$-statistic (for on-line monitoring of batch/continuous processes as well as for post-batch analysis), the degrees of freedom in the F-distribution considers the number of components ($R$) and the number of batches ($l$). This is a direct generalization of the degrees of freedom for Hotellings $T^2$ statistic derived from normally distributed variables. The $D$-statistic, however, uses $t$-scores, which are constructed variables based on the original measured variables in the dataset. It is unclear whether the generalization of the degrees of freedom for Hotellings $T^2$ to the $D$-statistic is valid. The $t$-scores can also be derived from other models of the three-way dataset $X$, e.g. using PARAFAC or Tucker models (H.G. Law et al. [120], R. Coppi & S. Bolasco [121]). The $t$-scores of,
e.g., a PARAFAC model might have completely different statistical properties as the $t$-scores from the unfold PCA model. These differences are not reflected in the direct generalization of degrees of freedom going from $T^2$ to $D$. This is still an open question. Note that for the $SPE$-statistics things are different. The residuals (as a result from either an unfold PCA or PARAFAC model) are used to calculate approximate degrees of freedom for the $\chi^2$-distribution and the problem as mentioned above does not exist. The observation is that indeed the degrees of freedom for the $SPE$ of an unfold PCA or PARAFAC model are different.

6.2.8.6 Further comments on the squared prediction error ($SPE$)

The second test statistic, which will be discussed in further detail, is the squared prediction error. E.N.M. Van Sprang et al. [41] already showed that the performance for this test statistic is generally good. The overall type I error is satisfying ($\alpha_{\text{upred}} = \alpha_{\text{actual}}$) and the detection power is good as will be shown in the following. Therefore, the performance of this chart is not alarmingly poor.

6.2.8.6.1 Non-normality

Assuming normality of the residuals, the $SPE$-statistic is distributed according to a weighted sum of $\chi^2$ distributions. This weighted sum can be approximated by a $\chi^2$ distribution $g\chi^2(b)$ where $g$ is a size factor and $b$ the degrees of freedom (G.E.P. Box [23]). Inspection of normality plots for the residuals showed a deviation from normality for all datasets. Therefore, the residuals are not normally distributed and the assumption of normality is violated.

The $g$ and $b$ parameters are obtained by fitting the $g\chi^2(b)$ distribution to the reference distribution of the $SPE$ from the NOC data. Experience has shown that, although the residuals are not exactly distributed according to a normal distribution, the $g\chi^2(b)$ distribution approximates the behaviour of the $SPE$-statistic quite well. Hence, it appears that in practice small deviations from normality are not troublesome.
### 6.2.8.7 Results adjustments to the control charts

The results of the previously discussed adjustments to the control charts are presented in Table 13.

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>Bonferroni</th>
<th>Bonferroni + Leave one out</th>
<th>Bonferroni + Leave one out + $S_k^*$</th>
<th>D/AP + Leave one out</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE</td>
<td>OT1</td>
<td>73.87</td>
<td>9.95</td>
<td>5.79</td>
<td>9.05</td>
</tr>
<tr>
<td>SPE</td>
<td>OT1 N&amp;M</td>
<td>7.85</td>
<td>-</td>
<td>5.23</td>
<td>-</td>
</tr>
<tr>
<td>SPE</td>
<td>AST</td>
<td>18</td>
<td>38</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>SPE</td>
<td>D</td>
<td>19.46</td>
<td>3.12</td>
<td>11.58</td>
<td>10.59</td>
</tr>
<tr>
<td>SPE</td>
<td>OT1 N&amp;M</td>
<td>2.68</td>
<td>-</td>
<td>3.85</td>
<td>-</td>
</tr>
<tr>
<td>SPE</td>
<td>AST</td>
<td>56</td>
<td>91</td>
<td>72</td>
<td>76</td>
</tr>
</tbody>
</table>

*Overall type I (OT1) error and action signal time (as the percentage of the batch length).*

The overall type I error is calculated using the definition of the null hypothesis by Nomikos and MacGregor (OT1-N&M), as well as the definition given in this study (OT1). Also, the overall type I error is calculated as the average for all six data sets. The action signal time (AST) for the faulty batches have been calculated as a percentage of the batch length. No detection of the fault results in an action signal time of 100%. The average AST for the 13 faulty batches is presented in Table 13.

The first column of Table 13 represents the results found for the standard approach for batch monitoring, that is the current deviations approach as described by Nomikos and MacGregor in their paper. The second column represents the results of the standard approach extended with the Bonferroni adjustment, the third column the standard approach extended with Bonferroni plus the leave-one-out method, the fourth column shows the results for the standard approach extended with Bonferroni, leave-one-out and $S_k^*$ and the fifth column describes the results for the standard approach extended with leave-one-out, $S_k^*$ and D/AP. It should be noted that the last extension, that is $S_k^*$, has no effect on the SPE-chart. The residuals are already...
calculated using the time varying scores as can be seen from Equation 112. Also, the Bonferroni and D/AP correction do not apply to OT1-N&

*Overall type I error*

It can be seen that the OT1-N&

M is slightly too high for SPE and too low for D. These results are in agreement with E.N.M. Van Sprang et al. [41]. The OT1 error in combination with the standard approach is far too high for both charts. This is expected considering the sequential nature of hypothesis testing: \( K \) hypothesis tests are performed each having a probability of 5% rejecting \( H_0 \). This results in high probability of a false alarm. Adjusting the limits according to the Bonferroni theory obviously improves the results for both the SPE and D-chart.

The results for the standard approach extended with the Bonferroni limits and leave-one-out approach improve the OT1 for the SPE even further. Using a time varying covariance matrix (column four in Table 13) to calculate the D-statistic improves the OT1 error to an acceptable number. Therefore, it may be concluded that the standard approach extended with Bonferroni limits, leave-one-out approach and a time varying covariance matrix gives satisfying results in terms of the OT1.

If the limits are based on the D/AP procedure, the leave-one-out approach and a time varying covariance matrix, the OT1 error increases. This is a disappointing result since the D/AP procedure is intuitively an appealing method. A reason might be that the OT1 error is calculated based on a total of 275 batches, which might be too limited, a number to obtain a reliable estimate of the OT1 error.

Also for the OT1-N&

M method the leave-one-out approach is helpful. The average value found for the actual \( \alpha \) is in good agreement with the expected value for \( \alpha \). Note that in this case, the OT1 error is calculated over approximately 23000 samples (sum over batches for all datasets \( \times \) sum over all time points for all datasets). Hence, these results are more reliable than the ones obtained for the D/AP procedure above.

*Action Signal Time*

It may be expected that the action signal time is related to the overall type I error. This is also what is concluded from the results given in Table 13. For example, the increment of the OT1 after applying the Bonferroni adjustment increases the average AST for both charts. In fact, three wrong batches were unnoticed after applying the Bonferroni adjustment. Also, it may be concluded that the average AST for the SPE-
Chapter 6 ♦ Statistics and mechanistics

chart is better than the average AST of the $D$-chart. This is in agreement with the theory described in Section 6.2.4, as well as the observations from several researchers:

E.B. Martin et al. [27]: "An abnormal batch will exhibit significant deviations in an SPE plot".

B.M. Wise et al. [62]: "It is our experience that most process faults show up in $Q$. Very few faults are detected by $T^2$ alone".

The AST for the $SPE$-chart using the D/AP procedure is hardly improved compared to the Bonferroni adjustment. This is caused by the weak correlation of the sequential $SPE$ test-statistic. In this case, the factor $m = K$ and there is hardly any difference between the Bonferroni limits and D/AP limits.

Sometimes an action signal in the $SPE$-chart is followed by an action signal in the $D$-chart. This is mainly due to the embedded error and therefore has no diagnostic value.

6.2.9 CONCLUSIONS

In this study the performance of control charts used for on-line monitoring of batch processes have been assessed and suggested improvements have been evaluated using six data sets. The mechanistics of the control charts are discussed. It seems that most faults break the correlation structure. This is due to the mechanistic properties of the $D$ and $SPE$-chart. The current deviations approach of Nomikos and MacGregor has been used for on-line batch process monitoring because it was found to work reasonable (also in earlier work), but the conclusions have a more general applicability. The main conclusions are the following:

- Most process faults show up in the $SPE$-chart and hardly any in the $D$-chart. This is mainly due to the construction of the model and mechanistic properties of the control charts where by definition all faults break the correlation structure and therefore are detected in the $SPE$. From an engineering point of view, all faults show up in the $SPE$, not only new events such as sensor failures. This hampers the diagnostic capability of the $SPE$-statistic.
The $D$-chart is hardly of practical use in detecting process faults. This is due to the construction of this test statistic. A small improvement of the $D$-statistic is obtained by using a time varying covariance matrix of the scores.

A new null-hypothesis has been formulated which improves the performance of the control charts in terms of false alarms. This new null-hypothesis necessitates the use of Bonferroni or similar types of corrections. A consequence of this adjustment is that the Action Signal Time (the time it takes to signal a fault) increases.

A leave-one-out approach is suggested to use in calculating the control limits. This works very well in practice. Moreover, on-line aap monitoring of all available batches using this leave-one-out approach is well suited for selecting NOC batches.
6.3 The effect of the size of the training set and number of components on the false alarm rate in statistical process monitoring

6.3.1 SUMMARY
This paper describes the sensitivity of false alarm rate to misspecification of the number of PCA components in multivariate statistical process control (MSPC) models. Using simulations, it is shown that choosing an incorrect number of components in monitoring models may seriously affect the false alarm rates of the control charts. Furthermore, the false alarm rate becomes worrisome when the size of the training set is small. Using a leave-one-out procedure for building the control charts partly solves this problem.

6.3.2 INTRODUCTION
One way of monitoring continuous or batch processes is to use multivariate statistical process control (MSPC) models (P. Nomikos & J.F. MacGregor [6], J.V. Kresta et al. [122]). To establish this monitoring tool, an empirical model is built on the data. From these models, control charts are derived: the $SPE$ and $D$ charts. The performance of these charts is reflected in probability of the type I and type II errors. In this paper the focus will be on the probability of the type I error or false alarm rate.

MSPC models are often based on principal component analysis. In PCA a number of components must be chosen and this number can therefore be considered as a metaparameter. The number of components is important with respect to the type I and type II errors (H.W. Tong & C.M. Crowe [87], H.H. Yue & S.J. Qin [110]). The choice of the number of components is often supported by cross validation. However, the goal of cross validation methods is to find an optimal model in terms of the model fit. This is not the objective of e.g. statistical batch process monitoring.

---

*HJR in: H.J. Ramaker, E.N.M. van Sprang, J.A. Westerhuis and A.K. Smilde, The effect of the size of the training set and number of principal components on the false alarm rate in statistical process monitoring, Chemometrics and intelligent laboratory systems, accepted.*
where the objective is to minimize the probability of the type II error for a given probability of the type I error. To select the number of components according to this objective, large datasets and numerous faulty batches are required. In practice, ad-hoc methods are used to select the number of components supported by process knowledge.

A. Nijhuis et al. [123] showed the robustness of MSPC to the number of significant PC's in the sense that out-of-control situations are correctly flagged going from one principal component to three components. For all choices of components, the out-of-control situations were flagged in the control charts. The robustness against the false alarm rate was not determined. However, the speed of detection of real out-of-control situations and the false alarm rate are closely related. That is, an increase in the speed of detection might be coupled to an increased false alarm rate. It is therefore worthwhile to investigate the possible causes for a high false alarm rate. In the paper by H.W. Tong & C.M. Crowe [87] it is mentioned that the false alarm rate can become unacceptably high even for moderate size problems. Furthermore, it is an often heard remark that multivariate charts such as the $SPE$ chart suffer from too many false alarms. A common way to overcome this problem is to set the limits of the control charts rather high e.g. 99.9% (P. Norvikos [45]). However, this reduces the capability of the control chart to detect wrong batches.

Therefore, the question is how sensitive the probability of the type I error is to misspecification of selecting the number of components. A simulation study is performed to show two things i) the sensitivity of the $D$ and $SPE$ charts to the number of components in terms of the false alarm rate ii) improvement of the false alarm rate using a leave-one-out procedure.

It will be shown that problems related to over- or underfitting can be substantially reduced for the $SPE$ chart by using a improved leave-one-out procedure. However, this leave-one-out procedure is not helpful for the $D$ chart.

6.3.3 SIMULATION OF A CONTINUOUS PROCESS
A continuous process is simulated with some special features. Fifteen process variables ($J = 15$) were measured each minute for a period of $K$ time points. The process has been designed in such a way that there are only three underlying phenomena. The data set has been constructed as follows:
\[ X = z_1 y_1 + z_2 y_2 + z_3 y_3 + S \]

where the vectors \( y_1 \), \( y_2 \) and \( y_3 \) form the columns of an arbitrary orthogonal matrix \( Y (15 \times 3) \). The vectors \( z_1 \), \( z_2 \) and \( z_3 \) have been constructed using a moving average model. The element of e.g. vector \( z_1 \) at time interval \( k \) is constructed as follows:

\[ z_1(k) = a_1(k) - \theta \cdot a_1(k-1) \]

where \( a_1 \) is a random shock with a standard deviation of 0.7. The \( \theta \) 's for \( z_1 \), \( z_2 \) and \( z_3 \) are 0.8, 0.5 and -0.2 respectively. These numbers for the random shock and \( \theta \) were chosen in order to reflect different types of dynamics in the data. The matrix \( S \) represents white noise with mean zero and standard deviation of 0.36. \( S \) comprises approximately 40% of the total variation in \( X \). An amount of 60% explained variance is commonly encountered in MSPC models (E.N.M. Van Sprang et al. [41]).

### 6.3.4 Multivariate Statistical Process Control

Once the data \( X \) has been obtained, MSPC can be applied to the data (B.M. Wise & N.L. Ricker [112], J.V. Kresta et al. [122]). The approach used in this paper is based on MSPC and will be explained in the following. Two different datasets are simulated. The first dataset \( X_{\text{training}} \) represents normal operating conditions (NOC) and is referred to as the training set. The NOC training set is statistically in control by definition and is used for model building. The other dataset \( X_{\text{test}} \) represents the test set and will be used for monitoring. This test set is independent from the NOC data but also represents NOC conditions.

First, a PCA model is built on the NOC data. A PCA decomposes the matrix \( X_{\text{training}} \) into the sum of \( R \) outer products of scores \( t \) and loadings \( p \) plus a residual part \( E \):

\[ X_{\text{training}} = \sum_{r=1}^{R} t_r p_r^T + E = TP^T + E \]
where $T(K \times R)$ is the score matrix, $P(J \times R)$ contains the loadings and $E(K \times J)$ is the residual matrix. Here, $K$ denotes the number of samples, $J$ the number of process variables and $R$ the number of principal components.

Secondly, the monitoring of a sample is performed. The scores $t_{new}$ and residuals $e_{new}$ for a new scaled measurement $x_{new}(J \times 1)$ from the test set $X^{test}$ are found by projection on the model plane. From these scores and residuals, the $D$ and $SPE$ statistics are calculated:

$$t_{new} = (P'P)^{-1} P' x_{new} = P' x_{new} \quad \Rightarrow \quad D = t_{new}' A^{-1} t_{new} \quad (124)$$
$$e_{new} = x_{new} - Pt_{new} \quad \Rightarrow \quad SPE = e_{new}' e_{new} \quad (125)$$

Here, $A$ is the variance-covariance matrix of the scores $T$. The $D$ statistic follows an $F$-distribution and the control limits (CL) for the control chart can be calculated according to (N.D. Tracy et al. [14]):

$$CL_{new} = \frac{R(K^2 - 1)}{R(K - R)} F(\alpha; R; K - R) \quad (126)$$

The $SPE$ statistic follows a weighted chi-square distribution ($\sim gX_h^2$). The parameters $g$ and $b$ are estimated according to J.F. Jackson & G.S. Mudholkar [24]. From this chi-square distribution, the CL for the $SPE$ chart can be calculated.

The probability of the type I error is calculated for the test set. The type I error describes the event that a NOC sample is flagged out-of-control while it is in control. The limits for $D$ and $SPE$ statistics are derived assuming in-control data. In the ideal case, the probability of the type I error reflects the choice of the significance level $\alpha$. In this work, the significance level $\alpha$ of the $SPE$ and $D$ charts is set to 0.05. In other words, there is a probability of 5% that the type I error occurs. The false alarm rate for the test set is calculated as follows. Every sample is projected to calculate the $SPE$ and $D$ statistic. Since in this simulation the test set is in control by definition, a crossing of the control limits in the control charts is regarded as a false alarm. The number of false alarms is counted and divided by the number of observations of the test set. This number is the false alarm rate.
In order to obtain reliable results, the construction of the training and test set is repeated hundred times. As a result, the average false alarm rate over the hundred test sets is calculated.

6.3.4.1 Size of the NOC data
It is expected that the statistical properties of the control charts are directly related to the number of samples in the training set. That is, the estimation of the reference distribution parameters will be increasingly accurate if the number of samples increases. The number of samples in the training set can be regarded as another meta parameter. To study this behavior, the training set is constructed for different lengths of the time $K$. The difference in length is as follows: 20, 30, 50, 75, 100, 150, 250, 500, 1000 and 2000 time points. That is, a training set is constructed for e.g. 20 time points. This is repeated 100 times and the average false alarm rate is calculated. Then, the next training set with a length of 50 time points is constructed, etc. The length of the test set is fixed and contains 250 time points.

6.3.4.2 Number of principal components
Due to the construction of the dataset, only three underlying phenomena are present in the data. If the number of principal components is less than three, the model underfits the data. The opposite is true if the number of principal components is higher than three: the model overfits the data. In order to study the effect of over and underfitting on the false alarm rate, every model is calculated for 1 to 14 principal components. The fifteenth component is omitted since there are no residuals in $SPE$ when fifteen components are selected.

