

Imperial College London
Department of Chemical Engineering

**On extending process monitoring and diagnosis to
the electrical and mechanical utilities: an advanced
signal analysis approach**

by Inês da Mata Cecílio

April, 2014

Supervised by Prof. Nina F. Thornhill

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering of
Imperial College London

Declaration of originality

I herewith certify that all material in this dissertation which is not my own work has been properly acknowledged.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

This thesis is concerned with extending process monitoring and diagnosis to electrical and mechanical utilities. The motivation is that the reliability, safety and energy efficiency of industrial processes increasingly depend on the condition of the electrical supply and the electrical and mechanical equipment in the process.

To enable the integration of electrical and mechanical measurements in the analysis of process disturbances, this thesis develops four new signal analysis methods for transient disturbances, and for measurements with different sampling rates. Transient disturbances are considered because the electrical utility is mostly affected by events of a transient nature. Different sampling rates are considered because process measurements are commonly sampled at intervals in the order of seconds, while electrical and mechanical measurements are commonly sampled with millisecond intervals.

Three of the methods detect transient disturbances. Each method progressively improves or extends the applicability of the previous method. Specifically, the first detection method does univariate analysis, the second method extends the analysis to a multivariate data set, and the third method extends the multivariate analysis to measurements with different sampling rates.

The fourth method developed removes the transient disturbances from the time series of oscillatory measurements. The motivation is that the analysis of oscillatory disturbances can be affected by transient disturbances.

The methods were developed and tested on experimental and industrial data sets obtained during industrial placements with ABB Corporate Research Center, Kraków, Poland and ABB Oil, Gas and Petrochemicals, Oslo, Norway.

The concluding chapters of the thesis discuss the merits and limitations of each method, and present three directions for future research. The ideas should contribute further to the extension of process monitoring and diagnosis to the electrical and mechanical utilities. The ideas are exemplified on the case studies and shown to be promising directions for future research.

Acknowledgements

I am grateful to the Portuguese Foundation for Science and Technology (FCT) under Fellowship SFRH/BD/61384/2009, and to the Marie Curie FP7-IAPP project 'REAL-SMART', Contract No: PIAP-GA-2009-251304, for the financial support to my PhD degree and industrial secondments with ABB.

Thanks to the collaboration with ABB, I was able to explore and discuss rich data sets from which I identified the challenges which needed to be addressed and that eventually led to this thesis. For this, I am grateful to P. Lipnicki, D. Lewandowski, M. Orkisz, and M. Wojcik of ABB Corporate Research, Kraków, Poland, to K. Rapp of ABB Oil, Gas, and Petrochemicals, Oslo, Norway, to M. Mercangoez of ABB Corporate Research Switzerland and J. Kuehnel of ABB Turbo Systems Switzerland.

During my secondments in ABB Corporate Research, Kraków, and in ABB Oil, Gas, and Petrochemicals, Oslo, I was impeccably guided by James Otewill, John Pretlove, and Harald Fretheim. They were both mentors, experts, and friends. I hope our paths will keep crossing.

I was also fortunate to always be surrounded by great friends in London, Oslo, and Kraków. Thanks to them, I am a different person now from what I was four years ago.

The greatest debt for this thesis I owe to Professor Nina Thornhill. I feel confident in the quality of my work thanks to the support and opportunities for growth she offered me. Nina believed in me and challenged me from the start. She never told me the answer, but always helped me find it. Nina had a great influence in the way I write and was a tireless reviewer of this thesis. I cannot thank her enough.

The greatest debts for being, however, I owe to Pedro and to my family. My parents and grandparents are responsible for where I am now and will always endeavour to see me happy throughout my life. Pedro, thank you for the past nine years. We shall see many more.

Contents

1	Introduction	21
1.1	Motivation	21
1.2	Challenges to consider	24
1.3	Signal analysis	28
1.4	Contributions of this thesis	30
1.5	Publications	31
1.5.1	Publications from this thesis	31
1.5.2	Other publications	33
1.6	Outline of the thesis	33
I	Background and Context	36
2	Process, electrical and mechanical systems	37
2.1	Overview of the process, electrical and mechanical systems	38
2.1.1	The chemical process	38
2.1.2	The electrical utility	39
2.1.3	Compressor, electric motor, and electrical drive	39
2.2	Control and monitoring in industrial processes	41
2.2.1	Regulatory process control	41
2.2.2	Monitoring and diagnosis	43
2.2.3	Data characteristics in control and monitoring systems	43
2.3	Detailed models of the compression system and control	44
2.3.1	Characteristics of the system analysed	44
2.3.2	Compressor model	45
2.3.3	Induction motor model	47
2.3.4	Drive control algorithms	50
2.4	Disturbances	52
2.4.1	Terminology	53
2.4.2	Time scales	55

<i>Contents</i>	6
2.4.3 Common root causes	57
2.5 Case studies analysed in this thesis	59
2.5.1 Compressor rig	59
2.5.2 Gas plant	61
2.5.3 Turbocharger tests	62
2.6 Chapter summary	63
3 State-of-the-art in process monitoring and diagnosis	65
3.1 Stages of process monitoring and diagnosis	66
3.1.1 Treatment of data	66
3.1.2 Disturbance detection	67
3.1.3 Disturbance diagnosis	68
3.1.4 Related tasks	69
3.2 Classification of methods of process monitoring and diagnosis	70
3.3 Analytical models	73
3.4 Qualitative models	75
3.4.1 Causality models	77
3.4.2 Precedence models	79
3.4.3 Connectivity models	80
3.4.4 Forms of representation	80
3.5 Qualitative trends	84
3.6 Numerical data	86
3.6.1 Data-based models	86
3.6.2 Time series properties	94
3.7 Chapter summary and discussion	101
II Contributions of the thesis	104
4 Univariate detection of transient disturbances	105
4.1 Background	108
4.1.1 Detecting anomalous segments with nearest neighbours	108
4.1.2 Similarity measures for time series	109
4.1.3 Anomaly index definitions	110
4.2 Method development	111
4.2.1 Development case study	111
4.2.2 Time series	112

4.2.3	Algorithm	112
4.2.4	Significance level	115
4.2.5	Outputs of the detection	117
4.2.6	Computational effort	119
4.3	Parameter settings and sensitivity	119
4.3.1	Relation between parameters and dynamics of the system	120
4.3.2	Recommendations for parameters and analysis of sensi- tivity	121
4.4	Application to test case study	128
4.5	Chapter summary and discussion	132
5	Multivariate detection of transient disturbances in uni-rate systems	133
5.1	Background	134
5.1.1	Multivariate analysis with SVD	134
5.2	Method development	135
5.2.1	Development case study	135
5.2.2	Limitation of univariate approach in measurements with strong oscillatory trends or noise	137
5.2.3	Algorithm	138
5.2.4	Significance level	147
5.3	Parameter settings and sensitivity	147
5.3.1	Influence of parameters in the detection results	148
5.3.2	Recommendations for parameters and analysis of sensi- tivity	149
5.4	Application to test case study	153
5.5	Chapter summary and discussion	154
6	Multivariate detection of transient disturbances in multi-rate systems	156
6.1	Background	157
6.1.1	Multi-rate systems	157
6.2	Method development	158
6.2.1	Development case study	158
6.2.2	Algorithm	159
6.2.3	Comparison of the performance of the multi-rate and uni-rate methods	163
6.3	Application to test case study	165

6.4 Chapter summary and discussion 168

7 Removal of transient disturbances from oscillating measurements 169

7.1 Background 170

7.1.1 Non-linear deterministic systems 170

7.1.2 Predictability in non-linear deterministic systems 171

7.1.3 Nearest neighbour imputation 172

7.2 Method development 172

7.2.1 Development case study 173

7.2.2 Algorithm 173

7.2.3 Adequacy of the replacing segment 179

7.3 Parameter settings and sensitivity 182

7.3.1 Influence of parameters in the removal task 182

7.3.2 Recommendations for parameters and analysis of sensitivity 184

7.4 Application to test case study 191

7.5 Chapter summary 192

III Conclusions 195

8 Summary and discussion 196

8.1 Answer to the research questions 196

8.2 Summary and discussion of the methods developed 198

8.2.1 Univariate detection of transient disturbances 198

8.2.2 Multivariate detection of transient disturbances in uni-rate systems 200

8.2.3 Multivariate detection of transient disturbances in multi-rate systems 202

8.2.4 Removal of transient disturbances from oscillating measurements 203

9 Future research directions 206

9.1 Propagation path of a persistent disturbance in multi-rate systems 206

9.1.1 Background 208

9.1.2 Illustrative example 211

9.1.3 Algorithm for uni-rate method 212

<i>Contents</i>	9
9.1.4 Idea for a multi-rate method	217
9.1.5 Application of the uni-rate and multi-rate methods . . .	220
9.1.6 Comments on the proposed research direction	222
9.2 Monitoring and diagnosis in irregularly sampled time series . .	223
9.2.1 Background	224
9.2.2 Idea to adapt embedded vectors and similarity to irregularly sampled time series	227
9.2.3 Application to univariate detection of transient disturbances	229
9.2.4 Comments on the proposed research direction	230
9.3 Large-scale systems: Systematic approach of integrated analysis and functional specifications for a semi-automated tool	232
9.3.1 Background	233
9.3.2 Idea for a systematic approach to integrated analysis . .	235
9.3.3 Functional specifications for semi-automated tool	241
9.3.4 Comments on the proposed research direction	243
9.4 Chapter summary	244
10 Conclusion	246
IV Appendices	272
A Appendices for Chapter 4	273
A.1 Fitting of anomaly index vector to gamma distribution with skewness smaller than 0.77	273
A.2 Probability of false detection in gamma distributions with skewness smaller than 0.77	275
B Appendices for Chapter 5	277
B.1 Confidence level of the detection threshold	277
B.2 Optimization of α and β for additional groups of measurements	279
C Experiments with the gas compression rig	281
C.1 Description of the experimental facilities	281
C.2 Experimental conditions	283
D FRECOL	286

List of Tables

1.1	Examples of time series properties explored by signal analysis techniques in the process monitoring and diagnosis literature. .	29
1.2	Challenges in signal analysis relevant to this thesis, and current state-of-the-art.	35
1.3	Contributions in this thesis framed by the conditions addressed.	35
2.1	Characteristics of the compressor-motor-drive system analysed.	45
2.2	Use of disturbance-related definitions and terminology in relevant research areas.	53
2.3	Common changes in a propagating disturbance due to dynamic characteristics of the process system.	54
2.4	Causes of electrical disturbances according to their location. . .	58
2.5	Case studies used in each technical chapter, distinguished by their purpose. The relevant characteristics are also indicated. .	59
2.6	Variables measured by the data acquisition module in the drive. Only variables used in this thesis are indicated.	60
3.1	References on analytical model-based methods for process monitoring and diagnosis.	74
3.2	References on qualitative model-based methods for process monitoring and diagnosis.	78
3.2	References on qualitative model-based methods for process monitoring and diagnosis.	79
3.3	References on qualitative trend-based methods for process monitoring and diagnosis.	85
3.4	References on methods using data-based models for process monitoring and diagnosis.	88
3.4	References on methods using data-based models for process monitoring and diagnosis.	89

<i>List of Tables</i>	11
3.5 References on advanced signal analysis methods for process monitoring and diagnosis - Part I: detection.	96
3.6 References on advanced signal analysis methods for process monitoring and diagnosis - Part II: diagnosis.	98
4.1 Definitions of anomaly index reported in the literature.	110
4.2 Numerical outputs of the detection method for the reference example.	118
4.3 Summary of the roles and recommended values for the parameters.	121
5.1 Parameters involved in the multivariate detection method, with recommended values.	148
7.1 Embedded vectors excluded from the neighbourhood assessment of the pre- and post-transient references. Excluded embedded vectors are at one of the extreme ends of the time series X , and next to and including a transient disturbance. $T = t_f - t_i$, and n is the total number of samples of X	177
7.2 Inconsistency indices D_k and s_k based on the pre- and post-transient references for the development case study.	182
7.3 Summary of the roles and recommended values for the parameters.	190
7.4 Inconsistency indices D_k and s_k for the new times series of the measurement in the test case study.	192
8.1 Contributions in this thesis framed by the conditions addressed.	197
9.1 Common changes in a propagating disturbance due to dynamic characteristics of the system.	208
9.2 Directionality measures $H_{X \rightarrow Y}$ for selected pairs of measurements in the reference case study calculated in four different tests. F and S refer to the fast and the slow-sampled measurement, respectively.	222
9.3 Root cause diagnosis with integrated analysis applied to an industrial case study.	237
9.3 Root cause diagnosis with integrated analysis applied to an industrial case study.	238

<i>List of Tables</i>	12
9.3 Root cause diagnosis with integrated analysis applied to an industrial case study.	239
9.3 Root cause diagnosis with integrated analysis applied to an industrial case study.	240
B.1 Measurements included in each sub-group used for the optimisation of α and β . N_V refers to the number of measurements in the group.	280
C.1 Independent conditions and the modes they assumed in the tests.	284

List of Figures

1.1	Process schematic for a section of a gas processing plant, also showing the electrical drive, electric motor and gear box for the compressors.	22
1.2	Four-day snapshot of the measurements tagged in Figure 1.1. .	24
1.3	Examples of transient disturbances	25
1.4	Example of measurements from the same system sampled at different rates.	27
1.5	Example of a measurement after compression: some samples have been eliminated. The measurement is from a real gas processing plant.	27
2.1	Generic schematic for a compressor, electric motor, and electrical drive system.	40
2.2	Block diagram of a feedback control loop.	41
2.3	Compression system with relevant state variables and parameters, adapted from Greitzer [1976].	45
2.4	Block diagram for a centrifugal compressor.	47
2.5	Cross-sectional view of an asynchronous induction motor showing the main components.	48
2.6	Time sequence showing, above, the rotation of the magnetic field generated by the stator (direction represented by the arrow). The rotation is due to variations in magnitude and direction, shown below, of the three-phase voltages supplied to the stator coils.	49
2.7	Block diagram for an asynchronous induction motor.	50
2.8	Main components of an a.c. voltage-source inverter drive. . . .	51
2.9	Block diagram of the DTC algorithm in an a.c. voltage-source inverter drive.	52

<i>List of Figures</i>	14
2.10 Example of a measurement with a slowly developing disturbance drifting from the normal operating value (dashed line).	55
2.11 Example of measurement with persistent disturbances.	56
2.12 Simplified schematic of the gas compression experimental rig.	60
2.13 Illustrative schematics for a turbocharger performance test.	62
3.1 Hierarchical trees to organize the methods of process monitoring and diagnosis. The shaded node indicates where the methods developed in this thesis fit.	72
3.2 Schematic of a tank with cross-sectional area A , level Z , inlet flow F_1 , and outlet flows, F_2 and F_3	76
3.3 Digraph and adjacency matrix for the example in Figure 3.2.	81
3.4 Measurement trend illustrating the identification of three qualitative trends.	84
3.5 Plant profiles from a wastewater plant. Left panel corresponds to normal operating conditions. Right panel corresponds to disturbed operation. Taken from Thornhill et al. [2006].	87
3.6 Sequence of values from a process measurement under normal operating conditions. The dashed lines mark the limit beyond which new values of the measurement will be considered abnormal.	90
3.7 Example of a neural network with one hidden layer. The activation functions are f_k , and weights w_i and v_k	93
4.1 Reference example: time series of the shaft speed of a compressor, with transients around 5 and 11 s. The compressor speed is normalized by its initial value.	106
4.2 Top panel: temperature measurement with normalized units. Bottom panel: corresponding coefficients of Haar wavelet at different times (horizontal axis) and scales (vertical axis). Lighter tones represent higher amplitude coefficients.	107
4.3 Time series with an anomalous segment, drawn in a thick grey line.	108
4.4 Representation of a time series of samples x_i and the selection of the first three embedded vectors, given parameters m , τ and δ	113

4.5	Selected embedded vectors from the development case study, with grey tones relating to their position in the original signal. Each embedded vector covers 1 s. Parameters are $m = 1001$, $\tau = 1$ and $\delta = 1$	114
4.6	Anomaly index vector for the reference example ($k = 3$).	115
4.7	Performance of the detection method as a function of the number of samples in a transient and of the ratio between embedded vector samples and samples in a transient. The performance index can vary between zero and three. Parameters fixed in the analyses were $\tau = 1$, $\delta = 1$ and $k = 3$	123
4.8	Computational time as a function of the number of samples in a transient. The computational time is an average over the different values of m	124
4.9	Performance of the detection method as a function of δ . The performance value is an average over five time series with different starting instants, and it can vary between zero and two. Parameters fixed in the analyses were the sampling interval, such that the number of samples in a transient was 30, $\tau = 1$, $m = 15$ and $k = 3$	125
4.10	Performance of the detection method as a function of k , for different values of δ . The performance index can vary between zero and three. Parameters fixed in the analyses were the sampling interval, such that the number of samples in a transient was 30, $\tau = 1$ and $m = 15$	126
4.11	Performance of the detection method as a function of k and δ . The performance index can vary between zero and three. Parameters fixed in the analyses were the sampling interval, such that the number of samples in a transient was 30, $\tau = 1$ and $m = 15$. The number of samples is $n = 300$	127
4.12	Process schematic for the test case study <i>Gas plant case 1</i>	128
4.13	Condition of the measurements in the test case study <i>Gas plant case 1</i>	130
4.14	Colour plot for the industrial case study, zoomed in to the first transient and reordered to highlight the different times of start of the transient.	131

5.1	Development case study: time series of measurements from the <i>Compressor rig case 2</i>	136
5.2	Simplified schematic of the gas compression experimental rig.	137
5.3	Univariate anomaly index vectors \mathbf{a}_i for the <i>Compressor rig case 2</i>	140
5.4	Orthonormal basis functions obtained from SVD of the anomaly index matrix. The values on top of the plots indicate the singular value associated with each basis function.	141
5.5	Anomaly index vectors after the first selection step, $\hat{\mathbf{a}}_r$, for the development case study.	143
5.6	Comparison of $(u_{r,j}s_j)^2$ for each variable r with the selection thresholds (dashed lines) derived from $\beta = 0.2$	145
5.7	Final anomaly index vectors $\tilde{\mathbf{a}}_r$ for the development case study.	146
5.8	Plant-wide anomaly index vector for the development case study.	147
5.9	Performance of the detection method as a function of parameters α and β . N_V denotes the number of measurements in the set. Lighter tones denote better performance.	150
5.10	Performance of the detection method as a function of the number of samples in a transient, N_t . The performance index can vary between zero and two.	152
5.11	Computational time for calculating one univariate anomaly index vector, shown as a function of the number of samples in a transient, N_t . The computational time is an average over the seven measurements in the set.	152
5.12	Time series of the measurements in the test case study <i>Gas plant case 2</i>	153
5.13	Anomaly index vectors for the measurements in the test case study <i>Gas plant case 2</i> . Grey lines are the univariate anomaly index vectors \mathbf{a}_i . Black lines are the final anomaly index vectors $\tilde{\mathbf{a}}_i$ reconstructed after the multivariate step.	155
6.1	Measurements from the reference example used to demonstrate the multi-rate detection method. Measurement P1 has been re-sampled with a sampling interval of 0.5 s to create a multi-rate data set.	159

6.2	Construction of the anomaly index vectors \mathbf{ai} for a fast-, X , and a slow-sampled measurement, Y . The black dots represent samples in the time series of the measurements. The rectangles show the samples included in the embedded vectors \mathbf{x}_i and \mathbf{y}_i . ai are the corresponding anomaly indices. The sampling intervals are related as $\Delta t^f / \Delta t^s = 3$, and the embedding dimensions are $m^f = 4$ and $m^s = 2$	161
6.3	Univariate anomaly index vectors for multi-rate measurements of the development case study. The first \mathbf{ai} is an expanded anomaly index vector, for a slow-sampled measurement.	163
6.4	Final anomaly index vectors for the reference example. Black lines refer to the multi-rate method. Grey lines refer to the uni-rate method with the fast-sampled measurements downsampled to the slow rate. Dashed lines represent the corresponding detection thresholds.	164
6.5	Performance of the multi-rate (black line) and uni-rate methods (grey line) as a function of the sampling interval of the slow-sampled measurements Δt^s . The sampling interval of the fast-sampled measurements Δt^f is fixed. The performance index can vary between zero and two.	165
6.6	Time series of the measurements in the test case study <i>Turbocharger case</i>	166
6.7	Anomaly index vectors for the measurements in the compressor tests case study. Grey lines are the univariate anomaly index vectors \mathbf{ai} . Black lines are the final anomaly index vectors $\tilde{\mathbf{ai}}$ after the multivariate step.	167
7.1	State space of dimension $M = 2$ for a sinusoidal time series. . .	170
7.2	A non-linear deterministic time series with five nearest neighbour cycles followed by five similar points (adapted from Thornhill [2005]).	171
7.3	Development case study: time series of measurements from the <i>Compressor rig case 2</i> case study.	173
7.4	Close-up of measurement S1 highlighting the transient (grey line) and the pre- and post-transient references (black thick lines).174	

7.5	Measurement S1 highlighting the $k = 2$ nearest neighbours (thick black lines) and the embedded vectors excluded from the neighbourhood assessment (grey lines), for the pre-transient reference of the first transient.	176
7.6	Generation of the reconstructed time series (grey line) by concatenating the replacing segments, and the transient-free segments of the original time series (black line).	178
7.7	Reconstructed time series (grey lines) after removal of transients, superimposed on the original time series (black lines).	180
7.8	Performance of the removal method as a function of τ and δ , when $\tau = \delta$. Parameters fixed in the analyses were $m = 800$ and $k = 2$. The number of samples per cycle is $T_p = 800$. Results are shown for each variable, and for the estimations based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).	186
7.9	Computational time as a function of T_p/τ . The computational time is an average over the three variables, and includes the removal of the two transients.	187
7.10	Performance of the removal method as a function of m/T_p . Parameters fixed in the analyses were $\tau = 1$, $\delta = 1$ and $k = 2$. Results are shown for each variable, and for the estimations based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).	188
7.11	Performance of the removal method as a function of k . Parameters fixed in the analyses were $\tau = 1$, $\delta = 1$ and $m = T_p = 800$. Results are shown for each variable, and for the estimations based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).	189
7.12	Original time series (top panel) and reconstructed time series based on the pre-transient reference (middle panel) and post-transient reference (bottom panel).	194
9.1	Close-up on the start of a disturbance induced in measurement S1.sp. The sequence of plots reflects, from top to bottom, the propagation path of the disturbance.	209

9.2	Illustration of predictability: three identical X samples, marked with crosses, can be indicated by three past identical segments in X (self-predictability) or in Y (non-linear mutual predictability). Adapted from Le Van Quyen et al. [1998].	210
9.3	Time series of measurements from the <i>Compressor rig</i> data set.	212
9.4	Predictability errors for measurements $S1.sp$ (represented by X) and $S1$ (represented by Y) of the reference case study. Top panels: self-predictability. Bottom panels: mutual predictability. Values on top of plots indicate the average predictability error.	215
9.5	Predictability errors for measurements $S1.sp$ (represented by X) and $S1$ (represented by Y) of the reference case study determined with the multi-rate method. Top panels: self-predictability. Bottom panels: mutual predictability. Values on top of plots indicate the average predictability error.	220
9.6	Measurements from the reference case study after downsampling by a factor of 5000.	221
9.7	Pairs of differences (represented by arrows) used in assessing the distance between segments p (black line and markers) and q (grey line and markers).	226
9.8	Construction of first three embedded vectors from the irregularly sampled time series X . Each embedded vector spans M time units and lags the previous by Δ time units.	228
9.9	Compressor speed measurement from <i>Compressor rig case 1</i> . The measurement was manipulated in order to have an irregularly sampled time series. The values are normalized by the initial value.	229
9.10	Close-up on the ompressor speed measurement to highlight the irregular spacing between samples.	230
9.11	Normalized anomaly index vector. The dashed line indicates the detection threshold.	231
9.12	Context diagram of the proposed tool representing its interactions with the main external entities.	241
A.1	Base time series for generation of cases representing normal operation scenarios.	274

A.2	Estimated gamma shape parameter (grey line) and derived skewness (black line) for three sets of clean signals.	275
A.3	Influence of gamma parameters a and b on the relation between the proposed threshold and the gamma quantile function, with $p = 0.999999$	276
B.1	Base time series for generation of cases representing normal operation scenarios.	278
B.2	Univariate anomaly index vector $\mathbf{a}i_2$ used in the statistical study of the detection threshold.	278
B.3	Distribution of the values of relative variance of the SVD terms with positive detections.	279
B.4	Performance of the detection method as a function of parameters α and β for five different sub-groups of measurements. N_V denotes the number of measurements in the group. Lighter tones denote better performance.	280
C.1	Simplified schematic of the gas compression experimental rig.	282
D.1	Single area power system frequency control.	288
D.2	Simulink model in FRECOL: first layer	290
D.3	Under mask of FCR block.	290
D.4	FCR generation, under <i>Multi-unit generation</i> mask.	291
D.5	Under mask of FRR block.	291
D.6	FRR generation, under <i>Multi-unit generation</i> mask.	292
D.7	Under mask of RR block.	293
D.8	RR generation, under <i>Overall generation</i> mask.	293
D.9	Area defining the constrained random disturbance $\Delta P_{L,max}$ (left), and a time realization of $\Delta P_{L,max}$ (right).	294
D.10	System response to a mild power imbalance DPL	295
D.11	System response to a more severe power imbalance DPL	295
D.12	System response to power imbalance DPL with variations close to worst-case scenario.	295

Chapter 1

Introduction

1.1 Motivation

Society depends on goods produced by industrial processes, and expects these processes to be reliable, safe and energy-efficient. These requirements increasingly depend on the condition of the electrical supply and electromechanical equipment used in the process. Two recent episodes, in March 2013, gave evidence for this tight link between society, industrial processes, and their electrical and mechanical utilities. A water pump failure in the first case [Pfeifer, 2013b] and a power cut in the second [Pfeifer, 2013a] disrupted two industrial gas processing plants and temporarily stopped the supply of natural gas to the UK. The result was an increase in the price of this commodity of more than 50% in one day, and a wide-spread fear of shortage. Figure 1.1 shows a section of an industrial process, in this case a process to condition natural gas into selling quality. Examples of other industrial processes include the production of chemicals, petrochemicals, pulp and paper, food and pharmaceutical products.

The process in Figure 1.1 is equipped with two gas compressors, shown at the bottom of the figure. They are powered by a single electric motor, which shows the tag V1 for a vibration measurement. The motor is connected to the electric grid via a variable-speed drive, which adapts the frequency from the grid to control the speed of the motor. Electrical machines, like the electric motor, are powered by electrical energy. They convert it into mechanical energy to drive rotating machinery, such as a compressor or a pump, which, in turn, convert this energy into internal energy in the process fluids. The use of elec-

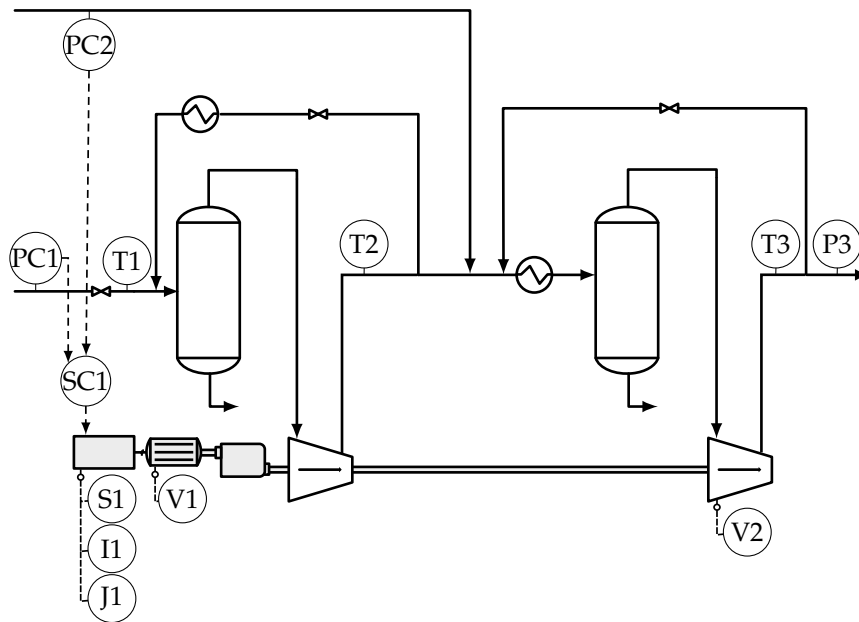


Figure 1.1: Process schematic for a section of a gas processing plant, also showing the electrical drive, electric motor and gear box for the compressors.

trical energy to drive rotating machinery is increasingly common due to the greater energy-efficiency and easier maintenance of electromechanical equipment compared with traditional drives, such as gas turbines [REAL-SMART, 2011].

Energy paths, thus, connect the operations of the process, and its electrical and mechanical utilities. Signal paths also do so. These signals are shown in the figure as dashed lines. Their purpose is to communicate the condition of the process to a computational device, and instructions from this device to the electromechanical equipment. This is part of a strategy to keep the operation of the process at some specified conditions, and is known as process control.

When the operation of any of these systems deviate from their desired conditions, that system is said to be affected by a disturbance [Gertler et al., 1999, Couper et al., 2005]. An example in the process system is a deviation in the temperature inside a reactor. An example in the electrical system is a deviation in the frequency of the supply from 50 Hz, in the case of Europe. When the temperature in the reactor deviates, the quality of the final product may be affected, and extra energy may be needed to bring the temperature back to its desired value. When the supply frequency deviates, electromechanical equipment may be damaged, hence risking an accident. Electromechanical

equipment or the machinery it drives may also trip to protect themselves from damage, which halts production.

These examples show how disturbances compromise the reliability, safety and energy efficiency that industry and society expect of the production process. Disturbances in the quality of the electrical supply, in particular, are becoming more frequent [Statnett, 2012]. In part, this is due to the increasing use of renewable energy sources and energy-efficient equipment [Bollen, 2000].

Disturbances are reflected in the measurements collected from points in the process, electrical and mechanical systems. As an example, Figure 1.2 shows a four-day snapshot of measurements from the three systems in Figure 1.1. These measurements have been scaled to the same minimum-to-maximum range. These measurements show disturbances in the form of sharp transient spikes and oscillations. The figure also shows that these features are distributed across the majority of the measurements. This reveals that once a disturbance has entered the plant it can spread between the process, mechanical and electrical systems, propagating through the energy and signal paths that connect them.

The interaction between these three systems, the increasing use of electromechanical equipment in industrial processes, and the increase in electrical disturbances mean that the electrical supply and electromechanical equipment can be expected to become increasingly important sources of disturbances to the process industry.

Process monitoring and diagnosis is an area of research and practice which seeks to detect disturbances in the process and diagnose the root causes [Venkatasubramanian et al., 2003a, Russell et al., 2000]. It does so by analysing the measurements of physical variables from the process system.

Similar practices exist for rotating machinery, aircraft engines, and automotive equipment [Jardine et al., 2006, Isermann, 2005], where operational conditions are also critical. In the case of industrial plants, both the process and its equipment are monitored, but the tasks are often performed by two different departments, and the relations between these interacting systems usually remain unexplored.

There have been calls from industrial commentators to integrate the information about the condition of the process and its equipment [Reeves, 2005, Schiltz, 2008]. Also, Cecílio et al. [2011] showed examples where analyses to

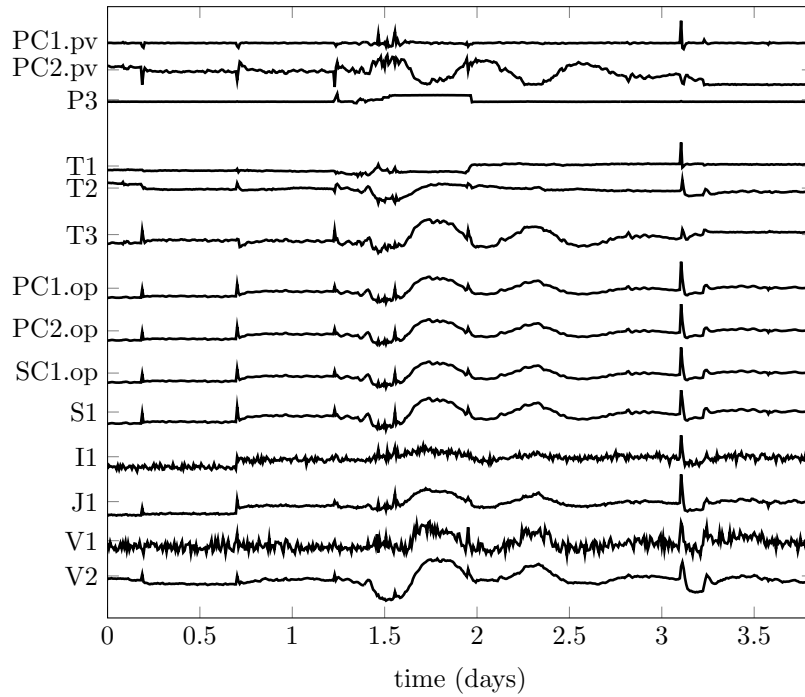


Figure 1.2: Four-day snapshot of the measurements tagged in Figure 1.1.

locate the origin of a disturbance were misleading if based on process measurements only. The disturbances in the examples crossed between the process and its electrical and mechanical systems.

In summary, this section aimed to show that there are relevant motivations to extend the analysis of process disturbances to measurements from the electrical and mechanical utilities. Whether or not this is possible, and how to do it, are research questions addressed in this thesis.

1.2 Challenges to consider

The integration of the electrical and mechanical utilities in the analysis of process disturbances entails specific challenges not addressed in previous work. These challenges include, but are not limited to, the following:

- analysis of transient disturbances,
- access to process and electromechanical measurements,

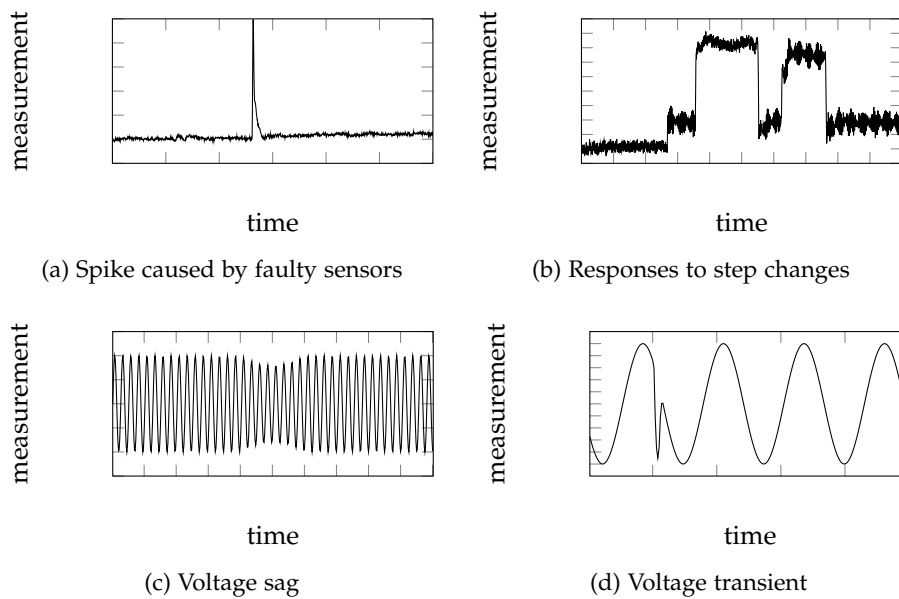


Figure 1.3: Examples of transient disturbances

- analysis of process and electromechanical measurements with irregularities such as
 - different sampling rates for each system,
 - time-misalignment between systems, and
 - irregular sampling rate within a system, and
- application of the methods to large-scale systems.

Transient disturbances are short-lived and infrequent deviations from normal operation. Figure 1.3 shows examples of this type of disturbance reflected in measurements from the system. The examples include a spike caused by a faulty sensor (Figure 1.3a), responses of the system to step changes (Figure 1.3b), a voltage sag (Figure 1.3c), and a voltage transient (Figure 1.3d).

The reason to consider transient disturbances is that the electrical utility is mostly affected by events of a transient nature. These are caused by imbalances between the power supplied to and consumed from the electric grid, which lead to momentary instabilities in the frequency and voltage of the supply [Bevrani, 2009]. In addition, the time scale of events in the electrical and mechanical systems is faster than events in the process. Even if a disturbance is long-lived from the perspective of an electromechanical system, it is likely to

be considered momentary compared to the time scale of events in the process system.