6.3.5 False alarm rate results
The results for the false alarm rate have been presented in Figure 103. In this Figure, only the datasets with a number of 20, 50, 100, 500 and 2000 observations are presented. The resulting curves of the other datasets (30, 75, 250, 1000) are in-between the curves as shown in Figure 103:
The following conclusions can be drawn for the false alarm rate for the SPE chart (Figure 103a):

- The false alarm rate is too high for the smaller datasets ($K = 20 - 100$)
- If the data is overfitted, the false alarm rate is too high. This is especially true for the smaller datasets ($K = 20 - 100$). The residuals that result from the projection of a new independent sample will on average always be larger than the residuals of the NOC model (H.-J. Ramaker et al. [71]). This is true because the new sample did not take part in fitting the NOC samples. The difference in size between the residuals of the new sample and NOC samples increases when the data is overfitted. The chi-square distribution does not capture this as reflected in the false alarm rate. Overfitting the large datasets ($K = 500 - 2000$) will not further increase the false alarm rate.
- For very large datasets ($K = 500 - 2000$), the false alarm rate will be slightly higher than $\alpha (0.05)$ in the case of overfitting the data, although not disturbingly
Chapter 6 • Statistics and mechanistics

- Underfitting is not a problem for reasonably sized datasets \((K > 50)\)
- The false alarm rate is in perfect accordance with \(\alpha (0.05)\) if the sample size is large enough \((K > 100)\) and the model is chosen correctly \((R = 3)\)

The false alarm rate calculated for the \(D\) chart leads to the following conclusion (see Figure 103b):

- Underfitting the data is not really a problem. In case of overfitting, the false alarm rate will always be too small. This problem is less pronounced for large datasets \((K > 500)\). While the residuals are on average larger for a new sample compared to the test set, the scores are on average smaller. This effect is reflected in the false alarm rate that is too small.
- The false alarm rate is in perfect accordance with \(\alpha (0.05)\) if the sample size is large enough \((K > 100)\) and the model is chosen correctly \((R = 3)\)

6.3.5.1 Discussion about the false alarm rate results

The \(D\) statistic as defined by Nomikos and MacGregor is different from the Hotelling \(T^2\) statistic. The difference is that the \(D\) statistic is based on scores obtained from a latent variable model while the \(T^2\) statistic is based on the originally measured process variables. For manifest variables the distributions of calibration samples and test samples are the same. This is not true for the \(D\) statistic where the scores and residuals obtained from a component model are different for test samples and calibration samples. A direct generalization of Hotelling's \(T^2\) statistic to component models is not valid.

The Hotelling \(T^2\) statistic uses the correlation between a set of manifest variables and these are stored in the covariance matrix \(B\). The \(T^2\) statistic is defined as follows, where \(x\) is a measurement vector \((J \times 1)\) of \(J\) manifest variables:

\[
T^2 = (x - \bar{x}) B^{-1} (x - \bar{x})
\]

(127)

Tracy et al (N.D. Tracy et al. [14]) derived exact distributions for \(T^2\) statistic values for calibration samples used to calculate the mean and covariance matrix and for new samples. The \(T^2\) values of the calibration samples follow a Beta-distribution while the \(T^2\) of the new samples follow an \(F\)-distribution. However, an important assumption is
that the calibration samples and the new samples come from the same multivariate normal distribution. This assumption is violated when a latent variable model is used.

If a latent variable model (e.g. PCA) is developed from the calibration data, the total variance of all $J$ process variables is distributed in a specific way over the scores. As an example, hundred PCA models are developed for hundred different sets of manifest variables (20 samples × 15 variables). The average variance of all scores (white bars) is plotted in Figure 104:

Hundred test sets (250 samples × 15 variables) obtained from the same distribution as the calibration sets are projected to the PCA models. The average variance of the test scores is plotted in the same figure (black bars). The first score value of the calibration set is much larger than the first score of the test set. Each sample in the calibration set tries to draw the PCA model and specifically the first principal component towards itself. Furthermore, the variance of the latter scores are low and go to zero. This is not the case for the test set scores. In this example, three scores were sufficient to describe the data. It is easy to imagine that the covariance matrix from the calibration scores does not represent the covariance of the test scores. Stated otherwise, the
expected value of the estimated covariance matrix of the calibrations scores ($\Sigma_{\text{train}}$) differs from its counterpart in the testset ($\Sigma_{\text{test}}$) where in both cases the expectations are over repeats of the calibration and test set. Wierda S.J. Wierda [119] already showed that the performance of the $T^2$ chart is very sensitive to violating the assumption that $\Sigma_{\text{true}} = \Sigma_{\text{test}}$.

In the same way as for the scores, the residuals of the calibration samples will be different from the test set samples. The residuals for the test set are expected to be higher because the test set did not take part in the fitting of the model in the data. For the same example set, the variance of the residuals is calculated and plotted in Figure 105:

![Figure 105: Variance of the residuals](image)

Clearly, the variance of the projected residuals (black bars) is higher compared to the variance of the model (white bars). This can be understood in the same way as explained above.
6.3.6.5 LEAVE-ONE-OUT METHOD

In order to overcome the difference in distribution between the calibration set and the test set, an additional test set should be used. This additional test set is projected on the calibration model and scores and residuals are obtained. The scores and residuals from this additional test set represent scores and residuals from new samples and therefore should be used to calculate the covariance matrix (for the $D$ statistic) and the control limits for $SPE$ statistic.

Unfortunately problems arise when the number of calibration samples is only small. In that case it is not possible to split the calibration samples over the calibration set and the additional test set. However, to simulate the additional test set, a leave one out projection procedure (D.J. Louwerse & A.K. Smilde [38]) is applied.

The leave-one-out method works as follows (H.-J. Ramaker et al. [25]). From the training set $X^{training}$ the first sample $k = 1$ is removed. A model $P_{k=1}$ is built for the remaining $K-1$ samples. The left out sample is projected onto the model $P_{k=1}$ to find the scores $t_{k=1}$ and residuals $e_{k=1}$. Then, the next sample ($k = 2$) is left from $X^{training}$ and in the same manner, the scores $t_{k=2}$ and residuals $e_{k=2}$ are calculated for the left out sample. This is repeated for all samples. The score matrix $\bar{T} = [t_{k=1} \ t_{k=2} \ldots \ t_{k=K}]$ and residual matrix $\bar{E} = [e_{k=1} \ e_{k=2} \ldots \ e_{k=K}]$ form the reference distributions for the $SPE$ and $D$ charts and are called the leave-one-out scores and residuals. The leave-one-out scores and residuals are better comparable to the scores and residuals of a completely new and independent sample since they are calculated in a similar manner. This is illustrated in Figure 104 and Figure 105 where the variance for the scores $\bar{T}$ and residuals $\bar{E}$ are presented (gray bars) using the same calibration and test set as discussed in Section 4.1. As a result of the leave-one-out procedure, the variance of the leave-one-out scores and residuals are better comparable to the monitored score and residuals.

6.3.6.1 Results for the leave-one-out method

The results found for the false alarm rate where the reference distributions from the control charts are based on the leave-one-out method are presented in Figure 106:
The following conclusion from the SPE chart based on leave-one-out residuals can be drawn (see Figure 106a):

- When overfitting the data, a far better false alarm rate is obtained as compared to the case with fitted residuals (see Figure 103a). Therefore the performance of the control chart based on the leave-one-out residuals is more robust against overfitting.
- The false alarm rate is slightly higher than $\alpha (0.05)$ for the smaller datasets ($K = 20 - 50$) in the case of underfitting. However, if a reasonable model is chosen ($2 < R < 10$), the false alarm rate is satisfying.
- The false alarm rate is in perfect accordance with $\alpha (0.05)$ for all sample sizes if the model is chosen correctly ($R = 3$).
The last case describes the false alarm rate for the $D$ chart calculated in a leave-one-out manner (see Figure 106b):

- Underfitting the data gives satisfying results for the false alarm rate for large datasets ($K = 500 - 2000$) while the smaller datasets ($K = 20 - 100$) have a false alarm rate that is slightly higher than $\alpha (0.05)$. Therefore the leave-one-out method is only more robust against underfitting for the largest datasets ($K > 100$).
- The false alarm rate is in perfect accordance with $\alpha (0.05)$ if the sample size is large enough ($K > 500$) and the model is chosen correctly ($R = 3$)
- The problems associated to overfitting and underfitting small datasets ($K < 2000$) are not solved (compare with Figure 103b).

It was already explained that the generalization going from the Hotelling $T^2$ statistic to the $D$ statistic is problematic. These problems are not solved by using a leave-one-out method. Researchers in the field of MSPC are challenged to investigate alternative test statistics for the scores that have better statistical properties.

The results found for continuous processes can be extrapolated to batch processes. The sample size of NOC data for batch processes lies typically around 20 – 50 batches. It was shown that such small datasets may suffer from poor false alarm rate properties. Overfitting this data will enlarge this problem of the false alarm rate even further. It is the experience of the authors that the false alarm rate for batch process monitoring models is on average too high. It is recommended to use a parsimonious model together with a proper leave-one-out procedure.

### 6.3.7 CONCLUSIONS

Various datasets have been constructed to examine the effect of over/underfitting to the false alarm rate. The results of these simulations lead to the following conclusions. First, the performance of the $SPE$ chart in terms of the false alarm rate is too high for small datasets ($K < 100$) in case of overfitting. This problem is substantially reduced by using leave-one-out residuals. Secondly, the false alarm rate for the $D$ chart is troublesome for small datasets ($K < 500$) if the data is overfitted. This effect is not
improved by using a proper leave-one-out procedure to calculate the scores. Finally, it is recommended to use a parsimonious model together with a proper leave-one-out. The performance of the $D$ chart is troublesome. More research is required to investigate alternative test statistics for the scores that have better statistical properties.
6.4 Fault manifestation in the $D$ and $SPE$-chart for different batch monitoring strategies*

6.4.1 SUMMARY
The manifestation of process faults in the control charts is studied for three methods of batch process monitoring. These methods are discussed in Chapter 4.3. For batch process monitoring, two test statistics are monitored in-control charts: the $SPE$ and $D$-statistic. The diagnosing capabilities of the $SPE$ and $D$-statistics are established by a mathematical analysis of their underlying structure. This analysis shows that the diagnosing capabilities are severely hampered by the problem of embedded error, always present in projection methods (like PCA). These problems are, to some extent, present in all the process monitoring models considered, i.e. global, local and time evolving models.

6.4.2 INTRODUCTION
Many products originating from the chemical, pharmaceutical and food industry are produced in a batch-wise manner. The flexibility of batch processes can be regarded as an important advantage as compared to e.g. continuous processes. For this reason, the usage of batch processes is expected to increase in the coming years. Batch processes can be characterized by their finite duration, non-linear behavior of the process variables and high conversions. Most batch processes are recipe driven and this recipe is often based on expert knowledge. Monitoring of batch processes is preferable for several reasons. Monitoring schemes may lead to e.g. shorter batch durations, improved product quality, better process understanding, more consistent batches, reduced waste streams or improved safety circumstances. One way of monitoring batches is to develop fundamental models and use these models to predict the ongoing of a batch process. However, the chemistry of most batch processes such as biochemical or polymerization reactions is very complex. Therefore, it is not

possible to construct a detailed model because too limited knowledge is available. Furthermore, model development of such complex processes is time-consuming.

One of the alternatives is to measure and collect important process variables during a batch run. These measurements from as many batches as possible are stored in a well organized database. Such a collection of data contains valuable information about the batch process under consideration. Multivariate statistical methods are well suited to deal with this type of data. These methods offer the possibility to capture the underlying phenomena present in the data in terms of easily interpretable latent variables.

The application of multivariate statistical methods to batch processes is referred to as statistical batch process monitoring. The concept of statistical batch process monitoring was introduced in the mid-ninety's by P. Nomikos & J.F. MacGregor [6]. Many applications and extensions of statistical batch process monitoring have followed since (S. Rannar et al. [31], S. Wold et al. [53], D. Neogi & C.E. Schlags [44], D.J. Louwerse & A.K. Smilde [38], J.A. Westerhuis et al. [42], A.K. Smilde & H.A.L. Kiers [39]). The statistical performance of a batch is monitored in two multivariate control charts, namely the $SPE$ and $D$-chart.

This chapter discusses alternative models, which can be used in statistical batch process monitoring. From these models, control charts are derived. This series focuses on the fault detection (Part I) and the diagnostic capabilities (Part II) of these charts. The goal of this chapter is twofold. First, the mechanics and non-complementarity of the control charts are shown. This is done by attempting to construct process faults that only manifest themselves in one of the two charts. Secondly, the diagnosing characteristics of typical faults are shown for all three models.

The chapter is organised as follows: in Section 6.4.3 a description of the $SPE$ and $D$-statistic is given. In Section 6.4.3.1, Section 6.4.4 and Section 6.4.5 the mechanistic of the $SPE$ and $D$-statistic are thoroughly discussed. Section 6.4.7 explains the problem of the embedded error. Section 6.4.9 describe the simulation of a series of faults. These faulty batches are then monitored using three models for statistical batch process monitoring. The diagnosing capability for each model is discussed.
6.4.3 A QUALITATIVE DESCRIPTION OF THE D AND SPE-STATISTIC

A description of the global, evolving and local model is given section 4.3. Section 4.3 describes the performance of these three models in terms of the overall type I error (OT1) and action signal time (AST). The OT1 is an indicator for the number of false alarms where the AST is a measure of the speed of detecting process upsets. It was concluded from Part I that in terms of detection all three models gave satisfying results. Another issue besides fault detection is the diagnosis of a fault. This part deals with the issue of the diagnostic properties of the three models. The distinction between the failure of a sensor or the drift of a process might be beneficial. The broken sensor might be easily repaired and there is no need to stop the reaction.

In part I it was explained that the statistical behavior of a batch is monitored in the SPE and D-chart. These two charts are the tools for diagnosing a fault. If e.g. a fault manifests itself only in the SPE-chart, certain characteristics can be described to this process fault. The same is true if a process fault manifest itself only in the D-chart. The construction of the SPE and D-statistic is explained briefly in the following. This has already been explained in part I, however, the mechanics are explained in more detail now. Also the physical representation of the SPE and D-statistic is discussed.

6.4.3.1 Mechanics of the SPE and D-statistic

The mechanics of the SPE and D-statistic is explained using a simple example. A set of 10 samples, denoted as \( \mathbf{X} \) (10x2), is used to perform a principal component analysis (PCA) on. First, the columns of \( \mathbf{X} \) are mean centered and scaled to unit variance. The pre-processed data is represented by the diamonds in Figure 107.
A PCA model with only one component is chosen to describe the data. The PCA model that describes $\mathbf{X}$ looks like:

$$\mathbf{X} = \mathbf{t}\mathbf{p}^\prime + \mathbf{E}$$

(128)

The scores are given by $\mathbf{t}$ (10×1), the model loadings by $\mathbf{p}$ (2×1) and the residuals by $\mathbf{E}$ (10×2). The graphical interpretation of the PCA model is given in Figure 107. The samples of the matrix $\mathbf{X}$ are orthogonally projected onto the line represented by the loadings. As can be seen from Figure 107, these projections onto this line are given by the dots. In this picture, two important distances can be distinguished.

First, consider the vector that connects e.g. sample 1 to the line. This connection is described by the vector given by the first row of the residual matrix $\mathbf{E}$. The sum of the squared elements of this vector makes up the $SPE$-statistic for this sample:
Statistical batch process monitoring

\[ \text{SPE}_i = \sum_{j=1}^{I} e_{ij}^2 \]  \hspace{1cm} (129)

Secondly, consider the distance from the projection of the sample (given by the dots) to the centre of the line given by [0,0]. The centre of the model represents the average of all samples. The coordinates of e.g. the first sample on the line are given by the first row of the score vector \(t\). The sum of the squared elements of this row of \(t\) scaled by the covariance matrix is referred to as the \(D\)-statistic. This covariance matrix is calculated from the scores. In this example, the covariance matrix of the scores \(t\) will result in a scalar \(s^2\) because \(t\) is a vector. The equation to calculate the \(D\)-statistic that represents the distance from the projected sample towards the model centre is given by:

\[ D_i = t_i \cdot s^{-2} \cdot t_i \]  \hspace{1cm} (130)

Summarizing, the samples can be described by scores and residuals. The residuals are a measure of the distance towards the model where the model is described by the loadings. This is expressed in the \(SPE\)-statistic. The scores are a measure of the distance within the model represented by the \(D\)-statistic. That is, the distance from the orthogonally projected sample towards the centre of the model. A sample can be assigned as non-representative if the \(SPE\) or \(D\)-statistic or both statistics are above control limits.

6.4.3.2 Relationship between real process disturbances and the \(SPE\) and \(D\)-statistic

In statistical batch process monitoring, fault detection plays an important role. The result of the projection of a sample is expressed in terms of the \(SPE\) and \(D\)-statistic. The question therefore is: what kind of relation is there between a real process fault and the result of this fault in the \(SPE\) and \(D\)-statistic? To answer this question, the process measurements are categorised in two groups: i) faults that break the correlation structure or ii) faults that obey the correlation structure but have a more than normal variation. As an example, think of the following situation. The temperature \(T_{im}\) and pressure \(P_{im}\) are measured in the inner tube of a car tyre, as well as the ambient temperature \(T_a\). Now, as \(T_a\) increases because of the sunny
weather, so will $T_{pm}$ and $P_{pm}$. Thus, the process variables are correlated and behave according to simple physical laws. This is depicted in Figure 108:

The white bars represent situations where the process behaves under normal conditions. The black bars show the reaction of $T_{pm}$ and $P_{pm}$ when $T$ is changing. This is according to the correlation between the process variables. Besides, the variation of the increment of the process variables is considered as acceptable.

The first group of faults is the following. In extreme situations where the surrounding temperature may become very high (e.g. the car has been parked in the burning sun), the pressure and temperature in the inner tube might become dangerously high. Although the process variables behave according to the law of physics, the variation of the process variables is more intense than under normal circumstances. This is illustrated in Figure 108 as the patterned black bar. Such a fault where the variation of the process variables is abnormal high but the correlation between the process variables remains intact will be denoted as intensified correlation throughout this chapter.

The other group of faults is of the following. Suppose the temperature in the inner tube $T_{pm}$ increases as a result of $T$, but because the valve of the car tube is leaking, the pressure $P_{pm}$ remains constant. This is not according to the correlation between the process variables, and therefore the correlation is broken. This can be seen from Figure 109. Such disturbances will be denoted as breakage of the correlation.
Often heard remarks in the literature (R. Duna & S.J. Qin [114], S. Albert & R.D. Kinley [43], J.V. Kresta et al. [10], P. Nomikos & J.F. MacGregor [6]) are that abnormal variation that still obeys the correlation structure of the process variables is described by the scores (D-statistic) while new events not present in the NOC data will represent itself in the residuals (SPE-statistic). It is believed that an intensified correlation (e.g. process shifts) is detected in the D-statistic while the breakage of the correlation (e.g. sensor failure) is detected in the SPE-statistic. This implies that the control charts are complementary. That is, certain type of process faults can be detected in either the SPE-chart of D-chart. Assessing this complementarity gives better understanding how the SPE and D-chart work together to detect faulty batches. Therefore, it is investigated what type of faults form the base of the complementarity. That is, faults are constructed that generate a pure SPE-signal or a pure D-signal. These faults can be regarded as extreme faults. If both type of faults cannot be constructed for a model, the complementarity of the control charts does not exist. It will be shown that for some models pure D- or SPE-signals cannot be constructed. Furthermore, the issue of embedded error is discussed. Due to the embedded error, most faults will be distributed over both charts.

6.4.4 CONSTRUCTION OF PURE D-FAULTS

The performance of monitoring charts based on the D-statistic are troublesome (E.N.M. Van Sprang et al. [41]). The reason for this poor performance can be understood from the nature of a D-fault. This section describes the construction of a fault that only affects the scores (and thus D-statistic) in an on-line monitoring scheme.
Consider the measurement of a pure sample \( \mathbf{x}^* \) depicted as the black dot in Figure 110:

![Diagram of Figure 110](image)

**Figure 110**

*Graphical interpretation of a pure D-signal.*

In the ongoing of this chapter a sample \( \mathbf{x}^* \) refers to the ideal measurements of the process variables from a batch process. This sample is free from any experimental noise and can be written exactly according to the model \( \mathbf{p} \). Therefore this sample can be expressed in the coordinate system described by the latent variables as:

\[
\mathbf{x}^* = \mathbf{p}t^*
\]  

(131)

There are no residuals because \( \mathbf{x}^* \) can be written exactly according to the model. The scores \( t^* \) found for \( \mathbf{x}^* \) are the true scores \( t^* = \mathbf{p}^* = (\mathbf{p}^* \mathbf{p}^*)^{-1} \mathbf{t}^* \). However, in a real situation the pure sample \( \mathbf{x}^* \) is affected by process faults denoted as \( \mathbf{e} \). A real sample \( \mathbf{x} \) affected by process fault is written as:

\[
\mathbf{x} = \mathbf{x}^* + \mathbf{e} = \mathbf{p}t^* + \mathbf{e}
\]  

(132)

In this Section, the process fault \( \mathbf{e} \) is chosen in such a way that it only affect the scores. This requires a specific direction of the process fault \( \mathbf{e} \). It can be seen from
Figure 110 that the direction of the process fault is given by $e^D$. The uppercase $D$ refers to the case where the process fault represents a pure $D$-fault. It can be seen that $e^D$ moves along the line given by $p$. This process fault $x^D (= x^* + e^D)$ only affects $D$, and thus, represents a pure $D$-fault and generates a pure $D$-signal. The required behaviour of the process variables to generate a pure $D$-signal is given by $x_1^D$ and $x_2^D$ from Figure 7. Only those combinations of $x_1^D$ and $x_2^D$ that sum up to $x^D$ will result in a pure $D$-fault.

For practical reasons, it was assumed in the preceding that $x^*$ could be written exactly according to the model. Therefore the residuals $e^*$, describing random error sources such as measurement error, were zero. This is not a very realistic assumption, therefore $e^*$ is non-zero in the remaining of this chapter. This has the following consequences: compared to the situation where the residuals $e^*$ are zero, the vector $e^D$ is not lying on the line described by $p$ anymore. Instead, it moves somewhere in the space described by $x$ in a direction perpendicular to the model $p$. This situation is depicted in Figure 111:

![Graphical interpretation of a pure D-signal with nonzero residuals.](image-url)
In this picture, \( x' = p' + e' \) is represented by the black dot. In this case the nature of \( x' \) is different compared to Equation 131 where the residuals \( e' \) are zero. The vector \( x'^D = x' + e' = p' + e' + e'^D \) represents the measurement of process variables that contain a process fault \( e'^D \).

The extracted residuals \( e^0 \) for \( x' \) are given by:

\[
e^0 = x' - x' = (p' + e') - p' = e' \tag{133}
\]

For a pure \( D \)-fault, the extracted residuals \( e^0 \) resulting from the projection of \( x' \) or \( x'^D \) should be the same if the process fault represents \( e'^D \) where the extracted residuals for \( x'^D \) are given by

\[
e^0 = x'^D - x'^D = (p' + e' + e'^D) - (p' + e'^D) = e' \tag{134}
\]

The extracted residuals are the same for both cases. This should hold for every pure \( D \)-signal.

From Figure 111, an important remark can be made. Suppose that the process fault \( e'^D \) was of opposite direction and moved towards the origin of the line. This will still result in a pure \( D \)-signal, but the value for the \( D \)-statistic will decrease. Such a process fault will never lead to detection in conventional \( D \)-charts. N.D. Tracy et al. [14] mentioned to use a control chart with upper and lower control limits. This, however, suggests a two sided test with null hypothesis of the form \( D = 0 \) on target (target > 0) and alternative hypothesis \( D \neq 0 \) on target. A \( D \)-value of zero might then indicate wrong behaviour, but this contradicts intuition since the best batch should have a \( D \)-value of zero (S.J. Wierda [129]). This suggestion is therefore not followed.

6.4.4.1 General construction of a pure \( D \)-fault:
This section describes the general way to construct a sample \( x'^D \). Therefore the theory and figures presented in the ongoing of this chapter are applicable to higher dimensional problems. The machinery to construct such a sample can be easily extrapolated to the more specific case where a sample \( x'^D \) is constructed using local,
evolving or global models. The general equation to construct such a sample $x^D$ reads as:

$$x^D = \frac{P t^* + e^* + e^D}{x^*}$$  \hspace{1cm} (135)$$

The process fault $e^D$ can be written according to:

$$e^D = \delta \cdot Pt^* \rightarrow x^D = Pt^* + e^* + \delta \cdot Pt^* = (1 + \delta)Pt^* + e^*$$  \hspace{1cm} (136)$$

Equation 136 means that $e^D$ moves along or parallel to the line in Figure 110. The positive value of $\delta$ will increase the $D$-statistic and therefore moves away from the model center. This process fault describes the situation where the process variables behave exactly according to the correlation. In other words, the projection of $x^D$ or $x^*$ results in exactly the same residuals $e^*$ according to Equation 133 and 134. This was referred to as an intensified correlation. The construction of such process faults that generate a pure $D$-fault in an on-line manner using local, evolving and global models will be discussed in the following sections.

6.4.4.2 Construction of a pure $D$-fault for the local model

For a local model, $K$ different models are constructed at every time interval. These models are denoted as $\hat{P}_k$ ($k = 1, \ldots, K$). The measurement of the process variables that is taken at every time interval is denoted as $\hat{x}_k^*$. The general Equations 135 and 136 applied to a local model to construct a pure $D$-signal at time interval $k$ are as follows:

$$\bar{e}^D_k = \delta \cdot \hat{P}_k \bar{t}_k^* \rightarrow \bar{x}^D_k = (1 + \delta) \cdot \hat{P}_k \bar{t}_k^* + \bar{e}_k^*$$  \hspace{1cm} (137)$$

An example of the process variables that represent this fault is given in Figure 112.
In this figure, the white bars represent the measurements of a new independent batch without a process fault ($\hat{\mathbf{x}}_{k}^*$) and therefore represents NOC conditions. At time interval $k + 3$, the process fault appears. This is represented by the black bars. It can be seen that at this time interval the process variables react according to the correlation. It must be stressed that the situation shown in Figure 112 represents very specific and unique process conditions. Only the slightest deviation from this situation will distribute the process fault over both the scores and residuals and the instantaneous correlation is broken. In this case, both the $SPE$ and $D$-statistic are affected. This will be further discussed in the Section about the embedded error.

It can be concluded that it is possible to construct a process fault that results in a pure $D$-signal in an on-line monitoring scheme for the local model. For this reason it is also possible that real process faults can be detected exclusively in the $D$-chart using a local model.
6.4.4.3 Construction of a pure D-fault for the evolving model
This section describes the construction of a fault that is detected exclusively in the $D$-chart for the time-evolving model. Let the batch operated at NOC be depicted in Figure 113a as the white bars.

This batch has a duration of seven time intervals. Suppose that the first measurement of the process variables is carried out and the future measurements of the batch are unknown. This measurement corresponds to the white bars that represent NOC conditions at time interval $k = 1$ and is denoted as $\bar{x}_{1}$. This part of the batch is modelled as follows:

$$\bar{x}_{1}^{*} = \tilde{P}_{1} t_{1}^{*} + \tilde{e}_{1}^{*}$$

(138)
Now, let the batch resulting in a pure $D$-signal after projection at the first time interval be denoted as $\bar{x}_1^D$. This fault is constructed according to the following equation (see general Equation 136):

$$\bar{x}_1^D = (\delta + 1) \cdot \bar{P}_1 \cdot \bar{i}_1^* + \bar{e}_1^*$$ (139)

The process fault is depicted as the black bars in Figure 113a. The evolving model at the first time interval equals a local model for the same time interval. Therefore, considering the first time interval, the same conclusions can be drawn as for a local model.

Now consider the measurement at the second time interval of the batch. The mathematical construction of a pure $D$-signal is discussed first. Let the vector $\bar{x}_2^*$ consist of the white bars measured at time interval $k = 1$ and $k = 2$ from Figure 113a. This vector can be partitioned and is given by:

$$\bar{x}_2^* = \begin{bmatrix} \bar{x}_{21}^* \\ \bar{x}_{22}^* \end{bmatrix}$$ (140)

where the first part corresponds to time interval one and the second part to time interval two. The fault that will result in a pure $D$-signal is given by $\bar{x}_2^D$ and is calculated as:

$$\bar{x}_2^D = (\delta + 1) \cdot \bar{P}_2 \cdot \bar{i}_2^* + \bar{e}_2^* = (\delta + 1) \cdot \begin{bmatrix} \bar{P}_{21} \\ \bar{P}_{22} \end{bmatrix} \bar{i}_2^* + \begin{bmatrix} \bar{e}_{21}^* \\ \bar{e}_{22}^* \end{bmatrix}$$

$$= \begin{bmatrix} \bar{x}_{21}^D \\ \bar{x}_{22}^D \end{bmatrix}$$ (141)

The only fault that results in a pure $D$-signal is represented by the grey bars in Figure 113b. Notice that the correlation is intensified over both time intervals in order to generate a pure $D$-signal. This is referred to as an intensification of the overall correlation.
However, this pure mathematical construction cannot be realized because of the following. The mathematical part of $\tilde{x}_2^D$ corresponding to the first time interval ($\tilde{x}_{21}^D$) will be replaced with the real fault given by $\tilde{x}_1^D$. These parts are different as can be seen from the following:

$$\tilde{x}_{21}^D \neq \tilde{x}_1^D \quad \text{because} \quad (\delta + 1) \cdot \tilde{i}_2^* \tilde{P}_{21} + \tilde{e}_{21}^* = (\delta + 1) \cdot \tilde{i}_1^* \tilde{P}_1 + \tilde{e}_1^* \quad (142)$$

since the loadings $\tilde{P}_{21}^D$ and $\tilde{P}_1^D$ are different. The vector that results is depicted in Figure 113c. In this picture $\tilde{x}_{21}^D$ is replaced with $\tilde{x}_1^D$. If the vector $\tilde{x}_2^D$ is projected, the extracted residuals found for $\tilde{x}_2^D$ will not be the same as $\tilde{x}_2^*$ according to Equation 133 and 134. This is required for a pure D-signal. The overall correlation pattern is broken instead. The same arguments hold for the following time points.

It was shown that a batch generating a pure D-signal couldn't be constructed for evolving models in an on-line monitoring scheme except for the first time interval. The overall correlation is always broken while it can still be true that the instantaneous correlation is intensified. The only exception to this is the first time interval of a batch where indeed it is possible to intensify the correlation in such a way that a pure D-signal is generated. From a practical point of view the first measurement can be used to check the initial conditions of a batch. A shift in these conditions might lead to an intensified correlation.

**6.4.4.4 Construction of a pure D-fault for the global model**

This section will show that for on-line batch monitoring the overall correlation modeled by a global model will always be broken. This situation is easier to explain as compared to the evolving models. Consider the normal batch $\mathbf{x}^*$ represented as white bars from Figure 114.
Assume this figure represents the whole batch run. To construct a faulty batch $\ddot{x}_k^D$ in such a way that only the $D$-chart is affected because the correlation is intensified, only one situation is possible. This situation is given by the following equation:

$$\ddot{x}_k^D = (\delta + 1) \cdot \dddot{y}_k^* + \dddot{z}_k^*$$  \hspace{1cm} (143)

The batch $\ddot{x}_k^D$ has been illustrated in Figure 114a. It is required that the overall correlation is intensified during the entire batch run from time point one until seven. Projection of this batch indeed will lead to the desired pure $D$-fault. Projection of $\dddot{x}_k^D$ resembles a post-batch analysis, and for such an analysis it is possible to detect a fault only in the $D$-chart. Another situation of a post batch analysis is depicted in Figure 114c. This picture represents a completed batch. The instantaneous correlation has been changed from $k = 4$ until the end of the batch. The post batch projection of this batch will break the correlation structure, because the first part of the batch did not take part in the intensification of the correlation. This batch does not resemble the
required correlation intensification from Figure 114a. Deviations from this ideal correlation intensification will result in a breakage of the overall correlation.

The situation where the correlation indeed is intensified but the batch has not finished yet is illustrated in Figure 114b. The batch has evolved until \( k + 3 \). The instantaneous correlation is intensified on the time intervals 1, 2, 3. This is illustrated as black bars. Because of the filling in procedure, the process variables at \( k = 4,5,6 \) and 7 resemble the last measurement from \( k = 3 \). These measurements are given in Figure 114b as grey bars. Clearly, this batch deviates from the batch depicted in Figure 114a although the first measurement \( k = 1 \) until \( k = 3 \) fulfil the requirements. The required overall correlation intensification is disturbed because of the filling in procedure. This disturbance will be higher if the batch has just started because relatively large future parts of the batch need to be filled in. In other words, the overall correlation will always be broken using the global model for on-line batch monitoring. This is not only true for the current deviations approach but for any filling-in procedure.

It can be stated that pure \( D \)-signals cannot be constructed for global models for on-line monitoring. This statement holds for every filling in method. The overall correlation is always broken when global models are used for on-line batch monitoring.

### 6.4.5 Construction of pure \( SPE \)-faults

This section describes the construction of process faults that generate a pure \( SPE \)-signal in an on-line monitoring scheme. The line of reasoning will be the same as for the construction of faults that generate pure \( D \)-signals. Like a pure \( D \)-signal, the direction of the vector \( \mathbf{e} \) that describes a process fault needs to be very specific in order to only affect the residuals. Assume that \( \mathbf{x}^* \) can be written exactly according to the model. To visualize this situation, the residuals \( \mathbf{e}^* \), describing random error sources such as measurement error, are zero for the moment. The vector of interest is denoted as \( \mathbf{e}^{SPE} \) and is illustrated in Figure 115:
The superscript $SPE$ refers to the case where the process fault represents a pure $SPE$-fault. It can be seen that $e^{SPE}$ moves perpendicular from the model given by $p$. The projection of the sample $x^{SPE}$ leads to unaffected scores compared to the scores calculated for the sample $x^*$. The uppercase $SPE$ refers to a vector or number that will generate a pure $SPE$-signal. The residuals after projection ($e^0$) of $x^{SPE}$ equal the process fault ($e^{SPE}$). In this situation, the $SPE$ is based upon a completely projected process fault.