To analyse the process, electrical and mechanical systems together, it is desirable to have simultaneous access to measurements from all. However, this can be difficult, according to the opinion of industrial collaborators. The reason is that different engineering domains in the industrial plant, such as the control of the process and the monitoring of the electrical drives, are often under the supervision of different departments and use technology from different vendors. As a result, the measurements from different systems are often stored apart and not brought together.

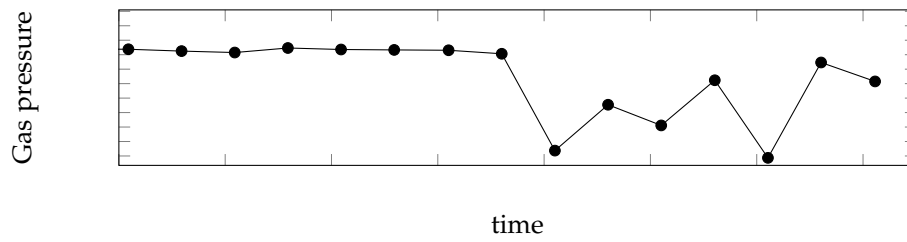
An additional challenge is related to the practices adopted in the sampling and storage of those measurements.

In general, process, electrical and mechanical variables vary continuously. For example, the temperature of a gas does not jump from 40°C to 41°C but instead passes through all infinitesimal values in between. The measurements of these variables, on the other hand, are discrete. The rate at which they are sampled is, in general, related to the time scales of the events that affect the variables. Therefore, common intervals between process samples are in the order of seconds, in contrast to the millisecond intervals found in electromechanical measurements. Figure 1.4 illustrates a slow-sampled process measurement and a fast-sampled electrical measurement from performance tests to a commercial turbocharger compressor. The disturbance affecting both variables is the same, but the way it is characterised in the measurements is different because of the different sampling rates.

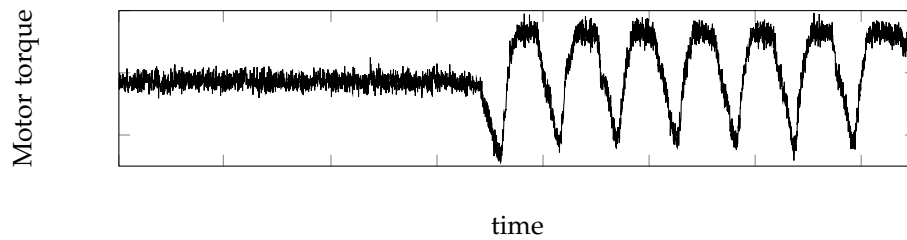
Process, electrical and mechanical measurements may also be misaligned. In other words, samples which were taken at the same time in the different systems may be assigned different time stamps. This happens when the clocks of the different data acquisition tools are not synchronised.

Additional irregularity in the measurements derives from the storage practices. To save memory, measurements are sometimes compressed, either by eliminating samples or by substituting the values of the samples by a constant value, for example, the average over a period. A result of such practices is that the interval between samples in a measurement is not constant. Figure 1.5 shows an example from a gas processing plant.

Finally, the inclusion of electrical and mechanical measurements in the analysis of process disturbances increases the dimension of the data set. Large-



(a) Slow-sampled measurement



(b) Fast-sampled measurement

Figure 1.4: Example of measurements from the same system sampled at different rates.

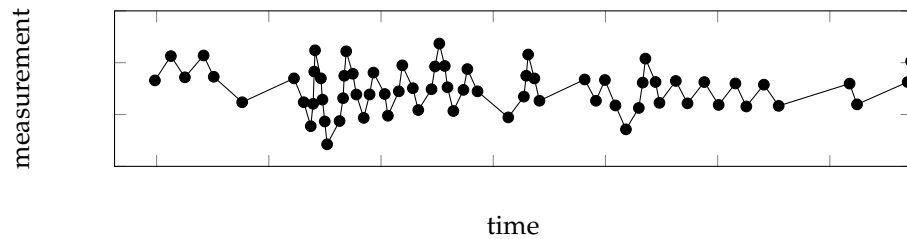


Figure 1.5: Example of a measurement after compression: some samples have been eliminated. The measurement is from a real gas processing plant.

scale industries such as oil and gas may have several thousands of measured variables from the process system alone. This large dimension is a known challenge in process monitoring and diagnosis [Kresta et al., 1991]. Including measurements from the electrical and mechanical systems will further increase this challenge.

The discussion in this section is relevant to the research questions. On the one hand, the integration of electrical and mechanical systems in the analysis of process disturbances would benefit from a change in some industrial practices of sampling, storage and sharing of data. However, this cannot be enforced. Therefore, most challenges of this integration should also be handled through appropriate techniques of process monitoring and diagnosis.

1.3 Signal analysis for process monitoring and diagnosis

Venkatasubramanian et al. reviewed in detail the use of data-driven [Venkatasubramanian et al., 2003c] and model-based [Venkatasubramanian et al., 2003a,b] approaches for process monitoring and diagnosis. One set of techniques is advanced signal analysis. Signal analysis applies to measurement data from process variables, and is independent of any physical understanding of the process, such as that captured by first-principles models.

The distinguishing feature of advanced signal analysis compared to other techniques of process monitoring and diagnosis is that it handles process measurement data as time series. Time series are also known as discrete-time signals in the fields of electrical engineering and signal processing. A time series is an ordered sequence of samples from one variable, obtained at consecutive points in time. Therefore, time series of process measurements not only capture the conditions of the process, but also how these conditions evolved in time. This is advantageous because disturbances in process variables include undesired values as well as undesired trends.

Advanced signal analysis takes the time order of measurements into account. As a result, it is able to identify complex features in the trend of a measurement, and relations between the trends of different measurements. Table 1.1 gives examples of time series features and relations which were explored by signal analysis to provide information about the presence and cause of process disturbances.

The signal analysis approach has been particularly useful to analyse disturbances that propagate between interacting variables [Thornhill and Horch, 2007]. Studies in the literature have so far been limited to the process system, but the concept of plant-wide disturbance can be extended to variables from the process, electrical and mechanical systems. As discussed, process variables interact through mass and energy flows, and control signals. As a result, one disturbance can upset a large number of measurements, even measurements distant from the root cause of the problem. Commonly affected processes are those with energy integration and recycle streams, as well as plants with gas compressor trains.

Table 1.1: Examples of time series properties explored by signal analysis techniques in the process monitoring and diagnosis literature.

Time series property	Information on disturbances	Example of references
correlation between spectral content of time series	identifies groups of measurements with common oscillating disturbances, and their contributions to those disturbances	[Jiang et al., 2007]
cross-correlation between process variable and controller output time series	recognizes stiction in a control valve	[Horch, 1999]
higher order statistics of time series spectrum	quantifies non-linearity in a measurement, which suggests proximity to the root cause of a non-linear disturbance	[Choudhury et al., 2004]
transfer entropy between time series	finds the direction of propagation of repetitive disturbances	[Bauer et al., 2007b]
nonlinear cross-correlation between time series	estimates time delay between measurements to suggest the direction of propagation of a disturbance	[Stockmann et al., 2012]

Thornhill and Horch [2007] comprehensively reviewed the use of signal analysis techniques for the analysis of plant-wide disturbances, while Shardt et al. [2012] examined some more recent developments. The plant-wide approach, and consequently signal analysis, grew significantly in the last decade because of cheap data storage. This allowed industry to keep data historians storing long histories of most process measurements. This approach has been effective in the analysis of large scale industrial plants, and is available in commercial monitoring tools [Horch et al., 2007] such as the Plant Disturbance Analyser tool (PDA) by ABB.

It should be recalled that one research question in this thesis is whether it is possible to extend the analysis of process disturbances to measurements from the electrical and mechanical utilities. This extension raises specific challenges not addressed in previous work. Signal analysis techniques are relevant to address those challenges, in view of the strengths discussed in this section. However, review of the literature on process monitoring and diagnosis shows

that there have been few solutions so far. Table 1.2 summarizes the state-of-the-art with regards to these challenges. Chapter 3 will review the literature in greater detail.

The observations in table 1.2 show that the current state-of-the-art in process monitoring and diagnosis is limited in some areas. This is particularly true in the analysis of transient disturbances, of measurements sampled with irregular rate, and of measurements with different sampling rates. Some of these open questions should be solved in order to meet the research aims of this thesis.

1.4 Contributions of this thesis

Table 1.2 summarized the challenges raised by the analysis of disturbances across process, electrical and mechanical systems. The table also showed that the current state-of-the-art offers limited solutions to this integrated analysis. The challenges which are less addressed in the literature are the analysis of transient disturbances, of measurements sampled with irregular rate, and of measurements with different sampling rates.

In order to extend the state-of-the-art and address an industrial problem, this thesis will provide contributions to signal analysis of transient disturbances, and of measurements with different sampling rates. The reason to focus on these two conditions is that they always exist, even if industry modifies inadequate practices in the sampling and storage of measurements. Disturbances of a transient nature are a result of the dynamics of the systems involved. Similarly, sampling rates in the process, electrical and mechanical systems are adapted to these dynamics, which inherently have different time scales.

In brief, the key contributions of this thesis are four methods. Table 1.3 frames these methods, and also indicates the thesis chapters and publications which address them.

The four entries in the table define the conditions that need to be handled by different signal analysis methods. The analysis of persistent disturbances in measurements with equal sampling rate has been solved, and commercial tools are available as briefly discussed in section 1.3. Persistent disturbances refer to deviations which repeat numerous times [Thornhill and Horch, 2007].

All four methods developed in the thesis address the analysis of transient disturbances. Two of the methods apply to a single measurement at a time, hence are independent of the sampling rates of each system. The other two methods apply to groups of measurements. In one of them, all measurements must have the same sampling rate, while in the other the measurements can have different sampling rates. In more detail, these contributions are:

- A method to detect transient disturbances in a single measurement. The statistical significance of the detection threshold is analysed and parameter guidelines are established. A colour map is also suggested to visualize the detection results. The method is demonstrated on experimental and industrial case studies. (Chapter 4)
- A method to detect transient disturbances in a multivariate set of measurements, sampled at equal rate. The statistical significance of the detection threshold is analysed and parameter guidelines are established. The method is demonstrated on experimental and industrial case studies. (Chapter 5)
- A method to detect transient disturbances in a multivariate set of measurements, sampled at different rates. The method is compared with the uni-rate method applied to the measurements downsampled to the lower sampling rate. The method is demonstrated on experimental and industrial case studies. (Chapter 6)
- A method to remove transient disturbances from the time series of an oscillating measurement. The segment replacing the transient disturbance agrees with the underlying dynamics of the original measurement. Guidelines for the parameters are established. The method is demonstrated on experimental and industrial case studies. (Chapter 7)

1.5 Publications

1.5.1 Publications from this thesis

Results from this thesis have been presented in international conferences or published in journal articles, as listed below. The publications which form the

core of the contributions in this thesis are also referred in the framework of table 1.3.

Journal articles and conference proceedings

- Inês M. Cecílio, Su-Liang Chen, and Nina F. Thornhill. Importance of auxiliary systems for process fault detection and diagnosis. In *Proceedings of the 19th Mediterranean Conference on Control & Automation (MED)*, pages 952–957, Corfu, 2011. IEEE
- Inês M. Cecílio, Knut Rapp, and Nina F. Thornhill. Process performance analysis in large-scale systems integrating different sources of information. In *Proceedings of the 8th IFAC International Symposium on Advanced Control of Chemical Processes*, pages 45–50, Singapore, 2012
- Inês M. Cecílio, James R. Ottewill, John Pretlove, and Nina F. Thornhill. Nearest neighbors method for detecting transient disturbances in process and electromechanical systems. Submitted to. *Journal of Process Control*, 2013. [Preprint]
- Inês M. Cecílio, James R. Ottewill, Harald Fretheim, and Nina F. Thornhill. Multivariate detection of transient disturbances for uni- and multi-rate systems combining nearest neighbors methods and SVD. Submitted to. *IEEE Transactions on Control System Technology*, 2014. [Preprint]

Oral and poster presentations

- Inês M. Cecílio and Nina F. Thornhill. Improving process performance looking at the whole picture. In *CPSE Industrial Consortium Meeting*, December 2011
- Inês M. Cecílio. Detecting and diagnosing disturbances in natural gas processes with signal analysis. In *SET for Britain national competition, Parliamentary and Scientific Committee, London UK*, March 2013a
- Inês M. Cecílio. Why is the heating off? – Finding the causes of disturbances in natural gas processes. In *Chemical Engineering PhD Research Symposium*, July 2013b

1.5.2 Other publications

The list below indicates other publications which resulted from work done in parallel to this thesis.

Reports and conference proceedings

- Inês M. Cecílio. Electrical interactions with process systems. WP8.1 – Use of causal signal analysis – Part I. Technical report, Project REAL-SMART Secondment Report, 2011
- Inês M. Cecílio. Electrical interactions with process systems. WP8.1 – Use of causal signal analysis – Part II. Technical report, Project REAL-SMART Secondment Report, 2012
- Inês M. Cecílio. Electrical interactions with process systems. WP8.1 – Use of causal signal analysis – Part II.2. Technical report, Project REAL-SMART Secondment Report, 2013
- Inês M. Cecílio, Anne Mai Ersdal, Davide Fabozzi, and Nina F. Thornhill. An open-source educational toolbox for power system frequency control tuning and optimization. In *Proceedings of the 4th IEEE European Innovative Smart Grid Technologies*, Lyngby, 2013

Oral presentations

- Inês M. Cecílio and Nina F. Thornhill. Data fusion in support of operational decision-making. In *CPSE Industrial Consortium Meeting*, April 2010
- Inês M. Cecílio and Moncef Chioua. Informatics in process operations – Academic algorithms, practical applications. In *CPSE Industrial Consortium Meeting*, June 2012

1.6 Outline of the thesis

The thesis is divided into three parts and consists of ten chapters. The first part presents the background to the work of this thesis. Chapter 2 explains the chemical process, its rotating machinery, electromechanical equipment and electrical supply. These are the systems that generate the measurements to

which this thesis applies. Chapter 3 provides a broad survey of the methods of process monitoring and diagnosis, and concludes on the suitability of advanced signal analysis methods to meet the objectives of the thesis.

Part II develops four new methods which extend the state-of-the-art in the analysis of transient disturbances, and of measurements with different sampling rates. Chapter 4 presents a method to detect transient disturbances in the measurement of a single variable. Chapter 5 extends the detection of transient disturbances to a multivariate set of measurements, sampled at equal rate. Chapter 6 further extends the multivariate analysis to measurements with different sampling rates. Chapter 7 develops a method to remove the transient disturbances from the time series of oscillatory measurements. All methods are developed on case studies which derive from experimental work with a gas compressor rig, and tested on industrial case studies.

Part III closes the thesis with a discussion about the main achievements, and suggestions for future research. Chapter 8 summarises the four main achievements of the thesis, and gives a critical discussion of their merits and limitations. Chapter 9 presents three ideas for future research which aim to address unsolved challenges in the extension of process monitoring and diagnosis to the electrical and mechanical utilities. Chapter 10 finishes with a conclusion of the whole thesis.

Table 1.2: Challenges in signal analysis relevant to this thesis, and current state-of-the-art.

Challenge	State-of-the-art
Transient disturbances	Most techniques for plant-wide disturbance analysis implicitly assume that the unwanted deviation repeats numerous times [Thornhill and Horch, 2007]. However, this is not the case for transient disturbances.
Groups of measurements with different sampling rates	Most contributions to multi-rate systems focus on problems other than the detection and diagnosis of disturbances [Li et al., 2003, Misra et al., 2002].
Groups of measurements with time-misalignment	Some authors have addressed analogous problems. One problem is aligning trends of measurements from batch processes [González-Martínez et al., 2013]. Another is determining time delays between distinctive events in process measurements [Bauer and Thornhill, 2008, Stockmann et al., 2012].
Measurements with irregular sampling rates	To the best of the author's knowledge, process monitoring and diagnosis techniques have not been applied to measurements sampled at irregular intervals. Nonetheless, there has been work done on the related areas of model estimation and predictive control [Sheng et al., 2002, Ding and Ding, 2010]
Large-scale systems	Several authors have addressed this problem. One strategy is reducing the set of measurements to a smaller set. This uses dimensionality reduction techniques, which take advantage of the physical dependencies between measurements [Maurya et al., 2005]. Another approach is to divide the measurements into conceptually meaningful blocks [Wold et al., 1996], for example, according to process sections [Qin et al., 2001].

Table 1.3: Contributions in this thesis framed by the conditions addressed.

Sampling rate	Time scale of disturbance	
	Persistent	Transient
Measurements with equal sampling rates	State-of-the-art [Thornhill and Horch, 2007]	Chapters 4, 5, 7 [Cecílio et al., 2013] [Cecílio et al., 2014]
Measurements with different sampling rates	Future work	Chapter 6 [Cecílio et al., 2014]

Part I

Background and Context

Chapter 2

Process, electrical and mechanical systems

The methods proposed in this thesis are developed and demonstrated on processes which handle gas mixtures. The reason why these processes provide good case studies is that they are driven by compressors, which are increasingly powered by electric motors and electrical drives. Therefore, gas handling processes and the compression system formed by compressor, electric motor and electrical drive generate the types of measurements to which this thesis applies.

This chapter introduces the chemical process, compressor, electric motor and electrical drive, as well as concepts of their operation which are relevant to this thesis. Section 2.1 starts with an overview of those systems. At the core of their industrial operation is automated control, hence section 2.2 follows with a description of this task, and the associated task of monitoring. Then, section 2.3 focuses on fundamental theory of the compressor-motor-drive system. It presents relevant mathematical models and control algorithms. Despite automated control, these systems are still affected by unwanted deviations. Thus, section 2.4 introduces the terminology of disturbances, and discusses types and causes. With the context set, section 2.5 then introduces the case studies, and the reasons why they were chosen. The chapter closes with a summary.

2.1 Overview of the process, electrical and mechanical systems

2.1.1 The chemical process

The term *chemical process* designates the system of unit operations used to transform feed materials into products [Seborg et al., 2010, Seider, 2011]. There is a broad variety of chemical products, and classifications can be found in several chemical engineering textbooks, such as Seider [2011] and Silla [2003]. An example of a chemical product is ethylene. It is produced by cracking heavier hydrocarbons, and is required in large quantities to produce other chemical products, like the polymer polyethylene. Detergents and pharmaceuticals are examples of chemical products which are used directly by consumers.

Unit operations and process equipment

Unit operations transform the feed materials chemically or physically. A chemical transformation is a change in the molecular structure of the materials. A physical transformation, on the other hand, may refer to a change in a physical condition, such as temperature, or to the separation of materials based on their physical properties, such as density. Unit operations are governed by fundamental principles, such as fluid mechanics, reaction kinetics, and thermodynamics.

Unit operations are carried out in process equipment. For example, heat exchangers are used to cool, or heat, liquids, or gases, or to boil liquids. A reactor is the tank where a chemical reaction occurs. A particular type of process equipment is rotating machinery. This type of equipment converts mechanical energy into internal or kinetic energy in the fluids it handles. An example is a compressor, which is responsible for raising the pressure and moving gas. Couper et al. [2005] describe a variety of process equipment in detail.

Continuous and batch processes

Chemical processes vary with regards to their operating mode. The two extremes are continuous and batch operation [Perry and Green, 1997].

Continuous processes are the most common in industry because they yield the highest volume of products. In these processes, materials flow continuously in streams, in and out of process equipment, and the processes should be at steady-state. Continuous processes are only stopped once or twice a year. As a result, their mix of products is inflexible. Examples of continuous processes are petrochemicals, refineries, steel and paper making. All case studies analysed in this thesis pertain to continuous processes.

Batch operation, on the other hand, allows for a wider variety of products. In batch processes, a piece of equipment is loaded with a batch of material, processes the batch for a specified time, and is then emptied. To produce a different mix of products, operators may change the next batch of materials, or the residence times in the equipment. These processes are inherently unsteady state. Batch processes are often found in the pharmaceutical industry.

2.1.2 The electrical utility

Some unit operations in chemical processes require services from auxiliary systems. These are known as utilities, and mostly exchange energy with the process streams. There is rarely a transfer of mass between process and utility streams. An example of a utility is cooling water, which can be used in the cooling of process streams and rotating equipment.

The electricity supply is another example of a utility. It can be used as an alternative source of thermal energy, through resistances in electrical heaters, and as source of power to instruments and controllers. Its greatest use, though, is to power electric motors [Waide and Brunner, 2011]. Electric motors are used in processes to produce mechanical energy, and drive a variety of process machinery, such as compressors, pumps, mixers and extruders.

The use of electrical energy can be significant in chemical processes due to heavy process machinery. For example, large-scale gas handling processes, like those producing and supplying natural gas, have around 10 compressors, and each compressor can require up to 80 MW of power [Devold, 2009]. When driven by electrical energy, this power is supplied through the electric motor, which connects to the electricity supply, sometimes through an electrical drive.

2.1.3 Compressor, electric motor, and electrical drive

Figure 2.1 shows the set-up of a compressor, with a motor and electrical drive.

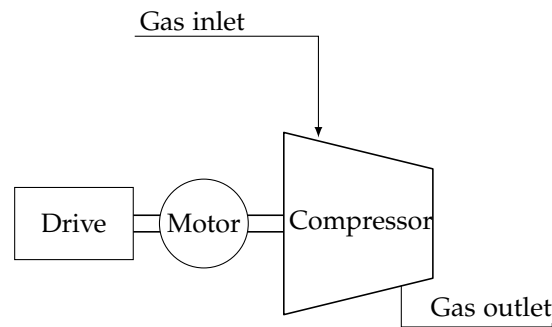


Figure 2.1: Generic schematic for a compressor, electric motor, and electrical drive system.

A compressor is a mechanical device with moving parts, which acts on a gas in order to increase its pressure, and to circulate it [Hanlon, 2001]. Compressors are differentiated by the operation of their moving parts. The case studies analysed in this thesis involve the centrifugal type. When gas enters a centrifugal compressor, rotating blades accelerate the gas while it moves forward along the compressor. The gas then flows through a chamber where it loses velocity. Bernoulli's principle explains that the reduction in rotational kinetic energy causes the pressure of the gas to rise. Descriptions of other types of compressors, their operation, and uses can be found in Hanlon [2001].

The moving parts of the compressor can be driven by an electric motor, specifically by means of its shaft. Most motors convert electrical energy into mechanical energy by electromagnetic principles [Hambley, 2010]. The fundamental principle is described by Lorentz's law, which explains that the interaction of a magnetic field and an electric current gives rise to a force. This force then compels the motor shaft to rotate. Motors draw current from a power supply. This current can be alternating (a.c.) or direct (d.c.). The former varies over time in magnitude and direction, typically in a sinusoidal waveform but also in square or triangular waveforms. The latter has a single direction.

Either the speed or the torque of the electric motor can be controlled. For instance, speed regulation is used to vary the amount of gas handled by the motor-driven compressor, in order to meet changing process targets. This control is achieved by power electronics units known as electrical drives. The electrical drive featured in this thesis varies the speed of a motor by changing the frequency of the applied voltage [Dorf and Svoboda, 2010].

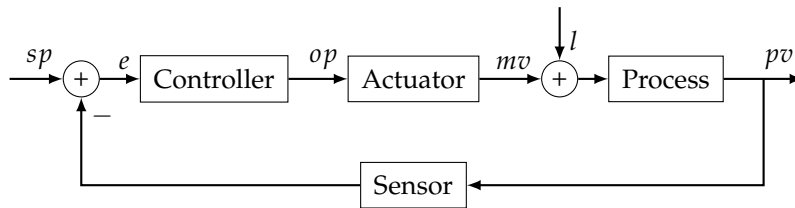


Figure 2.2: Block diagram of a feedback control loop.

Section 2.3 gives a detailed explanation of the principles of operation of these devices, as well as their theoretical models.

2.2 Control and monitoring in industrial processes

Chemical processes are automated so that they operate at or near optimum performance. At a lower level, this is achieved by regulatory process control. On a higher level, process monitoring supervises the process conditions to detect unacceptable deviations. This section first explains the basic principles of regulatory process control. Then process monitoring is introduced. This is a brief introduction because the discipline of process monitoring and diagnosis is explained in detail in chapter 3. Finally, this section discusses how measurement data is collected and stored in the control and monitoring systems.

2.2.1 Regulatory process control

The objective of regulatory process control is to maintain process conditions at or near target values, regardless of the adequacy of the targets. The targets are provided by operators, or by higher levels of control.

Closed-loop feedback control

Figure 2.2 shows the block diagram of a closed-loop feedback control system.

The process condition under control is named process variable (pv). This is a physical quantity such as flow, level, temperature, or pressure. Its target is referred to as set-point (sp). The process variable can deviate from the set-point because of a varying input into the controlled system, known as load (l), or because of changes to the set-point. In control terminology, deviations of

the process conditions from their target values are referred to as disturbances [Couper et al., 2005].

As represented in Figure 2.2, the process variable is measured by a sensor, and the measurement is converted into a signal by a transmitter (not shown in the figure). This signal is fed back and compared to its set-point. A controller then processes the resulting error (e) with a specific algorithm. The controller output (op) eventually commands an actuator, which is responsible for adjusting the manipulated variable (mv). The actuator is in most cases a control valve, but can also be a pump, a compressor or a heating element. The dynamic response of the process modifies the process variable, hence closing the loop. In this way, control transfers the deviations in the process variable to the manipulated variable, which should be associated with lower cost.

In a single control loop, as shown in Figure 2.2, the controller output is sent directly to the actuator. In cascaded control loops, the output signal is instead used as the set-point of a second controller. Some case studies analysed in this thesis include examples of cascade control. One example is that of the electrical drive that adjusts the speed or the torque of the electric motor. The electrical drive is a controller which receives its set-point from a process controller. The latter is usually a flow or pressure controller. The actuator in these cascaded loops is the compressor, which acts on the pressure of the process gas.

Control algorithms

For feedback control, the most widespread set of controller algorithms is that of P, PI, and PID controllers [Couper et al., 2005]. Equation (2.1) shows the function of a PID controller.

$$op = K_C \left(e + \frac{1}{\tau_I} \int edt + \tau_D \frac{de}{dt} \right) \quad (2.1)$$

It indicates that the controller output signal op is the result of three terms applied to the error variable e : a proportional (P), an integral (I) and a derivative (D) term. Constants K_C , τ_I and τ_D are known as controller gain, integration time constant, and derivative time constant. The P, and PI controllers include only the associated terms.

Comprehensive explanations of PID controllers and tuning of their parameters can be found in Seborg et al. [2010]. The same reference also discusses

more advanced control methods, such as feedforward control, model predictive control, control of batch processes. Section 2.3 will briefly explain the control algorithm used in electrical drives, which is also more advanced than PID control.

2.2.2 Monitoring and diagnosis

Process control can reduce the magnitude and the duration of disturbances in controlled variables, but it may not completely eliminate them. Therefore, industrial systems sometimes include monitoring and diagnosis tasks. Their aim is to detect remaining disturbances which are considered unacceptable, and identify their causes. In contrast to automated process control, process monitoring and diagnosis are mostly manual tasks done by operators, process and control engineers [Venkatasubramanian et al., 2003c].

Similar practices are also conducted for electrical and mechanical equipment [Jardine et al., 2006]. The objectives are to prevent machine breakdowns, schedule maintenance in a cost-effective manner, and create logs of failures. Examples of mechanical variables monitored are bearing vibration and shaft speed. Examples of electrical variables monitored are voltage and current [Martin, 1994].

Chapter 3 will address the state-of-the-art in methods for process monitoring and diagnosis. It will also comment on methods for monitoring machinery and electrical equipment.

2.2.3 Data characteristics in control and monitoring systems

Industrial controllers are computers and hence use digital technology. Commonly, these are programmable logic controllers (PLC) and distributed control systems (DCS) [Siemens Energy & Automation, Inc., 2007]. The former are mainly configured for operating individually, whereas DCS are configured as large networks of controllers, input and output modules, human-machine interfaces, and other applications [ABB, 2012].

Due to using digital technology, control algorithms and measurements have to be discretized. Discrete measurements have some time interval between samples. Measurements from physical variables are converted from analog to digital signals by the transmitter, or by a converter before the controller [Perry and Green, 1997]. This data is used by the regulatory control

system, but not stored. Common sampling intervals in process control systems are between 0.1 s and 1 s.

Data acquisition systems are built on top of the regulatory control system. They log the data used by the regulatory control system, display it in the control room, and store it in a data historian. Data is stored in historians at time intervals larger than the original sampling intervals. In process systems, this is normally above 5 s [Perry and Green, 1997].

Before storing, data may also be quantised, and compressed [Thornhill et al., 2004]. The values of quantised measurements are rounded, whereas compression discards certain samples. These procedures reduce the information available in the data. Kadlec et al. [2009] and Ge et al. [2013] have discussed other important characteristics of historical data in process systems.

Historical data is used at higher management levels, for example for periodic reports, short- and long-term planning. In the context of this thesis, historical data is relevant because most monitoring methods make use of it.

2.3 Detailed models of the compression system and control

This section focuses on fundamental theory of the system formed by the compressor, electric motor and electrical drive. As explained before, this system represents a good reference case for this thesis because it generates the types of measurements to which the thesis applies. This section presents the mathematical models of the compressor and motor under transient conditions, and the control algorithms of the electrical drive. The purpose is to have a qualitative model of this system, in order to interpret and compare the results which section 9.1 will present. The method discussed in that chapter aims to identify the propagation path of a disturbance. The theory in this section indicates the expected propagation path.

2.3.1 Characteristics of the system analysed

The results in section 9.1 were obtained with a system with the characteristics indicated in Table 2.1. These characteristics are relevant because they determine which models and algorithms to present.

Table 2.1: Characteristics of the compressor-motor-drive system analysed.

System	Characteristics
Compression system	Centrifugal compressor Downstream tank and outlet valve
Electric motor	Three-phase induction motor with two poles
Electrical drive	a.c. voltage-source inverter drive Direct torque control (DTC)

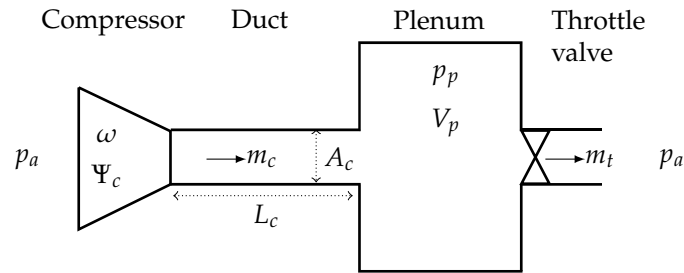


Figure 2.3: Compression system with relevant state variables and parameters, adapted from Greitzer [1976].

2.3.2 Compressor model

Figure 2.3 represents a basic compression system consisting of a compressor, a duct, plenum volume and a throttle valve, as used by Greitzer [1976].

Gravdahl et al. [2002] presented the set of equations that describe this system, for varying rotational speed. These derived from the works in compressor surge modelling by Greitzer [1976], Fink et al. [1992], Gravdahl and Ege-land [1999], and are the basis for most studies of compressor stability, surge and control.

$$\frac{dp_p}{dt} = \frac{a_a^2}{V_p} (m_c - m_t) \quad (2.2)$$

$$\frac{dm_c}{dt} = \frac{A_c}{L_c} (\Psi_c \cdot p_a - p_p) \quad (2.3)$$

Equation (2.2) represents the dynamics of the plenum pressure and derives from its mass balance under certain assumptions. The state variables p_p , m_c , and m_t represent respectively the plenum pressure, the mass flow through the compressor and the flow through the throttle valve. Parameters V_p and

a_a represent, respectively, the plenum volume and the sonic speed at ambient conditions.

Equation (2.3) represents the dynamics of the mass flow through the compressor and derives from the momentum balance over the length of the compressor. Variable Ψ_c is the compressor characteristic, and is described below (equation (2.5)). Parameters A_c and L_c represent, respectively, the compressor duct cross section and length. Parameter p_a represents the pressure at the compressor inlet. The model assumes this is the ambient pressure.

The flow through the throttle valve, m_t , can be considered a steady state function of p_p according to equation (2.4). Parameter $k_t > 0$ is proportional to valve opening.

$$m_t = k_t \sqrt{p_p - p_a} \quad (2.4)$$

The compressor characteristic Ψ_c describes the compressor pressure ratio, p_p/p_a , and can be considered a steady state nonlinear function of the compressor mass flow m_c and shaft speed ω . It is usually approximated to a third order polynomial on the mass flow (2.5a), with the coefficients being a second or third order polynomial of the speed (2.5b).

$$\Psi_c = c_0 + c_1 m_c + c_2 m_c^2 + c_3 m_c^3 \quad (2.5a)$$

$$c_i = c_{i,0} + c_{i,1} \omega + c_{i,2} \omega^2 + c_{i,3} \omega^3 \quad (2.5b)$$

The speed of the shaft of the compressor, ω , is a result of the balance of forces acting on that shaft. On the compressor side, a load torque τ_L is created by the fluid flow on the compressor impeller. This torque is approximated by Euler's pump equation (2.6). Parameter r denotes the impeller radius, while σ is known as slip factor and approximates how the fluid moves circumferentially relative to the rotating blades.

$$\tau_L = \sigma r^2 \omega |m_c| \quad (2.6)$$

The force rotating the shaft is induced by a motor, mounted on the same shaft as the compressor. This is indicated in the next section, in equation (2.8). The balance of forces in the shaft is indicated in equation (2.9).

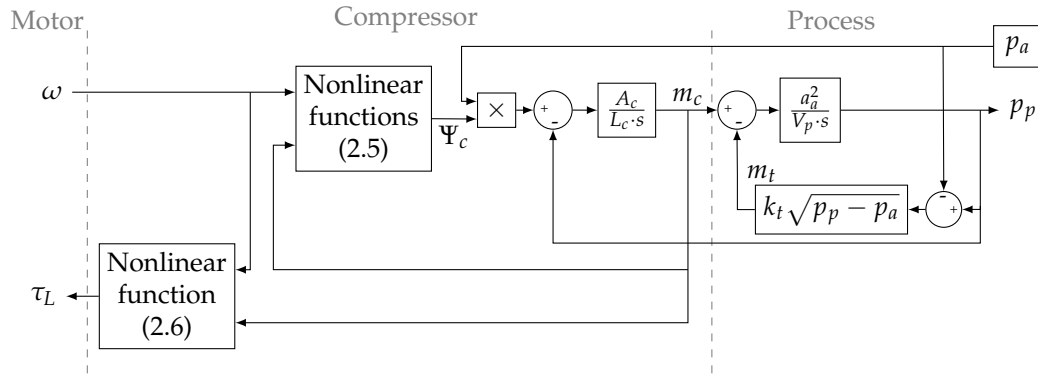


Figure 2.4: Block diagram for a centrifugal compressor.

Figure 2.4 shows the block diagram corresponding to the model of the compression system. The transfer functions are the Laplace transforms of the equations indicated in this section. The purpose of such diagram is to highlight the cause-and-effect relationships between the variables in the system.

2.3.3 Induction motor model

The drawing in Figure 2.5 represents a simplified cross-sectional view of an asynchronous induction motor. The figure shows the two main components of the motor: the stator, and the rotor [Hambley, 2010]. The stator is the stationary outer part of the motor. It is made of ferromagnetic materials and is wound with electric coils. The rotor is the rotating inner part of the motor, and is mounted on the motor shaft. It includes some type of conducting elements.

The dynamic behaviour of an asynchronous induction motor comprises its electromagnetic and mechanical subsystems. The set of equations presented here is comprehensively explained in Vas [1998] and Holtz [2002].

Electromagnetic subsystem

The coils in the stator are supplied with an a.c. voltage, \mathbf{u}_s . In three-phase induction motors, three separate inputs are supplied. Their waveforms lag each other by 120° . The supplied voltage produces an a.c. current, \mathbf{i}_s , which in turn generates a magnetic field.