The necessary behaviour of the process variables to generate a pure $SPE$-fault is given by $x_1^{SPE}$ and $x_2^{SPE}$. If $x_1^{SPE}$ and $x_2^{SPE}$ do not describe a situation such that $x^{SPE}$ lies somewhere in the line of $e^{SPE}$, the scores will be affected and the residuals are based upon the part of $e^{SPE}$ that is not projected. The construction of a vector $x^{SPE}$ will be discussed in the following using local, global and evolving models.
6.4.5.1 General construction of a pure SPE-fault:

Let the sample \( \mathbf{x}' = \mathbf{p}' + \mathbf{e}' \) represents NOC conditions. For the same reasons as for the pure \( D \)-signal, the residuals \( \mathbf{e}' \) are now chosen to be non-zero. The scores for \( \mathbf{x}'^{\text{SPE}} \) are calculated according to:

\[
\mathbf{t}'^{\text{SPE}} = \mathbf{P}' \mathbf{x}'^{\text{SPE}} = \mathbf{P}' \left( \mathbf{x}' + \mathbf{e}' + \mathbf{e}'^{\text{SPE}} \right) = \mathbf{P}' \mathbf{x}' + \mathbf{P}' \mathbf{e}' + \mathbf{P}' \mathbf{e}'^{\text{SPE}}
\]

(144)

Since these scores \( \mathbf{t}'^{\text{SPE}} \) are not allowed to be affected compared to the scores \( \mathbf{t}' \), it should hold that \( \mathbf{t}'^{\text{SPE}} = \mathbf{t}' \). Therefore the following should hold:

\[
\mathbf{t}' = \mathbf{t}'^{\text{SPE}} \quad \Rightarrow \quad \mathbf{P}' \mathbf{x}' + \mathbf{P}' \mathbf{e}' = \mathbf{P}' \mathbf{x}' + \mathbf{P}' \mathbf{e}' + \mathbf{P}' \mathbf{e}'^{\text{SPE}} = \mathbf{t}' + 0 + \mathbf{P}' \mathbf{e}'^{\text{SPE}} \quad \Rightarrow \quad \mathbf{P}' \mathbf{e}'^{\text{SPE}} = 0
\]

(145)

Since the residuals \( \mathbf{e}' \) are perpendicular to the loadings \( \mathbf{P}' \), the term \( \mathbf{P}' \mathbf{e}' \) equals zero. Equation 145 holds if the vector \( \mathbf{e}'^{\text{SPE}} \) lies in the null-space of the loadings \( \mathbf{P}' \) because

\[
N(\mathbf{P}') = \{ \mathbf{e}'^{\text{SPE}} : \mathbf{P}' \mathbf{e}'^{\text{SPE}} = 0, \mathbf{e}'^{\text{SPE}} \in \mathbb{R}^{l} \}
\]

(146)

The vector \( \mathbf{e}'^{\text{SPE}} \) is constructed as follows. Consider the singular value decomposition of some matrix \( \mathbf{X} \) that can be written according to:

\[
\mathbf{X} = \mathbf{USV}' = \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1' + \mathbf{U}_2 \mathbf{S}_2 \mathbf{V}_2' = \mathbf{TP} + \mathbf{E}
\]

(147)

where

\[
\mathbf{U} = [\mathbf{U}_1 ; \mathbf{U}_2], \mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & 0 \\ 0 & \mathbf{S}_1 \end{bmatrix}, \mathbf{V} = [\mathbf{V}_1 ; \mathbf{V}_2]
\]

(148)

Thus,

\[
\mathbf{P}' = \mathbf{V}_1 \quad \Rightarrow \quad \mathbf{P} = \mathbf{V}_1
\]

(149)

Since \( \mathbf{V} \) is an orthogonal matrix it holds that \( \mathbf{V}_1 \mathbf{V}_2 = 0 \). Hence, \( \mathbf{V}_2 \) is a basis for the null space of \( \mathbf{P}' \) and \( \mathbf{e}'^{\text{SPE}} \) can be constructed using that basis.
6.4.5.2 Construction of a pure SPE-fault for the local model

The construction of a pure SPE-fault for a local model is fairly easy. First, the null space of the loadings $\tilde{P}_k$ is calculated at time interval $k$. An arbitrary combination of the columns of the null space is taken to be the vector $\tilde{e}_k^{SPF}$. The measurement that describes the pure SPE-signal is like:

$$\tilde{x}_k^{SPF} = \tilde{x}_k^* + \tilde{e}_k^* + \tilde{e}_k^{SPF}$$

(150)

The scores at time interval $k$ are calculated according to:

$$\hat{\tau}_k^{SPF} = \hat{P}_k^{T} x_k^{SPF} = \hat{P}_k^{T} (\tilde{x}_k^* + \tilde{e}_k^* + \tilde{e}_k^{SPF}) =$$

$$\hat{P}_k^T \tilde{x}_k^* + 0 + 0 = \hat{\tau}_k^*$$

(151)

where the residuals $\tilde{e}_k^*$ are nonzero. The projection of the residuals $\tilde{e}_k^*$ on the model $\hat{P}_k$ is zero due to the model and the projection of the process fault $\tilde{e}_k^{SPF}$ is zero by construction. Obviously the scores $\hat{\tau}_k^{SPF}$ are not affected and equal the pure scores $\hat{\tau}_k^*$. The extracted error $e_k^0$ is calculated as:

$$\tilde{e}_k^0 = \tilde{x}_k^{SPF} - \tilde{x}_k = (\tilde{x}_k^* + \tilde{e}_k^* + \tilde{e}_k^{SPF}) - \hat{P}_k^T \hat{\tau}_k^* =$$

$$\left(\tilde{x}_k^* + \tilde{e}_k^* + \tilde{e}_k^{SPF}\right) - \hat{P}_k^T \left(\hat{P}_k^{T} \left(\tilde{x}_k^* + \tilde{e}_k^* + \tilde{e}_k^{SPF}\right)\right) =$$

$$\left(\tilde{x}_k^* + \tilde{e}_k^* + \tilde{e}_k^{SPF}\right) - \hat{P}_k \left(\hat{\tau}_k^* + 0 + 0\right) = \left(\tilde{x}_k^* + \tilde{e}_k^* + \tilde{e}_k^{SPF}\right) - \tilde{x}_k^* = \tilde{e}_k^* + \tilde{e}_k^{SPF}$$

(152)

This pure SPE-signal can be constructed at every time interval. Since the construction of such a fault is possible, it could also happen in a real life process. The process variables describing this fault break the correlation in a very specific way. The likelihood of meeting these process conditions in a real life situation can be considered as very low.

6.4.5.3 Construction of a pure SPE-fault for the evolving model

The SPE-statistic, using an evolving model, calculated at time interval $k$ is based upon the current part of the residuals $e_k^0$ (P. Nomikos & J.F. MacGregor [6]). For this reason, only the null space of the current part of the loadings $\hat{P}_k^'$ needs to be calculated. The current part (cp) of e.g. the loadings $\hat{P}_k^*$ is denoted as $\hat{P}_k^{sp}$ while the
past part (pp) is referred to as $\tilde{P}_{PP}^*$. Thus the matrix $\tilde{P}_{PP}^*$ can be partitioned according to:

$$
\begin{bmatrix}
\tilde{P}_{PP}^* \\
\tilde{P}_{PP}^{SP} \\
\tilde{P}_{PP}^{CP}
\end{bmatrix}
= 
\begin{bmatrix}
\tilde{P}_{PP}^* \\
\tilde{P}_{PP}^{SP} \\
\tilde{P}_{PP}^{CP}
\end{bmatrix}_{R \times j(k,1)}
\quad (\text{153})
$$

Consider the batch $\tilde{x}_k$ ($j \times 1$) at time interval $k$. The current part $\tilde{x}_{k}^{SP}$ ($j \times 1$) of this batch is replaced with:

$$
\tilde{x}_{k}^{CP} = \tilde{x}_{k}^{SP} + \tilde{e}_{k}^{CP} + \tilde{e}_{k}^{SP}
\quad (\text{154})
$$

where the vector $\tilde{e}_{k}^{SP}$ ($j \times 1$) is constructed from the null space of the loadings $\tilde{P}_{PP}^*$. The batch where the current part has been replaced according to Equation 154 will be referred to as $\tilde{x}_{k}^{SP}$ ($j \times 1$). The scores $\tilde{t}_{k}^{SP}$ for this particular batch are calculated according to:

$$
\tilde{t}_{k}^{SP} = \tilde{P}_{PP}^* \tilde{x}_{k}^{SP} = \left[ \tilde{P}_{PP}^* \quad \tilde{P}_{PP}^{SP} \right] \tilde{x}_{k}^{SP} = \\
\tilde{P}_{PP}^* \left( \tilde{x}_{k}^{PP} + \tilde{e}_{k}^{PP} \right) + \tilde{P}_{PP}^{SP} \left( \tilde{x}_{k}^{SP} + \tilde{e}_{k}^{CP} + \tilde{e}_{k}^{SP} \right) = \\
\tilde{P}_{PP}^* \tilde{x}_{k}^{PP} + \left[ \tilde{P}_{PP}^* \quad \tilde{P}_{PP}^{SP} \right] \left[ \tilde{e}_{k}^{PP} \quad \tilde{e}_{k}^{CP} \quad \tilde{e}_{k}^{SP} \right] + \tilde{P}_{PP}^{SP} \tilde{x}_{k}^{SP} + \tilde{P}_{PP}^{SP} \tilde{e}_{k}^{SP} = \\
= 0
$$

$$
\left[ \tilde{P}_{PP}^* \quad \tilde{P}_{PP}^{SP} \right] \tilde{x}_{k}^* = \tilde{P}_{PP}^* \tilde{x}_{k} = \tilde{t}_{k}^*
\quad (\text{155})
$$

where the projection of residuals $\tilde{e}_{k}^{PP}$ and $\tilde{e}_{k}^{SP}$ are zero due to the model properties and the process fault $\tilde{e}_{k}^{SP}$ is zero by construction. It can be seen from Equation 155 that the scores $\tilde{t}_{k}^{SP}$ are equal to $\tilde{t}_{k}^*$ and therefore the scores are not affected by the process fault $\tilde{e}_{k}^{SP}$. The extracted error and thus $SPE$-statistic at time interval $k$ is based upon the current part of $\tilde{x}_{k}^{SP}$.
Chapte rr 6

•  Statistic s  an d  mechanistics

\[ t^0 = \tilde{x}_k^{\text{SPE}} - \bar{x}_k^{\text{SPE}} = \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} - \bar{x}_k^{\text{SPE}} \right) - \bar{x}_k^{\text{SPE}} = 0 \]

\[ \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} - \bar{x}_k^{\text{SPE}} \right) - \bar{x}_k^{\text{SPE}} = \left( \bar{p}_k^{\text{PP}} \left( \bar{x}_k^{\text{PP}} + \tilde{e}_k^{\text{PP}} \right) + \bar{p}_k^{\text{CP}} \left( \bar{x}_k^{\text{CP}} + \tilde{e}_k^{\text{CP}} + \tilde{e}_k^{\text{SPE}} \right) \right) = \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} \right) - \bar{x}_k^{\text{SPE}} \left( \bar{p}_k^{\text{PP}} \left( \bar{x}_k^{\text{PP}} + \tilde{e}_k^{\text{PP}} \right) + \bar{p}_k^{\text{CP}} \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} \right) + 0 + 0 \right) = \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} \right) - \bar{x}_k^{\text{SPE}} \left( \bar{p}_k^{\text{PP}} \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} \right) + \bar{p}_k^{\text{CP}} \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} \right) + 0 + 0 \right) = \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} \right) - \bar{x}_k^{\text{SPE}} \left( \bar{p}_k^{\text{PP}} \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} \right) + \bar{p}_k^{\text{CP}} \left( x_k^{\text{SPE}} + \tilde{e}_k^{\text{SPE}} \right) + 0 + 0 \right) \]

It can be seen that the extracted residuals are based upon the complete projection of the process fault \( \bar{e}_k^{\text{SPE}} \) since they are in the same direction. Then, the extracted residuals are used to calculate the SPE-statistic. The construction of pure SPE-signals in an on-line monitoring scheme can be repeated for every time interval using an evolving model. Since a pure SPE-signal is possible by construction, it could also happen in practice. However, like the pure SPE-signal constructed for a local model, the process variables require a very specific breakage of the correlation. It is unlikely that the process conditions will be exactly similar to this required behaviour of the process variables.

6.4.5.4 Construction of a pure SPE-fault for the global model

The main difference between the global model on the one hand and the local and evolving model on the other hand is that for the global model the future part of the batch needs to be estimated. The filling-in procedure using the current deviations approach makes it impossible to construct a pure SPE-signal. It will be sufficient to show that the scores are affected. The loadings \( \bar{p} \) and the batch \( \bar{x}_k \) of a global model can be partitioned in three parts: past part (pp), current part (cp) and a future part (fp). The current part of the batch \( \bar{x}_k \) is replaced according to the following equation:
Statistical batch process monitoring

\[ \hat{x}^{SP} = \hat{x}^{SP} + e^{SP} + \tilde{e}^{SP} \]  \hspace{1cm} (157)

Here, the process fault given by \( \hat{e}^{SP} \) \((J \times 1)\) is constructed from the null space of the loadings \( \tilde{P}^{SP} (R \times f) \). The batch that is to be projected is given by \( \hat{x}^{SP}_k \) \((J \times 1)\) where the current part of this batch is replaced according to equation 157. The future part \( \hat{x}^{SP}_k \) \((J \times K-K) \times 1\) of this batch is estimated with the current deviations approach:

\[ \hat{x}^{SP}_k = \begin{bmatrix} \hat{x}^{SP}_1 \\ \vdots \\ \hat{x}^{SP}_{K-K} \end{bmatrix} \]  \hspace{1cm} (158)

The scores at time interval \( k \) are calculated according to:

\[ \tilde{t}_k = \tilde{P}^{'} \hat{x}_k = \begin{bmatrix} \tilde{P}^{PP} \\ \tilde{P}^{cp} \\ \tilde{P}^{fp} \end{bmatrix} \begin{bmatrix} \hat{x}^{SP} \\ \vdots \\ \hat{x}^{SP}_{K-K} \end{bmatrix} = \]

\[ \tilde{P}^{PP}\hat{x}^{SP} + \tilde{P}^{cp}\hat{x}^{SP} + \tilde{P}^{fp}\hat{x}^{SP} = \]

\[ \tilde{P}^{PP}\hat{x}^{SP} + \hat{e}^{SP} \]

\[ \tilde{P}^{PP}\hat{x}^{SP} + \hat{e}^{SP} \]

\[ \tilde{P}^{PP}\hat{x}^{SP} + \hat{e}^{SP} \]

It can be seen from Equation 159 that the scores found after projection of the batch are not the same as the true scores. The process fault was taken from the null space of the loadings \( \tilde{P}^{SP} \). The projection of the process fault to the future part of the loadings will not be zero according to Equation 145. As a result, it can be concluded that for on-line monitoring using a global model it is impossible to construct a pure \( SPE \)-signal. If this is not possible by construction, it will also not happen in a real process.

-280-
The only case where a pure SPE-signal can be constructed is for a finished batch. This is also true for the pure D-signal. A process fault $e^{SPE}_K (JK \times 1)$ can be constructed from the null space of the complete loading matrix $\tilde{P}$. The finished batch that results in a pure SPE-signal after projection looks like:

$$\frac{\bar{x}^{SPE}_K}{JK \times 1} = \bar{x}^*_K + \bar{e}^*_K + e^{SPE}_K$$

(160)

6.4.6 CONCLUSIONS FROM SPE AND D-MECHANICS

The following conclusions can be drawn from the preceding Sections. Pure SPE and D-faults were constructed to show the complementarity of the D and SPE-chart for the different models. In case of on-line monitoring, pure SPE and D-signals can be constructed only for the local model. For the evolving model, only pure SPE-signals can be constructed. No pure SPE or D-signal can be constructed for the global model. Therefore, for evolving and global models the control charts cannot be complementary due to model construction. This hampers the diagnostic value of the control charts. The control charts of the local model are indeed complementary by construction.

The pure D-signal and pure SPE-signal may be considered as the two extremes. The pure D-signal requires an intensified correlation between the process variables where the pure SPE-signal requires a very specific breakage of the correlation. The process conditions necessary for a pure D and SPE-signal can be regarded to be very rare and will not happen in practice. The consequence of this will be discussed in the following sections.

In case of post-batch analysis for the global and time evolving models ($k = K$), pure SPE and D-faults can be constructed by projection of the entire batch run to the model corresponding to time interval $K$. In fact, the time-evolving model is exactly the same as the global model for $k = K$, that is $\tilde{P}_k = \tilde{P}$. Such pure SPE or D-faults need to describe disturbances that occurred during the whole batch run. Process faults that occurred e.g. somewhere halfway the batch can never result in pure D or SPE-faults for a post-batch analysis. The faults representing pure D or SPE-faults for post-batch analysis are non-realistic and will not happen in practice.
There is no overall local model available at time interval $K$ as is the case for the global and evolving model. This makes the post-batch analysis different for the local model. The measurements at every time interval from a finished batch are projected onto the corresponding models. This is not different from on-line monitoring of a new batch as described for the local model. Therefore the same arguments as explained for on-line monitoring using local models hold for the pure $SPE$ and $D$-fault.

6.4.7 THE PROBLEM OF THE EMBEDDED ERROR

It was shown in the previous Section that the complementarity does not exist by construction in some cases. This is not true for e.g. the local model where the complementarity does exist. However, the process faults that represent pure $SPE$ or $D$-faults are considered to be rare. It is more likely that the process conditions deviate from this rare behaviour. This Section describes the effect of this deviation in terms of the embedded error.

The measurement of a sample will always be affected by a measurement error. In this chapter, this is referred to as a process fault. It will be shown that this process fault affects both the $SPE$ and $D$-statistic. This problem is known in factor analysis as the embedded error (E.R. Malinowski [124]). The embedded error problem will be illustrated with the same simple example from the previous section. Consider the measurement of a pure sample $x^*$ depicted as the black dot in Figure 116:

![Figure 116](image)

*Graphical interpretation of the embedded error.*
This sample can be written exactly according to the model and there are no residuals. Let the sample \( \mathbf{x} \), depicted as the checkerboard dot in Figure 116, be affected by a process fault. This sample is considered to be very likely in a real process. The next step is to project \( \mathbf{x} \) on the model \( \mathbf{p} \) to find the scores and the residuals. The projected sample is depicted in Figure 110 as the striped dot. The scores are calculated according to:

\[
\iota = \mathbf{p}' \mathbf{x} = \mathbf{p}' (\mathbf{x'} + \varepsilon) = \mathbf{p}' \mathbf{x'} + \mathbf{p}' \varepsilon = \iota' + \mathbf{p}' \varepsilon \neq \mathbf{t}'
\]  

(161)

The score \( \iota \) found for \( \mathbf{x} \) consists of two mixed parts. The first part represents the distance of the centre of the plane towards the black dot. This was given by the true scores \( \iota' \). The distance of the black dot to the striped dot gives the second part. This results from the projection of the process fault \( \varepsilon \).

It can be seen from Figure 110 that the extracted residuals for \( \mathbf{x} \) are given by \( \varepsilon^0 \). These extracted residuals represent the distance towards the model and result from the projection. Clearly, the extracted residuals are not the same as the process fault \( \varepsilon \). This can be shown mathematically from Equation 162:

\[
e^0 = \mathbf{x} - \mathbf{p'} = \mathbf{x} - \mathbf{pp'} \mathbf{x} = \mathbf{x} - \mathbf{pp'} (\mathbf{x'} + \varepsilon) = \mathbf{x'} + \varepsilon - \mathbf{pp'} (\mathbf{x'} + \varepsilon) \\
= (\mathbf{x'} - \mathbf{pp'} \mathbf{x'}) + (\varepsilon - \mathbf{pp'} \varepsilon) = (\mathbf{x'} - \mathbf{p'} \mathbf{x'}) + (\varepsilon - \mathbf{pp'} \varepsilon) = (\mathbf{x'} - \mathbf{x'}) + \varepsilon - \mathbf{pp'} \varepsilon \neq \mathbf{e} 
\]  

(162)

Summarizing, the calculated scores \( \mathbf{t} \) and residuals \( \varepsilon^0 \) are different from the true underlying scores \( \mathbf{t}' \) and \( \varepsilon \). Therefore, general faults will most likely be distributed over the scores and residuals. This is schematically depicted in Figure 117:
The process fault may differ in direction and size, and therefore will determine the magnitude of $e^0$ and $p\varepsilon$. However, there are directions of $\varepsilon$ (indicated in Figure 117 as the bold arrows) that define special cases as explained before. The directions of these vectors can be considered as rare. It is much more likely that the vector describing a process fault lies in other directions as depicted in Figure 117. It was shown that these vectors would result in the embedded error (Figure 116).

It can be concluded that the complementarity of the SPE and D-chart by the local model is very likely to be destroyed because of the embedded error. Most process faults will therefore be distributed over both control charts. This hampers the diagnostic properties of these charts.

6.4.8 CASE STUDY: TEMPERATURE CONTROL OF A PVC BATCH REACTOR
The data set used for illustration purposes consists of 67 NOC batches where 7 process variables have been measured over 58 time intervals. The main reason to use this dataset is that the correlation between the process variables is fairly predictable
from an engineering point of view. This makes it possible to simulate certain classes of faults that are realistic. The reactor has been illustrated in Figure 118 and the location of the measured process variables in the reactor is given:

PVC is produced on large scale at Pernis-Holland. At the start of a batch, a 50/50 w/w % mixture of water and monomer are mixed together with some stabilisers to affect the particle morphology. When the mixture reaches a specified temperature, initiator is added. Energy is provided to the mixture until the reaction has started to supply enough heat by itself. This polymerisation reaction is highly exothermic and therefore the withdrawal of heat is very important. This cooling is realised by a cooling jacket and condenser. At 70% conversion, the liquid monomer is no longer present and the remaining vinyl chloride is present in the gas-phase. Vinyl chloride is then absorbed from the gas phase and the reaction is killed when 80-90% conversion is reached. The process variables, which are considered to be of most importance, are the measured temperatures. Therefore, temperature control of the reactor is very important. The subset of process variables comprises the cooling system of the reactor. In order to give an impression of these process variables, the trajectories are given in Figure 119:
6.4.9 Process Faults for Batch Processes

In this section the different models are illustrated by examining the detection of a certain type of fault in the Shell PVC process. Besides the relative action signal time (RAST), diagnosing the fault will also be discussed. For this purpose, four different type of faults are constructed. A new batch operated at NOC is denoted as \( x \). The batch \( x \) contains no process faults and consists of the seven process variables measured at 58 time intervals as described in section 6.4.4.2. The measurements from time interval \( k \) until \( k + 6 \) are depicted in Figure 120.

The dotted line represents the average value for all the batches. The solid lines are \( 3\sigma \) limits and the dots represent the measurements of a new independent batch.
Every white bar in this picture represents the NOC value of a scaled process variable measured at a certain time interval $k$. This Figure will be used in the following to explain the construction of 4 different process faults. For every class of faults a description is given how the faults are constructed. The ability of the models to detect the faults will be discussed. A batch that contains a process fault will be denoted in the ongoing of this chapter as $x^*$.  

6.4.9.1 Breakage of the correlation: sensor faults

A distinction can be made for two different situations. The first situations deals with process variables that are integrated in a control loop. In this case a sensor fault will distort the controller action and process control of the reactor is affected.
The other situation is given by process variables that are not part of a control loop. A failure of such a sensor does not always have to disturb the process. The main disadvantage of such a fault is that complete information about the process under consideration cannot be obtained anymore. For simplicity, the case when the process variable is not part of the control loop is considered. Such a sensor fault is illustrated in Figure 121.

At time interval \( k + 2 \) the sensor measuring process variable five fails indicated by the black bar. The remaining process variables do not respond to this failure since the broken sensor is not incorporated in a control loop. At time interval \( k + 2 \) itself the correlation between the process variables is broken.

**6.4.9.1.1 Construction**

Since the action signal is defined as three consecutive points outside the control limits, the actual sensor fault was constructed slightly different as depicted in Figure 121. That is, the fault lasted four time intervals from \( k \) until \( k + 3 \) instead of one. The original value of the process variable was multiplied by a factor \( \delta \) and was tuned such that the process variable just exceeds the \( 3\sigma \) limits from Figure 119 and sometimes stays within these limits. Note that such faults would have been undetected using univariate charts like the ones in Figure 119. The sensor fault was constructed at
different periods during the batch run. This procedure is repeated for every process variable. For example, the first wrong batch $x^*$ contains a sensor fault for process variable number one at time interval 10 until 13 where the value for $\delta$ is 1.01. The actual values for $\delta$ and relevant time intervals are given in Table 14.

Table 14

<table>
<thead>
<tr>
<th>j = 1</th>
<th>$\delta$</th>
<th>Global model</th>
<th>Evolving model</th>
<th>Local model</th>
</tr>
</thead>
<tbody>
<tr>
<td>k = 10 : 13</td>
<td>1.01</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 30 : 33</td>
<td>1.004</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 50 : 53</td>
<td>1.0266</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>j = 2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k = 10 : 13</td>
<td>1.01</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 30 : 33</td>
<td>1.0045</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 50 : 53</td>
<td>1.03</td>
<td>-</td>
<td>$SPE$-chart</td>
<td>-</td>
</tr>
<tr>
<td>j = 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k = 10 : 13</td>
<td>1.01</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 30 : 33</td>
<td>1.0037</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 50 : 53</td>
<td>1.0265</td>
<td>-</td>
<td>$SPE$-chart</td>
<td>-</td>
</tr>
<tr>
<td>j = 4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k = 10 : 13</td>
<td>1.18</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 30 : 33</td>
<td>1.055</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 50 : 53</td>
<td>1.35</td>
<td>-</td>
<td>$SPE$-chart</td>
<td>-</td>
</tr>
<tr>
<td>j = 5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k = 10 : 13</td>
<td>1.18</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 30 : 33</td>
<td>1.06</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 50 : 53</td>
<td>1.38</td>
<td>-</td>
<td>$SPE$-chart</td>
<td>-</td>
</tr>
<tr>
<td>j = 6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k = 10 : 13</td>
<td>1.45</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 30 : 33</td>
<td>1.7</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 50 : 53</td>
<td>6.15</td>
<td>-</td>
<td>$SPE$-chart</td>
<td>-</td>
</tr>
<tr>
<td>j = 7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k = 10 : 13</td>
<td>1.02</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 30 : 33</td>
<td>1.011</td>
<td>-</td>
<td>-</td>
<td>$SPE$-chart</td>
</tr>
<tr>
<td>k = 50 : 53</td>
<td>1.08</td>
<td>-</td>
<td>$SPE$-chart</td>
<td>-</td>
</tr>
</tbody>
</table>

Detection results of faults that break the correlation.

From this table it can be seen that a total of 21 wrong batches containing a sensor fault are simulated.
6.4.9.1.2 Results
For every fault, it was checked i) which model detected the fault first and ii) in which of the two control charts. The detection of the sensor faults gives a very consistent outcome. The global model did not detect any fault. This model seems to suffer from the embedded error where the fault is spread over the scores and residuals. For all the faults at time interval 10 – 13 and 30 – 33, the faults are detected by the local model in the SPE-chart. The faults at time interval 50 – 53 are detected in the SPE-chart by the evolving model except for the sensor fault for process variable one. It must be stressed that the moment of detection in the SPE-chart is very close for the local and evolving model at the time intervals 50 – 53. A more realistic sensor fault would have been detected in the SPE-chart by both models with the same RAST.

It can be concluded that the local model is well suited to detect breakage’s of the correlation due to e.g. sensor faults over the whole batch run. These faults are detected exclusively in the SPE-chart enhancing the diagnostic properties of this model. The global model is not capable to detect the small sensor failures fast enough.

6.4.9.2 Instantaneous intensification of the correlation
One of the main advantages of using multivariate methods instead of univariate methods is the fact that multivariate methods can be utilized to model the correlation between the individual process variables. Intensification in this correlation can be caused by an upset in the reactor. As discussed in Section 2, the variance of the process variables increases but the correlation between the process variables remains intact (not the case for e.g. sensor failures). An example is given in Figure 112. At time interval $k + 3$ the process variables significantly change according to the correlation between the process variables. This may be referred to as an instantaneous intensification of the correlation since it only occurs in one point in time. This type of fault will be further discussed in the following.

6.4.9.2.1 Construction
The construction of the process fault has some specific properties. First, the new batch $\mathbf{x}$ from Figure 120 is monitored and for each time interval $k$ the scores $\mathbf{t}_k^*$ and residuals $\mathbf{e}_k^*$ are calculated. To simulate the faulty batch at time interval $k$ from Figure 112 (denoted by $\mathbf{x}_k^*$), the following calculation is used:
\[ \mathbf{x}_k^* = (\delta + 1) \cdot \mathbf{t}_k^* \mathbf{P}_k + \mathbf{e}_k^* \] (163)

Here, \( \delta \) is chosen in such a way that the raw process variables from Figure 119 stay within or move slightly outside the 3\( \sigma \) limits depending on the time interval. Then,\( \mathbf{x}_k^* \) is re-scaled using the scaling constants from the NOC data. Note that \( \mathbf{x}_k^* \) is constructed specifically to keep the correlation constant by using the loadings \( \mathbf{P}_k \).

If this batch \( \mathbf{x}_k^* \) is to be monitored using a local model, the embedded error is zero. This is true because \( \mathbf{e}_k^* \), \( \mathbf{P}_k \) and \( \mathbf{x}_k^* \) are simultaneously calculated. Projection of \( \mathbf{x}_k^* \) to the loadings \( \mathbf{P}_k \) in order to calculate the scores leads to:

\[ \mathbf{x}_k^* \mathbf{P}_k = (\delta + 1) \cdot \mathbf{t}_k^* \mathbf{P}_k + \mathbf{e}_k^* \mathbf{P}_k \]
\[ = (\delta + 1) \cdot \mathbf{t}_k^* \mathbf{P}_k + \mathbf{e}_k^* \mathbf{P}_k \]
\[ = (\delta + 1) \cdot \mathbf{t}_k^* \mathbf{I} + 0 = (\delta + 1) \cdot \mathbf{t}_k^* \] (164)

Which show that the embedded error is zero. As a result, the fault at time interval \( k \) is expected to be detected exclusively in the D-chart by the local model. For the same reasons as the sensor fault, the fault lasted four time intervals instead of just one from \( k \) until \( k + 3 \). The value of \( \delta \) is chosen in such a way that the process variables from Figure 119 stay within or move slightly outside the 3\( \sigma \) limits depending on the time interval. The actual values for \( \delta \) and relevant time intervals are given in Table 15.

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \delta )</th>
<th>Global model</th>
<th>Evolving model</th>
<th>Local model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k = 1 : 4 )</td>
<td>4</td>
<td>-</td>
<td>SPE-chart</td>
<td>-</td>
</tr>
<tr>
<td>( k = 10 : 13 )</td>
<td>6.5</td>
<td>-</td>
<td>SPE/D-chart</td>
<td>D-chart</td>
</tr>
<tr>
<td>( k = 20 : 23 )</td>
<td>10</td>
<td>SPE-chart</td>
<td>SPE-chart</td>
<td>D-chart</td>
</tr>
<tr>
<td>( k = 30 : 33 )</td>
<td>8</td>
<td>SPE-chart</td>
<td>SPE-chart</td>
<td>D-chart</td>
</tr>
<tr>
<td>( k = 40 : 43 )</td>
<td>14.5</td>
<td>SPE-chart</td>
<td>SPE-chart</td>
<td>D-chart</td>
</tr>
<tr>
<td>( k = 55 : 58 )</td>
<td>4.47</td>
<td>-</td>
<td>-</td>
<td>D-chart</td>
</tr>
</tbody>
</table>

Detection results of faults that intensify the correlation

From this table it can be seen that a total of 6 wrong batches containing a change of the correlation are constructed.
6.4.9.2.2 Results

For all the faults, except for time interval 1 – 4 that gave no detection, the fault was detected uniquely in the $D$-chart for the local models. This was expected since the fault is constructed in such a manner that the embedded error is zero. Apparently $\delta$ is not large enough to generate a signal at time interval 1 – 4.

The faults at time interval 1 – 4, 10 – 13, 20 – 23, 30 – 33 and 40 – 43 were detected by the evolving model in the $SPE$-chart. At time interval 10 – 13 also the $D$-chart detected the fault. The global model detected only the faults at time interval 20 – 23, 30 – 33 and 40 – 43 in the $SPE$-chart. The fault at time interval 55 – 58 goes unnoticed by both models. The detection of these faults mainly in the $SPE$-chart for the global and evolving models can be expected from the Theory Section because the overall correlation pattern is broken by definition and such faults are generally picked up by the $SPE$-fault.

Thus, it can be concluded that a fault, which intensifies the instantaneous correlation between the process variables, at a certain time interval, is not always detected by the models. For the global and evolving models this fault is detected in the $SPE$-chart, sometimes followed by a signal in the $D$-chart.

6.4.9.3 Realistic intensification of the instantaneous correlation between the process variables

This fault resembles a more realistic fault as compared to the previous fault. It was assumed in the previous fault that all process variables react instantaneously to a process upset. This is only possible if the dynamics of the process are fast or the time between the process measurements is sufficiently large to let all process variables respond. A more realistic situation is depicted in Figure 122.
At time interval $k$ the process is in-control. At time interval $k + 1$ three process variables are somehow affected by a process upset. This does not intensify the instantaneous correlation at $k + 1$ but breaks it instead. At time interval $k + 2$ almost all process variables respond to the process upset. Now, at time interval $k + 3$ all process variables are affected according to the correlation. This time interval resembles the instantaneous change of the correlation previously described. After this time interval the process returns gradually to the NOC region at $k + 7$. So, this fault breaks the correlation first until $k + 3$ where the process variables behave according to the correlation. After $k + 3$ the instantaneous correlation is broken again until the process is back on target.

6.4.9.3.1 Construction
The simulated fault describes the heating of the reactor as a result of an unwanted side reaction. The fault has a duration of 7 time intervals and reads as:
Time $k$: the temperatures and pressure in the reactor are increasing as a result of an unwanted side reaction that generates extra heat.

Time $k+1$: the temperatures and pressure have increased further. As a result of the controller action, the amount of refrigerant is increased. This decreases the temperature of the refrigerant medium and this is detected in the temperatures of the inlet and outlet of the cooling jacket.

Time $k+2$: the fault remains and the temperatures in the reactor and the pressure are becoming dangerously high. The amount of refrigerant water is further increased to cool the reactor.

Time $k+3$: the temperatures and pressure in the reactor have reached a maximum now. The control action of the opening valve is adjusted to further increase the amount of refrigerant.

Time $k+4$: the temperature in the reactor is decreasing again as a result of the control action. Therefore, the amount of refrigerant is decreased. This slightly increases the temperatures of the inlet and outlet of the cooling jacket.

Time $k+5$: The temperatures and pressure in the reactor have reached their nominal values while the temperature of the cooling medium still increases. The control action of the opening valve is adjusted to decrease the amount of refrigerant further.

Time $k+6$: the temperature drift in the reactor has been successfully controlled and the process is back on NOC.

The construction of this fault is almost similar to the fault described in the previous section (Equation 163). The reason to simulate the fault at time interval $31 - 37$ was supported by the fact that loadings $P_k$ of all variables are very similar on these intervals and only differ in intensity. The fault is constructed in such a way that not all the process variables contribute to the disturbance at some time intervals. Also, the magnitude of this contribution is not equal for all process variables but can be set individually. Therefore, the loadings of the local model at time interval $k$ are constructed:

$$\hat{P}_k^o = \Delta \hat{P}_k$$

(165)
where $\Delta (J \times J)$ is a diagonal matrix. For example, the matrix $\Delta$ at time interval $k = 31$ where only the temperatures and pressures (process variables 1, 2, 3, and 7) are increasing looks like:

$$
\Delta = \begin{bmatrix}
13 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 13 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 13 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 12 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

Then, the batch $\hat{x}^*_k$ is constructed using Equation 163 and monitored as described in the previous section. This batch is projected using all three models.