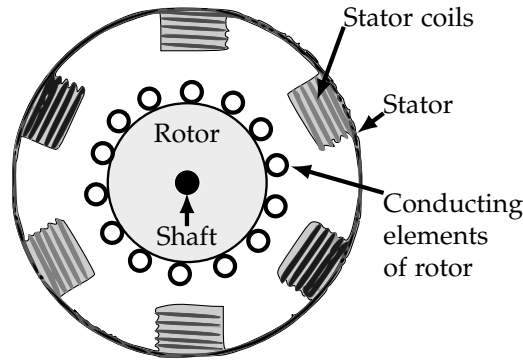


Figure 2.5: Cross-sectional view of an asynchronous induction motor showing the main components.

Figure 2.6 illustrates the fact that the direction of the magnetic field generated by the stator rotates. This is because the current in each coil varies in magnitude and direction, and the three currents have different phases. The angular speed of this rotating magnetic field is ω_s , and is an integer fraction of the supply frequency.

The motion of the stator magnetic field in relation to the rotor induces a current along the conducting elements of the rotor, \mathbf{i}_r . The rotor current, in turn, generates a second magnetic field. The interaction of the resulting magnetic field and rotor current creates a force. This force compels the rotor to rotate in the direction of the rotating stator magnetic field. The rotor rotates with angular speed ω , which is smaller than ω_s . This guarantees that the stator magnetic field keeps rotating in relation to the rotor.

The dynamic behaviour described above is modelled by the differential voltage equations of the stator (2.7a), and rotor (2.7b). The algebraic equations (2.7c) and (2.7d) complete the system.

$$\mathbf{u}_s = R_s \mathbf{i}_s + \frac{d\mathbf{\Psi}_s}{dt} + j\omega_s \mathbf{\Psi}_s \quad (2.7a)$$

$$0 = R_r \mathbf{i}_r + \frac{d\mathbf{\Psi}_r}{dt} + j(\omega_s - \omega) \mathbf{\Psi}_r \quad (2.7b)$$

$$\mathbf{\Psi}_s = L_s \mathbf{i}_s + L_m \mathbf{i}_r \quad (2.7c)$$

$$\mathbf{\Psi}_r = L_r \mathbf{i}_r + L_m \mathbf{i}_s \quad (2.7d)$$

Variables $\mathbf{\Psi}_s$ and $\mathbf{\Psi}_r$ are the stator and rotor flux linkages. Flux linkages are the total magnetic field passing through the surface of the stator coils, and

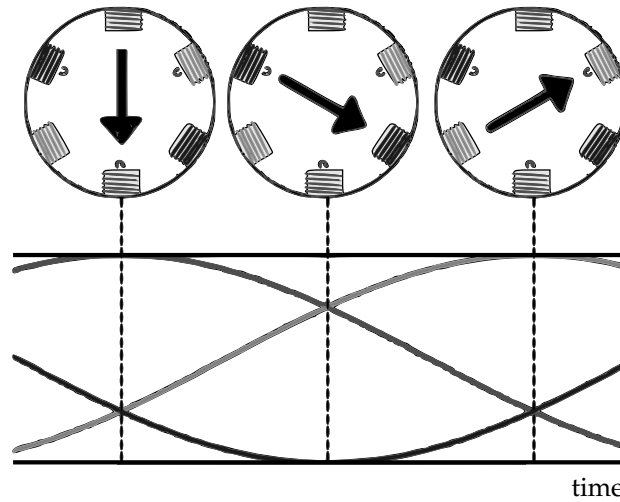


Figure 2.6: Time sequence showing, above, the rotation of the magnetic field generated by the stator (direction represented by the arrow). The rotation is due to variations in magnitude and direction, shown below, of the three-phase voltages supplied to the stator coils.

the conducting elements of the rotor. Parameters R_s , R_r are, respectively, the stator and rotor resistances. Parameters L_s , L_r , and L_m are the stator, rotor and mutual inductances.

The variables denoted by bold characters are space phasors. A phasor is a vector which represents a sinusoidal function. It rotates around the origin of a complex plane so that its projection onto the real or imaginary axes gives the value of the function at each moment in time. In equation (2.7b), j is the unit imaginary number. The norm and speed of rotation of the phasor represent, respectively, the amplitude and frequency of the sinusoidal function. Space phasors can combine as a sum the three sinusoidal components of the three-phase quantities. Therefore, they are compact representations for three-phase a.c. machines.

The complex plane where the phasors rotate is the reference frame. Equation 2.7 uses the frame rotating synchronously with the stator magnetic field.

As discussed, the interaction of the magnetic field with the rotor current creates a force which rotates the rotor. The tendency of that force to rotate the rotor is the electromagnetic torque, τ_e . Equation (2.8) is one form of describing the mechanism of production of τ_e , where P is the number of pole pairs in the machine. This relation links the electromagnetic subsystem of the motor with its mechanical subsystem.

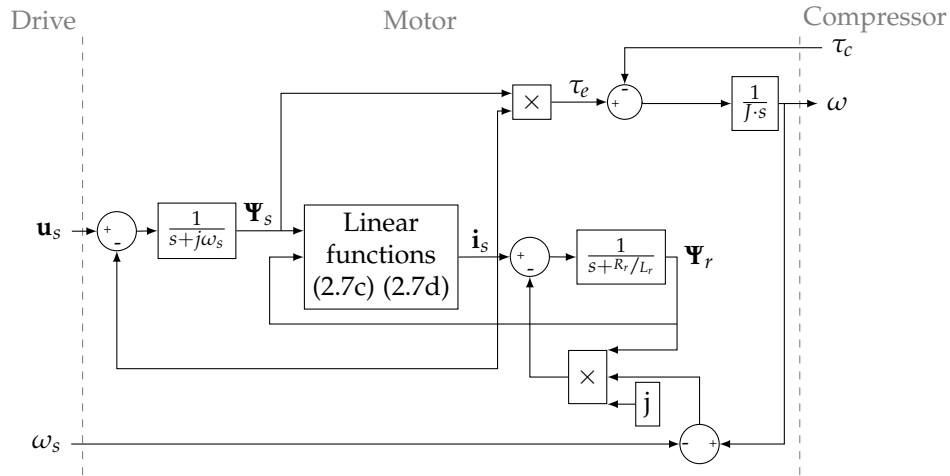


Figure 2.7: Block diagram for an asynchronous induction motor.

$$\tau_e = \frac{3}{2} P \Psi_s \times \mathbf{i}_s \quad (2.8)$$

Mechanical subsystem

The mechanical subsystem is modelled by the differential equation of motion (2.9), derived from the angular momentum balance of the rotating shaft. The angular speed of the shaft is a result of the angular momentum balance between the electromagnetic torque, and the torque produced by the load connected to the shaft, τ_L . Parameter J is the moment of inertia of the rotating parts.

$$\frac{d\omega}{dt} = \frac{1}{J} (\tau_e - \tau_L) \quad (2.9)$$

Figure 2.7 shows the block diagram corresponding to the model of this system. The transfer functions are the Laplace transforms of the equations indicated in this section. Multiplying constants were removed for clarity.

2.3.4 Drive control algorithms

An a.c. voltage-source inverter drive is a power electronics unit which converts the constant-frequency a.c. voltage from the line in to an a.c. voltage output of variable frequency and amplitude. The purpose of varying the frequency and

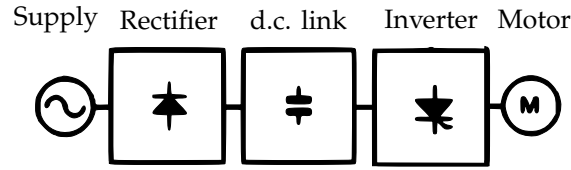


Figure 2.8: Main components of an a.c. voltage-source inverter drive.

amplitude of the voltage is to control the flux and the electromagnetic torque of the induction motor. The motor speed is controlled with an outer feedback loop, cascaded with the flux and torque controls.

Figure 2.8 shows the three main components of that electrical drive [Dorf and Svoboda, 2010].

A rectifier converts the supply voltage from a.c. in to a unidirectional waveform, hence d.c. Although unidirectional, this waveform is still pulsating. A d.c. link smooths the d.c. voltage until nearly constant. Finally, an inverter converts this d.c. voltage in to a quasi-sinusoidal a.c. voltage of desired amplitude and frequency. This is achieved through a series of switches. The final output is determined by the combination of switches used, or switching state, the frequency of switching, and the level of the voltage from the d.c. link. In the electrical drive featured in this thesis, switching is controlled by a Direct Torque Control (DTC) algorithm.

Direct Torque Control

DTC controls the stator flux linkage and torque directly and independently. Additionally, DTC can be part of a speed control loop, as mentioned before. Commonly, none of the controlled variables is measured, but instead estimated. The only measured variables are the stator current, i_s , and the d.c. link voltage, $U_{d.c.}$.

Using the d.c. link voltage and the switching states of the inverter k , the stator voltage is reconstructed according to equation (2.10).

$$\mathbf{u}_s = \frac{2}{3} U_{d.c.} e^{j(k-1)\pi/3} \quad (2.10)$$

The stator current and reconstructed stator voltage are then used to estimate the stator flux linkage $\hat{\Psi}_s$ and the electromagnetic torque $\hat{\tau}_e$ in real time. The estimator uses the model of the induction motor, which was discussed in section 2.3.3. Vas [1998] discusses in detail techniques of estimation.

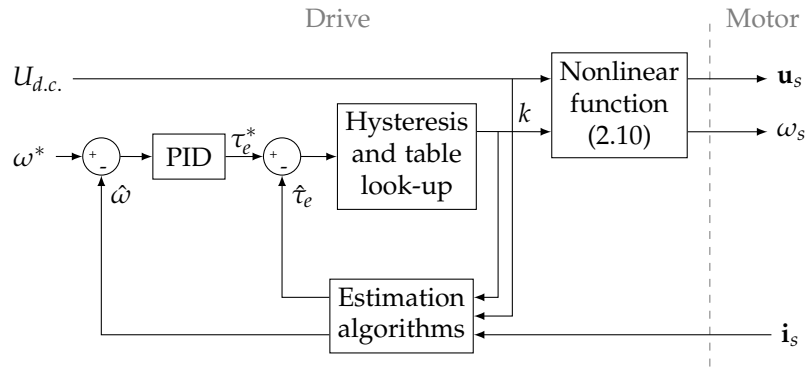


Figure 2.9: Block diagram of the DTC algorithm in an a.c. voltage-source inverter drive.

The torque control loop uses the estimated electromagnetic torque $\hat{\tau}_e$, and set-point τ_e^* . This is provided externally or by the outer speed control loop. The control algorithm aims at restricting the error within a hysteresis band. The manipulated variables are the switching states for the inverter, which are selected from an optimum switching table. The resulting stator voltage follows (2.10).

Speed control requires an additional PID controller based on the rotor speed error ($\omega^* - \omega$). The star superscript denotes the speed set-point, or reference. In the electrical drive analysed, the speed of the motor is estimated from the stator current \mathbf{i}_s , and estimated stator flux linkage $\hat{\Psi}_s$ and torque $\hat{\tau}_e$. Vas [1998] discusses also several techniques for speed estimation. The speed control loop outputs the torque set-point τ_e^* for the inner torque control loop. In the electrical drive analysed, the estimation cycles of both control strategies run each 25 ms.

Figure 2.9 shows the corresponding block diagram of the DTC algorithm.

2.4 Disturbances

Deviations of the process conditions from their target values compromise the productivity, safety, and energy efficiency of chemical processes. This section focuses on these deviations. It starts by introducing the terminology used in the literature, and in this thesis. Then, it discusses types and causes of these deviations.

Table 2.2: Use of disturbance-related definitions and terminology in relevant research areas.

Definition	Term	
	Disturbance	Fault
Deviation in variable	Process control [Couper et al., 2005]	Process monitoring [Venkatasubramanian et al., 2003a, Isermann, 2005, Himmelblau, 1978]
	Process monitoring [Gertler et al., 1999]	
	Plant-wide process monitoring [Thornhill and Horch, 2007]	
	Monitoring of electrical disturbances [Bollen, 2000]	
Input variable	Process control [Perry and Green, 1997, Seider, 2011]	Process monitoring [Isermann, 2005]
	Process monitoring [Isermann, 2005]	

2.4.1 Terminology

Disturbance

There are different definitions of disturbance in the literature. Table 2.2 indicates the two main concepts which have been used. The purpose of the table is to highlight the research areas and authors that have used each definition.

In the context of process control, Couper et al. [2005] defines disturbances as deviations of the process variables from their set-points. Conversely, Perry and Green [1997], Seider [2011] use the term disturbance to denote the input variable which causes the deviations of the process variables. This is referred to as load variable in Figure 2.2.

In the process monitoring literature, the term *disturbance* also refers to those two concepts. Additionally, the process monitoring literature also uses the term *fault*. This is used to refer to unpermitted deviations in variables of the system [Venkatasubramanian et al., 2003a, Isermann, 2005, Himmelblau, 1978].

This thesis uses the term *disturbance* as a deviation of the variables of the system from their desired values or trends. The reason is to follow the terminology in plant-wide disturbance analysis [Thornhill and Horch, 2007], which is the research area of this thesis. Furthermore, the term *disturbance* is used

Table 2.3: Common changes in a propagating disturbance due to dynamic characteristics of the process system.

Change	Underlying dynamic characteristic
Time lag between the disturbance in the measurements of two variables	Process dead time
Low pass filtering, <i>i.e.</i> smoothing of the disturbance trend	Process time constant
Decrease in the disturbance magnitude	Process gain smaller than one
Addition of noise	Measurement noise or outside influences

with equivalent meaning in electrical systems. Specifically, Bollen [2000] define power quality disturbances as deviations of the voltage or frequency in the power supply from their ideal behavior.

Root cause and propagation path

Root cause denotes the malfunction which originates the disturbance in the system [Himmelblau, 1978].

In a process system, when a disturbance originates at the root cause, it propagates widely through mass and energy streams, and signal paths [Thornhill and Horch, 2007]. The term *propagation path* denotes the directed succession of variables of the system which reflects the order of propagation. The path may not be uni-directional, for example if a recycle is involved in the propagation [Bauer and Thornhill, 2008].

The disturbance changes when it propagates along the process, due to the dynamic characteristics of the system. Table 2.3 indicates four changes which are commonly observed.

Plant-wide disturbances

Due to propagation, a disturbance from a single root cause can affect several variables throughout the plant. These disturbances are referred to as *plant-wide* disturbances [Thornhill and Horch, 2007].

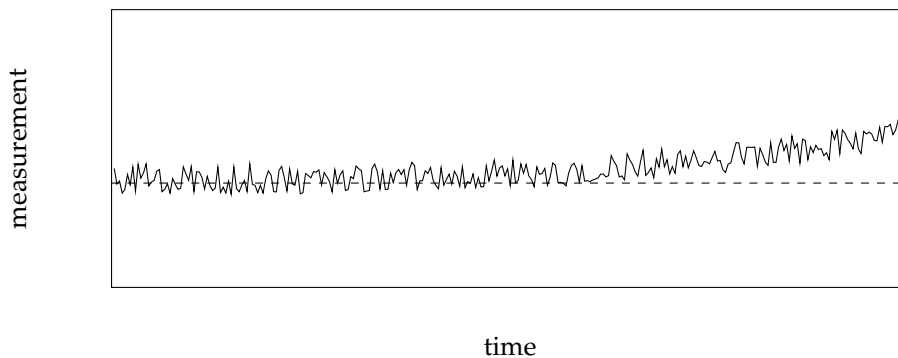


Figure 2.10: Example of a measurement with a slowly developing disturbance drifting from the normal operating value (dashed line).

2.4.2 Time scales

The review paper by Thornhill and Horch [2007] distinguishes three classes of time scales of process disturbances.

Slowly developing disturbances

One class comprises slowly developing disturbances. Figure 2.10 gives an example of a measurement with this type of disturbance. The measurement of a variable with such a disturbance shows a long-term trend, drifting from the desired values. The undesired condition is not so much the values of the measurement, as its trend. The reason is that the values are not significantly far from their target, especially in the initial stage of the disturbance. The drifting, however, is a symptom of problems which compromise safety, energy-efficiency, and productivity. Examples are, respectively, machinery wear, fouling in heat exchangers, and catalyst degradation.

Persistent disturbances

A second time scale of disturbances involves persistent, or repeating, disturbances. These are disturbances in which a deviation episode repeats tens to hundreds of times. There is no specification for the shape of the episode, or the frequency of repetition, as Figure 2.11 shows. Some episodes are oscillations (panel 2.11a), others have spiky features (panel 2.11b). Also, the episodes may repeat periodically (panel 2.11a), or at irregular intervals (panel 2.11b).

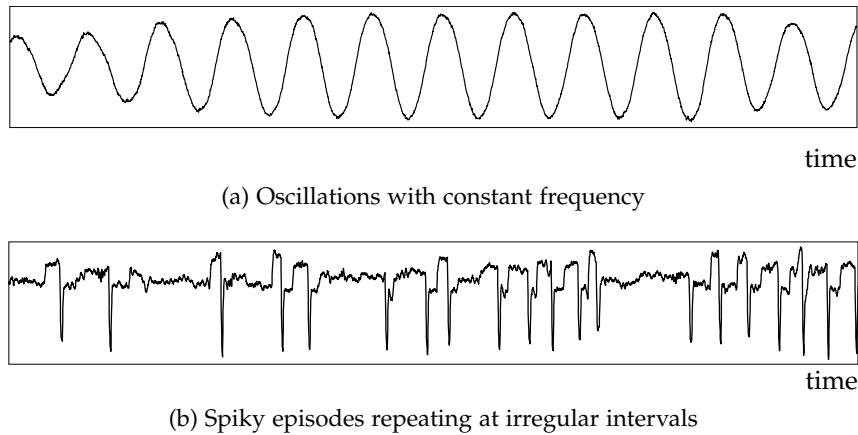


Figure 2.11: Example of measurement with persistent disturbances.

Transient disturbances

Thornhill and Horch [2007] suggest the third time scale to comprise abrupt disturbances, such as a compressor trip. This thesis considers a broader class, which involves transient disturbances.

Transient disturbances are here defined as short-lasting deviations of a measurement from its previous and subsequent trend. In addition, the deviation seldom repeats within the time horizon of analysis. The reason to use this definition is that the time scale and timing of the disturbance are more important than the shape of the disturbance.

The definition proposed includes less comprehensive types of disturbance mentioned in the literature. For instance, it includes spikes, as defined by Misra et al. [2002]. The authors define a spike as an abrupt change which maps to low-scale wavelet coefficients of significantly higher amplitude than the non-transient parts of the measurement. Other examples of transient disturbances were presented in Figure 1.3, in the Introduction.

Disturbances in the electricity supply

Standards EN 50160 [European Commission, 2002] and IEEE Standard 1159 [IEEE, 1995] classify disturbances in the electricity supply according to the duration, amongst other criteria. An electrical disturbance is in general considered of long duration when it lasts around one to three minutes, and of short duration when it lasts less than one second. In the time scale of process systems, these would all be considered short-lasting disturbances. This

suggests that most electrical disturbances would be considered as transient disturbances in the classification used for process disturbances.

2.4.3 Common root causes

Disturbances in the chemical process

Causes of disturbances in chemical processes may be grouped into three types: process, control system, and malfunctioning equipment.

Process malfunctions vary with the configuration of the process. These may derive from:

- the structure and design of the process: an example is a recycle stream, which is a physical feedback loop of mass and energy and thus can generate oscillating disturbances [Bauer and Thornhill, 2008];
- unstable operating conditions: one example is compressor surge, where pressure and flow oscillate with high frequency and amplitude [Gravdahl and Egeland, 1999]; another example is slug flow, in which pockets of gas are interspersed with liquid in a pipeline, creating a pulsating flow;
- degradation of operating conditions: one example is the degradation of the catalyst in a reactor [Maurya and Venkatasubramanian, 2007], which decreases the rate of conversion; another example is fouling in a heat exchanger, which compromises the desired outlet temperature.

The origins of a malfunctioning control system include [Thornhill and Horch, 2007]:

- poor tuning, which can cause persistent cyclic oscillations in feedback loops;
- controller interaction, or coupling, which happens when control loops compete for the same process variable;
- malfunctioning sensors and valves: the former may create noisy measurements, or assign values which are not the real state of the variable; the latter may lead to cyclic behaviour in the measurement of the following flow variable.

Table 2.4: Causes of electrical disturbances according to their location.

Location	Cause
Power generation	Loss of generation
Power distribution	Lightning strikes; bird strikes; Equipment failure on the distribution grid
Management of loads	Switching of heavy loads; Loads with poor power factor or lack of reactive power support; Load variations; Network overload

Chemical processes may also be affected by malfunctioning process equipment. The origins of malfunctions may be:

- mechanical wear [Jardine et al., 2006], which may result in cracks, misalignment and looseness of parts, fouling, insulation failures, and machine breakdowns;
- inadequate electric power conditions to electrically-driven equipment [Bullis, 1996], which may result in overheating, variations in torque produced, or the equipment being tripped for protection.

Disturbances in the electricity supply

Electrical disturbances may enter an industrial site from the generation and the distribution of power. They may also start internally, resulting from the management of electrically-driven equipment [Edomah, 2009], such as compressors and furnaces. Table 2.4 indicates some causes of electrical disturbances according to their location.

Symptoms of disturbances in the electricity supply include deviations in voltage and frequency, and distortions of the waveform by harmonics. The most common types of disturbances in industrial sites are voltage dips, and harmonic distortion [Bendre et al., 2004, Saksena et al., 2005]. Voltage dips are also known as sags in American documentation. Voltage dip is a reduction in voltage lasting around 10 ms to 1 s [European Commission, 2002].

Table 2.5: Case studies used in each technical chapter, distinguished by their purpose. The relevant characteristics are also indicated.

Chapter	Characteristics	Development case study	Test case study
4	Transient disturbances	Compressor rig case 1	Gas plant case 1
5	Transient disturbances	Compressor rig case 2	Gas plant case 2
7	Masked by other trends Measurements with equal sampling rate		
6	Transient disturbances Masked by other trends Measurements with different sampling rate	Compressor rig case 2 (with some variables downsampled)	Turbocharger case

2.5 Case studies analysed in this thesis

The methods proposed in this thesis were developed and tested on five case studies. Two of the case studies were used in developing the new methods, and three in testing. Different case studies are needed because the methods in this thesis address different types of disturbances, and data with different sampling conditions.

Table 2.5 indicates the case studies used in each technical chapter. It also indicates the characteristics of the case studies which justify their use in the chapter.

The names of the case studies in table 2.5 refer to the origin of the data. This includes (i) experimental work with a gas compressor rig, (ii) routine operation of an industrial process of gas processing, and (iii) performance tests of a commercial ABB turbocharger compressor. The remainder of this section describes these three origins.

2.5.1 Compressor rig

The *Compressor rig* case studies consist of measurement data from experimental work with a gas compressor rig located at ABB Corporate Research Center, Kraków, Poland. The rig is shown schematically in Figure 2.12, which also shows variables measured in the process system. A data acquisition module in the drive measured other variables related to the drive and the motor. Of

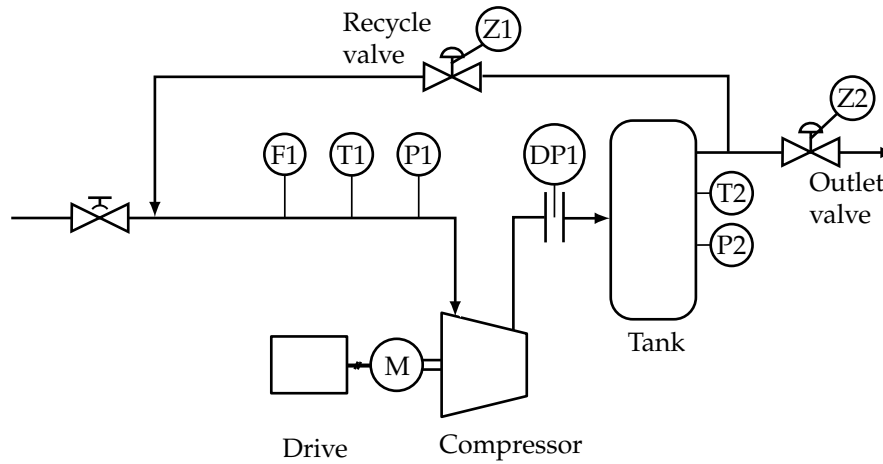


Figure 2.12: Simplified schematic of the gas compression experimental rig.

Table 2.6: Variables measured by the data acquisition module in the drive. Only variables used in this thesis are indicated.

Tag, as used in the thesis	Variable
S1.sp	shaft speed set-point
S1	shaft speed
N1	motor torque (estimated by drive)
Ia, Ib, Ic	Phase currents to motor stator

these variables, table 2.6 indicates those whose measurements are used in this thesis.

The main components of the process system are the centrifugal compressor, and the plenum volume which is provided by the tank. The process also includes valves and sensors. The main electrical and mechanical components include the transformer, the a.c. variable speed drive and the electric motor which is directly coupled to the compressor.

The reason to use this rig is that it reproduces the system compressor-motor-drive, which is a relevant link between chemical processes and the electricity supply. The experiments consisted of inducing a variety of electrical disturbances in the rig. Appendix C presents a full description of the experimental facilities and procedures.

All methods proposed in this thesis were developed with case studies from the *Compressor rig* data set. Furthermore, the tests allowed to demonstrate an idea for a future signal analysis method, which is presented in section 9.1. This method aims to identify the propagation path of a disturbance. With

this experimental set-up, the root cause of the disturbance is known, and established models of the system indicate the expected propagation path. In future work, the method suggested in section 9.1 may be validated with the *Compressor rig* data set.

Data characteristics

The monitored variables were measured and stored synchronously at a 5 kHz rate.

Disturbances

The data sets from the compressor rig include periods with persistent and transient disturbances. The persistent disturbances include persistent oscillations, and repeating pulses. The transient disturbances are step-like deviations.

2.5.2 Gas plant

The *Gas plant* case studies consist of measurement data and process schematics from a gas processing plant. These were provided courtesy of ABB Oil, Gas and Petrochemicals, Oslo, Norway.

The plant is fed through pipelines from a natural gas well [Devold, 2009]. The feed is a mixture of natural gas and heavier hydrocarbons. The latter may include ethane, propane, butane, iso-butane, and natural gasoline, and are collectively known as natural gas liquids (NGL), or condensate. Natural gas must be sold with high level of purity. Condensates are used for enhancing oil recovery in oil wells, and as diluents for heavy crude. Therefore, the gas processing plant separates the condensates from the natural gas, and pressurizes the gas to be exported.

The process can be classified as large-scale since it is formed by 35 areas, described by more than 300 schematics, and monitored by measurements from more than 7000 variables. Each area in the process is responsible for a main task. Examples are the gas-condensate separation area, and the gas recompression area.

No process schematics are presented in this section of the thesis. The reason is that not all case studies from the gas plant refer to the same area

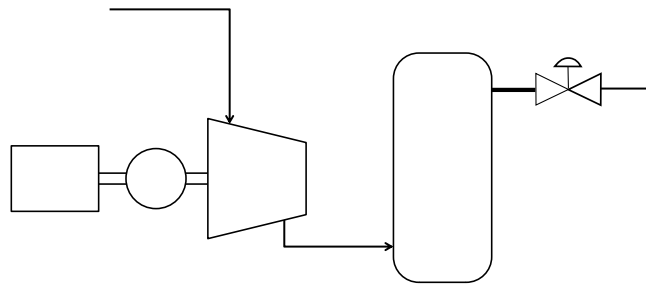


Figure 2.13: Illustrative schematics for a turbocharger performance test.

of the process. The relevant schematics will be shown in the sections of the thesis where the corresponding case studies are analysed.

Data characteristics

The measurements provided to the work of this thesis include process, mechanical and electrical variables. They originate from the same data historian, where all are stored with 30 s intervals between samples.

Disturbances

The data sets from the gas plant include periods with persistent and transient disturbances.

2.5.3 Turbocharger tests

The *Turbocharger* case study consists of measurement data from performance tests to a commercial turbocharger compressor. This data was provided courtesy of ABB's Corporate Research Centers in Kraków, Poland and in Baden-Dättwil, Switzerland, and ABB Turbo Systems, Baden, Switzerland. No schematics were provided, but Figure 2.13 illustrates the generic equipment layout for turbocharger tests.

The purpose of the performance tests was to obtain the compressor map. This map describes the performance of the compressor in steady-state operation, in terms of pressure rise across the compressor, mass flow and rotational speed [Hanlon, 2001].

The tests consist of reducing the gas flow step-wise by closing the valve downstream. The compressor is kept at constant speed. These conditions lead to an increase in the pressure ratio across the compressor. When steady state

is reached, the stable operation point is registered. This is repeated for several rotational speeds.

The motor used in the tests was a three-phase asynchronous motor, and was controlled by an a.c. variable speed drive.

Data characteristics

The electrical drive was equipped with a data acquisition module which collected measurements from the drive and the motor. Measurements from process variables were collected by a different module. The sampling rates in the two modules were different. Specifically, the drive module sampled each 0.1 s, and the process module each 5 s. Figure 1.4, in the Introduction, showed two measurements from this source to illustrate the difference in sampling rates. It should also be noted that the clocks of the two modules were not synchronized. Prior to the analyses presented in this thesis, the measurements were aligned by visual comparison of some abrupt episodes.

Disturbances

The data sets from the turbocharger tests include periods with transient disturbances. There are also periods when measurements were affected by oscillations. However, the number of cycles of oscillation is not enough to consider these as persistent disturbances.

2.6 Chapter summary

This chapter discussed the chemical process, its rotating machinery, electromechanical equipment and electrical supply. These are the systems that generate the measurements to which this thesis applies.

The chemical process was introduced in section 2.1 as a series of unit operations to transform feed materials into products. The materials flow between process equipment, in which reactions, separations and transfers of energy can occur. The electrical supply provides energy for chemical processes, and in particular rotating machinery such as compressors.

Section 2.2 explained that chemical processes are automated so that they operate at or near optimum performance. At the lower level, this is achieved by regulatory process control. Control attempts to maintain the conditions of

a system at some target values. On a higher level, process monitoring and diagnosis analyse deviations from those targets, and hence enable corrective actions. Control as well as monitoring and diagnosis are based on measurement data. Measurements are sampled from the process variables and stored in a data historian. The signal analysis methods proposed in this thesis exploit the measurements in the data historian.

The measurements used in this thesis derived from gas-handling processes, driven by compressors. The compressors were powered by electric motors, and the motors were controlled by electrical drives. Therefore, section 2.3 presented the fundamental theory of these systems. The section discussed the mathematical models of the compressor and motor under transient conditions, and the control algorithms of the electrical drive. The reason to do that was to have a qualitative model of the system, in order to interpret and compare the results which section 9.1 will present.

Section 2.4 defined disturbance as a deviation of the variables of the system from their desired values, or trends. Other terminology was also introduced. This section also distinguished three time scales of process disturbances, and observed that electrical disturbances would mostly be considered as transient disturbances in those time scales. The same section also presented common root causes of disturbances in the chemical process, and the electricity supply.

Section 2.5 introduced the five case studies used in this thesis, and the reasons why they were chosen. The case studies originated from (i) experimental work with a gas compressor rig, (ii) routine operation of an industrial gas plant, and (iii) performance tests of a commercial ABB turbocharger.

Chapter 3

State-of-the-art in process monitoring and diagnosis

Process monitoring and diagnosis consists of methods to detect disturbances in the operation of processes and to diagnose the root causes. Chapter 1 introduced the new challenges which derive from extending process monitoring and diagnosis to the electrical and mechanical utilities. The aim of the current chapter is to conclude on the types of methods which are most suitable to address those challenges. This is done by providing a broad survey of the field of process monitoring and diagnosis, and discussing the advantages and disadvantages of each type of method. The discussion in this chapter will argue that advanced signal analysis methods are the most suitable to meet the objectives of the thesis.

The chapter is laid out as follows. Section 3.1 describes the three stages of process monitoring and diagnosis: (i) treatment of data, (ii) disturbance detection, and (iii) disturbance diagnosis. Section 3.2 presents the hierarchical trees which structure the methods of process monitoring and diagnosis and which will guide the literature review in the next sections. The top level of the hierarchical trees distinguishes between methods based on (i) analytical models, (ii) qualitative models, (iii) qualitative trends, and (iv) numerical data. Sections 3.3, 3.4, 3.5, and 3.6 review each of the four types of methods. Additionally, the sections discuss the relevance of the methods to machinery and electrical equipment, as well as the suitability of the methods for the work in this thesis.

3.1 Stages of process monitoring and diagnosis

Process monitoring and diagnosis may be divided into three stages: (i) treatment of data, (ii) disturbance detection, and (iii) disturbance diagnosis. This section describes these stages, and gives references to research and review papers. This section also describes other tasks which are closely related to process monitoring and diagnosis.

3.1.1 Treatment of data

Process monitoring and diagnosis depends extensively on measurement data. Therefore, the objective of data treatment is to ensure that measurement data has adequate characteristics for the desired application.

Data scaling

Data scaling is a treatment to make measurements independent of the engineering units of the physical variables. Normalization is a common data-scaling procedure, and is given by equation (3.1). The normalized value z_i of a measurement is obtained by subtracting the mean \bar{x} of all measurement values from the original value x_i , and then dividing the result by the standard deviation s_x of all measurement values.

$$z_i = \frac{x_i - \bar{x}}{s_x} \quad (3.1)$$

Data selection

Data treatment may also involve the selection of appropriate variables and periods of measurement [Ge et al., 2013]. One objective of data selection is to match signatures in measurement data to normal operating conditions or different types of disturbances. This objective usually requires the identification of steady-state operation in the measurements. Identification of steady-state is often performed manually [Kadlec et al., 2009], but Jiang et al. [2003] discuss several automated methods, and also suggest an approach based on wavelet transforms.

Replacement of undesired values

Data treatment may also involve detecting and replacing undesired values in the measurements. Undesired values may result from missing samples, incorrect values, or disturbances.

Missing samples and incorrect values are often caused by faulty sensors. The objective is to detect and replace those values by estimates of the true ones. The detection is commonly called gross error detection, or sensor validation [Venkatasubramanian et al., 2003a, Khatibisepehr et al., 2013]. The replacement is known as data reconciliation, or data rectification.

The replacement of disturbance-related values is often known as fault reconstruction [Qin, 2012, Ge et al., 2013]. Chapter 7 will develop a method to replace transient disturbances in the time series of an oscillating measurement. The new values are estimates of what the measurement would have been had the disturbance not been present. The replacement of disturbance-related values is important because subsequent analyses may be distorted by the presence of the disturbance in the measurement data [Jelali and Scali, 2010, Ge et al., 2013]. Furthermore, the replacement can lead to information about the disturbance, which can be helpful in methods that match signatures in measurement data to types of disturbances.

3.1.2 Disturbance detection

Disturbance detection involves determining that a disturbance is present in a measurement, and the time when that occurred [Russell et al., 2000]. More recently, disturbance detection has also included the task of clustering [Thornhill and Horch, 2007].

Disturbance detection is also known as fault detection, both in the process monitoring literature [Venkatasubramanian et al., 2003a], and in applications to machinery [Jardine et al., 2006] and electrical equipment [Nandi et al., 2005].

Presence and time

The earlier approach to determining the presence and time of a disturbance is known as Statistical Process Control (SPC). SPC is based on having upper and lower limits on the values of important measured variables. These limits define normal operating conditions, and beyond the limits a disturbance is

detected and alarms are raised for the operators [Wetherill and Brown, 1991]. The values of these limits are based on statistics which will be discussed in more detail in section 3.6. Sections 3.3 to 3.6 will discuss other approaches to determining the presence and time of a disturbance.

Clustering

More recently, disturbance detection has also included the task of clustering [Thornhill and Horch, 2007]. Clustering refers to identifying and grouping all measurements affected by the same disturbance. Clustering reveals the number of different disturbances present, and hence the number of root causes to be investigated. Additionally, clustering narrows down the candidate variables at the root cause of a disturbance to those variables in the corresponding cluster.

Clustering is associated with a plant-wide approach to disturbance analysis. Its importance to the process industry was highlighted in a study of industrial control systems by Desborough and Miller [2002]. A comprehensive review of this task can be found in Thornhill and Horch [2007].

3.1.3 Disturbance diagnosis

Disturbance diagnosis includes determining the root cause of a disturbance [Russell et al., 2000], the type and magnitude of the disturbance [Isermann, 2005, Ge et al., 2013], and its propagation path [Thornhill and Horch, 2007]. Disturbance diagnosis is also known as fault diagnosis in the process monitoring literature [Venkatasubramanian et al., 2003a], machinery [Jardine et al., 2006] and electric motor applications [Nandi et al., 2005].

Root-cause

Locating the root cause of a disturbance is known as root-cause analysis [Thornhill and Horch, 2007] or fault isolation [Gertler et al., 1999]. At times, the precise root cause is not located, and instead the diagnosis suggests a small area of the process, or the closest variable among those with measurements available [Thornhill and Horch, 2007, Bauer et al., 2007a]

Type and magnitude

Diagnosis may also involve matching signatures in measurement data to types of disturbances, and characterising the magnitude of the disturbance [Isermann, 2005, Ge et al., 2013]. Determining the type and magnitude of a disturbance is often known as fault identification. In particular, determining the type of a disturbance is useful when types are associated with root causes and maintenance strategies.

Propagation path

More recently, research has also focused on identifying the propagation path of a disturbance [Thornhill and Horch, 2007, Yang and Xiao, 2012]. When a disturbance originates at the root cause, it propagates along the process through mass and energy streams, and signal paths. The propagation path shows variables of the process in a directed succession which reflects the order of propagation. The path may not be uni-directional, for example if a recycle is involved in the propagation [Bauer and Thornhill, 2008].

3.1.4 Related tasks

This section describes and compares other tasks which are closely related to process monitoring and diagnosis.

Alarm management

When an unpermitted deviation is detected during real-time operation, an alarm is raised [Ge et al., 2013]. Alarm management aims to rationalize the alarms presented to the operators in order to reduce the workload of correcting the alarm condition, especially when multiple variables are affected simultaneously. Some successful contributions involve combining the alarm log with information of the process topology. In particular, Schlegel et al. [2013] extract the process topology automatically from a process schematic.

Control loop performance monitoring

As with process monitoring and diagnosis, control loop performance monitoring (CLPM) is concerned with detecting undesired behaviour, and locating root causes. However, CLPM focuses on detecting under-performing control

loops, and on locating root causes within the control system. Some methods are common to the tasks of CLPM and process monitoring and diagnosis, as section 3.6.2 will show. However, one distinguishing feature is that CLPM investigates the set-point, controller output, and manipulated variable, while process monitoring and diagnosis often use only process variables. Furthermore, CLPM handles each of these variables differently, whereas in process monitoring and diagnosis it is more common that all variables are handled in the same manner. Huang and Shah [1999], Qin and Yu [2007] and Shardt et al. [2012] review the topic of CLPM and important contributions.

A common root cause within the control system is stiction in a control valve. Stiction is excessive static friction, which causes the valve to respond non-linearly to the controller output. As a result, the process variable and controller output may oscillate in a limit cycle. These oscillations cause disturbances in the control loop, and possibly in all the process. Jelali and Huang [2010] review the topic of valve stiction, and compare several methods for its detection and diagnosis.

Soft Sensors

Soft sensors determine values of physical variables, as do actual *sensors*, but instead use *software* programs to infer those values from other measurements. The programs are models based on first-principles or past data from the process. Soft sensors are commonly used for process variables which can only be measured at low rate, or in off-line analyses. Applications of this technology include process control and management, as well as process monitoring and diagnosis. Other applications of this technology, and the most relevant methods are reviewed by Kadlec et al. [2009].

3.2 Classification of methods of process monitoring and diagnosis

In a three part series of papers, Venkatasubramanian et al. [2003a,b,c] proposed a hierarchical structure to organize the methods of process monitoring and diagnosis. These methods are the application of theories from other fields of science, and hence are often far apart from one another. At the top level, the structure distinguished between methods using analytical models, qualitative

models, or process history. Those papers also reviewed in detail the contributions to these three approaches. Other authors also have presented comprehensive reviews, and proposed hierarchical structures. However, the papers were more focused on specific topics. For example, Qin [2012] reviewed methods based on the statistics of the process historic data, while Isermann [2005] presented a survey on analytical model-based methods.

The literature review presented in the next sections follows the hierarchical trees shown in Figure 3.1, which are based on the structure proposed by Venkatasubramanian et al. [2003a,b,c]. The trees and literature review give a broad overview of process monitoring and diagnosis in order to place the work of the thesis in context. In particular, the shaded node in Figure 3.1d is where the methods developed in this thesis most appropriately fit.

The top node in each panel of Figure 3.1 represents the source of the information used by the monitoring and diagnosis methods. For example, Statistical Process Control (SPC) uses numerical data from routine operation (panel 3.1d). The nodes in the second level of each panel represent the information that the method seeks. Methods of SPC, for example, extract signatures of the data under normal operation. These signatures are often denoted as data-based models (node in the left). Finally, the third level of nodes distinguishes between families of methods. SPC uses statistical methods.

The hierarchical trees in Figure 3.1 update the structure proposed by Venkatasubramanian et al. [2003a,b,c]. One modification was to separate the nodes of qualitative trends (panel 3.1c) and numerical data (panel 3.1d), which were previously grouped under process history. Another modification, in panel 3.1d, was to create the node for Bayesian methods and the nodes of time series properties and advanced signal analysis. The main reason for these two modifications is to review in more detail the methods based on numerical data, which are those closer to the work in this thesis. Another reason for the update is that Bayesian methods and advanced signal analysis became relevant more recently. For the same reason, the tree in panel 3.1b also adds nodes for models of precedence and connectivity.

The hierarchical trees in Figure 3.1 will be revisited in the next four sections. Each of these sections discusses the methods in one of the four panels.

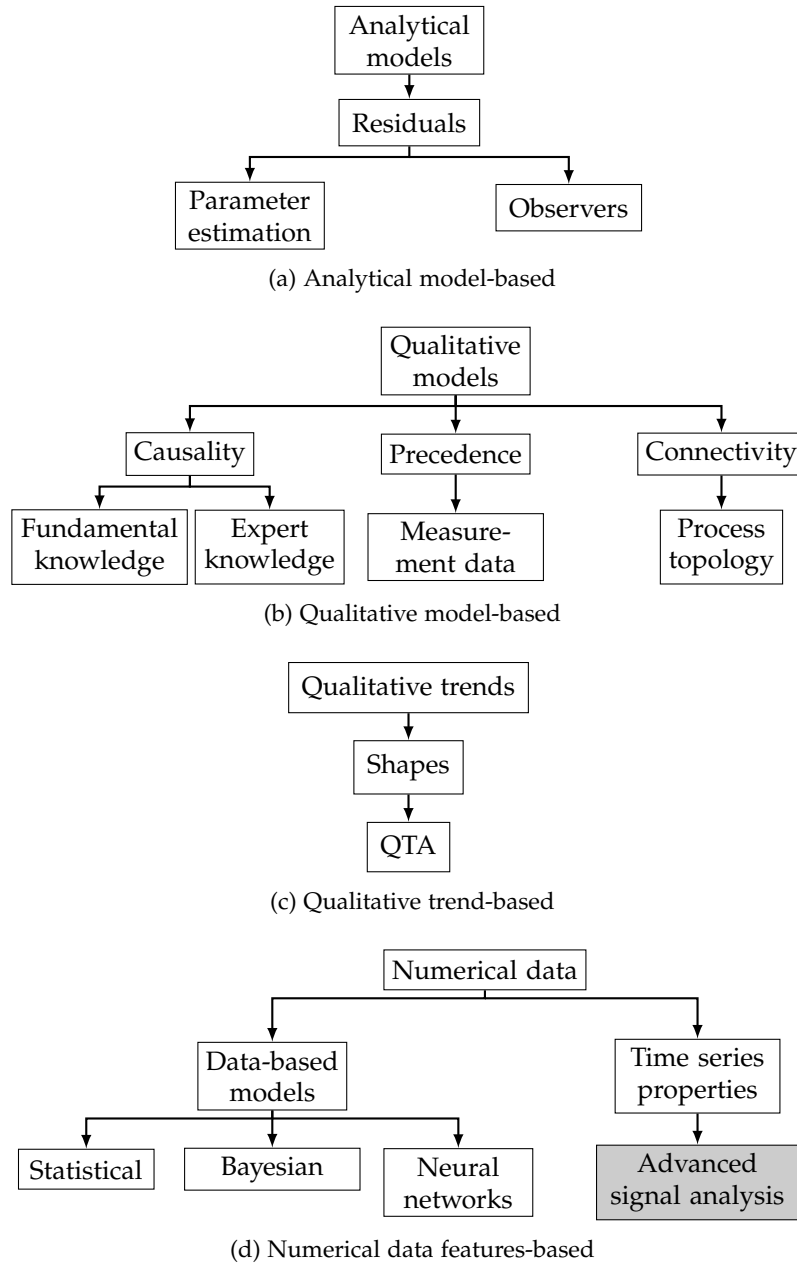


Figure 3.1: Hierarchical trees to organize the methods of process monitoring and diagnosis. The shaded node indicates where the methods developed in this thesis fit.

3.3 Analytical models

If one has access to analytical models of the system, then the detection and diagnosis of disturbances can use one of the analytical model-based methods. This approach was one of the earliest in process monitoring and diagnosis [Venkatasubramanian et al., 2003a], as evidenced by the several-decades-old review papers on the topic [Isermann, 1984, Frank, 1990]. These papers explain also the principles behind the approach and families of methods. A more recent review can be found in Isermann [2005].

Analytical models are those describing the physics or chemistry of the system. They are generally dynamic models in the form of a mathematical equation with known parameters. Examples include first-principles, frequency response, and state-space models [Isermann, 2005, Venkatasubramanian et al., 2003a]

Analytical model-based methods work by confronting the actual behaviour of the system with the prediction of the explicit mathematical model. The inconsistencies are known as residuals, or analytical symptoms [Isermann, 2005]. For disturbance detection, residuals are compared with a threshold [Venkatasubramanian et al., 2003a]. Diagnosis makes use of a known association between residuals and root causes [Isermann, 2005].

The methods for residual generation are of two general types, as indicated in Figure 3.1a: parameter estimation and observers.

Parameter estimation

Parameter estimation methods use analytical or numerical optimisation techniques. The objective is to minimize the error of the model by adjusting its parameters. The estimated parameters depend on physical coefficients of the system, thus changes in parameters are explicitly associated with root causes. An example is a fouling problem reflected in the heat transfer coefficient.

Observers

Observer methods estimate state or output variables from knowledge of a model and its parameters. Disturbances are traced to root causes associated with the modeled variables. Parity equations are often indicated as a separate method [Venkatasubramanian et al., 2003a], but have been shown to be a par-

Table 3.1: References on analytical model-based methods for process monitoring and diagnosis.

Parameter estimation	Observers
Application to nonlinear static process models [Isermann, 1993]	Pioneer works [Basseville, 1988]
Multiple sensor fault diagnosis combining parameter estimators and parity equations [Song and Zhang, 2002]	Generation of diagnostic observers for non-linear systems [Frank, 1990]
On-line parameter estimation algorithm for time-varying continuous systems [Chen et al., 2011]	Sensitivity reduction under noisy conditions [Gertler et al., 1999]
	Diagnosis of nonlinear systems using multiple-linear-models implemented in Kalman filters [Deshpande and Patwardhan, 2008]
	Unknown input observer applied in estimating the valve position for valve stiction diagnosis [Chitrlekha et al., 2010]
	Detection in hybrid systems, <i>e.g.</i> processes switching between operating modes [Hu and El-Farra, 2011]

ticular case of observer methods [Gertler, 1991]. Kalman filters are a stochastic version of observers.

Table 3.1 summarizes important developments in analytical model-based methods for process monitoring and diagnosis.

Analytical model-based methods for machinery and electrical equipment

Mechanical and electrical engineering applications have used extensively analytical model-based methods for monitoring and diagnosis. One example using parameter estimation is given by Nikranjbar et al. [2009], who applied a particle swarm optimisation technique to the diagnosis of induction motors. One example using observer techniques is given by Simani and Patton [2008], who used the Kalman filter to diagnose industrial gas turbines. The reason why analytical model-based methods are popular for machinery and electrical equipment is that a single model can be used for large numbers of identical pieces of equipment, and hence the cost of producing an accu-

rate model is compensated by its extensive use. This is not true for process systems, given that an industrial process tends to be unique [Venkatasubramanian et al., 2003c, Thornhill, 2007].

Comments on the use of analytical models

The main strength of analytical model-based methods is capturing the fundamental relationships between the variables of the system. In theory, they can trace the cause of a disturbance to an exact variable of the system.

However, analytical model-based methods are not used in this thesis for two reasons. One reason is that these methods require an accurate model of the system. Accurate models are costly for industrial processes, particularly first-principles models [Venkatasubramanian et al., 2003a]. The second reason is that these models cannot be reused because, as discussed, industrial processes tend to be unique.

Although analytical model-based methods are not used in this thesis, they were presented in this chapter to provide a broad context for the work of the thesis.

3.4 Qualitative models

Qualitative relationships between variables of a system can be exploited for process monitoring and diagnosis. The reason is that they provide information about the potential, or actual, propagation paths of a disturbance, and thus enable the diagnosis of its root cause. The nature of the relationships between variables can be, in growing order of influence:

- connectivity,
- precedence, and
- cause-and-effect.

The example in Figure 3.2 aims to clarify the concepts in the list above. The system represented is formed by two tanks in series. The first tank has one inlet flow F_1 , and two outlet flows, F_2 and F_3 . The tanks are linked by flow F_3 , which is also the inlet flow to the second tank. Flow F_4 is the outlet flow of the second tank. The levels in the tanks are Z_1 and Z_2 . The system of

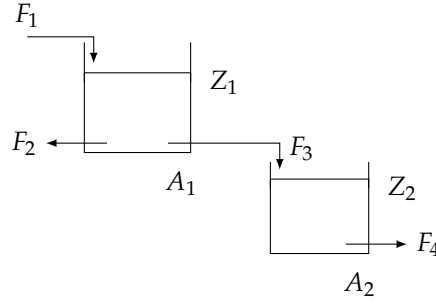


Figure 3.2: Schematic of a tank with cross-sectional area A , level Z , inlet flow F_1 , and outlet flows, F_2 and F_3 .

equations (3.2) relates the six variables. Parameters A_1 and A_2 are the cross-sectional areas of the tanks. Parameters R_2 , R_3 , and R_4 represent resistances to the flows F_2 , F_3 and F_4 , respectively.

$$A_1 \frac{dZ_1}{dt} = F_1 - F_2 - F_3 \quad (3.2a)$$

$$A_2 \frac{dZ_2}{dt} = F_3 - F_4 \quad (3.2b)$$

$$F_2 = \frac{1}{R_2} \sqrt{Z_1} \quad (3.2c)$$

$$F_3 = \frac{1}{R_3} \sqrt{Z_1} \quad (3.2d)$$

$$F_4 = \frac{1}{R_4} \sqrt{Z_2} \quad (3.2e)$$

Connectivity relationships refer to connections between the variables. These connections can be physical, through mass and energy streams, and electrical, such as control signals. In the example, the four variables in the first tank, that is, Z_1 , F_1 , F_2 and F_3 , have direct physical connections with each other, and the three variables in second tank also have direct physical connections with each other. Furthermore, variables Z_2 and F_4 are indirectly connected to the variables in the first tank through F_3 .

If a disturbance enters this system through the inlet flow F_1 , it will gradually affect all other variables. Variable F_4 will take longer than F_2 and F_3 to respond to the onset of the disturbance. This is due to the hold-up volume of the second tank, which represents an additional time constant between F_4 and the root cause. Because of this lag, variables F_2 and F_3 will both precede F_4 in

the propagation path of the disturbance. The relationships from F_2 to F_4 , and from F_3 to F_4 are examples of precedence relationships.

Although F_2 precedes F_4 , it is not the change in F_2 that causes F_4 to change. The fundamental equations (3.2) indicate that the dynamic behaviour of F_1 will first cause a change in Z_1 , which in turn will directly cause changes in F_2 and F_3 . The disturbance propagates to the second tank because F_3 affects Z_2 . Finally, it is the change in Z_2 that causes a change in F_4 . These relationships captured by the fundamental equations are of cause-and-effect.

Researchers have proposed different methods which can extract connectivity, precedence, or cause-and-effect relationships. As indicated in the hierarchical tree 3.1b, this section reviews those methods according to the nature of relationship they extract. Yang and Xiao [2012] present a review of the topic which also covers the three relationships. In the review series by Venkatasubramanian et al. [2003a], the authors discuss only the methods to generate causal relationships. Methods to extract connectivity and precedence relationships have grown since then.

Table 3.2 summarizes important developments in qualitative model-based methods for process monitoring and diagnosis.

3.4.1 Causality models

Causality models represent the cause-and-effect relationships between process variables. Cause-and-effect relationships can be inferred from fundamental knowledge, or from empirical knowledge.

Fundamental knowledge refers to first-principles understanding of a system, which is often in the form of differential equations (DE), algebraic equations (AE), and differential algebraic equations (DAE). Maurya et al. [2003] have derived separate algorithms to convert DE, AE, and DAE into causality models in the form of graphs. A graph is a structure with nodes connected by arcs, which may show the direction of influence from one node to the others. Section 3.4.4 will explain in more detail ways of representing qualitative models by means of graphs, matrices, *if-then* rules, and qualitative equations. In the literature, causality models have been mostly represented in the form of a graph.

Empirical knowledge refers to the understanding of experts about the process, equipment, and relations between disturbances and root causes. Angeli

Table 3.2: References on qualitative model-based methods for process monitoring and diagnosis.

Causality models	Precedence models	Connectivity models
SDG generation from steady state and dynamic model equations [Oyeleye and Kramer, 1988]	Measures sensitive to time delays, added noise, and attenuation between measurements [Bauer et al., 2007a, Bauer and Thornhill, 2008]	Adjacency and reachability matrices derived manually from plant schematics [Mah, 1990]
Digraph and SDG can be used for deciding the location of sensors [Bhushan and Rengaswamy, 2000]	Estimates of time delays for linearly and nonlinearly correlated measurements [Bauer and Thornhill, 2008, Stockmann et al., 2012]	Automated generation of an asset monitor for a heat exchanger from its electronic description [Schmidberger et al., 2006]
Systematic SDG generation from system of differential-algebraic equations [Maurya et al., 2003]	Directed transfer function method sensitive to energy transfer between frequencies of linearly correlated measurements [Gigi and Tangirala, 2010]	Automated extraction of connectivity from process schematics represented with the CAEX standard [Yim et al., 2006]
Dynamic SDG with time parameters for the graph arcs [Yang and Xiao, 2007]		Reachability matrix focused on the control system [Jiang et al., 2009]
Combined QTA with SDG-based models to reduce spurious solutions [Maurya and Venkatasubramanian, 2007, Gao et al., 2010]		Automated conversion of process connectivity in XML files into adjacency matrices. Combination with data-driven precedence models [Thambirajah et al., 2009]
Combined causality models in SDG form with precedence models from two data-driven methods [Yang and Xiao, 2012]		Combination of connectivity models, rules and alarm logs for alarm management [Schleburg et al., 2013]

Table 3.2: References on qualitative model-based methods for process monitoring and diagnosis.

Causality models	Precedence models	Connectivity models
Hierarchical SDG structure to analyse fault propagation with higher efficiency [Yang et al., 2013]		Automated generation of quantitative models combining intelligent P&ID and process modelling libraries [Barth and Fay, 2013]

[2008] reviews methods to capture this unstructured, and often incomplete, knowledge in a structured causal model. Commonly, a causality model based on empirical knowledge is represented as *if-then* rules, but uses of graphs have also been reported. For disturbance diagnosis in chemical processes, Iri et al. [1979] were the first to use expert knowledge to derive a type of graph known as signed digraph (SDG).

3.4.2 Precedence models

Precedence models represent the propagation path of a disturbance along the variables of the system. Section 2.4.1 discussed dynamic characteristics of the propagation, such as dead times and time constants.

The extraction of precedence models is a recent topic, and is done from measurement data. In the literature, these models are often said to show cause-and-effect relationships [Bauer and Thornhill, 2008, Duan et al., 2012]. However, a variable which precedes another does not necessarily cause it, as discussed with the example of the tanks in Figure 3.2. The reason is that there may be more than one direction of propagation, as with variables F_2 and F_3 in the example.

The methods that extract precedence models use advanced signal analysis to search for features in measurement data, such as time delays, attenuation, transfer of information, and conditional probability relations. As a result, these methods can also fit in the node of advanced signal analysis in the hierarchical tree 3.1d. A general overview of precedence models can be found in Yang and Xiao [2012].

3.4.3 Connectivity models

Connectivity models represent connections between process variables. These connections can be physical, through mass or energy streams, or electronic, such as control signals. Connectivity models are derived from process schematics. The extraction of connectivity from a process schematic can be done manually [Mah, 1990, Jiang et al., 2009], but recently there has been progress in automating this extraction [Yim et al., 2006, Thambirajah et al., 2009].

Automatic extraction is enabled by digital representations of the process schematic. Several vendors now offer intelligent P&ID tools which allow the export of a text file (XML type) describing equipment items, their properties, the connections between them and the directions of those connections. Thambirajah et al. [2009] explain the automated generation of a connectivity model from an XML file. The authors use the programming language C# to read the XML file, and a set of rules to convert the parsed information into the connectivity model. Barth and Fay [2013] further extended the use of intelligent P&ID tools to develop quantitative models of the process with a level of detail as required for control code tests. To that end, the authors developed an algorithm which maps the equipment items in the intelligent P&ID onto objects from a fluid mechanics modelling library.

3.4.4 Forms of representation

Qualitative relationships between variables can be represented in different forms. These include:

- graphs,
- matrices,
- rules, and
- qualitative equations.

Several review papers emphasize the form of the representation rather than the causal, precedence, or connectivity nature of the relationship extracted. For instance, Maurya et al. [2003] focus on signed directed graphs. However, the text below will argue that the forms of representation are interconvertible. Therefore, section 3.4 has emphasized instead the nature of the relationship.

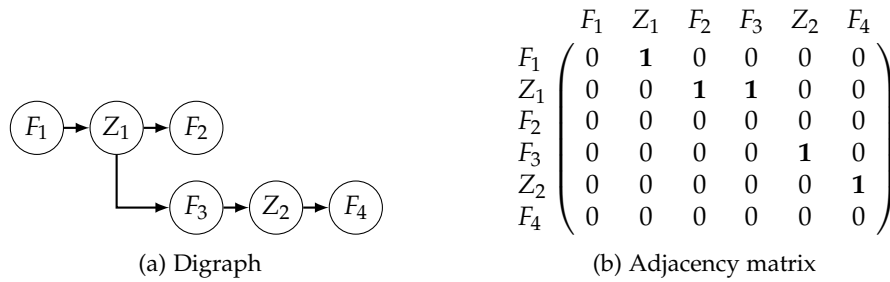


Figure 3.3: Digraph and adjacency matrix for the example in Figure 3.2.

Graphs

A graph is a mathematical abstraction of the structural relationships between discrete objects [Mah, 1990]. The objects are represented as nodes. In the context of process monitoring and diagnosis, some authors have used nodes to represent process variables, events [Venkatasubramanian et al., 2003b], controllers [Jiang et al., 2009], and process equipment [Mah, 1990].

Two related objects are connected by an arc, or edge. A digraph is a particular type of graph in which the arcs have a direction. This is because there is a direction in the influence from one object to the other. Figure 3.3a shows the digraph for the cause-and-effect relations of the example described in Figure 3.2. The digraph shows that the changes in flow F_1 have an effect on the level Z_1 of the first tank, which in turn affects F_2 and F_3 . The flow F_3 , which enters the second tank, affects the level Z_2 , which in turn affects the outlet flow F_4 .

Signed digraphs (SDG) are digraphs which have a positive or negative sign attached to the directed arc. For process variables, the sign on the arc can indicate whether the value of the influenced variable varies in the same, or opposite, direction as the influencing variable. SDGs have been widely used to represent causality relationships between process variables. The generation of SDG to represent causality relationships was comprehensively described by Maurya et al. [2003]. Recent contributions aim to devise hierarchical structures that improve the efficiency of using the qualitative model represented by the SDG [Yang et al., 2013].

A Bayesian belief network is another type of graph. Its initial nodes represent root causes, and arcs are assigned probabilities instead of positive and negative signs. An arc from a root cause X to a variable Y describes the con-

ditional probability, $p(Y|X)$, of the variable Y exhibiting a certain behaviour given the root cause X [Yang and Xiao, 2006].

Matrices

In graph theory, a digraph can be converted into an adjacency matrix [Mah, 1990]. Figure 3.3b shows the adjacency matrix equivalent to the digraph of the tank example. The rows and columns of the matrix are equivalent to the nodes of the graph. Entries in the matrix are equivalent to the arcs of the graph. Specifically, entry (i, j) represents an existent ('1') or non-existent ('0') direct relationship from node i to node j . For instance, the direct causal relationship from F_1 to Z_1 is indicated the number one in the entry $(1, 2)$ of the matrix.

In the context of process monitoring and diagnosis, adjacency matrices have been used to represent direct connections between variables [Mah, 1990, Jiang et al., 2009]. A reachability matrix additionally shows the indirect connections, that is, connections through other, intermediate, variables. It can be derived from the adjacency matrix using matrix algebra [Mah, 1990, Jiang et al., 2009].

Rules

Rules are *if-then* logical statements. Often, expert operators have some empirical knowledge which relates disturbances in the system with root causes. Rules can represent this knowledge in a structured form, and be reasoned upon systematically [Angeli, 2008]. Kramer and Palowitch [1987] showed how to convert SDGs into a set of rules.

Qualitative equations

Qualitative equations are equation-like expressions which relate the directions of variation of the variables, instead of their numerical values [De Kleer and Brown, 1984, Kuipers, 1994]. Equation (3.3) is the qualitative equation for the tank equation 3.2b in steady-state. The square brackets function gives the direction of variation of the enclosed variable, that is, positive, negative, or constant. Thus, a positive variation in F_1 has to be matched by positive variations in F_2 and F_3 .

$$[F_1] - [F_2] - [F_3] = 0 \quad (3.3)$$

Qualitative model-based methods for machinery and electrical equipment

The qualitative models discussed in this section have had little application in the monitoring and diagnosis of machinery and electrical equipment. In fact, in a comprehensive review of machinery diagnosis Jardine et al. [2006] make no references to a qualitative model-based approach. The exception are the rule-based models. Rules are reasoned upon by inference mechanisms, such as hypothesis testing [Uraikul et al., 2007]. The rules and inference mechanism are commonly implemented in a software known as expert system. Recent surveys on expert systems are provided by Liao [2005] and Angeli [2008]. Early applications of expert systems in the monitoring and diagnosis of machinery and electrical equipment include the first Soviet expert system for monitoring the vibration of a large turbo generator [Biber et al., 1990]. More recent applications include a decision support tool for experts monitoring British Energy turbine generators [Todd et al., 2007].

Comments on the use of qualitative models

The main contributions of this thesis are not based on qualitative models. However, section 9.1 discusses a future research direction which involves deriving the propagation path of a disturbance from measurement data. The main strength of qualitative model-based methods is the ability to explain the suggested root cause, which is useful especially when analytical models are inexistent or incomplete. In the method suggested in section 9.1, the measurement closest to the root cause is inferred by tracking the disturbance up the propagation path.

A qualitative model on its own may generate spurious solutions [Venkatasubramanian et al., 2003b,c]. Hence, some authors have proposed to integrate qualitative models with other methods. Examples are the integration of causality models with qualitative trend analysis [Maurya and Venkatasubramanian, 2007], explained later in section 3.5, and of connectivity models with the signal analysis method of transfer entropy [Thambirajah et al., 2009], explained later in section 3.6. Future developments from the method suggested

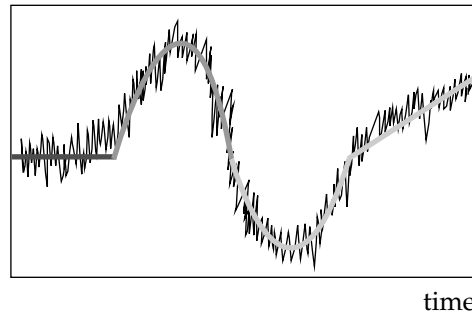


Figure 3.4: Measurement trend illustrating the identification of three qualitative trends.

in section 9.1 could also focus on the integration of the precedence model with other results. Section 9.3 will discuss some ideas towards such developments.

3.5 Qualitative trends

Qualitative trends are the shapes formed by trends in measurement data. In the measurement shown in Figure 3.4 three different shapes may be identified: a constant line, a sine wave, and a line with positive slope.

Shapes in measurements can be matched to types of disturbances, and may identify measurements affected by a common disturbance [Maurya et al., 2010]. The methods to analyse those shapes are known as Qualitative Trend Analysis (QTA).

Qualitative trend analysis

Methods of QTA extract basic shapes from measurement of process variables, compare those shapes to a database, and in that way diagnose the cause of the disturbance. Maurya and Venkatasubramanian [2007] explain in detail the principles of this methodology, and also review some methods for extracting the trends and comparing them with the database.

Table 3.3 summarizes important developments in QTA methods, as well as recent contributions. It shows that research in QTA focused on the definition of the basic shapes, how to match measurement trends with the shapes in the

Table 3.3: References on qualitative trend-based methods for process monitoring and diagnosis.

Qualitative trend analysis	
Definition of basic shapes	[Love and Simaan, 1988, Cheung and Stephanopoulos, 1990, Janusz and Venkatasubramanian, 1991]
Extraction of trends by projection onto wavelets	[Bakshi and Stephanopoulos, 1994]
Automated detection of valve stiction and dead-band based on the qualitative shapes in the pv and op measurements	[Rengaswamy et al., 2001, Yamashita, 2006]
Included additional shapes for the detection valve stiction	[He et al., 2007a, Scali and Ghelardoni, 2008]
(i) Interval-halving method to handle different time-scale events; (ii) fuzzy shape recognition method	[Maurya and Venkatasubramanian, 2007]
Algorithm for on-line trend extraction	[Maurya et al., 2010]
Algorithm with self-adaptive parameters according to on-line estimation of background noise	[Charbonnier and Portet, 2012]
QTA framed as optimisation problem of fitting of spline functions, constrained to the measurement trends	[Villez et al., 2013]

database, and the on-line implementation of QTA. The table also gives references for the use of QTA in the diagnosis of valve stiction. The reason to use QTA is that stiction manifests with particular patterns in the measurements of the process variable (pv) and controller output (op). Specifically, these exhibit non-sinusoidal oscillatory trends.

Comments on the use of qualitative trend analysis

The main strength of QTA is that it resembles the analyses intuitively drawn by the human eye. As a result, QTA can recognize shapes even if these are masked by other trends such as noise.

However, QTA is not suitable for the challenges addressed in this thesis. One reason is that QTA depends on a comprehensive database which relates

shapes to types of disturbances and root causes. For example, to detect a transient disturbance, that transient would have to exhibit one of the particular shapes in the database. Furthermore, the same transient disturbance may have different shapes in different measurements, as shown in Figure 5.1 of Chapter 5. Therefore the thesis will need a more general approach than is offered by QTA.

3.6 Numerical data

In industrial processes, measurements from past operation of the system are readily available in the data historian. Several methods can use these historical measurements for monitoring and diagnosing the conditions of a system, without the need for fundamental understanding. For these two reasons, these methods are widely used in industrial process systems [Qin, 2012, Ge et al., 2013]. These methods are also known as data-driven.

The hierarchical tree in Figure 3.1d divides data-driven methods into those extracting data-based models and those extracting properties from time series. Methods using data-based models focus on plant profiles, that is, the values of all process measurements at a particular instant in time. These methods aim to match plant profiles to normal operating conditions or a type of disturbance. Other methods focus on time series, that is, the time-ordered sequence of samples of a particular measurement. Their aim is to extract properties from the time series of a measurement or from the relations between time series of different measurements. These properties relate to the presence of a disturbance, and to the possibility of a certain measurement being the root cause.

3.6.1 Data-based models

Data-based models are descriptions of measurement data under normal operating conditions, and possibly also under the effect of different types of disturbances. The measurement data are plant profiles, that is, the values of all process measurements at a particular instant in time. Figure 3.5 shows two examples, each with 23 plant profiles spanning 40 process measurements. The plant profiles on the left-hand side correspond to normal operating condi-

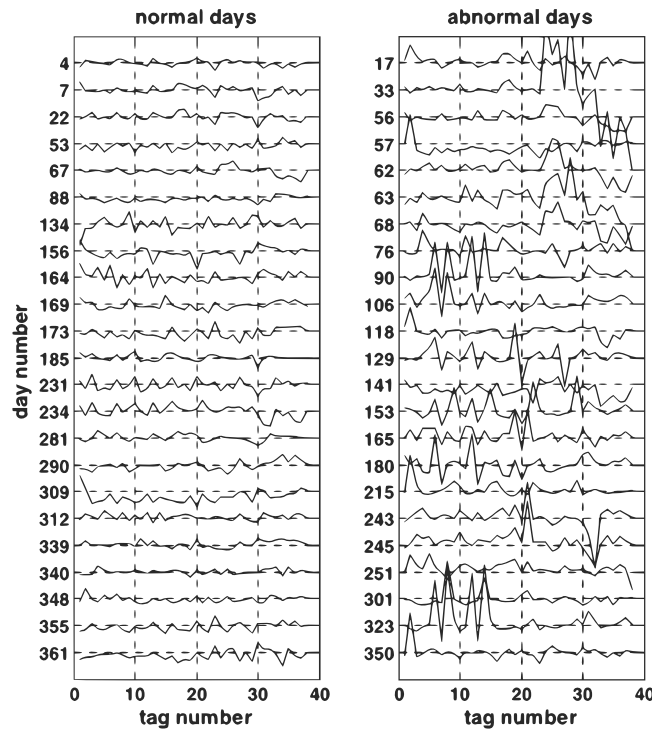


Figure 3.5: Plant profiles from a wastewater plant. Left panel corresponds to normal operating conditions. Right panel corresponds to disturbed operation. Taken from Thornhill et al. [2006].

tions, while the plant profiles on the right-hand side correspond to disturbed operation.

The detection and diagnosis of disturbances with data-based models normally follows a training step and an implementation step, as described below.

- In the training step, an expert selects and assigns past plant profiles to normal operation or to a type of disturbance. A model is built to describe common characteristics of the plant profiles under the various conditions. The hierarchical tree in Figure 3.1d shows three main divisions within the methods to build data-based models. These three divisions will be discussed in the next subsections.
- The implementation step uses the data-based model to match new plant profiles to the condition of the process at that instant. This can be done online or offline. When only the normal condition has been modelled, the root cause is diagnosed by assessing the contribution of each measurement to the departure from the normal condition.

Table 3.4: References on methods using data-based models for process monitoring and diagnosis.

Statistical	Bayesian	Neural networks
Pioneer works in multivariate SPM [Kresta et al., 1991]	First application of Bayesian networks to process fault diagnosis [Rojas-Guzman and Kramer, 1993]	First applications in process diagnosis [Venkatasubramanian, 1985]
Multi-way PCA for batch processes [Nomikos and MacGregor, 1994]	Computational efficient Bayesian algorithm for process diagnosis [Pernestal, 2007]	Multi-resolution neural network using wavelets as activation function [Bakshi and Stephanopoulos, 1993]
Introduction of ICA for non-Gaussian process monitoring [Kano et al., 2004]	Integration of GMM and Bayesian inference for processes with multiple operating modes [Yu and Qin, 2008]	non-linear PCA methods based on neural networks [Dong and McAvoy, 1996, Jia et al., 2000]
Introduction of kernel-PCA for non-linear process monitoring [Lee et al., 2004]	Probabilistic framework to combine various monitoring models for control loop assessment [Huang, 2008]	Particle swarm optimization algorithm for selecting parameters of the neural network [Samanta and Nataraj, 2009]
Improvements to the original ICA algorithm [Lee et al., 2006, Zhang and Zhang, 2010]	Estimation of probabilities of different problems in a control loop in the presence of missing data [Qi et al., 2010]	Hierarchical structure for neural networks [Eslamloueyan, 2011]
Enhanced results with SVM compared to kernel-PCA in non-linear systems [Mahadevan and Shah, 2009]	Adaptation of Bayesian methods to consider auto-correlation between measurement samples [Qi and Huang, 2011]	
Extension of classical methods to multiple operating mode systems with GMM [Chen and Zhang, 2010]	Real-time detection and quantification of instrument gross errors [Gonzalez et al., 2012]	

Table 3.4: References on methods using data-based models for process monitoring and diagnosis.