### 6.4.9.3.2 Results

For the local model this fault is detected in the $D$-chart only (RAST = 3). This is a promising result considering the fact that not all the process variables take place in the intensification of the modelled correlation at time intervals 31 and 36. In other words, although the instantaneous correlation is broken and the embedded error is not zero, the fault leads to detection only in the $D$-chart. Probably the effect of the embedded error is very small. This enhances the diagnostics of this fault that clearly is more related to an increase of variance of constant correlation instead of a breakage of the correlation. For the time evolving model as well as the global model the fault is detected in the $SPE$-chart (RAST = 3) while the $D$-chart does not detect at all. This detection in the $SPE$-chart is just as fast as for the detection in the $D$-chart for the local model. These types of fault that break the overall correlation structure seemingly manifest themselves in the $SPE$-chart for the global and time evolving model. The RAST for all the models is the same. The fact that the process fault was detected only in the $D$-chart by the local model enhances the diagnostic properties of such models. Fault detection in this chart can be ascribed due to an intensified correlation, which is indeed true at a certain time interval.
6.4.9.4 Small change breaking the correlation over time

This typical fault describes a minor drift of the process where the instantaneous correlation among the process variables remains intact, however the correlation over time is broken. This time correlation is implicitly modelled and captured in the time evolving and global models. Therefore these models are believed to detect these type of faults. A figure representing the disturbed variables is not helpful in understanding and is therefore not shown like the other faults.

6.4.9.4.1 Construction

The fault that describes this behaviour is modelled as follows. In contrast to the earlier constructed faults, the latent variables from the global models instead of the local are used to construct this faulty batch because the time correlation is captured by a global model. A normal operating batch \( x \) is projected onto the model resulting in scores \( t^* \) and residuals \( e^*_k \). These will be used to construct a fault as described above by calculating:

\[
\delta_k \cdot \tilde{P}^p \tau + e^*_k, \quad \text{where} \quad \delta_{k-1} < \delta_{k-2} < \ldots < \delta_{k-R}
\]  

\( \delta \) is steadily increasing over time with small steps of 0.3. After time interval 30 the process is assumed to be in-control again. At time point 2, the current part (cp) of \( \tilde{P} \) differs from the part at time point 1. Because of this, the correlation in time is changing. The above described procedure is repeated for every time interval until \( k = 30 \). The direction within the model plane in which the scores will evolve is kept constant but the model plane changes at each time interval.

\[ \text{-296-} \]
6.4.9.4.2 Results

The fault was detected by the time evolving model and global model in the SPE-chart both with an RAST of 21 time intervals. It is expected that this fault breaks the correlation between the process variables when using global and time evolving models.

The local model also detected the fault in the D-chart with an RAST of 21 time intervals. This fault is detected in the D-chart because the loadings of the local model $\tilde{P}_y$ at time interval 19, 20 and 21 are very similar to corresponding current parts of the loadings $\tilde{P}_y^{ep}$ from the global model. It was already shown that these faults are detected in the D-chart by the local model.

It can be concluded that faults, which break the correlation over the time, while at the same time the correlation between the process variables is maintained, are detected by all models with the same RAST. The local model is also capable of detecting this fault although time correlation is not explicitly modelled for this type of model.

6.4.10 Conclusions

In this chapter the diagnosing performance of the SPE and D-chart based on local, global and evolving models is assessed. The required behaviour of the process variables to generate either an pure SPE or D-signal is thoroughly discussed. These pure signals are represented by very specific process conditions that are not realistic in a real process. The conclusions can be split into two parts concerning the control charts and the models.

For the control charts, a summary of the results is given in Table 16:

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|}
\hline
 & 6.4.10.1.1 On-line monitoring & 6.4.10.1.2 Post batch analysis \\
\hline
Global & - & SPE & D \\
Evolving & SPE & SPE & D \\
Local & SPE & D & SPE & D \\
\hline
\end{tabular}
\end{table}

Detection of Realistic Process Faults

\begin{tabular}{|l|c|c|}
\hline
 & 6.4.10.1.1 On-line monitoring & 6.4.10.1.2 Post batch analysis \\
\hline
Global & SPE & - \\
Evolving & SPE & - \\
Local & SPE & D & - \\
\hline
\end{tabular}

Complementarity properties for the various models.
It can be seen from the upper part of this table that pure SPE and $D$ signals can only be constructed for local models in the case of on-line monitoring. This is favourable for diagnosing faults. Considering the post-batch analysis, pure SPE and $D$ can be constructed for all three models. However, the process conditions representing these pure signals are very unlikely to occur. This results in embedded error that complicates the diagnostics of the control charts. It can be seen from the lower part of Table 16 that despite the construction and embedded error problem, the local model still detects faults in either the $D$ or SPE-chart. This is not true for the global and evolving model where realistic process faults are detected in the SPE-chart. This hampers the diagnostic properties of these control charts.

The conclusions concerning the models are as follows. It is better to use local or time evolving models for on-line monitoring of batch processes instead of global models. The local models are able to detect pure SPE and $D$-faults, where the time evolving model only can detect pure SPE-faults. Therefore, in terms of diagnosing capabilities the local model is slightly favoured.
6.5 Correlations in batch process data*

6.5.1 SUMMARY
A repeatedly returning discussion in modelling batch process data using component models is the effect of the matricizing direction of the three-way array $X$ and the selection of a model for capturing the auto-and cross-correlations of the process variables. To study the capture of the correlations by the models, a benchmark dataset is matricized and modelled accordingly. Next, an independent batch run is fitted to the models and the recovery of the correlations is analysed. The results show that all models are well able to capture the correlations in the data and the matricizing direction only affects the distribution of the correlations among the model parameters.

6.5.2 INTRODUCTION
Since the introduction of multivariate statistical batch process monitoring by Nomikos and MacGregor in 1992 (J.F. MacGregor & P. Nomikos [22]), there has been a lot of research done in this field. This has led to numerous applications and many extensions of the original proposed ideas (S. Rannar et al. [31], S. Wold et al. [53], D. Neogi & C.E. Schlags [44], D.J. Louverse & A.K. Smilde [38], J.A. Westerhuis et al. [42], A.K. Smilde & H.A.L. Kiers [39]).

A discussion, which regularly returns is the effect of the matricizing directions of the three-way array $X$ ($I \times J \times K$) and which model to use to capture the auto- and cross-correlations present in the data (T. Kourtì [18]). In the following, a preliminary assessment is given of correlations in batch process data and how the different models capture these correlations.

Batch processes can be thought of as non-linear, non-stationary multivariate time series of limited length. By removing the average trajectory of a process variable during a batch run, the main non-linear behaviour of the process data is removed and the remaining variation around this average batch trajectory can be modelled using a linear model such as PCA. The model captures auto- and cross-correlations present in

the data. In statistical batch process monitoring the aim is to detect deviations from normal operating conditions such as process upsets, sensor failure or impurities. This reflects itself in correlation breakage or intensification of the existing correlations. With an intensified correlation is meant a process, which has similar correlations as under normal operating conditions, with the difference, that the correlations have a much higher or lower magnitude.

The procedure is to collect a historical reference set of batches from which it is known that these batches operated under normal operating conditions (NOC). Next, a model is constructed of the historical dataset, which captures the correlations under normal operating conditions. After that, a new batch is monitored by checking if the new batch has similar correlations as the reference data. Therefore, understanding how the models capture these correlations is relevant.

After removal of the average trajectory of the process variables, the main non-linearity in the data is removed. Now, the data can be seen as non-stationary linear multivariate time series with a zero mean. From this, auto and cross-correlations between the process variables can be computed and analysed. The recovery of correlations by the models is studied by fitting new data to the models, compute auto and cross-correlations from the fitted data and compare these with the auto- and cross-correlations from the original data.

6.5.3 Theory

6.5.3.1 Structure of the data
A single batch run is represented by a matrix $X (K \times J)$ where $J$ process variables are measured during $K$ time intervals. Since a batch process is recipe driven, it is assumed that a batch behaves according to a certain trajectory. Batches that are repeated according to the recipe are stacked in a three-way array $X (I \times J \times K)$ where $I$ is the number of batches. Having three-way data in $X$, there are three possible ways to arrange the data. The first is to keep the three-way array as is and use multi-way models such as PARAFAC and Tucker3. Next, there are two sensible directions of matricizing the three way array (P. Nomikos [21]), leading to two different Tucker1 models. The first is matricizing in the batch direction and forming matrix $X_{NM} (I \times JK)$. The second is matricizing in the process variable direction and form the matrix $X_{PI} (KI \times J)$. Note that the mean time trajectory of all process variables has been
removed before hand. This makes matricizing to $\mathbf{X}_{PV}$ rather different from the approach described by S. Wold et al. [53], where the mean trajectories are not removed. As a result, the non-linearity is still present in the data. This will have an effect on the correlations in the data (T. Kourt ’ [18]). Nevertheless, both approaches have shown to work well in practice (E.N.M. Van Sprang et al. [41]).

In order to compare the different models, a similar preprocessing is required for each model. For that reason, the average batch trajectory of the process variables is removed prior to matricizing and modelling. That is, each column in $\mathbf{X}$ is mean centered and scaled to unit variance (auto scaling).

6.5.4 Correlation in batch process data

To understand how the models capture the correlations, an overview of correlations present in the data is given. Within a batch, there are time dependent correlations due to the underlying chemistry of the process. Furthermore, seasonal influences or process history may cause correlations between the batches e.g. carry-over effects. The following correlations can be distinguished (see Figure 123):

1. Correlation within a batch
   a) Autocorrelation of process variables
   b) Correlation between process variables at the same time interval
   c) Cross-correlation between process variables over time

2. Correlation between batches
   d) Correlation between batches

Figure 123
Correlations in batch process data.
1. **Within batch run correlations**
   a) Autocorrelation of a process variable over time
   b) Correlation between the process variables at the same time interval
   c) Cross-correlation between the process variables over time

In addition, in the three-way array there are between correlations such as:

2. **Between batch run correlations**
   d) Correlation between the batches

Since batch processes are recipe driven, it is expected that all the batches have a similar within correlation structure over time (1). In principle, since each batch is considered as an independent sample from a population of batch runs, there should be no correlation between the batches.

**6.5.5 MATRICIZING AND MODELLING THE THREE-WAY ARRAY**

The two sensible directions of matricizing also define two extremes in modelling the correlations in the data. The first way is matricizing in the batch direction and forming matrix $X_{NM}$. The second ways is matricizing in the process variable direction and form the matrix $X_{Pv}$. The arrangement of the three-way array also defines the distribution of the auto- and cross-correlations in the data and therefore, also defines the distributions of the correlations among the model parameters. In here, the distributions of the correlations among the model parameters of three modelling strategies will be discussed: the global model, the local model and the $pv$-model. The first strategy is proposed by P. Nomikos & J.F. MacGregor [6]. With this strategy $X$ is matricized in the batch direction and a model is constructed on $X_{NM}$.

$$X_{NM} = T_{NM} P_{NM}^T + E^{l_e/k} \quad (168)$$

This model forces all within batch-run variations (1.a, 1.b, 1.c) into the model loadings $P_{NM}$ ($J K x R$). The differences between the batches are captured by the scores $T_{NM}$ ($I x R$).
A geometrical representation of the global model is given in Figure 124. In this figure, it is shown that a point in the JK dimensional space represents a batch (Batch 1, Batch 2 and Batch 3). The model plane ($P_{NM}$) captures the principal direction of the process, which is defined by the underlying physics and chemistry. It is easy to understand that the differences between the batches are given by the projections (reflected in the scores $T_{NM}$) of the batches on the model plane. Actually, in process monitoring, it is tested how much within variation is present in a new batch against a historical reference. If the new batch has a similar within variation as the reference batches, the new batch will be in the model plane somewhere between the reference batches. The position in the model plane is monitored using a $D$-chart (P. Nomikos & J.F. MacGregor [6]). If a batch does not have a similar within variation, the correlations break down and the batch will deviate from the model plane. The level of disimilar within variations is monitored with Q/SPE-chart (P. Nomikos & J.F. MacGregor [6]).

If this model is used for on-line process monitoring, it is well known that assumptions have to be made about the future unknown observations (P. Nomikos & J.F. MacGregor [6]). As a result of these assumptions, the projection of a batch will "walk" a certain trajectory on the plane during the evolution of the batch (see the running batch in Figure 124). Moreover, because assumptions are made, the batch data will not have a similar within variation compared to historical reference batches and the distances to the model will vary a lot for each time interval $k$. To overcome this feature, a historical reference distribution is estimated for each time interval $k$. As
the batch approaches the end of its trajectory, the projection reaches its true point on
the model plane.

The second strategy is suggested by P. Nomikos [21] and further elaborated
by H.-J. Ramaker et al. [125]. With this strategy the process is approximated by $K$
independent local models and does not include previous process measurements in the
model. With batch processes, it is assumed that each batch run is operated in an
identical manner according to a recipe. Therefore, each batch run has the same
underlying phenomena as a driving force implying that the present auto and cross-
correlations in the data are not arbitrary but fixed for each time interval. Given that
the correlations in the data are fixed, also the direction of the process for each time
interval is fixed and can be captured with a multivariate model valid for time interval
$k$.

$$X^{ij} = T_k P_k^T + E^{ij} \quad (169)$$

The auto-and cross correlations (1.a, 1.b and 1.c) are implicitly captured by the plane
direction of the loadings $P_k$. A geometrical interpretation is given in Figure 125. The
underlying dynamics will determine the principal direction of the process for time
interval $k$ and the model plane will be different for each time interval.

![Figure 125](Correlations captured by a local model.)
The last strategy is matricizing in the batch direction and model $X_{pv}$ with a bilinear model. This was suggested by P. Nomikos [21] and elaborated by Wold and co-workers (S. Wold et al. [53]). This model will be referred to as the pv-model.

$$X^{K_{df}} = T_x P_{x}^T + E^{K_{df}}$$

(170)

This strategy uses a common model $P_{pv} (J \times R)$ consisting of the overall correlations between the process variables. The behaviour of the process is now expressed on a common base defined by the model $P_{pv}$. That is, the path followed by the projections determines the dynamic behaviour of the process. This is different from the other approaches where the model plane captured the dynamics of the process. A geometric representation is given in Figure 126. A new batch follows a path in the model plane. In process monitoring, the historical batches follow a similar path in the model plane. Next, control limits are placed around this trajectory and new batches are compared if they follow the same path.

Figure 126
Correlations captured by a pv-model.
6.5.6 CASE STUDY
To show that all previously mentioned models capture the correlations in the data, a benchmark dataset is modelled using the previously mentioned models. Next, a new independent batch is fitted and the obtained model parameters are used to reconstruct the original data. From this, auto- and cross-correlations of the process variables are computed and compared with the correlations computed from the original data.

6.5.6.1 Description of the data
The dataset is described by P. Nomikos & J.F. MacGregor [126] and consists of 50 batches for which 9 process variables are measured during 200 time intervals. The following process variables are measured: two flow-rates of the monomers, four temperatures, the density of the latex, the conversion and an estimate of the rate of energy release.

6.5.6.2 Preprocessing
To compare the models, the main non-linear behaviour is removed by column centering in the $I$ mode and scaled to unit variance by dividing each column by its standard deviation after centring. Next, the data cube is matricized in the desired direction and modelled by the three models. The number of components for each model is determined using cross-validation and scree-plots (S. Wold [137]).

6.5.6.3 Results
Because the results were similar for all batches, one batch is selected for analysis. As a result, 49 batches are used to construct a reference model and the selected batch (normal batch) is used for analysis (batch 43). Furthermore, to avoid an abundance of results, a selection is made of two process variables based on their different nature of correlation during the batch run.

For this analysis, the normal batch is preprocessed and rearranged accordingly. Next, the normal batch is fitted to the different models. The first step is to reconstruct the original signal from the models.
The explained variance by each model is given in Table 17. The global model explains a relative small amount of the variance in the data compared to the pv-model and local model. This is easy to understand since the pv-model and local model use much more model parameters to model the batch data. Using a similar number of parameters with a global is not possible since this will lead to over fitting.

The low amount of explained variance by the global model does give information about the nature of the process data, namely that the data does not have an ideal batch structure. That is, the underlying common phenomena suffer from relative large variations, which are different from batch-to batch.

The recovered signals after fitting the new batch are given in Figure 127. Each graph represents one process variable with the results from the models and the original signal. From the reconstructed signals auto-and cross-correlations are computed.
The autocorrelations of two process variables are given in Figure 128 and Figure 129: Process variable 2 (pv2) and process variable 6 (pv6). Process variable 2 is a flow rate and is expected to be less autocorrelated than process variable 6, which is a temperature measurement. From top to bottom, the graphs correspond to the autocorrelation of original (preprocessed) signal, the local model, the pv-model and the global model.
The cross correlations between the two variables are given in Figure 130. In the original signal of process variable 2 an autocorrelation is observed for the first seven lags. The same correlation structure is observed for the fitted local and pv-model. The global model shows a more correlated structure compared to the original correlation structure, implying that the reconstructed signal is more smoothed than the original. The autocorrelation for process variable 6 is well captured by all models except for the global model, which shows a slightly different correlation structure.
Figure 130

Cross-correlation of process variable 2 and 6.