Statistical	Bayesian	Neural networks
Extension of the SVM approach to batch processes [Ge et al., 2011]		
Kernel-PCA and kernel-ICA adapted to monitoring disturbances at different scales [Zhang and Ma, 2011]		

Table 3.4 summarizes important developments in methods to build data-based models for process monitoring and diagnosis.

Statistical methods

Statistical methods are commonly known as Multivariate Statistical Process Monitoring (MSPM). These methods are widely used, and have been reviewed in survey papers [Qin, 2012, Ge et al., 2013] and text books [Russell et al., 2000].

The data-based model is the probability distribution of some statistic when the process is under normal operating conditions. The statistic is calculated from the plant profiles. A simple model is that derived for a single process measurement, X . Figure 3.6 shows the sequence of values of the example measurement X under normal operating conditions. The mean \bar{x} and standard deviation s_x of this group of values can be used to define the statistic

$$\frac{x - \bar{x}}{s_x} \quad (3.4)$$

which is assumed to follow a standard normal distribution.

In the implementation step, the statistical methods test the hypothesis that a new value of the statistic could be generated from the distribution based on normal operation. In the example above, if the new value of the measurement is beyond $\bar{x} \pm 6s_x$, then the probability that it was generated from the distribution based on normal operation is less than one in a million. Therefore, such a value is considered to be affected by a disturbance.

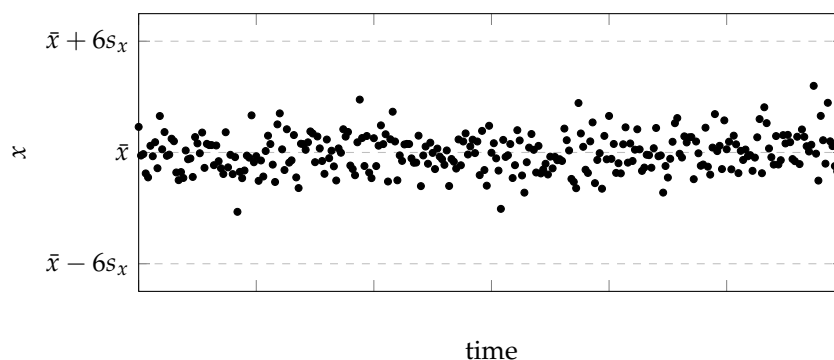


Figure 3.6: Sequence of values from a process measurement under normal operating conditions. The dashed lines mark the limit beyond which new values of the measurement will be considered abnormal.

The example of Figure 3.6 uses a univariate statistic, that is, a statistic derived from the values of a single measurement. However, in chemical processes it is also important to model the relations between measurements, hence statistical methods mostly use multivariate statistics, which comprise values from all measurements [Kourti and MacGregor, 1996, Russell et al., 2000, Qin, 2012].

The two main statistics used within MSPM are the T^2 and Q statistics. These are derived from a Principal Component Analysis (PCA) of the group of plant profiles under normal operating conditions. PCA handles each plant profile as a data vector, and extracts common components between several data vectors [Jolliffe, 2005]. These components have two important properties: (i) the first few components capture most of the statistical variance of the data vectors, and (ii) a linear combination of the components reconstructs the original data vectors. The formulation of PCA and the distributions of T^2 and Q implicitly assume that the data vectors have *Gaussian* distributions and are *linearly* correlated. Such properties do not always occur between variables of the process system. The following list gives an overview of other statistical methods used in MSPM which relax the tight assumptions of PCA.

- Process variables do not always follow Gaussian distributions for reasons which include non-Gaussian noise, feedback control loops, and multiple operation modes [Qin, 2012, Ge et al., 2013]. Independent Component Analysis (ICA) and Gaussian Mixture Models (GMM) have been the most used methods to avoid this assumption. ICA has a similar

principle as PCA, but uses statistics of higher order than the variance [Hyvärinen and Oja, 2000]. On the other hand, GMM divides the data set into subgroups of plant profiles which show Gaussian distributions [Bilmes, 1998].

- For process variables to be linearly correlated, the process has to operate in a small region of stable conditions. However, if the process has multiple products or operating conditions, then the correlations are not linear. In this case, GMM can be used due to dividing the data set into subgroups of plant profiles [Ge et al., 2013]. PCA and ICA have also been extended to the non-linear case by means of kernel functions. The principle is that the data vectors are mapped onto a higher dimensional space where their relationships are linear, and where PCA and ICA can be applied [Ge et al., 2013]. Support Vector Machines (SVM) [Vapnik, 1998] also use a kernel function to map the data vectors onto a higher dimensional space. In that space, SVM defines optimum boundaries for the normal and abnormal operations.
- In batch processes, each batch usually has different operating conditions. Therefore, the resulting data has three dimensions: variables, time and batch. The standard MSPM methods can only model two-dimensional data, and hence had to be extended into multi-way methods [Qin, 2012]. The basic idea is to transform the three-dimensional, or three-way, array into a two-dimensional matrix, to which PCA and the other MSPM methods can be applied. Yao and Gao [2009] give comprehensive explanations of the topic, and review recent contributions.
- Process variables are normally auto-correlated, that is, the present value of a variable is dependent on its past values. However, the standard MSPM methods assume otherwise. Dynamic extensions to these methods augment each plant profile with time-lagged data samples of each variable [Ku et al., 1995], hence modelling also the auto-correlation between samples. Qin [2012] give an overview of recent contributions to the topic.

Statistical methods have also been combined with other methods. For example, non-linear processes were initially monitored by combining PCA with neural networks [Dong and McAvoy, 1996], which are reviewed later in this

section. Another example is the use of signal analysis methods to detect disturbances with different time scales. To that end, Bakshi and Stephanopoulos [1993] and Misra et al. [2002] combined wavelet analysis with PCA. The authors suggest that the method can also handle measurements sampled at different rates.

Bayesian methods

Bayesian methods are used in process monitoring and diagnosis to quantify explicitly the probability $p(R_i|X)$ of each root cause R_i given the current measurement values X [Huang, 2008]. The posterior probability $p(R_i|X)$ is derived from the Bayes' theorem

$$p(R_i|X) = \frac{p(X|R_i) p(R_i)}{p(X)}, \quad (3.5)$$

which combines (i) the current measurement values X , (ii) the likelihood $p(X|R_i)$ of the measurement values given each root cause, and (iii) the probability $p(R_i)$ of each root cause occurring in the system.

The prior probabilities $p(R_i)$ of each root cause and the likelihoods $p(X|R_i)$ of the measurement values given each root cause can be derived from measurements from past operation of the system. Therefore, Bayesian methods are shown the hierarchical tree 3.1d. The probabilities $p(R_i)$ and $p(X|R_i)$ are the data-based models which must be derived in the training step of the Bayesian methods. The prior probabilities $p(R_i)$ can be determined from performance data of the equipment, for instance, their tendency to fail, or from historical data. The likelihoods $p(X|R_i)$ can be derived from historical data of operation under different types of disturbances. To determine $p(X|R_i)$, the root-cause R_i of each disturbance must be known.

Huang [2008] presented a framework and described the use of Bayesian methods in process monitoring and diagnosis, with an emphasis on control loops. The author also highlighted the ability of Bayesian methods to combine results from different sources of priori knowledge, such as analytical models, and to handle missing data. Khatibisepehr et al. [2013] have reviewed the use of Bayesian methods in the related area of soft sensors, and Patwardhan et al. [2012] reviewed the use in state estimation.

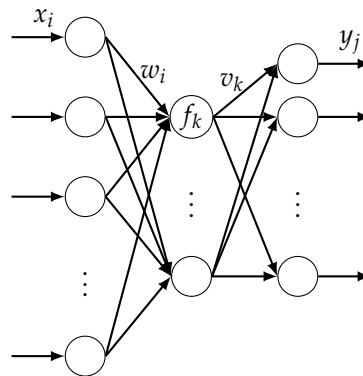


Figure 3.7: Example of a neural network with one hidden layer. The activation functions are f_k , and weights w_i and v_k .

Neural networks

Neural networks have often been used in process monitoring and diagnosis combined with the statistical methods. The reason is the ability of neural networks to model non-linear relationships between process variables.

As illustrated in Figure 3.7, neural networks are implemented as a sequence of layers, and each layer has a number of nodes which connect to the previous and subsequent layer. Each node in a inner, or hidden, layer is a function f_k which receives a w_i -weighted sum of the outputs of its parent nodes, and sends the result to the next layer via its own weight v_k [Kohonen, 2001].

In process monitoring and diagnosis, the parameters of the node function and the weights are estimated in the training stage from past measurements of the system. The choice of the node functions [Bakshi and Stephanopoulos, 1993], network structure and number of nodes [Samanta and Nataraj, 2009] have been some of the challenges addresses in the literature. The final data-based model is then able to relate non-linearly measurement values with types of disturbances.

Data-based models for machinery and electrical equipment

Mechanical and electrical engineering applications have used extensively data-based models for monitoring and diagnosis [Jardine et al., 2006]. The reason is that a single model can be used for large numbers of identical pieces of equipment. Therefore, the extensive use of the data-based model compensates the

cost of determining all relevant types of disturbances in a piece of equipment. Examples of work with data-based models include the use of kernel PCA [He et al., 2007b], SVM [Widodo and Yang, 2007], and neural networks [Saravanan et al., 2010].

Comments on the use of data-based models

The main strength of methods using data-based models is that they are easy to implement and lead to fast detection of abnormal conditions. As a result, MSPM methods are particularly popular in industry [Perceptive Engineering, 2014, Umetrics, 2014].

However, the methods developed in this thesis do not use data-based models. One reason is that data-based models require measurements for the training step which

- have been correctly selected and matched to each type of operation,
- include all types of operation modeled, and
- have numerous samples for each type of operation.

If these requirements are not met, then false alarms and missed disturbances are likely to happen. Another reason why this thesis does not use data-based models is that these models are also limited to the types of operation on which they were trained [Venkatasubramanian et al., 2003c].

3.6.2 Time series properties

Processes have dynamic behaviour. However, most of the methods that use data-based models assume that the value of a process variable at one instant is independent of all other instants [Ge et al., 2013]. One way to take the process dynamics into account is with the dynamic versions of MSPM methods, as discussed in section 3.6.1. Another way is to handle the measurements from process variables as time series. The time series from one variable is its ordered sequence of samples, which are obtained at consecutive time instants.

Some properties of a time series, or of the relation between time series, are related to the presence of a disturbance or to the proximity to the root cause. Examples are the non-linearity of a time series, the correlation between the power spectra of time series, and the transfer entropy between two time

series. Tables 3.5 and 3.6 give other examples of time series properties which provide information about the presence and cause of process disturbances.

The methods which exploit the properties of time series are known as *advanced signal analysis*. These methods are represented by the grey node in the hierarchical tree 3.1d. Advanced signal analysis methods may analyse measurements in the time domain, the frequency domain, or the time-frequency domain. The choice depends on the property they aim to investigate. It is common in the literature to refer to analyses in the time domain as time series analysis, and analyses in the frequency and time-frequency domains as signal processing. In contrast to data-based models, the methods which exploit the properties of time series do not need a training step to learn the model of the process.

The use of advanced signal analysis grew from the related discipline of control loop performance monitoring (CLPM), which was introduced in section 3.1.4. CLPM drove the development of the advanced signal analysis approach because control loops are often disturbed by persistent oscillations [Horch, 2007, Jelali and Scali, 2010]. As a result, the developments in advanced signal analysis have been focused on persistent disturbances. CLPM also drove the developments in advanced signal analysis towards plant-wide analysis [Thornhill and Horch, 2007], because disturbances in control loops tend to propagate throughout the whole plant.

Tables 3.5 and 3.6 summarize important developments in advanced signal analysis methods for process monitoring and diagnosis. For clarity, the methods are categorized as detection or diagnosis. This overview updates the extensive discussion presented in Thornhill and Horch [2007]. Shardt et al. [2012] also review some recent contributions, but with a focus on the control system.

Detection and clustering

As introduced in section 3.1.2, detection involves determining that a disturbance is present in a measurement, and the time when that occurred, while clustering refers to identifying and grouping all measurements affected by the same disturbance. The advanced signal analysis methods for detection and clustering are indicated in table 3.5, which distinguishes between univariate and multivariate analyses.

Table 3.5: References on advanced signal analysis methods for process monitoring and diagnosis - Part I: detection.

Univariate approaches		Multivariate approaches
Single-frequency oscillations	Multiple-frequency oscillations	Multiple-frequency oscillations
Regularity of the zero-crossings of the time series [Thornhill and Hägglund, 1997]	Empirical mode decomposition [Srinivasan et al., 2007, Babji and Rengaswamy, 2012]	Spectral PCA decomposition [Thornhill et al., 2002]
Regularity of the zero-crossings of the auto-covariance function [Thornhill et al., 2003, Karra and Karim, 2009]	Discrete cosine transform [Li et al., 2010, Wang et al., 2013]	Spectral ICA decomposition [Xia et al., 2005]
		Spectral NNMF decomposition [Tangirala et al., 2007, Babji and Tangirala, 2010]
		Implementation of NNMF with genetic algorithm optimization method [El-Ferik et al., 2012]
		Cross-correlation-based spectral envelope [Jiang et al., 2007]

Univariate detection was initially limited to measurements disturbed by oscillations of a single frequency. To detect that type of disturbance, the first methods analysed the regularity of the zero-crossings of the time series [Thornhill and Hägglund, 1997]. To improve the robustness of the method to random noise, Thornhill et al. [2003] and more recently Karra and Karim [2009] used instead the zero-crossings of the auto-covariance function of the measurement.

Using wavelet methods, which operate in the time-frequency domain, Matsuo et al. [2004] could also analyse oscillations whose magnitude varies in time, or are intermittent.

More recently, two alternative methods were proposed to handle one time series with multiple frequencies of oscillation. These methods are shown in the middle column of table 3.5. Srinivasan et al. [2007] proposed, and later improved [Babji and Rengaswamy, 2012], a method based on empirical mode decomposition, while Li et al. [2010] and Wang et al. [2013] developed a method based on the discrete cosine transform. The two methods are robust to the presence of slowly varying trends and noise in the measurement. However, they identify only the dominant frequency of oscillation.

Multivariate methods aim to identify all frequencies present in measurements disturbed by oscillations, and to determine clusters of measurements having each frequency. These aims are sought in the frequency domain because each frequency of oscillation appears as a separate peak in the spectrum of a measurement. The right-most columns of table 3.5 summarize relevant multivariate methods.

Some of the multivariate methods decompose the multivariate set of spectra into basis shapes. Each basis shape should capture one of spectral peaks. Different decomposition methods used include principal component analysis [Thornhill et al., 2002], independent component analysis [Xia et al., 2005], and non-negative matrix factorization (NNMF) [Tangirala et al., 2007, Babji and Tangirala, 2010]. These methods have different statistical assumptions on the basis shapes. NNMF is the most relaxed, and yields basis shapes which resemble the most with a spectrum. Recently, El-Ferik et al. [2012] proposed to implement the NNMF decomposition and search for basis functions using an optimisation method known as Genetic Algorithm. The purpose was to avoid converging to local optima.

In contrast to the spectral decomposition methods, Jiang et al. [2007] achieve detection and clustering using the spectral envelope. This is based on the cross-correlation between the spectra of the various measurements. Thus it is able to find common frequency components between all spectra.

Diagnosis and propagation path

As introduced in section 3.1.2, diagnosis involves determining the root cause of a disturbance, while the propagation path refers to the directed succession of process variables according to the order of propagation of the disturbance. Table 3.6 distinguishes between methods which investigate the time series

Table 3.6: References on advanced signal analysis methods for process monitoring and diagnosis - Part II: diagnosis.

Univariate approaches	Multivariate approaches	
Non-linear root causes	Linear relationships	Non-linear relationships
Non-linearity index based on surrogate testing [Thornhill, 2005]	Time lags determined from linear cross-correlation [Bauer and Thornhill, 2008, Yang et al., 2012]	Nearest neighbours to determine predictability [Bauer et al., 2007a]
Non-linearity index based on high-order statistics of the spectrum [Choudhury et al., 2004, Zang and Howell, 2007]	Transfer of energy between frequencies determined from directed transfer function [Gigi and Tangirala, 2010]	Nearest neighbours to determine time lags [Stockmann et al., 2012] Transfer entropy measure to determine predictability [Bauer et al., 2007b] Direct transfer entropy to distinguish direct and indirect pathways [Duan et al., 2012] Mutual information combined with transfer entropy to reduce computational requirements [Naghoosi et al., 2013]

of each measurement separately, and methods which analyse relationships between pairs of measurements.

The methods under the univariate heading focus on disturbances which start due to some non-linear instability in the system. Non-linear root causes include control valves with stiction, and hydrodynamic instabilities such as compressor surge and slug flow. The methods that diagnose disturbances with this type of root cause measure the non-linearity in the time series of each measurement. The reason to do that is that process equipment acts as a mechanical low pass filter [Stephanopoulos, 1984]. Hence, disturbed measure-

ments farther from the root cause have time series which are less non-linear [Thornhill and Horch, 2007].

Non-linearity indices have been developed based on surrogate testing [Thornhill, 2005], and high-order statistics in the spectrum of the measurement [Choudhury et al., 2004, Zang and Howell, 2007]. These indices have been often applied to the diagnosis of valve stiction [Choudhury et al., 2006, Zakharov et al., 2013].

The methods under the multivariate heading determine the propagation path of a disturbance through the system. As discussed in section 2.4.1, a variable which follows another in the propagation path may show time lag, low pass filtering, amplitude attenuation, and added noise. Advanced signal analysis methods exploit these properties. When the propagation path is determined, the first variable can be declared as the closest to the root cause.

There has been much debate about the nature of the relationships between variables in the propagation path. Several authors have argued that these relations are of causality, that is, that the dynamic behaviour of the lagging variable is caused by the preceding variable [Bauer et al., 2007b, Naghoosi et al., 2013]. Other authors have pointed out that precedence in a pairwise relation does not imply causality [Yang and Xiao, 2012]. This was also argued with the tanks example in section 3.4. The example showed that when there is more than one direction of propagation, then variables along different directions will have precedence relationships but not causal relationships.

Some methods to determine the propagation path are limited to pairs of variables which have a linear relationship. These methods are indicated in the middle column of Table 3.6. The cross-correlation method by Bauer and Thornhill [2008] determines time lags between time series. This idea was also used by Yang et al. [2012] for the management of alarms. Another method, which operates in the frequency-domain, uses directed transfer functions. These capture the transfer of energy between frequencies in different variables [Gigi and Tangirala, 2010].

Other methods are applicable to pairs of variables with linear or non-linear relationships, and are indicated in the right-most column of Table 3.6. These methods either use the concept of mutual nearest neighbours, or the concept of transfer entropy. Both concepts quantify if one time series helps predicting future values of the second time series.

Mutual nearest neighbours were first used by Bauer et al. [2007a], and have recently also been used to determine time lags between time series with a non-linear relationship [Stockmann et al., 2012].

Transfer entropy was also first used by Bauer et al. [2007b]. More recently, Naghoosi et al. [2013] proposed to initiate the search for precedence relationships with another measure, mutual information, which has lower computational requirements. Mutual information identifies which variables are related, and then transfer entropy can be used on the shortlisted variables. Duan et al. [2012] proposed a variant called direct transfer entropy. The authors argue that this measure distinguishes between a direct pathway without any intermediate variables, and an indirect pathway through some intermediate variables, such as F_1 and F_3 in Figure 3.3a.

Advanced signal analysis methods for machinery and electrical equipment

Mechanical and electrical engineering applications have often used methods based on analysing the time series of measurements. Jardine et al. [2006] present an overview of these applications, and more recent contributions include Widodo and Yang [2008], Ottewill and Orkisz [2013].

The analyses are mostly carried in the frequency, or time-frequency domain. The reason is that several variables in the mechanical and electrical systems are oscillatory. Changes in the normal frequencies of oscillation indicate the presence of disturbances, and often particular frequencies can be associated to a type of disturbance. As a result, it is common that the results of advanced signal analyses are used to build the data-based models which were described in section 3.6.1.

A difference from the process application is the lack of system-wide analyses. In fact, the focus is on analysing individual pieces of equipment, and not on comparing the behaviour of different machines placed in the same system. A reason to compare the behaviour of different machines would be to recognize when similar disturbances propagate between machines in the same system.

Comments on the use of advanced signal analysis methods

The distinguishing characteristics of advanced signal analysis methods can be summarized as follows:

- The measurements from process variables are handled as time series.
- The methods are not trained on a specific system.
- Most methods take a plant-wide approach. Examples are the methods for clustering, and for determining the propagation path.
- All methods were developed for persistent disturbances.

The first three characteristics are particular strengths of the methods, and justify the choice of advanced signal analysis for this thesis. The use of time series allows the detection of undesired trends in addition to undesired measurement values. It also allows the identification of complex features in the measurement trend, as well as relations between the trends in different measurements. The plant-wide approach allows a top-down analysis of the disturbances that propagate throughout the process. Instead of analysing the measurement of one variable at a time, the plant-wide methods can analyse a large number of measurements and then narrow down to the most interesting ones.

The main gap in advanced signal analysis is the absence of methods for slowly-developing and transient disturbances. These two types of disturbances induce changes in the trends of measurements, as do persistent disturbances. Therefore, their detection and diagnosis could also benefit from the time series approach.

3.7 Chapter summary and discussion

This chapter has provided a broad survey of the methods of process monitoring and diagnosis, and concluded on the suitability of advanced signal analysis methods to meet the objectives of the thesis.

The methods of process monitoring and diagnosis can be divided into those based on (i) analytical models, (ii) qualitative models, (iii) qualitative trends, and (iv) numerical data. Advanced signal analysis methods are a sub-division of methods based on numerical data. Section 3.3 to section 3.6 analysed in detail the methods in each of the four divisions.

Section 3.3 argued that analytical model based methods can take advantage of fundamental understanding of the system, and hence can identify exactly the cause of a disturbance in the system. However, these methods

are not suitable for the work in this thesis because accurate models are costly for large-scale systems, such as those formed by industrial processes together with their electrical and mechanical equipment. Furthermore, the analytical models cannot be reused because industrial processes tend to be unique.

Section 3.4 presented methods based on qualitative models which describe relationships between variables of the system, such as connectivity, precedence, and causality relationships. Qualitative models are useful to describe the propagation path of a disturbance, and can be derived from fundamental knowledge of the system or from measurement data. The work on this thesis focuses on using measurement data, but not on determining propagation paths. Therefore, the main contributions of this thesis do not use qualitative model based methods. Nonetheless, the future developments section of this thesis discusses a research direction which involves deriving a qualitative model of precedence from measurement data.

Section 3.5 argued that methods which use qualitatively the trends in measurement data are able to recognize relevant trends even in the presence of other trends or noise. However, these methods are not suitable for the work in this thesis because they depend on a comprehensive database which relates trends to types of disturbances and root causes. Therefore, to detect a transient disturbance, that transient would have to exhibit one of the particular trends in the database. Furthermore, a single transient disturbance may have different shapes in different measurements, and hence the method would not recognize that both measurements are affected by a common disturbance.

Section 3.6 reviewed the methods which take advantage of the numerical data available in the data historians. Some methods build data-based models which can associate the current state of the process with its being normal or affected by some type of disturbance. Data-based models are easy to implement and lead to fast detection of abnormal conditions. However, these methods were not chosen for the work in this thesis because their success depends on an accurate characterization of the data under normal and disturbed operations. Furthermore, the data-based models are limited to the systems and types of operation for which they were built. Preferably, the methods developed in this thesis should be as generic as possible.

Section 3.6 also reviewed methods which handle measurement data as time series. These methods are known as advanced signal analysis methods, and they can extract properties from the time series which are related to the

presence of a disturbance or to the proximity to the root cause. This section argued that these methods are the most suitable for the work in this thesis. One reason is that, by considering time series, advanced signal analysis methods take into account the dynamic behaviour of the system. This is advantageous because disturbances in process variables include undesired values as well as undesired trends. Furthermore, advanced signal analysis methods are not developed for a particular system or operating conditions, as with analytical models and data-based models. Finally, most advanced signal analysis methods take a system-wide approach. This approach is particularly useful in this thesis because the analysis can relate the behaviour of process variables with the behaviour of electrical and mechanical variables.

The next chapters of the thesis will develop methods of advanced signal analysis. Those chapters will introduce additional background and related work which are relevant for the context and understanding of the method developed.

Part II

Contributions of the thesis

Chapter 4

Univariate detection of transient disturbances

The extension of process monitoring and diagnosis to electrical and mechanical measurements makes it essential to detect transient disturbances. The reason is that disturbances related to the electrical utility are mostly of a transient nature, caused by power imbalances in the grid which lead to momentary frequency and voltage instabilities [Bevrani, 2009].

The aim of this chapter is to develop a method to detect transient disturbances in the measurement of a single variable. As discussed in section 2.4.2, a transient disturbance is defined in this thesis as a short-lasting deviation of a measurement from its previous and subsequent trend. In addition, the deviation seldom repeats within the time horizon of analysis. After the transient, the measurement may return to its previous trend or follow a different trend. Examples of the former are voltage spikes and deviations caused by sensor faults [Misra et al., 2002]. Examples of the latter are the responses of the system to step changes, as in Figure 4.1.

A method to detect transient disturbances is needed because the developments in advanced signal analysis have focused on persistent disturbances, as discussed in section 3.6.1. The methods to analyse persistent disturbances are not appropriate for transient disturbances because they rely on the repetition of the abnormal deviation episode [Xia et al., 2005, Thornhill, 2005, Jiang et al., 2007, Bauer et al., 2007a].

Methods of Multivariate Statistical Process Monitoring (MSPM) have been applied to some transient disturbances. For example, the traditional statis-

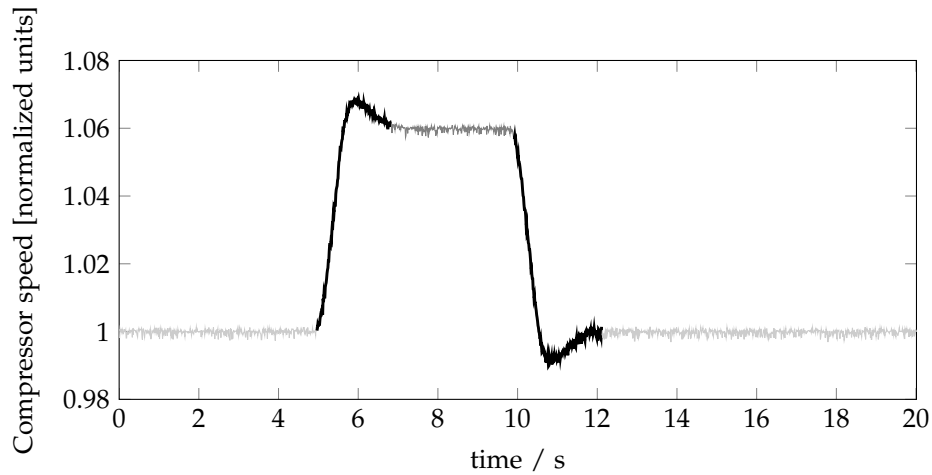


Figure 4.1: Reference example: time series of the shaft speed of a compressor, with transients around 5 and 11 s. The compressor speed is normalized by its initial value.

tical process control methods can detect transients whose magnitude goes beyond $\bar{x} \pm 6s_x$, where \bar{x} and s_x are the mean and standard deviation of a measurement X during normal operation. However, these methods are not appropriate if the system generating the measurement is oscillatory or cyclical in nature. Another example is a wavelet decomposition method [Misra et al., 2002] which detects transient disturbances which map to wavelet coefficients of high amplitude in the lower scales. To detect the transient-related coefficients, the non-transient parts of the signal must not produce other high amplitude coefficients at those same scales. However, this may not always be the case, as Figure 4.2 exemplifies. The top panel of the figure shows a temperature measurement from an industrial gas processing plant. The measurement has an oscillatory trend and a pulse at hour 4 caused by a transient disturbance. The bottom panel of Figure 4.2 shows the wavelet coefficients for the measurement at different times (horizontal axis) and scales (vertical axis). The light tones represent high amplitude coefficients. The plot shows that the transient disturbance maps to high amplitude coefficients from scales one to four. However, the measurement shows various steep increases and decreases which are not part of transient disturbances. The steep variations also map to high amplitude coefficients at the same scales as the transient, and hence the transient cannot be distinguished.

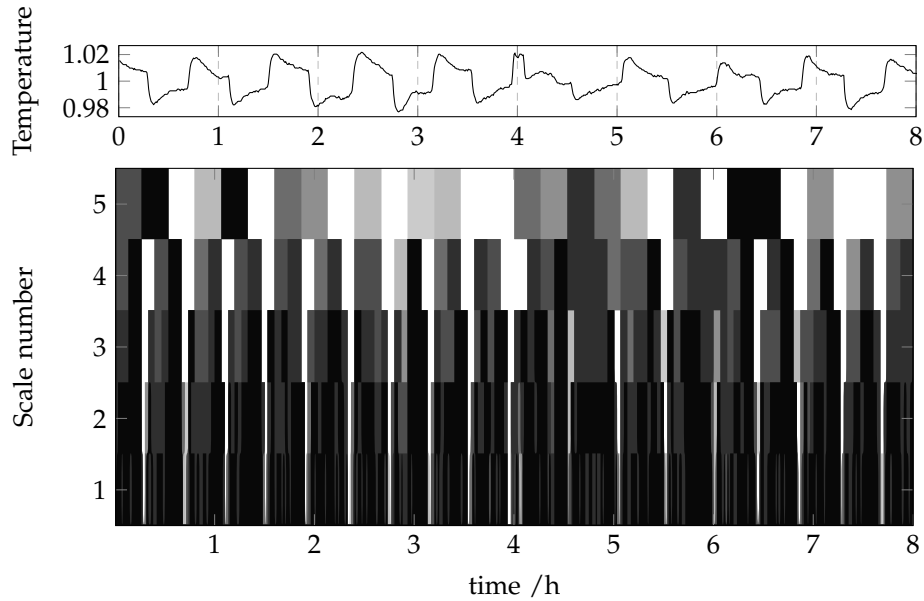


Figure 4.2: Top panel: temperature measurement with normalized units. Bottom panel: corresponding coefficients of Haar wavelet at different times (horizontal axis) and scales (vertical axis). Lighter tones represent higher amplitude coefficients.

Methods of Qualitative Trend Analysis (QTA) have also been applied to some transient disturbances. However, as discussed in section 3.5, these methods can only detect transients whose shapes match those in the QTA database. This thesis aims to develop a more general approach than is offered by QTA.

Following the advanced signal analysis approach, the method developed in this chapter handles the measurement of a variable as a time series. As a result, the transient disturbance can be seen as an unusual segment. The problem of detecting an unusual segment is framed as an anomaly detection problem, and solved with a nearest neighbours technique. One advantage of this method is that it can detect any rare and short-term deviation of the measurement from its overall trend, regardless of the frequency components or relative amplitude of that deviation. Another advantage is that the method does not require the development of data-based models, and hence is not specific to a particular system or mode of operation.

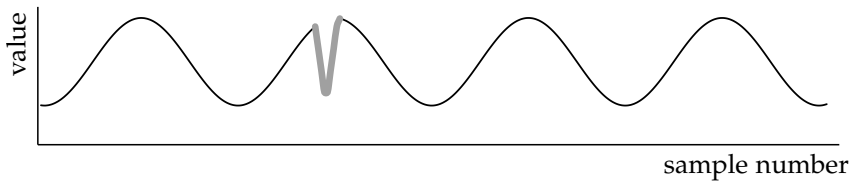


Figure 4.3: Time series with an anomalous segment, drawn in a thick grey line.

4.1 Background

4.1.1 Detecting anomalous segments with nearest neighbours

If the measurement affected by a transient disturbance is handled as a time series, then the sequence of samples corresponding to the transient will be distinct from the normal trend of the time series. In other words, this sequence of samples will be considered anomalous. A sequence of samples from a time series X which are ordered in time is known as a segment x . Figure 4.3 illustrates a time series with an anomalous segment drawn in a thick grey line.

Finding anomalous segments in a time series is a known problem in data mining and is classified as a particular case of anomaly detection, an area which has been reviewed by Chandola et al. [2009]. The techniques proposed for the detection of anomalous segments fall into one of three categories: (i) information theoretic, (ii) classification-based or (iii) based on nearest neighbours. All techniques rely on the assumption that the normal behaviour of a time series follows a defined pattern hence any segment not conforming to this pattern is an anomaly.

Nearest neighbours techniques are widely applied for anomalous segment detection [Chandola et al., 2009]. The basic idea is to use a similarity measure to evaluate the similarity between each segment in the time series and all other segments. The most similar segments are known as nearest neighbours. In Figure 4.3, the similarity between the anomalous segment and any of the other segments in the time series is low because its sequence of samples is significantly different. Therefore, even the nearest neighbour to the anomalous segment will be significantly different. The similarity of a segment to its nearest neighbours is used to define its anomaly index ai . Nearest

neighbours should be distinguished from near-in-time neighbours, which are adjacent segments with overlapping samples [Thornhill, 2005].

The effectiveness of nearest neighbours techniques in detecting anomalous segments has been demonstrated on time series with trends which resemble those of process measurements with transient disturbances. Examples include electrocardiogram signals [Chuah and Fu, 2007, Keogh et al., 2005] and current measurements from a Space Shuttle control valve [Keogh et al., 2005]. Additional reasons to choose the nearest neighbours techniques are their autonomy from models of normal and anomalous data, in contrast to classification-based techniques, and the use of a measure that is sensitive even when only one anomaly is present, in contrast to information theoretic techniques [Chandola et al., 2009].

4.1.2 Similarity measures for time series

Similarity measures for time series depend on the desired notion of identity. Measures commonly reported in the literature include:

- the Euclidean distance metric [Keogh et al., 2005],
- the cosine similarity [Thornhill et al., 2006] and related correlation measure [Lhermitte et al., 2010], and
- Dynamic Time Warp (DTW) [Fabozzi and Van Cutsem, 2011].

The Euclidean distance between two segments \mathbf{x}_1 and \mathbf{x}_2 with m samples considers each segment as a point in an m -dimensional space, and is defined as the 2-norm of the displacement vector between the two points.

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^m (x_{1,i} - x_{2,i})^2} \quad (4.1)$$

Zero distance indicates maximum similarity, and occurs only between segments which are equal in all samples, that is $\mathbf{x}_1 = \mathbf{x}_2$. Except when $\mathbf{x}_1 = \mathbf{x}_2$, the Euclidean distance is affected by the variance of the segments and by their units of measurement.

In contrast to the Euclidean distance, the cosine similarity and correlation are insensitive to variance and units of measurement. However, sensitivity to variance is desirable for the detection of transient disturbances since changes in variance may be considered a disturbance.

Table 4.1: Definitions of anomaly index reported in the literature.

Anomaly index	References
Distance to k^{th} nearest neighbour	[Chuah and Fu, 2007, Keogh et al., 2005]
Sum of k distances to k nearest neighbours	[Angiulli and Pizzuti, 2002, Zhang and Wang, 2006]
Number of nearest neighbours within distance D (inverse)	[Knorr et al., 2000, Palshikar, 2005]
Fraction of k nearest neighbours having the segment in their own k nearest neighbours (inverse)	[Hautamäki et al., 2004]
Density of k nearest neighbours relative to own density	[Breunig et al., 2000, Pokrajac et al., 2007]

DTW is a modified version of the Euclidean distance which attempts to maximize the matching of segments x_1 and x_2 by deforming the time axis. For the detection of transient disturbances, however, this property leads to less distinction between a short-duration transient and, for example, an oscillation in the background.

For these reasons, the method developed in this chapter uses the Euclidean distance metric to measure similarity between the segments of a measurement.

4.1.3 Anomaly index definitions

Anomaly detection is based on the anomaly index ai of each segment x of the time series. With the nearest neighbours technique, the anomaly index ai is defined from the similarity assessment. Table 4.1 summarizes different ways in which the results of the similarity assessment have been used.

The distance between a segment and its k^{th} nearest neighbour is a widely used approach, and is also the approach chosen for this method. The other approaches mentioned in table 4.1 are less adequate for dealing with industrial process measurements with transient disturbances. The reasons are listed below.

- Summing the distances from segment x_i to its k nearest neighbours is less sensitive to quantized measurements due to the possibility that a large part of k nearest neighbours have the same distance. Quantized measurements are common in industrial data.