The cross-correlation between the two process variables reveals that the local and pv-model capture more or less the same correlation structure whereas the global model shows a more smoothed structure which is expected since the reconstructed signal of process variable 2 is smoothed by the global model. The smoothed signal obtained with the global model can simply be understood since the model considers the entire process at once whereas with the W-model and local model the process is approximated by successive multiple models. The global model captures the principal directions in a JK-dimensional space. That is, only the main directions are considered. Because of this smoothed loadings are obtained and certain local variations are less described than others. This is also reflected by the explained variance in Table 17.
In short:
- The global model captures the general direction of the auto-and cross-correlations in the model loadings. However, because of its construction and nature of the data, it has a smaller amount of explained variance.
- With the pv-model, the correlations are captured by a combination of scores and a common model plane. The path of scores on the model plane represents the dynamics of the process.
- The local model is a fingerprint of the process correlations at time interval $k$.

### 6.5.7 CONCLUSIONS

It is shown that the data arrangement or the matricizing direction of $X$ determines the distribution of the correlations among the model parameters. Hence, there is no loss of correlation structure as a result of the model choice. However, the important difference between the different modelling strategies is the number of acceptable model parameters necessary to capture the underlying phenomena. Therefore, it is a priori not clear which model will give the best monitoring performance.
FUTURE DIRECTIONS*

In the second chapter of this thesis a framework (the I.T.A.-trajectory) is presented for statistical batch process monitoring. Although this framework is far from complete, the authors of this thesis believe that it is necessary to present the ideas of statistical batch process monitoring in such a manner. The implementation of similar statistical tools in total quality programs, own their success because the techniques are encapsulated in such management programs. It is often the plant manager that needs to be convinced of the advantages that statistical batch process monitoring has to offer. It is not the mathematics that are interesting for the plant manager, but the advantages that can be reached by using the philosophy of statistical batch process monitoring. Presenting statistical batch process monitoring in a framework such as the I.T.A-trajectory makes it easier to communicate with non-technicians.

Another important issue is the lack of proven examples in industry. If rough numbers would be available for the savings that statistical batch process monitoring can accomplish for every kilogram of produced product, another argument can be brought into the discussion with the plant manager.

In the initial phase of the I.T.A-trajectory, batch data is collected. In the past, collecting batch process data was a serious stumbling block. Many potential applications failed because of the time consumption that data collection took. However, it is our experience that through the years of this research, this barrier seems easier to take. The platforms used in production plants are increasingly improved for proper data warehousing. Nowadays, it is quite easy to log in from a remote location to the main database system, and retrieve the batch data in well-organized data sheets. This trend definitely increases the potential of statistical batch process monitoring and is in accordance with the expectation that the availability of data will exponentially increase in the next decade.

The third chapter deals with synchronizing batch data prior to modelling. The problem of batches having unequal run length is evident for almost all datasets. There are hardly any industrial applications where the operation of batches guarantees equal run length. The results for dynamic time warping showed that, dynamic time warping,
as a synchronizing tool seems to work well over a broad range of applications. However, the issue of on-line synchronization is not crystallized and is still in a research-stage. Also, the proposed method for synchronizing a set of batches has the disadvantage that intensities of the signals are changed. It is unclear how this affects the performance of SBPM models. Ideas exist to extent the dynamic time warping-algorithm from warping two pairs of signals to warping \( n \)-pair of signals at the same time. It can be concluded that the issue of synchronization is not only hot in the area of statistical batch process monitoring, but also in other fields of chemistry where the same problems can be recognized. Further research in this area is therefore worthwhile to undertake.

Black models for statistical batch process monitoring have been discussed and existing approaches are compared to new approaches. Using component models for statistical batch process monitoring gives rise to many new ideas for other black models. Besides the suggested approaches in this thesis, many more black models were used along the research period. Still, there is more research required to investigate the applicability of such models and moreover, their properties.

It became very clear that many complex questions came up from using relatively ‘simple’ black models. Therefore, it is important to understand ‘simple’ models first before more complex models are used.

The ideas about grey models and their application are discussed in chapter five. It is very appealing to combine the advantages of black and white models into grey models. Although the application of real first principle models for batch process monitoring is not considered in this thesis, some aspects of such approaches are captured in the application of grey models. Statistical batch process monitoring is typically a data-driven approach that originates from the field of chemometrics and data analysis. This might clarify the fact that engineers, at first sight, are hesitant to apply these techniques. The concept of grey models definitely brings the views of engineers and chemometricians together. It is shown that grey models can be very advantageous especially for diagnosing purposes. It is expected that the application of grey models will gain much interest in fields such as bio-technology where interpretation of the model parameters are highly desired.
Chapter six of this thesis discusses the mechanisms and statistical properties of the control charts derived from various SBPM models. Much effort is put into assessing the performance for the various models by introducing performance indices and studying the statistical characteristics. Studying the performance of the control charts is mainly induced because of the poor detection performance of the D-chart. It is common to quantify the performance of SBPM models by the detection power of faulty batches. In this research, the performance is based also on another important chart characteristic: the type I error. The results for the type I error show that the performance can be problematic. More research is required to set the control limits according to the type I error to overcome this problem. It seems that the statistics and mechanisms for this control chart are not without problems. It is our feeling that another statistic for the scores is required. Therefore, testing the performance of control charts needs to be based on either more robust statistics or non-parametric tests.

Testing the underlying statistical assumptions highly depends on the validity of the used statistics, which can be quite conservative. The fact that drawing such conclusions about statistical properties are related to the number of samples in the dataset is not hopeful considering the fact that statistical batch process monitoring models suffer from this. Therefore, it would be interesting to investigate batch-monitoring tools that require only few samples. Ideally, models are developed on a laboratory scale, scaled up to industrial reactors and then applied to various plants worldwide.

An important issue in quality improvement programs is the relation of the process operation with the end product quality. Hence, the relation between process performance and the end product specifications, expressed, as capability indices should be studied. The focus should be on the product observability of the system and the use of multivariate methods in combination with capability indices.

Another important issue in process monitoring is fault diagnosis. Although contribution plots are very successful in diagnosing process upsets, the way of diagnosing process upsets can be improved by fault reconstruction. That is, understanding how historical, known faults can be used to diagnose future batch runs.
Batch processes are complex with respect to the correlation that exists between batches and process variables over time. It is very interesting to study these correlations and how these correlations are being captured by statistical batch process monitoring models. However, because of the complexity of real process data and lack of samples, detailed and extensive simulations of batch processes with prior information about these correlation structures are required.

Batch processes are commonly monitored using typically engineering variables. In this thesis, various spectroscopically monitored batch processes are studied. Spectroscopic measurements have the ability to gain insight information about the chemistry of the process. An interesting question is how this spectroscopic data can be used together with physical or/and biological data. Combining different sources of process measurements gives a total view of the process. Such data fusion is believed to be necessary in order to relate process variation to end-product quality.

Despite all the questions, it is clear that statistical batch process monitoring is very useful for monitoring batch processes and is therefore well suited to be a part of a quality improvement program. Various applications from literature clearly show the excellent performance of statistical batch process monitoring in detecting process upsets.

However, it is now important to evaluate practical applications of those models that are well understood and continue to develop models that are better capable of dealing with the disadvantages that are inherent to batch processing such as changing recipes and a relative low number of historical samples.
♦ SYMBOLS AND NOTATION

AST  Action Signal Time
CL   Center Line
CUSUM Cumulative Sum
cp   current part
D    $D$-statistic
$D$-chart Control chart for the $D$ statistic
DCS  Distributed Control Systems
EWMA Exponentially Weighted Moving Average
ITA  Initial $\rightarrow$ Training $\rightarrow$ Application trajectory
LCL  Lower Control Limit
MSPC Multivariate Statistical Process Control
NM   Approach by Nomikos & MacGregor (1994)
NOC  Normal Operating Conditions
PCA  Principal Component Analysis
pp   past part
Q-plot Plot for the Q-statistic
RMW  Approach by Rännar, MacGregor & Wold (1998)
SCADA Supervisory Control And Data Acquisition
Scatter plot Plot for the scores
$SPE$-chart Control chart for the $SPE$ statistic
SPC  Statistical Process Control
$SPE$ Squared Prediction Error
UCL  Upper Control Limit
WKFH Approach by Wold, Kettaneh, Fridén & Holmberg (1998)
**Definitions:**

- **B**: consensus matrix
- **b**: block scores vector
- **B_k**: consensus matrix at time interval \( k \)
- **b_k**: block scores vector at time interval \( k \)
- **d**: adaptive parameter for RMW approach
- **E**: residual matrix
- **E_k**: residual matrix at time interval \( k \)
- **e**\(^*\)**: model residuals
- **e**: extracted error
- **e_i**: residual vector for sample \( i \)
- **e_k**: residual vector at time interval \( k \)
- **e_{new}**: residuals obtained for a new independent sample
- **F**: \( F \)-distribution
- **g**: weighting factor for the \( \chi^2 \) distribution
- **h**: degrees of freedom for the \( \chi^2 \) distribution
- **i**: \( i^{th} \) batch
- **I**: number of batches
- **j**: \( j^{th} \) process variable
- **J**: number of process variables
- **k**: \( k^{th} \) time interval
- **K**: number of time intervals
- **P**: loading matrix
- **p**: loading vector
- **P_k**: loading matrix at time interval \( k \)
- **p_k**: loading vector at time interval \( k \)
- **R**: number of components
- **r**: \( r^{th} \) component
- **S**: variance-covariance matrix
- **S_k**: variance-covariance matrix at time interval \( k \)
- **T**: score matrix
- **t**: score vector
- **t\(^*\)**: pure scores
- **T\(^2\)**: Hotelling statistic
- **t_i**: score vector for sample \( i \)
- **T_k**: score matrix at time interval \( k \)
$t_k$  score vector at time interval $k$
$t_{wuv}$ scores obtained for a new independent sample
$t_{wuv,k}$ score vector of a new batch at time interval $k$
$X$ two-way data matrix
$X$ three-way data array
$X^{(I \times JK)}$ matrix with size $(I \times JK)$
$x_{(I \times JK)}$ observation vector with size $(I \times JK)$
$x^*$ pure data
$x_a$ Target value or grand mean
$x_i$ observation vector for sample $i$
$X_k$ frontal slice: two-way data matrix at time interval $k$
$x_{k}$ observation vector at time interval $k$
$x_{\text{new}}$ new independent sample
$x^{*{\text{SPE}}}$ / $\tilde{x}^{*{\text{SPE}}}$ / $\hat{x}^{*{\text{SPE}}}$ batch that results in pure $SPE$-signal after projection using local/evolving/global model
$\tilde{x}^D$ / $\hat{x}^D$ / $\bar{x}^D$ batch that results in pure $D$-signal after projection using local/evolving/global model
$\Sigma$ true population covariance
$\hat{\Sigma}$ estimated covariance matrix concerning phase 1
$\Sigma_f$ estimated covariance matrix concerning phase 2
$a$ confidence level
$e^D$ process fault describing pure $D$-signal
$e^{*{\text{SPE}}}$ process fault describing pure $SPE$-signal
$\tilde{e}^D$ / $\hat{e}^D$ / $\bar{e}^D$ process fault for pure $D$-signal constructed with a local/evolving/global model
$e^{*{\text{SPE}}}$ / $\tilde{e}^{*{\text{SPE}}}$ / $\hat{e}^{*{\text{SPE}}}$ process fault for pure $SPE$-signal constructed with a local/evolving/global model
$\chi^2$ Chi-square distribution

-318-
APPENDICES
Appendix A

Is the model $\mathbf{C B}^T$ unique if the columns of matrix $\mathbf{B}$ are fully constrained to equal the pure compound spectral profiles? That is, is it possible to find a non-singular matrix $\mathbf{Q}$, other than a scaling matrix, such that

$$\mathbf{C B}^T = \mathbf{C Q}^{-1} \mathbf{Q B}^T = \tilde{\mathbf{C B}}^T$$

with the restriction that the spectral profiles in $\mathbf{B}^T$ are equal to the spectral profiles in $\tilde{\mathbf{B}}^T$.

Thus, $\tilde{\mathbf{B}} = \mathbf{B Q}^T = \mathbf{B} \mathbf{A}$, where the spectral profiles are preserved only if $\mathbf{A}$ is a diagonal matrix. Since $\mathbf{B}$ is full column rank, premultiplying with $\mathbf{B}^*$ (the Moore-Penrose) gives $\mathbf{B}^* \mathbf{B Q} = \mathbf{B}^* \mathbf{B} \mathbf{A}$, with $\mathbf{B}^* \mathbf{B} = \mathbf{I}$ (J.R. Schott [127]). Hence $\mathbf{Q}^T = \mathbf{A}$, and only scaling differences are allowed.
Appendix B

In the following it is shown that column mean centering for the Tucker3 model does not affect the model loadings in the noiseless case. The column centred matrix $X_c$ is a projection of $X$ which removes the constant terms in the data.

$$X_c = P^\perp X \quad \text{and} \quad P^\perp = (I - 11^T / I)$$  \hspace{1cm} (172)

where $P^\perp$ is an orthogonal projection matrix. Take the noiseless case where $X$ $(I \times JK)$ is modelled using a Tucker3 model given in equation.

$$X = AG(C \otimes B)^T$$  \hspace{1cm} (173)

and the orthogonal projection of $X$ is given by

$$P^\perp X = X_c = P^\perp AG(C \otimes B)^T$$  \hspace{1cm} (174)

From equation 174 it can be seen that model loadings $B$ and $C$ are not affected by the centering operation.
Appendix C

In general contribution plots are computed to study which process variable has the largest contribution to the D statistic. However, in the current monitoring problem it is not so interesting to know which process variable has the largest contribution since the process variables are the wave channels, but to know which chemical component contributes to the D statistic. The current model is a grey Tucker3 model with \( P, Q, R = 1, 3, 3 \).

The score \( a_k \) is computed by

\[
a_k = x^T_k v (v^T v)^{-1} \quad \text{and} \quad v = \left( G \left( C^T \otimes B^T \right) \right)^T
\]

(175)

Where \( x_k \) (JK x 1) is the observation vector of the whole batch. Furthermore, the "current deviations" approach is used to anticipate for the missing values. The D statistic for time interval \( k \) is computed by

\[
D_k = (a_k - \bar{a}_k) s_k^{-1} (a_k - \bar{a}_k)
\]

(176)

Where \( s_k \) is the variance of score \( a_k \) on time interval \( k \).

\[
D_k = \left( x^T_k v (v^T v)^{-1} - \bar{x}_k^T v (v^T v)^{-1} \right) s_k^{-1} \left( x^T_k v (v^T v)^{-1} - \bar{x}_k^T v (v^T v)^{-1} \right)
\]

(177)

\[
D_k = \left\| s_k^{-1/2} \left( (x_k^T - \bar{x}_k^T) \cdot v (v^T v)^{-1} \right) \right\|^2
\]

(178)

\[
D_k = \left( s_k^{-1/2} \sum_{\mu=1}^{L} \left( (x^T_{\mu} - \bar{x}^T_{\mu}) \cdot v (v^T v)^{-1} \right) \right)^2 \quad \text{with} \quad l = 1..L \quad \text{and} \quad L = K.
\]

(179)

\[
D_k = \left( s_k^{-1/2} \sum_{j=1}^{L} \sum_{l=1}^{L} \left( (x^T_{j\mu} - \bar{x}^T_{j\mu}) \cdot \sum_{r=1}^{R} \sum_{p=q}^{P} g_{pqr} b_{j\mu} c_{lqp} (v^T v)^{-1} \right) \right)^2
\]

(180)

Take \( b_k = s_k^{-1/2} (v^T v)^{-1} \)
\[ D_k = \left( b_k \sum_{j=1}^{l} \sum_{i=1}^{t} \left( \left( x_{ji} - \bar{x}_{ji} \right) \cdot g_{pr} \epsilon_{i,j} b_{j,r} \right) \right)^2 \]  

(181)

With the method of current deviations an observation at time interval \( k \) will be assumed to be valid throughout the rest of the batch run. Because of this, an unrealistic contribution is found for individual compounds. Therefore, the local contribution is computed. The local contribution is computed by means of the local observation vector \( \mathbf{x}_{i,k}(J \times 1) \).

The first component,

\[ \text{contribution}_{k}^{111} = \left( b_k \mathbf{g}_{111} \mathbf{b}_{1}^{T} \left( \mathbf{x}_{i,k} - \bar{\mathbf{x}}_{i,k} \right) \right)^2 \]

(182)

The second component

\[ \text{contribution}_{k}^{122} = \left( b_k \mathbf{g}_{122} \mathbf{b}_{2}^{T} \left( \mathbf{x}_{i,k} - \bar{\mathbf{x}}_{i,k} \right) \right)^2 \]

(183)

The third component

\[ \text{contribution}_{k}^{133} = \left( b_k \mathbf{g}_{133} \mathbf{b}_{3}^{T} \left( \mathbf{x}_{i,k} - \bar{\mathbf{x}}_{i,k} \right) \right)^2 \]

(184)
Appendix D

The batch consistency number (BCN) is a single index summarising the batch behaviour compared to historical batches obtained under normal operating conditions. Since the BCN is a summary for the behaviour of a single batch run, the BCN should be function of the D and Q statistic or \( BCN = f(D,Q) \). The definition of the BCN is given below:

\[ x \text{ is the observation vector and can be written as:} \]
\[ x = \hat{x} + \hat{x}^\perp \quad (185) \]
where \( \hat{x} = \tilde{V}x \) and \( \hat{x}^\perp = (I - \tilde{V})x \) with \( \tilde{V} = V(V^TV)^{-1}V^T \).

The D-statistic is given by
\[ D_i = (a - \bar{a})^T S^{-1} (a - \bar{a}) \frac{I(p)}{p(I^2 - 1)} \sim F(I, I - p) \quad (186) \]
where \( a = x^T V(V^TV)^{-1} \) and \( D \) is approximated by an \( F \) distribution with \( (I, I - p) \) degrees of freedom.