- Counting the number of neighbours of segment x_i within a certain distance requires a parameter D which is less intuitive to set than the integer k . Furthermore, the resulting anomaly index is less discriminating because the number of neighbours is an integer.
- Finding the k nearest neighbours of segment x_i , and counting how many of these neighbours have x_i in their own k nearest neighbours is also less discriminating due to being a ratio of integers.
- The density ρ_i of the neighbourhood of x_i is defined as the inverse of the distance to its k^{th} nearest neighbour, that is, $\rho_i = 1/d_k$. Similarly, $\rho_{j,i}$ denotes the density of the neighbourhood of the j^{th} nearest neighbour of x_i , where $j = 1 \cdots k$. Comparing density ρ_i to the average of the k densities $\rho_{j,i}$ is suitable if the anomalous segment has a significantly less dense neighbourhood. However, this is not the case when there are more than one identical anomalous segments, or when the underlying trend of the measurement is neither periodic nor steady.

4.2 Method development

This section explains the method proposed to detect transient disturbances in the time series of a measurement. The development case study is first presented and then used to illustrate the explanation. This section also analyses the statistics of the threshold for detection, and presents a colour map to visualize the detection results.

4.2.1 Development case study

Figure 4.1 presented the time trend of the shaft speed of a compressor from the case study *Compressor rig case 1*. Section 2.5.1 explained the origin of the data and the set-up of the system. The measurement lasts 20 seconds and is available at 1 kHz.

Two step changes, around 5 and 11 s, were imposed in the drive of the compressor by changing its speed set-point, and resulted in the two transients highlighted in the figure by the black lines. The objective of the proposed method is to detect those transients.

4.2.2 Time series

The method proposed handles the data from a measurement as a time series. A time series X is a finite sequence of n samples, taken at strictly increasing time instants (4.2).

$$\{x(t_1), x(t_2), \dots, x(t_n)\} : t_1 < t_2 < \dots < t_n \quad (4.2)$$

When the interval Δt between each sample is constant, the time instants t_i are strictly increasing multiples of that interval, that is $t_i = i \cdot \Delta t$. In this case, replacing t_i with its index i becomes a valid representation of time, and the time series can be simply represented by its sequence of samples in the form $X = \{x_1, x_2, \dots, x_n\}$.

A segment refers to any sequence of samples from the time series that are ordered in time. An embedded vector is a particular type of segment, with m samples τ instants apart, that is $\mathbf{x}_i = [x_i \ x_{i+\tau} \ \dots \ x_{i+(m-1)\tau}]$. Parameter m is known as embedding dimension and τ as embedding granularity.

4.2.3 Algorithm

The basis of detecting transient disturbances is the detection of anomalous segments in a time series. Detection of anomalous segments is a data-driven procedure based on computing an anomaly index ai_i for each embedded vector \mathbf{x}_i of the time series.

Embedding matrix

Equation (4.3) shows the embedding matrix \mathbf{X} constructed from the time series $X = \{x_1, x_2, \dots, x_n\}$.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{N_E} \end{bmatrix} = \begin{bmatrix} x_1 & x_{1+\tau} & \dots & x_{1+(m-1)\tau} \\ x_{1+\delta} & x_{1+\delta+\tau} & \dots & x_{1+\delta+(m-1)\tau} \\ \vdots & \vdots & & \vdots \\ x_{1+(N_E-1)\delta} & x_{1+(N_E-1)\delta+\tau} & \dots & x_{1+(N_E-1)\delta+(m-1)\tau} \end{bmatrix} \quad (4.3)$$

Each row in the embedding matrix is an embedded vector, defined as above, and each adjacent embedded vector in the matrix lags the previous by δ

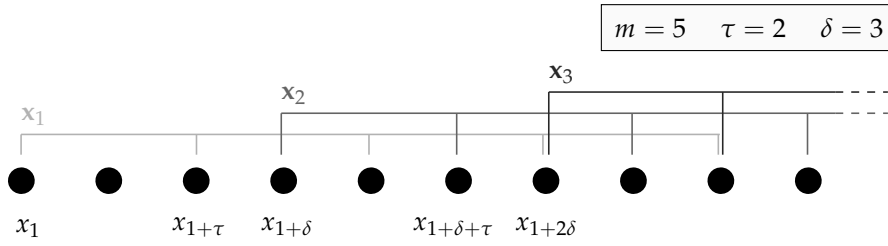


Figure 4.4: Representation of a time series of samples x_i and the selection of the first three embedded vectors, given parameters m , τ and δ .

samples, a parameter known as embedding step. Figure 4.4 illustrates the selection of three embedded vectors from a symbolical time series represented by dots. The selected embedded vectors will be adjacent in the embedding matrix.

The number of embedded vectors N_E in an embedding matrix is limited by the total number of samples n of the time series according to equation (4.4). The incomplete brackets indicate a floor function, which maps a real number to the largest previous integer.

$$1 + (N_E - 1)\delta + (m - 1)\tau \leq n \quad \Leftrightarrow \quad N_E = \left\lfloor \frac{n - (m - 1)\tau - 1}{\delta} + 1 \right\rfloor \quad (4.4)$$

Figure 4.5a plots some embedded vectors from the reference example. The grey tones relate the embedded vectors to their location in the original time series. The horizontal axis refers to the index of a sample in the embedded vector.

Before the similarity assessment, each embedded vector in the matrix is mean-centred. The effect of this step is that embedded vectors which have identical trends but different ranges of their numerical values will become similar. This is illustrated in Figure 4.5b, in which the embedded vectors previously at lower and higher ranges, dark and light grey lines respectively, are now close. This step is optional and used when different levels of operation and long-duration ramps should be considered normal.

Similarity

Each embedded vector is then compared to every other embedded vector, using the Euclidean distance metric. A simple implementation of the similarity

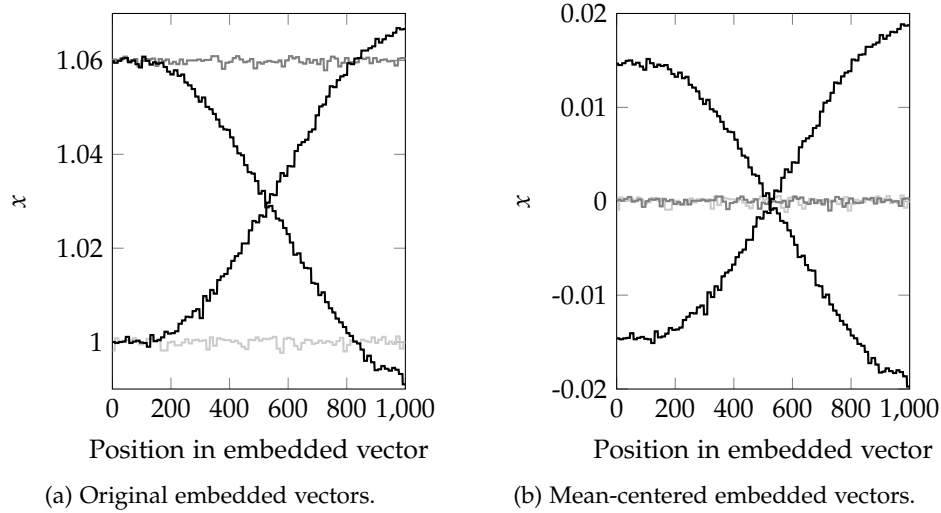


Figure 4.5: Selected embedded vectors from the development case study, with grey tones relating to their position in the original signal. Each embedded vector covers 1 s. Parameters are $m = 1001$, $\tau = 1$ and $\delta = 1$.

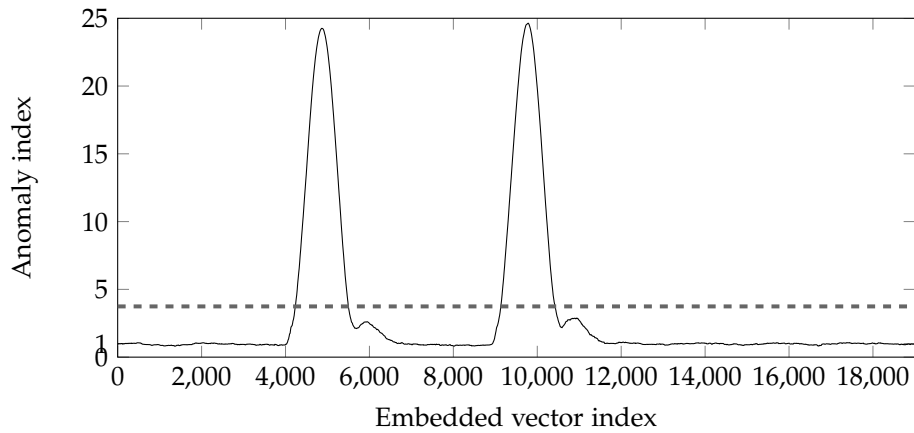
assessment is with two nested loops, running along the rows of the embedding matrix, which will generate a $N_E \times N_E$ matrix of distances.

Anomaly index vector

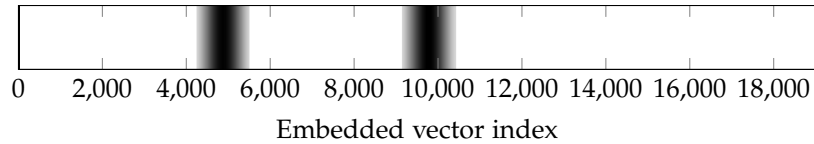
An anomaly index ai is attributed to each embedded vector. It is defined as the distance to the k^{th} nearest neighbour, that is, the k^{th} smallest distance between that embedded vector and every other.

It should be noted that distances to near-in-time embedded vectors are excluded from the neighbourhood assessment. The reason is to avoid that near-in-time embedded vectors be falsely detected as near neighbours, which can happen because near-in-time embedded vectors have overlapping samples. Specifically, when embedded vector \mathbf{x}_i is considered, its near-in-time embedded vectors are those which include one or more samples in common with it. For example, when one considers $\mathbf{x}_1 = [x_1 \ x_2 \ \dots \ x_m]$, the embedded vector $\mathbf{x}_m = [x_m \ x_{m+1} \ \dots \ x_{2m-1}]$ is the last of its near-in-time embedded vectors, for $\tau = 1$ and $\delta = 1$.

Attributing an anomaly index ai to each embedded vector generates an anomaly index vector \mathbf{ai} of N_E anomaly indices. Figure 4.6a shows the anomaly index vector for the reference example, with $m = 1001$ and $k = 3$. The vector was normalized by its median so that $ai = 1$ now approximates the average



(a) Normalized anomaly index vector. The dashed line indicates the detection threshold.



(b) Colour plot obtained after applying the threshold and mapping the detected indices onto a colour scale. Darker tones indicate higher anomaly indices.

Figure 4.6: Anomaly index vector for the reference example ($k = 3$).

anomaly index of non-anomalous embedded vectors. The next section will justify this statement.

Figure 4.6a shows that the anomaly indices of the embedded vectors with numbers around 5000 and 10000 stand out in the anomaly index vector. These embedded vectors are those which include the transient disturbances.

4.2.4 Significance level

A threshold on the value of the anomaly indices can be used to distinguish normal from anomalous embedded vectors. The threshold is based on the probability distribution of the anomaly indices because anomaly indices of anomalous embedded vectors are significantly high and are a small fraction of all anomaly indices. This is true because, by definition, transient disturbances are infrequent in the time series analysed.

In statistical literature, values numerically distant from the rest of the data are treated as outliers [Rousseeuw and Leroy, 2003]. The cutoffs to isolate

those values are based on robust statistics, which are less influenced by the outlier values than classical statistics and are, thus, more representative of the rest of the data [Huber and Ronchetti, 2009].

The threshold proposed here is based on two robust statistics, the median and the interquartile range, IQR, according to equation (4.5). The median is defined as the 50th percentile (Q_2), and IQR as the difference between the 75th and 25th percentiles. When the anomaly index vector is normalized by its median, then the term $Q_2(\mathbf{ai})$ in equation (4.5) is one. Figure 4.6a shows the threshold for the development case study with a dashed line.

$$ai > Q_2(\mathbf{ai}) + 6 \times \text{IQR}(\mathbf{ai}), \quad (4.5)$$

The performance of the detection threshold can be quantified by analysing its behaviour under null hypotheses of time series with no anomalies. The remainder of this section shows that the probability of false positives with the proposed threshold is less than one in a million. This is demonstrated for the cases of (i) steady state operation with only random noise, (ii) operation with non-random variability, and (iii) oscillatory operation.

The underlying assumption is that the anomaly index vector generated from a time series without anomalous embedded vectors approximately fits to a gamma distribution with a skewness of less than 0.77.

Skewness is a measure of asymmetry of a frequency distribution. The anomaly index vector in Figure 4.6a will have a skewed distribution because there are a few large anomaly indices and many small ones. The statistical gamma distribution commonly describes waiting times between events and, specifically, it models sums of exponentially distributed random variables [Weisstein, 2010]. It is reasonable to assume that the gamma distribution can model the anomaly index vector because anomaly indices are distances, hence square roots of the sums of squared differences. Several authors also refer to the use of the gamma function to model the frequency distribution of nearest neighbour distances [Chandrasekhar, 1943, Bansal and Ardell, 1972, Evans et al., 2002]. Furthermore, the gamma distribution presents a number of properties that agree with anomaly indices, such as being defined only for positive real numbers, being positively skewed and converging to a Gaussian distribution when its skewness tends to zero.

Appendix A.1 confirms the validity of assuming that the anomaly index vector generated from a time series without anomalous embedded vectors approximately fits to a gamma distribution with a skewness smaller than 0.77. To represent time series without anomalous embedded vectors, the demonstration includes (i) steady state operation with only random noise, (ii) operation with non-random variability, and (iii) oscillatory operation. Appendix A.2 proves that, when that assumption holds, then the probability that the proposed threshold causes false detections is less than one in a million.

4.2.5 Outputs of the detection

The numerical outputs of the transients detection method are:

- the number of transient disturbances,
- the initial and final time indices of each disturbance, and
- a measure of the severity of each disturbance.

A fourth output is a colour plot. This allows the visualization of the detection results in a compact way that suggests the propagation of the disturbance through the system.

Time indices and number of transient disturbances

Consecutive embedded vectors classified as anomalous identify one transient disturbance. The indices I and F identify, respectively, the initial and final anomalous embedded vectors in a group of consecutive anomalous embedded vectors. Indices I and F can be used to estimate the initial and final time indices of the corresponding transient. To this end, it is assumed that an embedded vector is anomalous if at least half of it includes the transient. Therefore, indices I and F estimate the initial \hat{I} and final \hat{F} time indices of the transient disturbance according to

$$\begin{aligned}\hat{I} &= \left\lfloor 1 + (I - 1)\delta + \frac{(m - 1)\tau}{2} \right\rfloor \\ \hat{F} &= \left\lceil 1 + (F - 1)\delta + \frac{(m - 1)\tau}{2} \right\rceil\end{aligned}\tag{4.6}$$

Table 4.2: Numerical outputs of the detection method for the reference example.

Number of transients	initial time index \hat{I}_t	final time index \hat{F}_t	severity \mathcal{S}_t
2	4733	5996	15.4
	9630	10923	15.1

The incomplete brackets in equation (4.6) indicate floor and ceiling functions which, respectively, map a real number to the largest previous and the smallest following integer.

The number of transient disturbances is given by the number of groups of consecutive anomalous embedded vectors.

Severity of a transient disturbance

The severity \mathcal{S}_t of transient disturbance t is given by equation (4.7), which represents the average of the anomaly indices ai_i from the anomalous embedded vectors with indices between I_t and F_t .

$$\mathcal{S}_t = \frac{\sum_{i=I_t}^{F_t} ai_i}{F_t - I_t + 1} \quad (4.7)$$

Table 4.2 exemplifies the numerical outputs of the detection method for the development case study.

Colour plot

A colour plot shows how the analysed measurement evolved in time with regards to the presence of transient disturbances. Figure 4.6b shows the colour plot for the development case study. Anomaly indices corresponding to anomalous embedded vectors are mapped onto a two-colour scale. In Figure 4.6b, the maximum anomaly index is mapped onto black, the anomaly index corresponding to the detection threshold is mapped onto white, and the anomaly indices in between are linearly mapped onto a weighted sum of the two colours. Anomaly indices which are below the detection threshold are mapped to the minimum limit of the colour scale, for a better contrast.

As shown in Figure 4.6b, the resulting colour plot compactly describes the periods affected by transients, and the evolution of the magnitude of the

anomaly index. Section 4.4 will also show that the colour plot can suggest in a compact way the propagation of the disturbance through the system.

4.2.6 Computational effort

The basic nearest neighbours technique is $\mathcal{O}(N_E^2)$, where N_E is the number of embedded vectors. The reason is that every embedded vector has to be compared to every other embedded vector. This yields a total of $N_E(N_E - 1)/2$ distance computations.

Figure 4.8 indicates the time taken in the computation of the distance matrix, using the development case study with different downsamplings. To reduce the computation time, some authors have suggested modifications to the basic nearest neighbours algorithm based on approximation [Keogh et al., 2005] or pruning [Ramaswamy et al., 2000].

4.3 Parameter settings and sensitivity

To generate the anomaly index vector \mathbf{ai} , the following parameters have to be selected:

- embedding granularity τ ,
- embedding dimension m ,
- embedding step δ , and
- number of nearest neighbours k .

This section starts by relating the physical meaning of the parameters to the dynamics of the system, and then recommends optimal parameter values and analyses the sensitivity of the detection results to those values. The optimisation is not done by mathematical programming, but instead by exploring the parameter space over a range of values. The objective is to understand how the parameters affect the performance of the method, and determine values that (i) are good recommendations as default values, (ii) are robust to change, and (iii) are meaningful relative to the sampling rate or the process dynamics.

4.3.1 Relation between parameters and dynamics of the system

The embedding granularity τ is the number of sampling intervals Δt between each sample included in an embedded vector. It should, therefore, be small enough to characterize the main trends of the transient disturbances. For example, if the interesting feature were an oscillation, τ should at least satisfy the Nyquist sampling criterion for that oscillation.

The embedding dimension m is the number of samples in an embedded vector and, together with τ and Δt , defines the duration of the embedded vector according to $(m - 1)\tau\Delta t$. This duration should match the time scale of the transient disturbances, otherwise the embedded vectors will capture events at other time scales. An example in the case study measurement in Figure 4.1 would be to detect the whole prominence between 5 s and 12 s as one disturbance if m is large enough. Another example would be to capture noise as a disturbance if m is too small.

The embedding step δ is the number of sampling intervals Δt between consecutive embedded vectors. Therefore, δ can influence the accuracy of the estimated start \hat{I} and end \hat{F} indices of the transient disturbances, and should be set as $\delta = 1$. Specifically, the accuracy in \hat{I} and \hat{F} will be within the range $\pm\delta/2$ samples. This means that if δ is greater than twice the width of the transient, the reduced time accuracy may lead to a failure to detect the transient. The actual accuracy in \hat{I} and \hat{F} depends on the elapsed time between the start of the measurement and the transients. On the other hand, larger values of δ accelerate the computation of the anomaly index vector since the method is $\mathcal{O}(N_E^2)$ and N_E is proportional to $1/\delta$.

The meaning of k , the number of nearest neighbours, can be described in the following way: if there is a group of at least $k + 1$ identical embedded vectors, these will be considered non-anomalous. Therefore, the minimum value for k has to be the number of identical transients, to avoid not detecting the transients. At the same time, k has to be smaller than the number of identical non-anomalous embedded vectors, in order to avoid these being considered anomalous.

Table 4.3 summarizes the roles of the parameters, as discussed in this section.

Table 4.3: Summary of the roles and recommended values for the parameters.

Parameters	Role	Recommended value
number of samples in a transient	characterize the main trends of the transient disturbances	30
τ		1
m	defines the time scale of the transient disturbance	15
δ	defines the time accuracy for the transient disturbance	$\delta \leq 5$
k	defines the maximum recurrence of identical transients	$3 \leq k \leq n/10$

4.3.2 Recommendations for parameters and analysis of sensitivity

This section determines the best values of the parameters, and analyses the sensitivity of the detection results in the range of those values. To these ends, anomaly index vectors are generated from the development case study with different parameter values, and the performance of the results is evaluated.

The performance is defined by noting that embedded vectors fall into one of two categories: those which include a transient, \mathbf{x}_j , and those which do not include a transient, \mathbf{x}_l . In the following, the total number of \mathbf{x}_j is denoted as N_E^{trans} , and the total number of \mathbf{x}_l is denoted as N_E^{norm} . Clearly, $N_E^{trans} + N_E^{norm} = N_E$, where N_E is the total number of embedded vectors. For parameter optimisation, the indices of \mathbf{x}_j and \mathbf{x}_l are identified visually beforehand. In this section, this is done with the development case study.

If the detection procedure is perfect, then all \mathbf{x}_j are classified as anomalous and all \mathbf{x}_l are classified as normal. However, if the detection is not perfect, some embedded vectors will not be correctly classified. Classification of the embedded vectors as anomalous or normal is based on the threshold proposed in section 4.2.4.

Metric FN (false negatives) (4.8) assesses the case of anomalous embedded vectors being incorrectly classified as normal. To that end, it counts the embedded vectors covering a transient, \mathbf{x}_j , which were correctly classified as anomalous, N_{TP} , where TP stands for true positives, and compares this num-

ber with the total number of \mathbf{x}_j , N_E^{trans} . This metric is one if the detection procedure is perfect, and zero if all \mathbf{x}_j were incorrectly classified as normal.

$$FN = \frac{N_{TP}}{N_E^{trans}} \quad (4.8)$$

Metric *FP* (false positives) (4.9) assesses the case of normal embedded vectors being incorrectly classified as anomalous. It counts the embedded vectors not covering a transient, \mathbf{x}_l , which were correctly classified as normal, N_{TN} , where *TN* stands for true negatives, and compares this number with the total number of \mathbf{x}_l , N_E^{norm} . This metric is one if the detection procedure is perfect, and zero if all \mathbf{x}_l were incorrectly classified as anomalous.

$$FP = \frac{N_{TN}}{N_E^{norm}} \quad (4.9)$$

As well as minimizing false negatives and false positives, the default parameter values should also maximize the ratio between the anomaly indices from embedded vectors covering a transient, ai_j , and from embedded vectors from normal operation, ai_l . This metric is denoted as *ANR* (anomaly-to-normal ratio) and defined by (4.10).

$$ANR = \frac{\sum_j ai_j}{\sum_l ai_l} \quad (4.10)$$

The next sections combine the above metrics into a performance index to evaluate the detection results.

Number of samples in a transient, embedding granularity τ and embedding dimension m

Parameters τ and m are directly dependent on the sampling interval, Δt , used to collect the data, hence values for these parameters must be given relative to that sampling interval.

A suitable sampling interval for the detection of transients must lead to enough samples in the transients to guarantee their characterization. The optimal number of samples in a transient was analysed by subsampling the case study measurement and considering the number of samples in the first transient, which is the shortest. Figure 4.7 shows, with the horizontal axis, how the number of samples in the shortest transient influences the performance

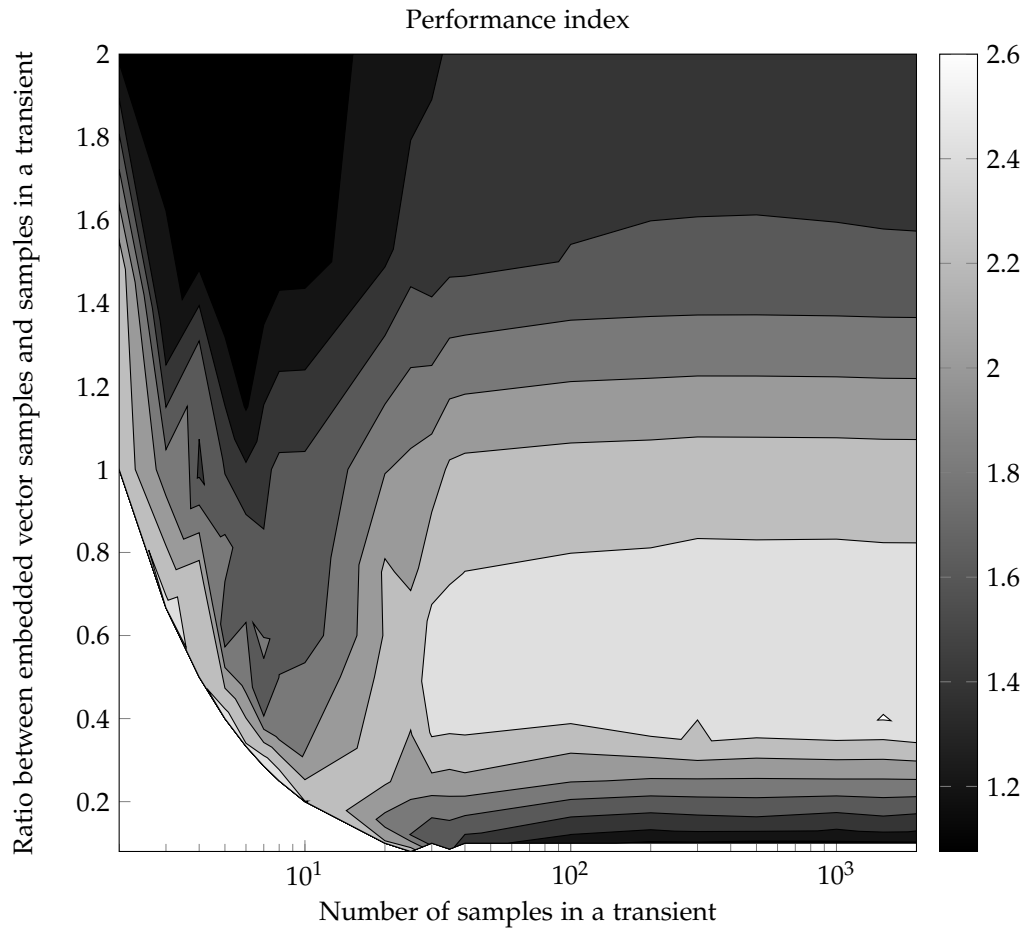


Figure 4.7: Performance of the detection method as a function of the number of samples in a transient and of the ratio between embedded vector samples and samples in a transient. The performance index can vary between zero and three. Parameters fixed in the analyses were $\tau = 1$, $\delta = 1$ and $k = 3$.

index. For this optimisation, the performance index is the sum of the three objectives, FP , FN and ANR , with ANR normalized by its maximum value.

Figure 4.7 shows higher values of the performance index the more samples characterize the transient, as expected. More importantly, it also shows that above 30 samples the performance is constant. This is important because the fewer samples needed in the time series, the faster is the method, as shown in Figure 4.8. Since most transient disturbances are abrupt, the use of 30 samples would characterize the transient as well as some of its details.

Users can set such sampling interval based on their knowledge or past experience with the system. For instance, if the experience of the site is that

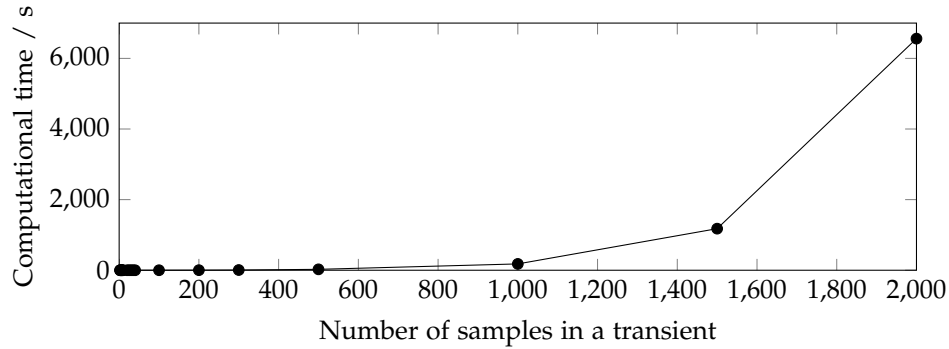


Figure 4.8: Computational time as a function of the number of samples in a transient. The computational time is an average over the different values of m .

electrical transients have a typical duration of 1 s, then the sampling interval should be 0.033 s. With the suitable sampling interval defined, parameter τ can be set to one.

Figure 4.7 also analyses, with the vertical axis, the optimal number of samples in the embedded vectors in relation to the number of samples in the transient. It shows that the detection method achieves highest performances when the ratio between the number of samples in the embedding vector and the number of samples in the transient is between 0.4 and 0.75. Choosing the ratio of 0.5 implicitly defines the optimal value for parameter m as 15. Another observation from Figure 4.7 is that the optimal range 0.4 to 0.75 is consistent above the value of 30 samples in the shortest transients. This is important because it is robust. For instance, if the time scale of the real dynamics of the system is up to 1.5 times slower than expected, and the transients are characterized by 40 samples instead of 30, $m = 15$ will still lead to similar performance indices.

Embedding step δ

Figure 4.9 shows the influence of different values of δ on the performance index. For this optimisation, the performance index is the sum of the objectives FP and FN because δ influences the accuracy of detection of the start and end of the transient. The values of performance index shown are the average of those from five time series. These are identical to the subsampled case study measurement, but start at different time instants, specifically with a delay of one sample. The reason to calculate performance indices for different starting

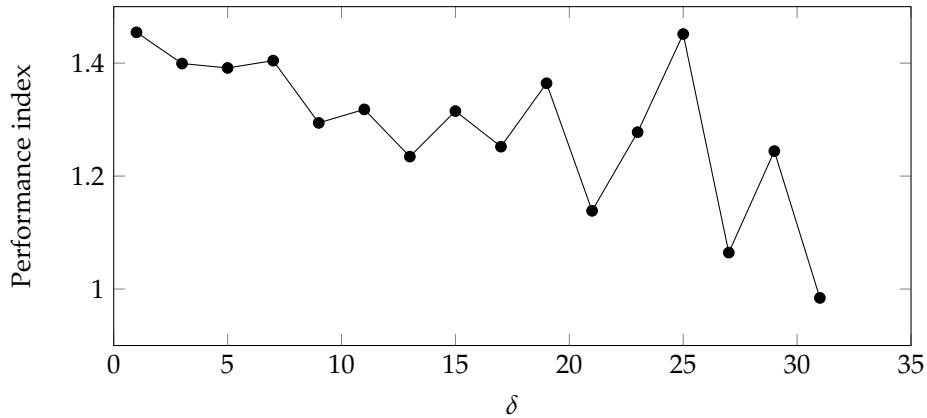


Figure 4.9: Performance of the detection method as a function of δ . The performance value is an average over five time series with different starting instants, and it can vary between zero and two. Parameters fixed in the analyses were the sampling interval, such that the number of samples in a transient was 30, $\tau = 1$, $m = 15$ and $k = 3$.

instants is because the actual accuracy of the results depends on the chosen section of data, in particular the elapsed time between the transients and the start of the time series.

Figure 4.9 shows that the average performance of the method decreases approximately linearly with δ , as expected from the discussion in section 4.3.1. Also, the decrease in performance becomes more inconsistent as δ increases. As expected, these results suggest that δ should be set to one. In this example, with 300 samples, the computational time with $\delta = 1$ is below 0.02 s. However, in time series with more samples it may be important to increase δ , given that the method is $\mathcal{O}(1/\delta^2)$. In such cases, it is recommended to keep δ below 5, otherwise Figure 4.9 shows that the average performance of the method may fall below 90% of the performance achieved with $\delta = 1$. The value of $\delta = 5$ is equal to $m/3$, given that $m = 15$. This means the uncertainty in the time of occurrence of the transients is a sixth of the duration of the transient. Ultimately, the uncertainty in the time of occurrence of the transients is a user-specific choice.

Number of nearest neighbours k

Parameter k must be at least equal to the number of identical transients. If, at the time of preparation of the data or from historical operation experience, users have an idea of typical numbers of transient disturbances occurring in

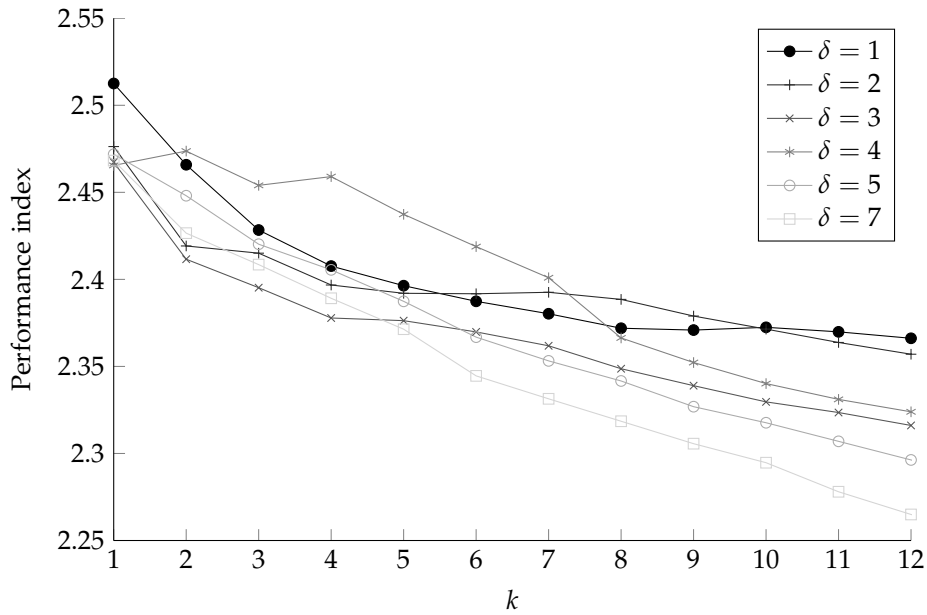


Figure 4.10: Performance of the detection method as a function of k , for different values of δ . The performance index can vary between zero and three. Parameters fixed in the analyses were the sampling interval, such that the number of samples in a transient was 30, $\tau = 1$ and $m = 15$.

the site, they may set k accordingly. In principle, however, such numbers are not known a priori, so k should be set high to avoid false negatives.

Figure 4.10 and Figure 4.11 show the influence of different values of k on the performance index, which for this optimisation is the sum of the three objectives, FP , FN and ANR , with ANR normalized by its maximum value. The seven lines correspond to different values of δ .

Figure 4.10 focuses on small values of k to show that, for most values of δ , the performance in the development case study increases with decreasing values of k . In particular, $k = 1$ leads to the highest performances in this example, which was expected since the case study measurement has only one transient of each shape. However, the minimum value recommended is $k = 3$, even when only one identical transient is expected. The reason is to allow some margin of error in case the number of identical anomalous embedded vectors is higher than expected. The reduction in performance, meanwhile, is small, specifically in this example it is less than 3%.

The maximum value of k recommended depends on the dynamics of the system. If time series from routine operation have some periodicity, k should

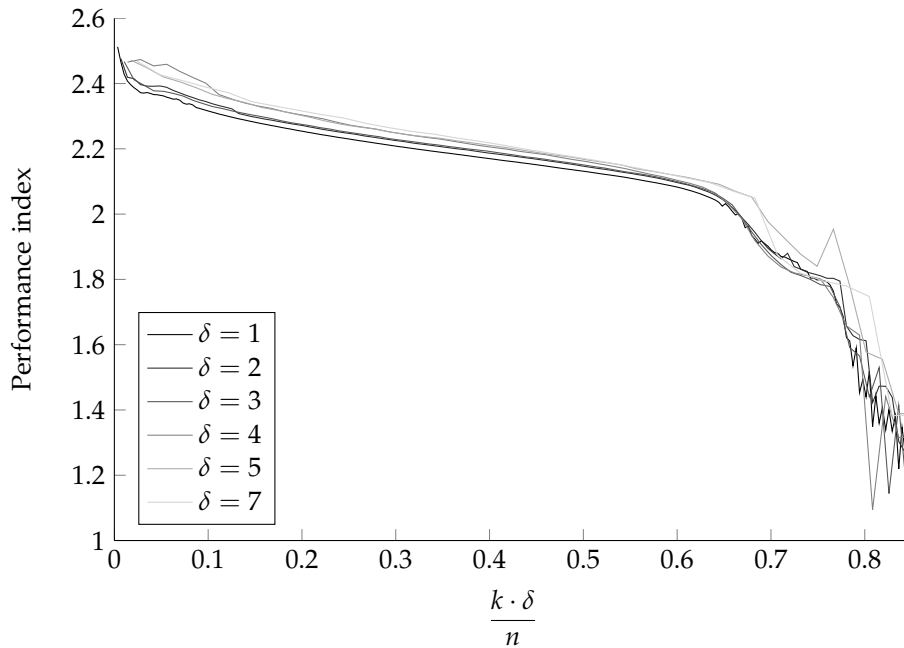
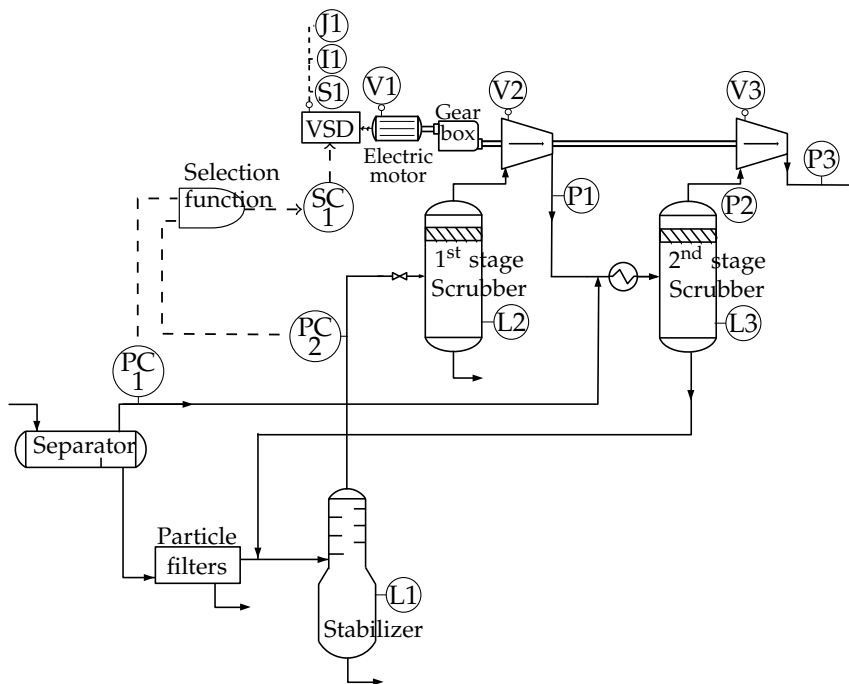


Figure 4.11: Performance of the detection method as a function of k and δ . The performance index can vary between zero and three. Parameters fixed in the analyses were the sampling interval, such that the number of samples in a transient was 30, $\tau = 1$ and $m = 15$. The number of samples is $n = 300$.

be less than n/T_p , where n is the number of samples in the time series and T_p is the number of samples in one cycle. If, on the other hand, routine time series derive from steady-state operation, then the value of k should not exceed $n/2\delta$. This value is approximately half the number of embedded vectors and derives from the fact that, by definition, the number of non-anomalous embedded vectors needs to be greater than the number of anomalous ones.

Figure 4.11 supports this last statement. The figure now plots the performance index as a function of $k \cdot \delta/n$, which is approximately the ratio between k and the number of embedded vectors. Ratios above 0.6 show a marked decay in performance, which agrees with the recommendation in the previous paragraph of using ratios below 0.5. For $\delta = 5$, for example, this means that k should not exceed $n/10$.

Table 4.3 summarizes the values recommended for all the parameters. These values assume the choice of a sensible sampling interval, as discussed in the beginning of the section. If the original sampling interval is smaller than necessary, data can be subsampled or the parameters adjusted proportionally.

Figure 4.12: Process schematic for the test case study *Gas plant case 1*.

4.4 Application to test case study

The *Gas plant case 1* is the case study used to test the transients detection method. The case study derives from routine operation of an industrial gas processing plant, as explained in section 2.5. Figure 4.12 shows the selected part of the process, which includes a gas-condensate separation section, with a separator, filters and stabilizer, and a gas recompression section, with scrubbers and compressors. The speed of the compressors is used to adjust the pressure in the system, either at the outlet of the separator or at the outlet of the stabilizer.

Figure 4.13a shows the time trends of 15 measurements affected by two sharp spikes occurring around minutes 500 and 2000. Each time trend has 5400 samples taken with a sampling interval of 30 s. The measurements include process variables, *i.e.* pressure and level, electrical variables, *i.e.* drive current and power, mechanical variables, *i.e.* speed and vibration, and control signals. Most time trends of the measurements are not constant. Some measurements show a ramp, namely L2, noise, namely I1, V1 and V3, and

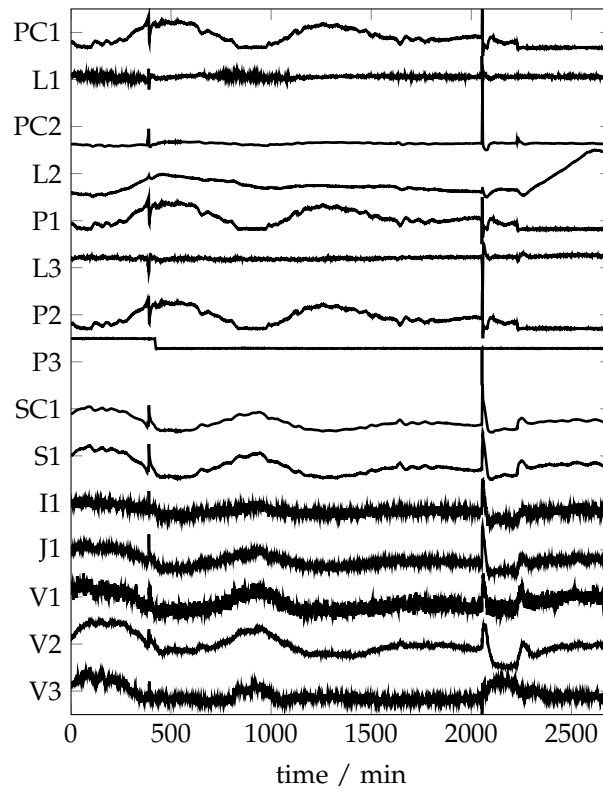
oscillations, namely PC1, P1, P2, SC1, S1, J1, V1 and V2. This thesis excludes ramps from the definition of transients due to being a long-term event.

The expectation for this test case study is that the nearest neighbours method correctly identifies the transient disturbances, while being insensitive to the other behaviour, that is, the oscillations, the noise and the ramp. The time scale of the shortest transient disturbances is of around 10 to 15 min, which is common for process disturbances. The original sampling interval of 30 s means these transients are described by 20 to 30 samples. The recommended number of samples in a transient is 30, which means the data should not be downsampled. The other parameters are set to their recommended values, that is $\tau = 1$, $m = 15$, $\delta = 1$ and $k = 3$. The embedding dimension of $m = 15$ covers 0.75 of the transients with 20 samples, which according to Figure 4.7 is in the region of highest performances for this number of samples.

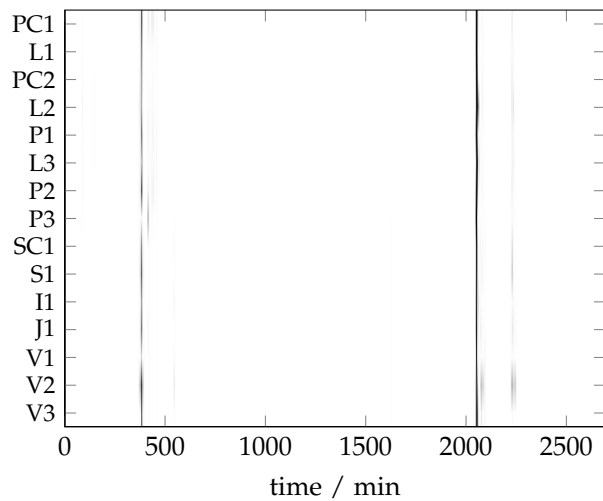
The results are shown in Figure 4.13b using the proposed colour plot. The upper limit in the colour scale corresponds to the maximum anomaly index in each measurement. The transient spikes are identified in all measurements, while the other noise, oscillations and ramp are ignored.

The colour plot also allows visualization of the order by which the disturbances appeared in each measurement. This is more easily visualized by reordering the measurements according to the start of the disturbance. Figure 4.14 reorders the measurements for the first disturbance, and zooms in to its start.

In each measurement, the transient disturbance starts when the colour changes from white to light grey. The darker tones reflect the transient becoming more distinct from the rest of the trend. The colour plot suggests that the transient first appeared in the gas streams between the gas outlet of the separator (PC1) and the compressors (P1 and P2). The results show that the disturbance is picked up by the control loop which acts on the compressors (SC1). The transient is consequently transmitted to the pressure at the inlet of the first compressor (PC2), and to some electro-mechanical measurements which reflect the operation of the compressors (S1, V2 and J1). Eventually the liquid condensate in the scrubbers is affected (L2 and L3) and the transient recycles back to the stabilizer (L1). This propagation path is suggested by the colour plot and it agrees with the physics of the system.



(a) Time trends.



(b) Results of the transient detection shown in the colour plot. Darker tones indicate the presence of a transient disturbance.

Figure 4.13: Condition of the measurements in the test case study *Gas plant case 1*.

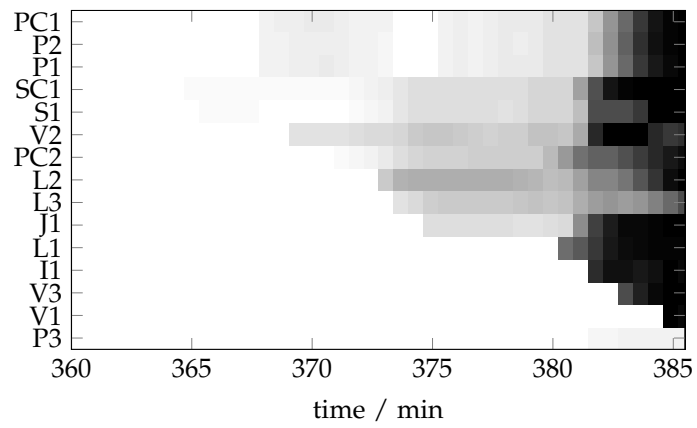


Figure 4.14: Colour plot for the industrial case study, zoomed in to the first transient and reordered to highlight the different times of start of the transient.

4.5 Chapter summary and discussion

This chapter has demonstrated the use of a nearest neighbours technique in the detection of transient disturbances in measurements from chemical process systems and associated electromechanical equipment. The extension of process disturbance analysis to electromechanical measurements is justified by the increasing use of electrically-driven machinery in process industries and the rising number of power quality incidents.

The method proposed in section 4.2.3 is based on the concept of anomaly detection, and was shown to be effective in detecting transients and distinguishing these from other dynamics such as oscillations, ramps, noise and change in operation level. The method requires a detection threshold and parameters. Therefore, section 4.2.4 and section 4.3.2 presented recommendations for both and analysed these recommendations. Section 4.2.5 also proposed a colour plot to visualize the detection results clearly and in a way that suggests the propagation of the disturbance through the system.

The transients detection method is a new contribution to the field of process monitoring and diagnosis. Besides its direct application in detection, this method is also useful as a preceding step for diagnosis and for the removal of transients in order to enhance the analysis of oscillating measurements. The method, and particularly the visual output, are also industrially applicable for plant-wide monitoring, for example to understand the plant-wide impact of a certain transient, to identify variables most often affected, and for identifying delays, which suggest a propagation path.

The results show that the method offers informative insights when combined with thorough knowledge of the process.

Chapter 5

Multivariate detection of transient disturbances in uni-rate systems

Chapter 4 developed a method to detect transient disturbances in the measurement of a single variable. The basis of the method is the identification of anomalous segments in the time series of the measurement. Segments of the time series with transient disturbances generate anomaly indices above the detection threshold, while transient-free segments have anomaly indices below the detection threshold. Chapter 4 also explored the performance of the nearest neighbours method in depth, and proposed default parameters which optimised the method for a wide range of cases.

However, the method in Chapter 4 takes a univariate approach, which means that measurements from different variables are analysed separately. A multivariate analysis, on the other hand, is able to exploit the presence of the same disturbance in several measurements. As this chapter later shows, the identification of a transient disturbance can be difficult in measurements with strong oscillatory trends or noise. In such cases, exploiting the presence of the same disturbance in other measurements can lead to an improved outcome.

The aim of this chapter is to extend the nearest neighbour method optimised in Chapter 4 into a multivariate analysis. The multivariate extension is solved using singular value decomposition on the set of anomaly index vectors of the various measurements. The method will be illustrated with experimental data from the *Compressor rig* data set, and with industrial data from the *Gas plant* data set. The chapter will demonstrate superior results with

the multivariate analysis in comparison to the univariate approach, when the analysed measurements have strong oscillatory trends or noise.

5.1 Background

5.1.1 Multivariate analysis with SVD

In this chapter, the objective of a multivariate analysis is to identify representative features across a set of anomaly index vectors. Representative features capture the largest proportion of variance of the anomaly index vectors, which happens if the feature has itself a large variance or is present in many of the anomaly index vectors.

In a generic multivariate analysis of a set of variables, the variables can be arranged as row vectors in a matrix \mathbf{A} , and the representative features across the set of variables can be identified through a Singular Value Decomposition (SVD) of matrix \mathbf{A} . SVD is a mathematical technique which factorizes the matrix \mathbf{A} into a product of three other matrices according to

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\top}. \quad (5.1)$$

Matrices \mathbf{U} , \mathbf{S} , and \mathbf{V}^{\top} show the following properties [Golub and Van Loan, 2012, Trefethen and Bau III, 1997]:

- The columns of matrix \mathbf{U} are orthonormal vectors which are basis functions for the columns of \mathbf{A} . In other words, the columns of matrix \mathbf{U} are linearly independent and, by linear combination, can represent every column of \mathbf{A} .
- Similarly, the rows of matrix \mathbf{V}^{\top} are orthonormal basis functions for the rows of \mathbf{A} .
- Matrix \mathbf{S} is a diagonal matrix of singular values of \mathbf{A} . A singular value s_j gives the total projection of the rows, or columns, of \mathbf{A} along the direction of the associated orthonormal basis, namely, row vector \mathbf{v}_j^{\top} or column vector \mathbf{u}_j . Furthermore, singular values are ordered by decreasing value.

When the rows of matrix \mathbf{A} are mean-centred, SVD has also a statistical interpretation: the square of a singular value s_j is directly related to the total

variance of \mathbf{A} along the direction defined by the corresponding row. Specifically, it equals the variance times $n - 1$, where n is the number of elements in the row vector. This means that the first few row vectors \mathbf{v}_j^\top are the directions which capture the largest proportion of variance of the rows of \mathbf{A} . This is the reason why SVD can be used for multivariate analysis.

SVD is closely related to the statistical technique of Principal Component Analysis (PCA) [Jolliffe, 2005]. PCA is formulated as

$$\mathbf{A} = \mathbf{T}\mathbf{V}^\top, \quad (5.2)$$

where $\mathbf{T} = \mathbf{U}\mathbf{S}$, from equation (5.1). In the context of PCA, matrices \mathbf{T} and \mathbf{V}^\top are known as score and loading matrix, respectively. PCA has been extensively used in process monitoring [Bakshi, 1998, Thornhill et al., 2002]. For example, in Multivariate Statistical Process Monitoring (MSPM) the interest is to identify representative features across plant profiles [Qin, 2012]. As explained in section 3.6.1, a plant profile is the ensemble of the values of all process measurements at a particular instant in time. Representative features across plant profiles are retrieved as the rows of matrix \mathbf{V}^\top .

5.2 Method development

This section explains the method proposed to detect transient disturbances in a multivariate context. The development case study is first presented and then used to illustrate the explanation. To argue the advantages of the multivariate analysis, the development case study is also used to show that univariate detection can be difficult in measurements with strong oscillatory trends or noise. This section also analyses the statistics of the threshold for detection in the multivariate context.

5.2.1 Development case study

Figure 5.1 presents the time series of the seven measurements in the case study *Compressor rig case 2*. The time series correspond to the measurement tags in the schematic of Figure 5.2, and are available at a rate of 1 kHz.

Two step changes were imposed in the set-point of the compressor speed around 6 and 18 s. These resulted in transient disturbances, which are visible in Figure 5.1 around these times in the first five measurements. Additionally,

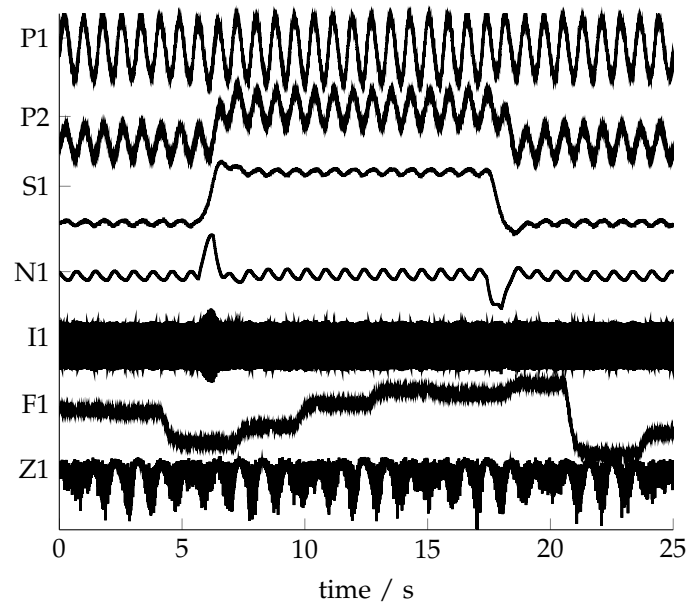


Figure 5.1: Development case study: time series of measurements from the *Compressor rig case 2*.

the first five measurements are also affected by oscillations, which are due to unstable operation of the compressor close to the surge limit. Of course, it is not desirable to operate an industrial compressor so close to its surge limit, but the experimental rig in question was able to explore such operation. The measurements of flow, F1, and valve position, Z1, present a different behaviour. These seven measurements were chosen because they reflect four distinct situations with regards to the presence of transients.

In measurements of speed, S1, and torque, N1, the transient disturbances can be seen clearly, whereas in measurements of suction pressure, P1, and discharge pressure, P2, the transients are present but masked by the oscillations. The measurement of current, I1, is also affected by the two transients, but only the first is clearly visible in its time series.

The interest of F1 is that it does not show the transient disturbances at 6 and 18 s. The reason is that the compressor was operating unstably in this experiment, and the flow rate was low and pulsating. However, the flow instrument is a vortex-shedding flow meter, not suited to disturbed flows especially at the lower end of its measurement range. Even though the flow measurement shows artefacts, it is included in the data set because it does show abrupt decays around 4 and 20 s, as well as several smaller positive steps which are

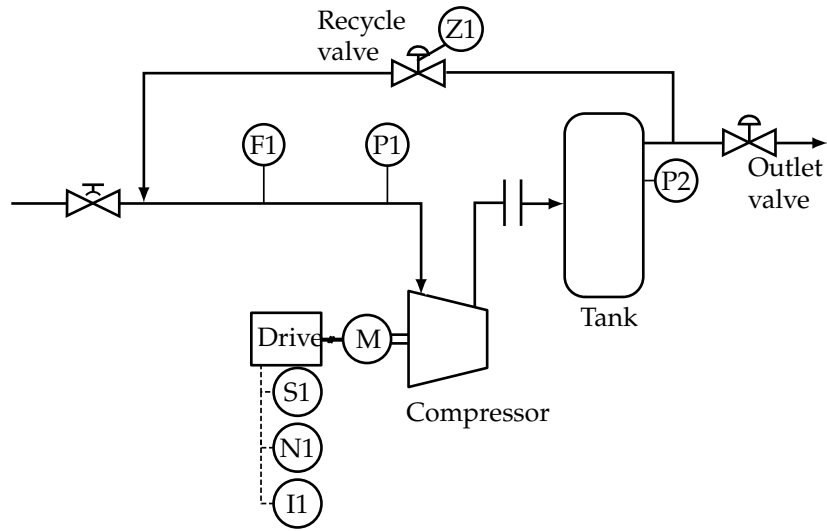


Figure 5.2: Simplified schematic of the gas compression experimental rig.

not synchronized with the transient features of the other measurements in the data set. As a result, F1 poses a useful challenge to the development of the method.

The interest of Z1 is being free of transient disturbances. The valve position, Z1, was fixed during the experiment. The persistent variations visible in its measurement were an electronic artifact, and thought to be produced by interference between sensors or in the communication between the sensors and the data acquisition module. The trend of Z1 is normalized and the persistent variations are only a small percentage (2%) of the mean value. Although Z1 has dips of short duration, these are not considered as transients because they are frequent and repeating.

The expectations for the method developed in the current chapter are (i) to detect the two transient disturbances in measurements P1, P2, S1, N1, and I1, (ii) to detect the abrupt decays in F1 around 4 and 20 s, and (iii) to acknowledge Z1 as free of transients.

5.2.2 Limitation of univariate approach in measurements with strong oscillatory trends or noise

As proposed in Chapter 4, the detection of transient disturbances in a measurement r is based on computing an anomaly index $ai_{r,i}$ for each embedded vector x_i of the time series of that measurement. An anomaly index $ai_{r,i}$ is

defined as the k^{th} smallest distance between embedded vector \mathbf{x}_i and all other embedded vectors. The sequence of anomaly indices $ai_{r,i}$ of each of the N_E embedded vectors of measurement r defines the anomaly index vector \mathbf{ai}_r .

The application of the univariate method of Chapter 4 to the seven measurements in Figure 5.1 generates the seven anomaly index vectors \mathbf{ai}_r shown in Figure 5.3. As discussed in Chapter 4, each anomaly index vector is normalized by dividing its anomaly indices $ai_{r,i}$ by the median value of all the anomaly indices in the vector. Transient disturbances are detected in measurement r if anomaly indices $ai_{r,i}$ are above the detection threshold marked by dashed lines. The anomaly index vectors in Figure 5.3 were computed using $\tau = 1$, $\delta = 1$, $m = 1001$, and $k = 5$, which are the values suggested in Chapter 4 after optimisation.

As Figure 5.3 shows, the univariate detection method yields the desired results for all measurements except P1. The detection of the transients is noteworthy in measurements P2 and I1 because in these measurements the transients are not clearly distinct from the rest of the time series. In measurement P1 the transients are even less distinct from the rest of the time series, and hence the corresponding anomaly indices did not protrude above the threshold. This means that in the univariate approach the presence of the transients in P1 is not recognized. The next section describes a multivariate method which strengthens the evidence that the transients occur in P1 by exploiting the presence of the same disturbances in the other measurements.

5.2.3 Algorithm

The multivariate detection method starts from the univariate anomaly index vectors \mathbf{ai}_r of N_V measurements.

Pre-treatment and anomaly index matrix

The multivariate analysis requires the normalization and mean-centring of each vector \mathbf{ai}_r . Normalization makes the univariate anomaly index vectors independent of their engineering units, and of the number of samples in the embedded vectors. Otherwise, it would be meaningless to compare the numerical values. This can be done by dividing the anomaly indices in vector \mathbf{ai}_r by the median value of all the indices in the vector, as done in section 5.2.2. The normalized anomaly index vectors are then mean-centred. The reason

is to take advantage of the statistical interpretation of SVD, as discussed in section 5.1.1.

The anomaly index vector of measurement r is now treated as a row vector, $\mathbf{a}\mathbf{i}_r^\top$, and arranged in row r of an anomaly index matrix, \mathbf{A} , of size $N_V \times N_E$. This arrangement is shown in the left-hand side of equation (5.3).

SVD of the anomaly index matrix

The anomaly index matrix \mathbf{A} is factorized with a singular value decomposition as $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$. Equation (5.3) shows the SVD, and puts the orthonormal basis functions \mathbf{v}_j^\top in evidence. The aim of SVD is to identify representative features across the various anomaly index vectors. These vectors are in the rows of \mathbf{A} , hence the features of interest are captured by the rows of matrix \mathbf{V}^\top , as discussed in section 5.1.1.

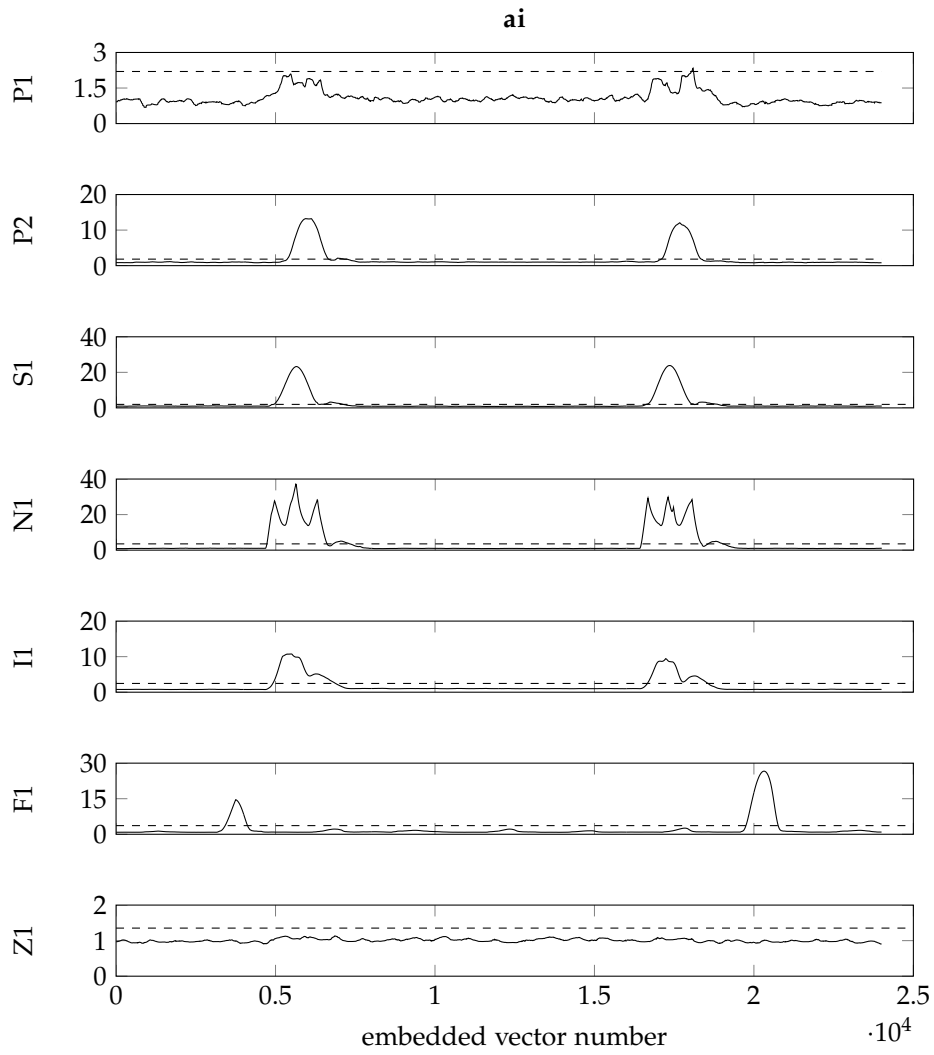
In equation (5.3), $u_{r,j}$ are elements of \mathbf{U} , and s_j are entries in the diagonal of \mathbf{S} .

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}\mathbf{i}_1^\top \\ \mathbf{a}\mathbf{i}_2^\top \\ \vdots \\ \mathbf{a}\mathbf{i}_{N_V}^\top \end{bmatrix} = \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ \vdots \\ u_{N_V,1} \end{bmatrix} s_1 \mathbf{v}_1^\top + \begin{bmatrix} u_{1,2} \\ u_{2,2} \\ \vdots \\ u_{N_V,2} \end{bmatrix} s_2 \mathbf{v}_2^\top + \cdots + \begin{bmatrix} u_{1,N_V} \\ u_{2,N_V} \\ \vdots \\ u_{N_V,N_V} \end{bmatrix} s_{N_V} \mathbf{v}_{N_V}^\top \quad (5.3)$$

Figure 5.4 shows the basis functions \mathbf{v}_j^\top derived from the $N_V = 7$ measurements in the development case study. They are ordered by the magnitude of the associated singular value, s_j , which is indicated on top of each plot.

As desired, the first few basis functions show entries with values which are distinctively higher than the rest of \mathbf{v}_j^\top . In the first and third basis functions, the higher entries correspond to embedded vectors which are associated with the transient disturbances in measurements P1, P2, S1, N1, and I1. In the second basis function, the higher entries are associated with the transient disturbances in measurement F1.

Some of the basis functions \mathbf{v}_j^\top , along with the corresponding singular values s_j and column vectors \mathbf{u}_j , will be discarded from the univariate anomaly index vectors $\mathbf{a}\mathbf{i}_r$. The two following selection steps determine which terms to discard. The detection of transients will be done on the final anomaly index vectors $\tilde{\mathbf{a}}\mathbf{i}_r$.

Figure 5.3: Univariate anomaly index vectors **ai** for the *Compressor rig case 2*.

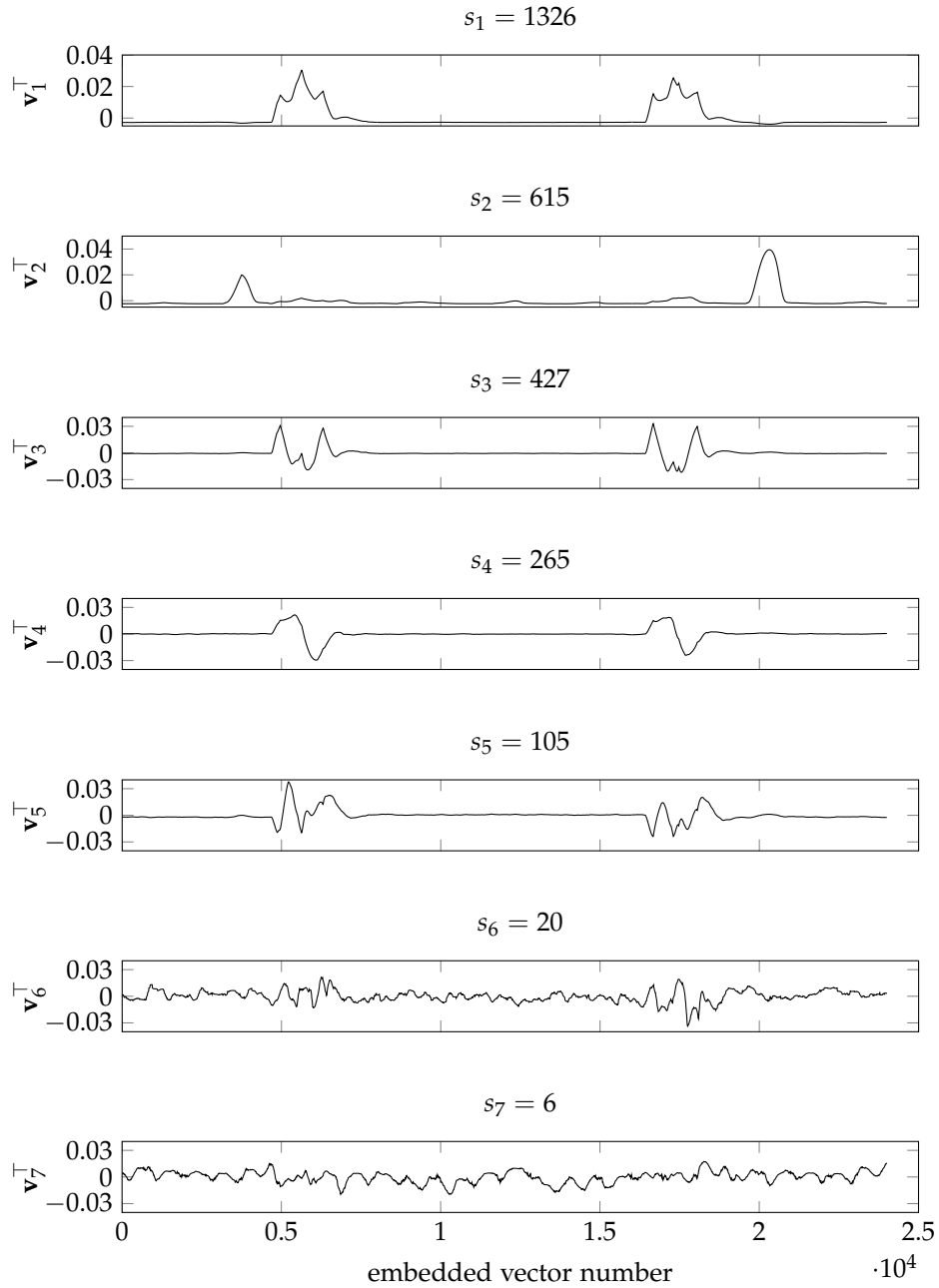


Figure 5.4: Orthonormal basis functions obtained from SVD of the anomaly index matrix. The values on top of the plots indicate the singular value associated with each basis function.

Selection of basis vectors relevant on a multivariate level

The first step discards basis functions \mathbf{v}_j^\top which are not representative across the various anomaly index vectors. This step also discards the corresponding singular values s_j and column vectors \mathbf{u}_j . Specifically, a basis function is retained if its associated singular value, s_j , satisfies the inequality below.

$$s_j^2 \geq \alpha \cdot \sum_{r=1}^{N_V} s_r^2 \quad (5.4)$$

As discussed in section 5.1.1, s_j^2 directly relates to the total variance of the rows of \mathbf{A} along the direction of the basis function \mathbf{v}_j^\top . Accordingly, $\sum_{r=1}^{N_V} s_r^2$ relates to the variance of \mathbf{A} along all directions \mathbf{v}_r^\top .

The selection criterion (5.4) is based on Kaiser's rule [Kaiser, 1960] for selecting components in principal component analysis. This was shown to be a reliable method in the study by Valle et al. [1999].

The development case study uses $\alpha = 0.05$. As a result, the selection retains the first three basis function shown in Figure 5.4, along with the corresponding singular values and column vectors. The anomaly index vectors which result after this step, $\hat{\mathbf{a}}_r$, are formed by the retained terms $u_{r,j}s_j\mathbf{v}_j^\top$, as shown in equation (5.5) for the development case study.

$$\hat{\mathbf{A}} = \begin{bmatrix} \hat{\mathbf{a}}_1^\top \\ \hat{\mathbf{a}}_2^\top \\ \vdots \\ \hat{\mathbf{a}}_{N_V}^\top \end{bmatrix} = \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ \vdots \\ u_{N_V,1} \end{bmatrix} s_1 \mathbf{v}_1^\top + \begin{bmatrix} u_{1,2} \\ u_{2,2} \\ \vdots \\ u_{N_V,2} \end{bmatrix} s_2 \mathbf{v}_2^\top + \begin{bmatrix} u_{1,3} \\ u_{2,3} \\ \vdots \\ u_{N_V,3} \end{bmatrix} s_3 \mathbf{v}_3^\top \quad (5.5)$$

Figure 5.5 shows the intermediate anomaly index vectors $\hat{\mathbf{a}}_r$ resulting from the first selection step. The new detection thresholds, marked with dashed lines, are based on the intermediate anomaly index vectors $\hat{\mathbf{a}}_r$. The numerical values of $\hat{\mathbf{a}}_r$ can be negative as well as positive, and this is due to the mean-centring step at the start of the method.