The Q-statistic is given by
\[ Q_i = \left\| (I - \tilde{V})x \right\|^2 = x^T(I - \tilde{V})^T(I - \tilde{V})x - g\chi_i^2 \quad (187) \]
which is approximated by a weighted chi-square distribution. Combining both the \( D_i \) and \( Q_i \) results in
\[ BCN = \frac{x^T V(V^TV)^{-1} S^{-1}(V^TV)^{-1} V^T x (I - p)}{F(I, I - R) P(I^2 - 1)} + \frac{x^T(I - \tilde{V})^T(I - \tilde{V})x}{g\chi_i^2} \quad (188) \]
\[ BCN = x^T \theta x - g\chi_i^2 \quad (189) \]
with
\[
\theta = \frac{\mathbf{V}(\mathbf{V}^T \mathbf{V})^{-1} \mathbf{S}^{-1} (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T (\mathbf{I} - \mathbf{P})}{F(I, I - R)} + \frac{(\mathbf{I} - \mathbf{V})(\mathbf{I} - \mathbf{V})^T}{P(I^2 - 1)} \chi_i^2
\]  \tag{190}

The BCN is a quadratic function of \(x\), assuming \(x\) follows a multinormal distribution the BCN can be approximated by a weighted chi square distribution (H.H. Yue & S.J. Qin [110]).
Appendix E

First, a model is built using the NOC data leaving out the first NOC batch. The remaining NOC data is denoted as $\tilde{X}$ and the left out NOC batch as $\tilde{x}'$ as can be seen in the following figure.

\[
\begin{array}{ccc}
\tilde{x}' & \vdots & x' \\
\tilde{X} & \vdots & \tilde{X}
\end{array}
\]

Figure 131

Schematic representation of the leave-one-out approach.

Here, the wiggle used in the notation refers to a left-out-batch and the superscript indicates the batch number left out from $X$. Then, a PCA model is built using $\tilde{X}$ resulting in the loadings $\tilde{P}$. The batch $\tilde{x}'$ is projected on $\tilde{P}$ to find the scores:

\[
\tilde{t}_k' = (\tilde{P}'\tilde{P})^{-1}\tilde{P}\tilde{x}_k'
\]  

\[ 191 \]
The scores calculated for this batch are denoted as $\tilde{t}_k'$. Here, the wiggle refers to the leave one out approach and the superscript indicates the batch number left from $X$. The residuals are calculated as follows,

$$
\hat{x}_k' = \tilde{\mathbf{P}} \tilde{t}_k'
$$
$$
\tilde{e}_k' = \tilde{x}_k' - \hat{x}_k'
$$

and are denoted in the same manner as $\tilde{e}_k'$. This procedure is repeated for each NOC batch and $\tilde{t}_k'$ and $\tilde{e}_k'$ are stored in respectively the matrices $\tilde{T}_k$ and $\tilde{E}_k$. The score matrix $\tilde{T}_{k=K}$ is used to calculate $\tilde{S}$ and the residuals $\tilde{E}_k$ are used to estimate the parameters $\tilde{g}$ and $\tilde{b}$. 
SAMENVATTING VOOR NIET INGEWIJDEN*

Een groot aantal producten die worden gebruikt in het dagelijkse leven (bijvoorbeeld plastics, antibiotica of bier) zijn afkomstig uit de chemische, farmaceutische of voedingsmiddelenindustrie. Deze producten kunnen op twee manieren worden geproduceerd, continue of batch. Een continue productie houdt in dat er dag en nacht reactanten in een reactor bij elkaar worden gevoegd om tot het gewenste product te reageren. Een batch productie houdt in dat er een vooraf vastgestelde hoeveelheid reactanten in een reactor worden gedaan en deze onder vastgestelde procescondities te laten reageren tot het eindproduct.


Een belangrijk kenmerk van een batchproces is dat dit proces volgens een recept uitgevoerd wordt. De procesoperator heeft een recept van bijvoorbeeld temperaturen en drukken die gedurende de reactie nauwkeurig moet worden uitgevoerd. Hoe nauwkeuriger de procesoperator dit recept navolgt, hoe zekerder het is dat er een kwalitatief goed eindproduct zal zijn. Er bestaan echter altijd bronnen van variatie in een proces die ervoor zorgen dat de uiteindelijke kwaliteit van het eindproduct zal verschillen van batch tot batch. Denk hierbij bijvoorbeeld aan een plotselinge verstoring van de hoeveelheid koelwater naar de reactor toe. Door een dergelijke procesfout loopt de chemische reactie niet meer volgens het recept en kunnen er ongewenste bijproducten worden gevormd. Het is echter belangrijk voor

*HJR/EVS

-328-
de producent om de spreiding in de productkwaliteit zo klein mogelijk te houden teneinde een zekere kwaliteit van het eindproduct te kunnen garanderen. Echter, er zijn andere redenen dan alleen de productkwaliteit om een batch gedurende de reactie te bewaken. Een processtoring kan namelijk ook tot gevaarlijke situaties leiden. In dat geval is een tijdige detectie van een temperatuurverstoring gewenst in verband met de veiligheid. Ook kan de productie van ongewenste bijproducten kan om milieutechnische redenen ongewenst zijn. Kortom, batchbewaking is wenselijk om verschillende redenen waaronder veiligheid, milieu en productkwaliteit.

De traditionele manier van batchbewaking gaat meestal aan de hand van het volgen van belangrijke procesparameters, zoals de druk en temperatuur. Deze procesvariabelen worden gemeten door middel van sensoren die in de reactor zijn geplaatst. De uitkomst van elke afzonderlijke meting wordt weergegeven in een zogeheten univariate controlekaart. De uitkomst van elke meting mag een bepaalde limietwaarde niet overschrijden. Is dit wel het geval, dan volgt er een waarschuwing in de desbetreffende controlekaart. In dit geval is er hoogstwaarschijnlijk een procesverstoring aanwezig.

Echter, deze manier van monitoren heeft nadelen. Zo moet er voor elk gemeten procesparameter een univariate kaart worden bewaakt. Als er bijvoorbeeld meer dan twintig van deze procesparameters worden bewaakt, is het bijna al een ondoenlijke zaak om al deze kaarten afzonderlijk te bewaken. Verder hebben univariate controlekaarten de eigenschap om valse signalen te genereren, dit wil zeggen, een procesparameter overschrijdt zijn controlelimiet maar in werkelijkheid is er niets aan de hand. Een ander belangrijk punt is dat procesparameters vaak een onderling verband hebben. Bijvoorbeeld, als de druk in de reactor omhoog gaat, dan gaat de temperatuur vaak ook omhoog. Als nu door een procesverstoring dit verband, of deze correlatie, verbroken wordt, dan gaat dit onopgemerkt in univariate controlekaarten.

De aanpak die in dit proefschrift beschreven staat, houdt rekening met het verband tussen de procesparameters en heeft veel minder last van deze nadelen. Om een batch te bewaken worden zogeheten multivariate kaarten gebruikt. In tegenstelling tot de univariate kaarten zijn er nu slechts twee controlekaarten nodig.

Om deze multivariate kaarten te kunnen construeren, is een model nodig. Dit model vormt de basis van de controlekaarten. Een statistisch batchbewakingsmodel wordt in grote lijnen als volgt verkregen. Allereerst moeten er data worden verzameld.
Het idee achter statistische batchbewaking is dat er wordt geluisterd naar wat de reactor te vertellen heeft door simpelweg procesparameters te meten. Een ander alternatief zou zijn om fysische modellen van de reactor te ontwikkelen waarmee het tijdsverloop van de procesparameters kan worden voorspeld.

De chemie van een batchreactor is echter vaak dusdanig complex dat deze fysische modellen geen betrouwbare resultaten opleveren. Hoe gaat het verzamelen van de data in zijn werk? Stel nu dat er methyl-methacrylaat (polymeer) wordt gemaakt in een batchreactor. Van deze batchreactie worden elke tien seconden twintig procesparameters gemeten. Het batchproces zelf duurt 3 uur. Aan het einde van de batch is een grote hoeveelheid aan data verzameld. Een extra kwaliteitsmeting aan het einde van deze batch vertelt dat het eindproduct van een goede kwaliteit is. Dit betekent dat het recept correct is uitgevoerd door de procesoperator, ook al zijn er hier en daar wat variaties opgetreden ten opzichte van het recept. Deze variaties zijn onvermijdelijk, maar klaarblijkelijk toegestaan om toch nog een product af te leveren van een voldoende kwaliteit. Het is deze variatie in de gemeten procesdata waar het om draait als er een statistisch model wordt gebouwd. Omdat we nu weten hoeveel variatie er is toegestaan om een goed eindproduct af te leveren, kunnen we net als bij de univariate kaarten controle limieten berekenen. Omdat er gebruik wordt gemaakt van statistiek, is het belangrijk om zoveel mogelijk data te gebruiken. Dus procesdata van slechts 1 batch is niet voldoende. Daarom worden de data van meerder batches verzameld (> 30) om de zogeheten referentiedistributie van mogelijke procesvariatie zo groot mogelijk te maken. Hoe meer procesdata er verzameld wordt, des te betrouwbaarder de controlelimieten worden. Als er eenmaal genoeg procesdata zijn verzameld, dan wordt de procesvariatie in kaart gebracht met een model. Met behulp van dit model kunnen er er twee controlekaarten worden geconstrueerd. Het proces kan nu worden bewaakt aan de hand van deze controlekaarten. Het verloop van een nieuwe batch wordt gevolgd en samengevat in twee teststatistieken, die op hun beurt worden weergegeven in de twee controlekaarten. Blijft de teststatistiek binnen de controlelimieten, dan operereet het proces volgens wens. Is dit niet het geval, dan is er een procesverstoring aanwezig.

Er bestaat een ruime keus aan modellen, in dit proefschrift wordt een aantal van deze modellen vergeleken en getoetst aan de hand van industriële procesdata. Er wordt onder andere getoetst hoe vaak een vals alarm wordt gegenereerd in de controlekaarten voor een specifiek model. Verder wordt er getoetst hoe snel een
model in staat is om een procesfout te detecteren in deze controlekaarten. De controlekaarten zelf en de statistiek die erachter zit worden aan de tand gevoeld. Traditioneel worden drukken en temperaturen gemeten aan batchprocessen. Deze procesvariabelen meten de fysische eigenschappen van een proces, maar geven weinig inzicht in de onderliggende chemie. De onderliggende chemie is veel beter te meten door gebruik te maken van spectroscopische instrumenten zoals nabij-infraroodinstrumenten (NIR). Een bundel licht met een bepaalde golflengte wordt dan door het reactiemengsel gestuurd met behulp van een speciale probe. Dit licht heeft een bepaalde energie, en moleculen in het reactiemengsel absorberen een gedeelte van deze energie. De bundel licht wordt vervolgens weer gedetecteerd. De hoeveel energie die is geabsorbeerd door de moleculen is een maat voor de hoeveelheid van een bepaalde component. Hiermee kan de conversie in de reactor worden gevolgd. Deze metingen zijn dus rijk aan chemische informatie en maken het batchbewakingsmodel een stuk krachtiger. Immers, nu kunnen ook chemische veranderingen worden gedetecteerd. Dit proefschrift beschrijft hoe batchprocessen die zijn uitgerust met NIR-apparatuur kunnen worden bewaakt.

Uiteindelijk is het de bedoeling dat de procesoperator in de controle kamer slechts twee controle kaarten hoeft te bewaken. Vroegtijdige en betrouwbare detectie van procesverstoringen is indirect een kostenbesparing op het proces. Het is daarom zaak een zo goed mogelijk model te kiezen in combinatie met de juiste statistiek.

Uiteindelijk is het de bedoeling dat de procesoperator in de controle kamer slechts twee controle kaarten hoeft te bewaken. Vroegtijdige en betrouwbare detectie van proces verstoringen is indirect een kostenbesparing op het proces. Het is daarom zaak een zo goed mogelijk model te kiezen in combinatie met de juiste statistiek.
REFERENCES


9. Jackson J.E., Quality control methods for several related variables, *Technometrics*, 1959, **1**, 359


-335-


53. Wold S., Kettaneh-Wold N., Friden H., Holmberg A., Modelling and diagnostics of batch processes and analogous kinetic experiments, *Chemometrics and Intelligent Laboratory Systems*, 1998, **44**, 331


55. Gollmer K., Posten C., Supervision of bioprocesses using a dynamic time warping algorithm, *Control Eng Practice*, 1996, **4**, 1287


59. Sakoe H., Chiba S., Dynamic programming algorithm optimization for spoken word recognition, *IEEE transactions on acoustics, speech, and signal processing*, 1978, **26**, 43


63. Dahl K.S., Piovoso M.J., Kosanovich K.A., Translating third-order data analysis methods to chemical batch processes, *Chemometrics and Intelligent Laboratory Systems*, 1999, **46**, 161


76. Wold S., Antti H., Lindgren F., Ohman J., Orthogonal signal correction of near-infrared spectra, *Chemometrics and Intelligent Laboratory Systems*, 1998, 44, 175

77. Wold S., Kettaneh-Wold N., Tjessem K., Hierarchical multiblock PLS and PC models, for easier model interpretation, and as an alternative to variable selection, *Journal of Chemometrics*, 1996, 10, 463


-338-


98. Tauler R., Barcelo D., Multivariate curve resolution applied to liquid chromatography-diode array detection, *Trends in analytical chemistry*, 1993, 12, 319


-340-


J.A. Westerhuis, E.N.M. van Sprang, H-J. Ramaker, A.K. Smilde and D. Wienke, Industrial process analysis using in-line NIR spectroscopy and grey models, **PLS 2003, 3rd international symposium on PLS and related methods**.

E.N.M. van Sprang, H-J. Ramaker and H.F.M. Boelens, Procesanalyse, beter begrip en inzicht, **Automatie**, 8 (2003) 4-9


E.N.M. van Sprang, H-J. Ramaker, J.A. Westerhuis and A.K. Smilde; Correlations in batch process data, submitted, 2003


RESEARCH ACTIVITIES

Visited congresses and symposia
Chemometriedag (1999), UvA/Amsterdam
MVA workshop (1999), TNO/Zeist
Industry-University symposium (2000), KvL Kopenhagen/Denemarken
ADCHEM 2000 (2000), Pisa/Italy
Symposium beheersing van variatie in productieprocessen (2000), TuE/Eindhoven
Chemometriedag (2000), KUN/Nijmegen
Gordon Research Conference (2001), Williamstown/USA
Chemometriedag (2001), UVA Amsterdam
KNCV werkgroep in-proces analyse (2002), AKZO Nobel/Arnhem
ICRM 2002 (2002), Veldhoven
CAC 2002 (2002), Seattle/USA
Industrial Kemometri (2002), Kopenhagen/Denmark
7th chemometrics symposium new topics in chemometrics (2002), Mechelen/België
Werkgroep Analytische chemie (2003), Shell/Amsterdam
SSC8 (2003), Aland Islands/Finland

Poster presentations
Chemometriedag (1999), UvA/Amsterdam
ADCHEM (2000), Pisa/Italy
Chemometriedag (2000), KUN/Nijmegen
Gordon Research Conference (2001), Williamstown/USA
ICRM (2002), Veldhoven
CAC (2002), Seattle/USA
SSC8 (2003), Aland Islands/Finland

Oral presentations
Industry-University symposium (1999), Kopenhagen/Denmark
Johnson Polymers (2000), Heerenveen
Van data naar kennis (2002), Bodegrave
DSM (2001), DSM/Geleen
Analytisch symposium (2002), Lunteren
Discussant ICRM (2002), Veldhoven
ITS colloquium (2002), UvA/Amsterdam
CAC (2002), Seattle/USA
Industrial Kemometri (2002), Kopenhagen/Denmark
7th chemometrics symposium new topics in chemometrics (2002), Mechelen/België
KNCV werkgroep in-proces analyse (2003), Shell/Amsterdam

-345-
Met dit dankwoord is er een einde gekomen aan een dynamische periode. Een periode waarin wij zonder meer veel hebben geleerd op vele vlakken. Ook al zijn wij zelf de hoofdverantwoordelijke, dit proefschrift is mede tot stand gekomen door de steun en hulp die wij van andere mensen hebben ontvangen. In de eerste plaats willen wij Age bedanken voor zijn bereidwilligheid en vertrouwen om ons een gezamelijk promotieonderzoek te laten verachten. Als discipelen van Age hebben wij altijd een grote mate van vrijheid gekregen en dat waarderen wij ten zeerste. Vele pittige, lastige, lange maar nuttige en leervolle discussies hebben wij gevoerd met Johan en Steve. We willen onze mentoren hiervoor hartelijk bedanken! De motor achter onze meesterlijke Matlab algoritmen was vele malen Hans: Bedankt voor deze inspanningen! In dit avontuur werden wij ook bijgestaan door de leden en studenten van de PAC groep. Allemaal bedankt voor jullie discussie, inzet, onzinnige opmerkingen en potjes Quake. Relaxen deden we natuurlijk op vrijdag in de vele hoeken en gaten die het B-gebouwrijk is. De vaste bewoners van deze corners willen we bij deze bedanken voor de gezelligheid. Rasmus Bro en zijn groep willen wij bedanken voor de gezellige brainstorm sessies in Kopenhagen. De mensen van DSM Research in Geleen worden speciaal bedankt voor het aanleveren van procesdata, het geven van suggesties en inhoudelijk commentaar. Verder hebben wij natuurlijk ook mental support ontvangen van onze naaste vrienden en familie, bedankt daarvoor. Ook willen wij Bill Gates bedanken voor de schitterende software die hij heeft gemaakt om snel en efficiënt grote documenten te maken. Tot slot bedanken wij alle mensen op de 4e, de 6e en anderen voor de leuke tijd die wij hebben doorgebracht aan de UvA.

Vergissingen zijn de poorten naar ontdekkingen.
James Joyce, Iers schrijver, 1882-1941