The first plot, for P1, is evidence that the first selection step facilitates the detection of transients in measurements with strong oscillatory trends or noise. Figure 5.3 had shown that the univariate anomaly index \mathbf{a}_i for P1 did not recognize the presence of the transients. On the other hand, the intermediate anomaly index vector $\hat{\mathbf{a}}_i$ of P1 already shows values above the detection

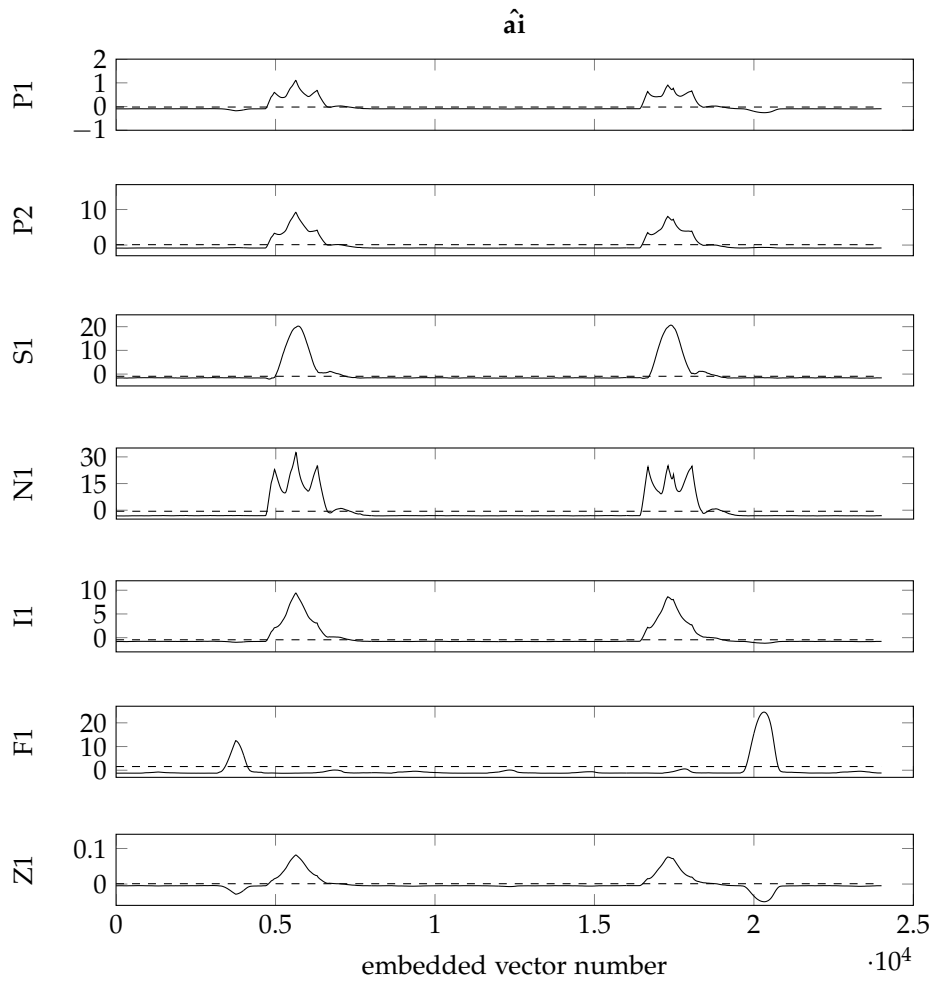


Figure 5.5: Anomaly index vectors after the first selection step, $\hat{\mathbf{a}}_i$, for the development case study.

threshold. This is because the first selection step discarded terms of the univariate anomaly index vector which were associated with the last four basis functions.

Selection of terms relevant on an individual level

The last plot of Figure 5.5 also shows that the anomaly index vector $\hat{\mathbf{a}}_i$ of Z1 after the first selection step erroneously presents values above the detection threshold. The reason is that the basis functions \mathbf{v}_j^T retained in the previous step show little similarity to the behaviour of the univariate anomaly index

vector \mathbf{a}_i of Z1. Therefore, a second selection step discards from each anomaly index vector $\hat{\mathbf{a}}_r$ the terms $u_{r,j}s_j\mathbf{v}_j^\top$ which are not relevant to the univariate anomaly index vector \mathbf{a}_r . Specifically, a term $u_{r,j}s_j\mathbf{v}_j^\top$ from anomaly index vector $\hat{\mathbf{a}}_r$ in equation (5.5) is only retained if its variance, given by $\frac{(u_{r,j}s_j)^2}{(N_E - 1)}$, satisfies

$$\frac{(u_{r,j}s_j)^2}{N_E - 1} \geq \beta \cdot \text{var}(\mathbf{a}_r). \quad (5.6)$$

As a parallel with PCA, β compares the contribution of each score $t_{r,j} = u_{r,j}s_j$ of anomaly index vector \mathbf{a}_r with all its scores.

The second selection step can be visualized in Figure 5.6. Each plot in this figure compares the values of $(u_{r,j}s_j)^2$ for each variable r . The dashed lines are the selection thresholds dependent on β , that is, $(N_E - 1)\beta \cdot \text{var}(\mathbf{a}_r)$. These thresholds were calculated using $\beta = 0.2$. Only terms $u_{r,j}s_j\mathbf{v}_j^\top$ such that $(u_{r,j}s_j)^2$ is above the threshold are retained in the final anomaly index vector.

Equation (5.7) shows that the final anomaly index vectors $\tilde{\mathbf{a}}_r$ are formed by the terms $u_{r,j}s_j\mathbf{v}_j^\top$ which are retained after this second selection step. Terms are discarded by setting $u_{r,j}$ to zero. The zeros in equation (5.7) follow the development case study for illustration.

$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{\mathbf{a}}_1^\top \\ \tilde{\mathbf{a}}_2^\top \\ \tilde{\mathbf{a}}_3^\top \\ \vdots \\ \tilde{\mathbf{a}}_{N_V}^\top \end{bmatrix} = \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ u_{3,1} \\ \vdots \\ 0 \end{bmatrix} s_1 \mathbf{v}_1^\top + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} s_2 \mathbf{v}_2^\top + \begin{bmatrix} 0 \\ 0 \\ u_{3,3} \\ \vdots \\ 0 \end{bmatrix} s_3 \mathbf{v}_3^\top \quad (5.7)$$

Figure 5.7 shows the final anomaly index vectors $\tilde{\mathbf{a}}_r$ for the development case study. It should be recalled that the objective is to detect the presence of two transient disturbances in all measurements except Z1, which must be considered as transient-free. In addition, the pair of transients in measurement F1 must be recognized as occurring at different times than the other measurements. Figure 5.7 shows that the proposed method achieves these objectives. The dashed lines represent the detection thresholds, which are based on the final anomaly index vectors.

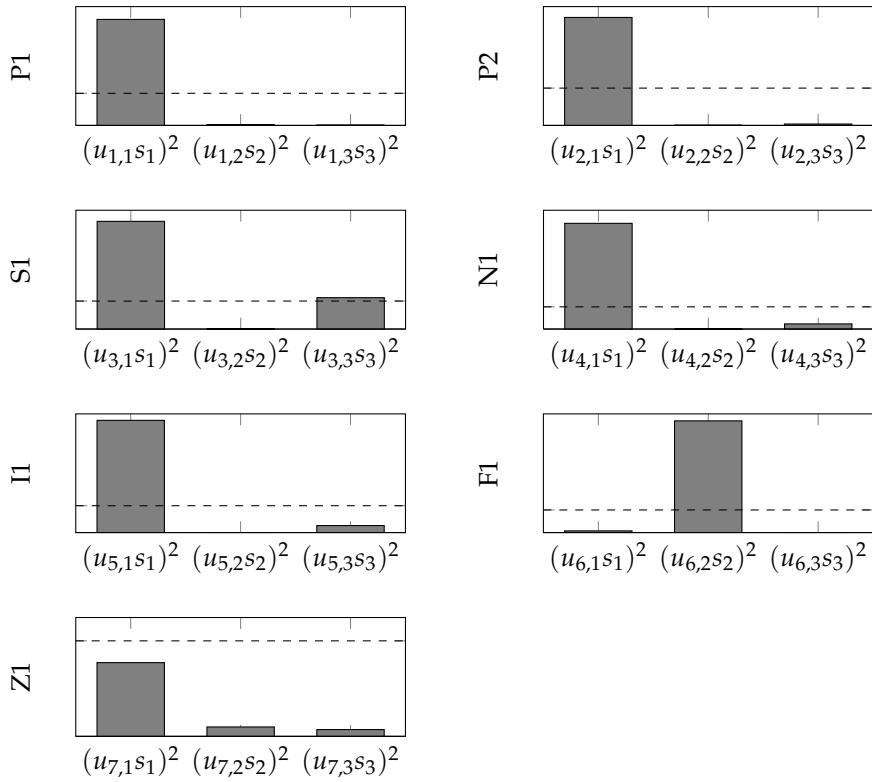


Figure 5.6: Comparison of $(u_{r,j} s_j)^2$ for each variable r with the selection thresholds (dashed lines) derived from $\beta = 0.2$.

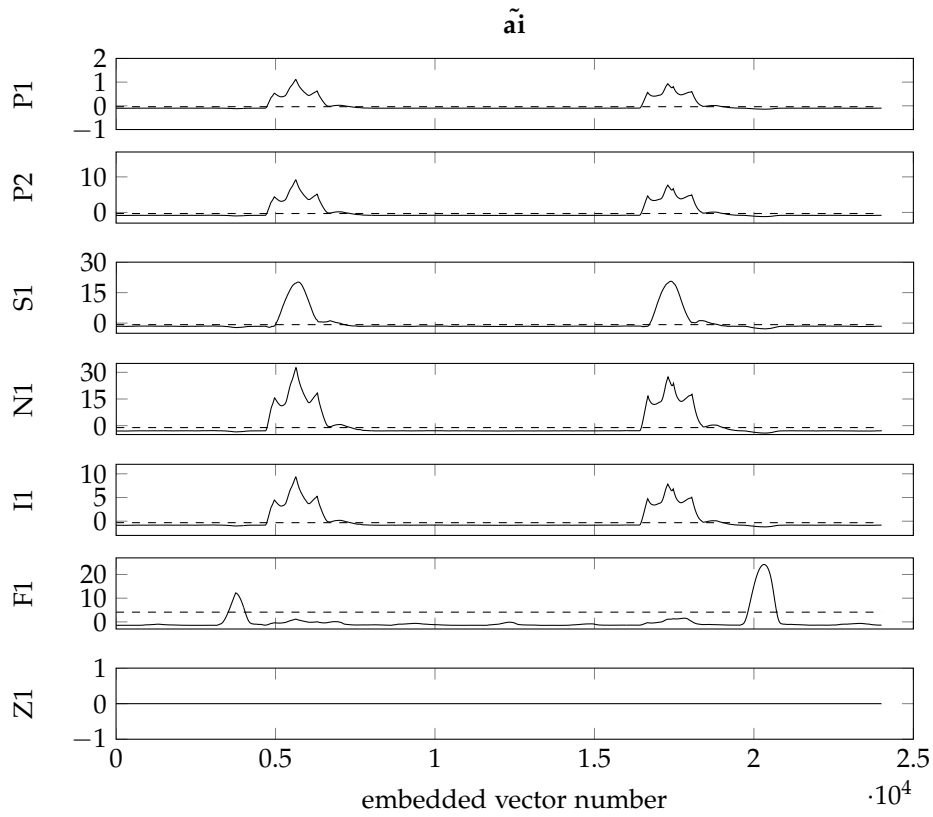


Figure 5.7: Final anomaly index vectors $\tilde{\mathbf{a}}_r$ for the development case study.

Plant-wide anomaly index vector

A global characterization of the group of variables with regards to transient disturbances can be given by a plant-wide anomaly index vector, $\tilde{\mathbf{a}}_{PW}$. This is given by the arithmetic average of the N_V final anomaly index vectors, as shown in equation (5.8).

$$\tilde{\mathbf{a}}_{PW} = \frac{1}{N_V} \begin{bmatrix} \sum_{r=1}^{N_V} \tilde{a}_{r,1} & \sum_{r=1}^{N_V} \tilde{a}_{r,2} & \cdots & \sum_{r=1}^{N_V} \tilde{a}_{r,N_E} \end{bmatrix} \quad (5.8)$$

Figure 5.8 shows the plant-wide anomaly index vector for the development case study. It clearly captures the occurrence of the two groups of transient disturbances. The two protusions of larger magnitude correspond to the transients which are present in most measurements, while the two protusions of smaller magnitude correspond to the transients which are present only in measurement F1, and hence are less important plant-wide.

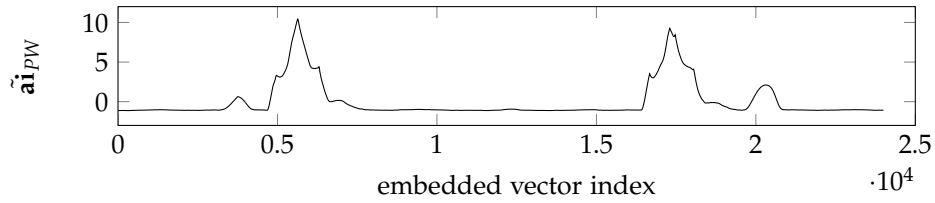


Figure 5.8: Plant-wide anomaly index vector for the development case study.

5.2.4 Significance level

A transient disturbance is detected when embedded vectors generated from the time series of a measurement are classified as anomalous. Classifying an embedded vector as anomalous happens when its anomaly index \tilde{a}_i is greater than a threshold.

The threshold used is indicated in the right-hand side of equation (5.9). It is the same as proposed in section 4.2.4, but operating on the probability distribution of $\tilde{\mathbf{a}}_i$ instead of \mathbf{a}_i .

$$\tilde{a}_i > Q_2(\tilde{\mathbf{a}}_i) + 6 \times \text{IQR}(\tilde{\mathbf{a}}_i), \quad (5.9)$$

The threshold is based on the median and interquartile range, IQR, of $\tilde{\mathbf{a}}_i$. The median is defined as the 50th percentile (Q_2), and IQR as the difference between the 75th and 25th percentiles. As discussed in section 4.2.4, the median and IQR are robust statistics because they are less influenced by outlier values than classical statistics [Huber and Ronchetti, 2009].

Appendix B.1 analyses the behaviour of the threshold under the null hypothesis of a time series with no anomalies. The analysis uses three cases representative of time series with no anomalies: (i) steady state operation with only random noise, (ii) operation with non-random variability, and (iii) oscillatory operation. The conclusion is that the significance level of this detection threshold depends on parameter β . With $\beta > 0.15$, the probability of false positives is less than one in a thousand.

5.3 Parameter settings and sensitivity

Multivariate detection of transient disturbances involves the selection of (i) the number of samples to characterize a transient, (ii) the parameters of the

Table 5.1: Parameters involved in the multivariate detection method, with recommended values.

Method	Parameters	Recommended value
	N_t , number of samples in a transient	$N_t \geq 40$
NN	τ , embedding granularity	1
	m , embedding dimension	$0.5 N_t$
	δ , embedding step	$\delta \leq 5$
	k , number of nearest neighbours	$3 \leq k \leq n/10$
SVD	α	$0.3/N_V$
	β	0.2

nearest neighbours method (NN in Table 5.1), and (iii) the parameters for the SVD selection steps. Table 5.1 indicates all parameters. Symbols n and N_V were defined earlier.

The recommended values for the parameters in the nearest neighbours method are those suggested in Chapter 4. Those values were optimised to yield the best univariate detection results. The multivariate detection method proposed in the current chapter builds on univariate anomaly index vectors, and aims to improve the univariate detection results. Thus, it should start from the optimal univariate anomaly index vectors.

Parameter N_t and the parameters of the SVD are specific to the multivariate detection method. Thus, the objective of this section is to find their optimal values. Before this, the section analyses how each parameter influences the multivariate detection task.

5.3.1 Influence of parameters in the detection results

The number of samples in a transient, N_t , is inversely related to the sampling interval of a measurement. Hence, it affects the characterization of the main trends of that disturbance. The more samples, the better will be the characterization. On the other hand, a larger N_t implies, in general, an increase in computational time with a power law relationship. The reason is that N_t is directly related to the number of samples n in the time series and the nearest neighbours step is $\mathcal{O}(n^2)$.

Parameter α is the adjustable parameter in inequality (5.4) and determines which basis functions \mathbf{v}_j^\top are retained. Thus, α influences the detection results

for all variables in the set. Parameter β is the adjustable parameter in inequality (5.6) and determines which terms $u_{r,j}s_j\mathbf{v}_j^\top$ are retained. Thus, β influences the detection results of each variable individually.

5.3.2 Recommendations for parameters and analysis of sensitivity

This section determines the best values of the parameters, and analyses the sensitivity of the detection results in the range of those values. To these ends, the final anomaly index vectors $\tilde{\mathbf{a}}_r$ are generated from each measurement r in the development case study with different parameter values. Each $\tilde{\mathbf{a}}_r$ is then evaluated individually using the metrics based on false negatives and false positives defined in section 4.3.2.

In brief, metric *FN* (false negatives) assesses whether embedded vectors that cover a transient are incorrectly classified as normal. In the comparison, N_E^{trans} is the total number of embedded vectors that include a transient, while N_{TP} is the number of embedded vectors correctly classified as anomalous. The metric is *FN* defined in equation (5.10), and has a value between zero and one.

$$FN = \frac{N_{TP}}{N_E^{trans}} \quad (5.10)$$

Metric *FP* (false positives) assesses whether embedded vectors not covering a transient were incorrectly classified as anomalous. In the comparison, N_E^{norm} is the total number of embedded vectors that do not include a transient, while N_{TN} is the number of embedded vectors correctly classified as normal. The metric is *FP* defined in equation (5.11), and has a value between zero and one.

$$FP = \frac{N_{TN}}{N_E^{norm}} \quad (5.11)$$

For parameter optimisation, the correct classification of embedded vectors into normal and anomalous is defined visually beforehand. The actual classification is based on applying the threshold proposed in section 5.2.4 to the final anomaly index vector. The measurement-specific metrics are then averaged over all measurements in order to obtain global *FN* and *FP* metrics for the whole set.

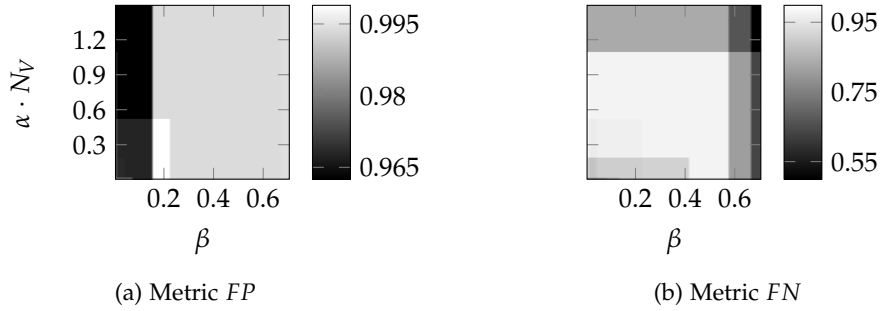


Figure 5.9: Performance of the detection method as a function of parameters α and β . N_V denotes the number of measurements in the set. Lighter tones denote better performance.

Selection of basis functions with α and terms with β

Figure 5.9 shows the influence of α and β on the detection results from the example in Figure 5.1. The performance is assessed by the global *FP* and *FN* metrics. The colours represent the magnitude of the metrics according to the scales shown in the figure. The vertical axis in the plots refers to $\alpha \cdot N_V$. The reason is that Kaiser [1960] suggests that the best value for α varies inversely with N_V .

The highest value of *FP* is attained in the range $\alpha < 0.5/N_V$ and $0.15 < \beta < 0.25$. The highest value of *FN* is attained in the range $0.2/N_V < \alpha < 1.1/N_V$ and $\beta < 0.6$. These results suggest that optimal values for the parameters are $\alpha = 0.3/N_V$ and $\beta = 0.2$.

The optimal values for the parameters are confirmed by additional analyses presented in appendix B.2. These analyses use five other sub-groups of measurements from the development case study. Each sub-group has one more measurement than the previous. The argument is that the measurements and the size of the sub-group could affect how the performance responded to α and β . The analyses show that the influence of $\alpha \cdot N_V$ and β on the performance of the detection is generally consistent between the groups of measurements. The results also justify the use of $\alpha \cdot N_V$ as vertical axis.

The effects of α and β can be observed in more detail from Figure 5.9. In particular, the effects of α are the following:

- The increase in *FN* at $\alpha \cdot N_V = 0.2$ reflects the need for a lower limit for α . Above this limit, the selection step can discard basis functions \mathbf{v}^\top which hinder the detection of transients in measurements such as P1.

- The small decrease in FP at $\alpha \cdot N_V = 0.5$ shows one of the disadvantages of increasing α beyond certain limits. This disadvantage is the removal of basis functions \mathbf{v}^\top which capture details that define the precise start and end times of the transients.
- The decrease in FN at $\alpha \cdot N_V = 1.1$ shows the other disadvantage of increasing α beyond certain limits. This effect in performance is greater than in the previous paragraph. In this case, the disadvantage is the removal of basis functions \mathbf{v}^\top which capture transients that are present in few measurements, such as the transients in measurement F1.

The effects of β are the following:

- The increase in FP at $\beta = 0.15$ reflects the need for a lower limit for β . Above this limit, the selection step can discard from the anomaly index vector of transient-free measurements, such as Z1, the terms $u_{r,j}s_j\mathbf{v}_j^\top$ which are associated with transients. This lower limit for β was expected from the discussion in section 5.2.4, and guarantee a probability of false positives of around one in a thousand.
- The small decrease in FP at $\beta = 0.5$ shows one of the disadvantages of increasing β beyond certain limits. Similarly to α , this disadvantage is the removal of terms $u_{r,j}s_j\mathbf{v}_j^\top$ which capture details that define the precise start and end times of the transients.
- The decrease in FN at $\beta = 0.6$ shows the other disadvantage of increasing β beyond certain limits. For measurements with masked transients, such as P1, transient-related terms $u_{r,j}s_j\mathbf{v}_j^\top$ capture a limited fraction of variance of the univariate anomaly index vectors. The value of β should be limited in order to retain such terms.

Number of samples in a transient, N_t

The number of samples in a transient was varied by downsampling the time series. The detection based on the downsampled time series used $\tau = 1$, $\delta = 1$, and $k = 5$ as NN parameters. Parameter m was reduced with the downsampling according to $m = 0.5 \cdot N_t + 1$. Parameters α and β were set to 0.05 and 0.2, respectively, following the discussion in the previous subsection.

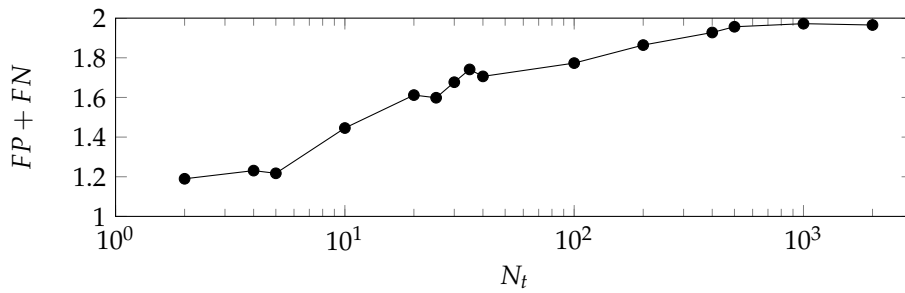


Figure 5.10: Performance of the detection method as a function of the number of samples in a transient, N_t . The performance index can vary between zero and two.

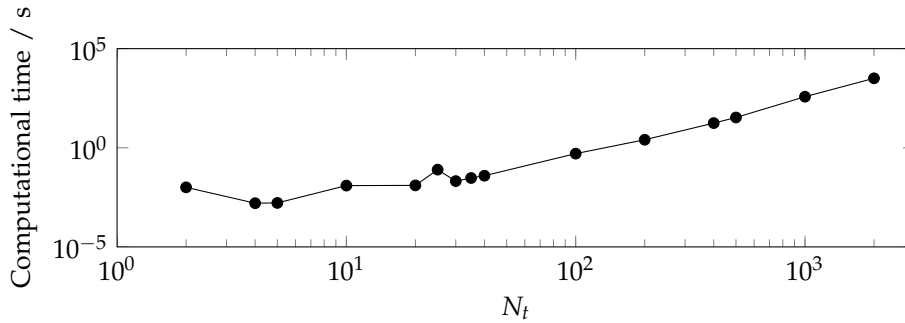


Figure 5.11: Computational time for calculating one univariate anomaly index vector, shown as a function of the number of samples in a transient, N_t . The computational time is an average over the seven measurements in the set.

Figure 5.10 shows the influence of N_t on a performance index consisting of the sum of metrics FP and FN . As expected, the performance index increases when more samples characterize the transient.

Two conclusions can be drawn from the figure. One is that reducing N_t from the original 2000 to 200 samples causes a limited reduction in performance. The plot shows that this reduction is around 5%. This is important because the computational time of the algorithm grows as a power of N_t , as shown in Figure 5.11. Thus, if the sampling rate is high, and the computational time prohibitive, the time series can be downsampled until attaining 200 samples in the transient, with a limited loss in performance. A second observation concerns measurements with slow sampling rate. The figure shows that even if there are only $N_t = 40$ samples in the transient, the performance is still around 85% of its maximum.

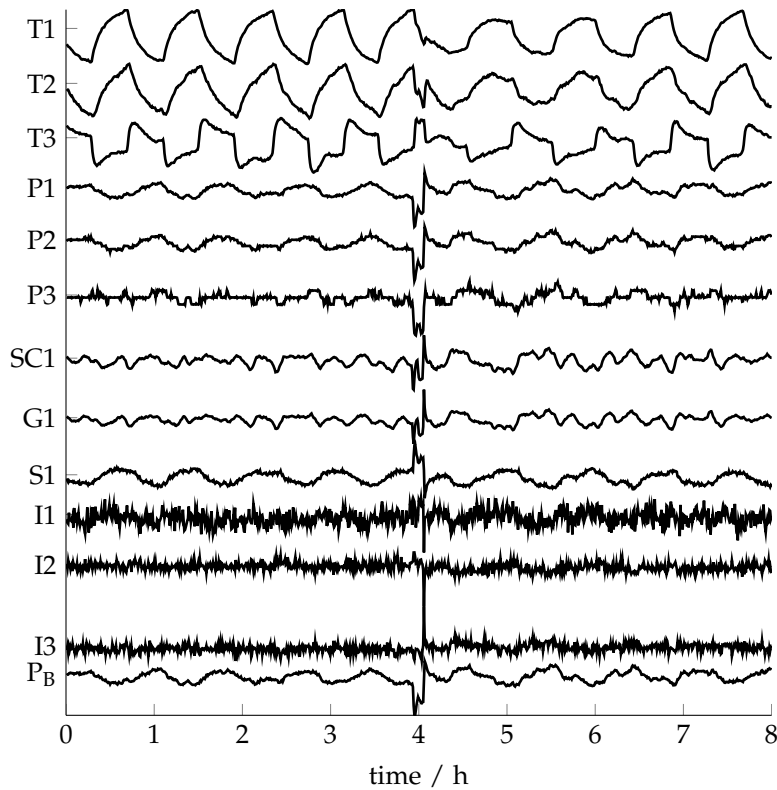


Figure 5.12: Time series of the measurements in the test case study *Gas plant case 2*.

5.4 Application to test case study

The *Gas plant case 2* is the case study used to test the multivariate detection method. The case study derives from routine operation of an industrial gas processing plant, as explained in section 2.5. Figure 5.12 shows the time series of 13 measurements from a gas expansion-recompression section of the process. These include gas temperatures (T1, T2, and T3), and pressures (P1, P2, and P3), control signals (SC1), valve position (G1), turbine speed (S1), drive currents (I1, I2, and I3), and pressure on the bearings of the turbine (P_B).

The transient affecting these measurements occurs around hour 4 in the figure. The transient manifests itself in different ways in each measurement. It must be noted that the underlying trends of most measurements are not constant. Several measurements show oscillations, namely T1, T2, T3, P1, P2, S1, and P_B, and the measurements of current I1, I2 and I3 are noisy.

The expectation for this case study is that the method combining nearest neighbours and SVD correctly detects the transient disturbance in all measurements. The multivariate step with SVD is expected to increase the robustness of the method to the other behaviour in the trends, that is, the oscillations and the noise. All measurements are sampled synchronously at 30 s intervals.

The time scale of transient disturbance is approximately 15 min, which is common for process disturbances. The sampling interval of 30 s means these transients are described by 30 samples. The recommended value is 40, however this is not possible in this case study. Therefore the nearest feasible value is used. Other recommended parameter values for this case are $\tau = 1$, $m = 20$, $\delta = 1$ and $k = 3$, for the nearest neighbours step. For the multivariate step, the parameters are chosen according to section 5.3.2 as $\alpha = 0.03$ and $\beta = 0.2$.

Figure 5.13 shows, in black lines, the final anomaly index vectors $\tilde{\mathbf{a}}_i$ for each measurement. These can be compared to the anomaly index vectors \mathbf{a}_i , obtained in the univariate analysis, which are represented in Figure 5.13 with grey lines. The comparison shows that the multivariate analysis is able to retrieve the feature between embedded vector numbers 400 and 500 which is common to all the anomaly index vectors \mathbf{a}_i . The features of the individual \mathbf{a}_i which have been discarded include, most notably, the smaller protusions in the \mathbf{a}_i of measurements T1 and T3, and the noisy trend in the \mathbf{a}_i of measurement I1. By discarding such features, the anomaly indices of embedded vectors covering the transient become more distinct.

The outcome of the analysis is that a transient disturbance was detected starting at 3.9 h and ending at 4.15 h.

5.5 Chapter summary and discussion

This chapter has presented a contribution to the detection of transient disturbances in a multivariate data set comprising process, electrical, and mechanical measurements. The new method extends the univariate transient detection method in Chapter 4 to a multivariate method. The results in this chapter have demonstrated that the univariate method can fail in measurements with strong oscillatory trends or noise. In contrast, the multivariate method correctly detects the transients in such measurements because it exploits the presence of the same disturbance in other measurements.

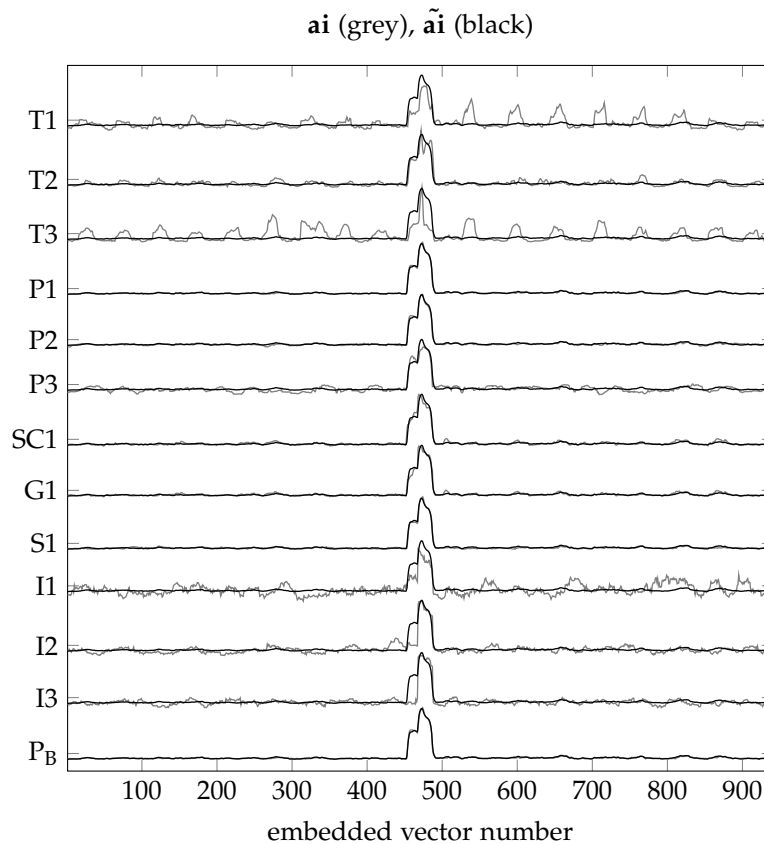


Figure 5.13: Anomaly index vectors for the measurements in the test case study *Gas plant case 2*. Grey lines are the univariate anomaly index vectors \mathbf{ai} . Black lines are the final anomaly index vectors $\tilde{\mathbf{ai}}$ reconstructed after the multivariate step.

The multivariate analysis was implemented as a singular value decomposition of a multivariate set of features known as anomaly index vectors. The values in these vectors are associated with segments of a measurement. Segments exhibiting a transient disturbance lead to high anomaly index values. The multivariate method requires a detection threshold and parameters, thus section 5.2.4 and section 5.3.2 presented recommendations for both and analysed these recommendations.

This chapter demonstrated that the multivariate method is effective in detecting transients in measurements where the univariate method is limited.

Chapter 6

Multivariate detection of transient disturbances in multi-rate systems

Chapter 1 of this thesis highlighted the importance of extending the analysis of process disturbances to include the measurements from the electrical and mechanical equipment which service the process. Examples of these measurements are voltage levels in the motor power supply, and motor speed. An example of a process measurement is flow rate.

Transient disturbances are particularly relevant when considering electrical and mechanical measurements, and thus Chapter 5 proposed a method to detect transient disturbances in a multivariate context. The method is based on an advanced signal analysis technique known as nearest neighbours, and on multivariate statistics implemented as a Singular Value Decomposition (SVD). The method requires that all measurements of the multivariate set be available with the same sampling rate. However, process, electrical, and mechanical measurements in industry are commonly available with different sampling rates. Specifically, the sampling rate for process measurements is typically 100 to 1000 times slower than that for electromechanical measurements.

The reason why the multivariate detection method in Chapter 5 is not applicable to multi-rate systems is that SVD needs to be done on the anomaly index matrix \mathbf{A} . Each row in matrix \mathbf{A} is the anomaly index vector \mathbf{ai}_r of one measurement, and each element in \mathbf{ai}_r is the anomaly index for one embedded vector of that measurement. For the matrix implementation, the number of embedded vectors for each measurement must be the same. However,

the slow-sampled measurements cannot have as many embedded vectors as fast-sampled measurements because of the longer intervals between samples. Therefore, the method of Chapter 5 can include electromechanical measurements only if these are downsampled to the process rate. Downsampling would lead to loss of information and, potentially, incorrect detection results.

The aim of this chapter is to extend the multivariate method of Chapter 5 to the case of a multi-rate system. This is defined as a discrete-time system with two or more operating sampling rates [Ding and Chen, 2005]. Other authors have addressed multi-rate systems, however the majority focused on problems other than the detection of disturbances [Misra et al., 2002, Li et al., 2003].

The multi-rate detection method is based on expanding the anomaly index vectors of slow-sampled measurements to match those of fast-sampled measurements. The method will be illustrated with experimental data from the *Compressor rig* data set, and with industrial data from the *Turbocharger* data set. The chapter will demonstrate that the multi-rate method successfully incorporates measurements with different sampling rates, and improves the detection results by comparison to a uni-rate method, for which the fast-sampled measurements had to be downsampled.

6.1 Background

6.1.1 Multi-rate systems

Multi-rate systems have been identified as a challenge in areas related to process disturbance analysis. Review of the literature shows that a significant part of research has been in the fields of control [Li et al., 2003] and soft sensors [Prasad et al., 2002], and the aim is usually to estimate the outputs of the system at the faster sampling rate. Significant contributions have addressed the development of state space models at the faster sampling rate [Li et al., 2003], and parameter and state estimation [Ding and Chen, 2005]. Multi-rate data has also been used to calibrate models by applying Bayesian methods [Shao et al., 2011]. These techniques have been mostly applied on petroleum, chemical and biological systems, in which the measurements with lower sampling rate derive from quality variables which require laboratory analysis.

Multi-rate systems have also been addressed in the field of process monitoring. In this case, the objective was to develop data-based models which describe process data during normal operation, so that new incoming measurements can be compared to these models. To that end, Bakshi [1998] and Misra et al. [2002] combined wavelet analysis and principal component analysis. The authors suggested that these methods could handle multi-rate systems because the integration of the multi-rate variables is done over different time scales, into which the measurements are previously decomposed by the wavelet analysis.

As justified in section 3.6, this thesis does not use data-based models and uses instead advanced signal analysis methods. With these methods, the analysis of process disturbances in multi-rate systems remains an open question.

6.2 Method development

This section explains the method of multivariate detection of transient disturbances in multi-rate systems. The objective of this method is to be able to use the advantages of the multivariate approach when some measurements are sampled at a slower rate than others, without having to downsample the fast-sampled measurements. The development case study is first presented and then used to illustrate the explanation. Section 6.2.2 shows the formulation of the method, and section 6.2.3 demonstrates that the multi-rate method improves the detection of transients in a multi-rate data set compared to the alternative approach of downsampling to the lower sampling rate.

6.2.1 Development case study

Figure 6.1 presents the time series of three measurements selected from the case study *Compressor rig case 2*. Measurements P1, N1, and I1 were selected to facilitate the visualization of the details of the multi-rate method.

The data set of this case study was not multi-rate because all the measurements including the process measurements were available at 1 kHz. However, in practice the sampling rate for process measurements is slower than for electrical and mechanical measurements. Therefore the development data set has to be manipulated to construct a realistic multi-rate data set. This is done by resampling the pressure measurement P1 with a sampling interval of 0.5 s,

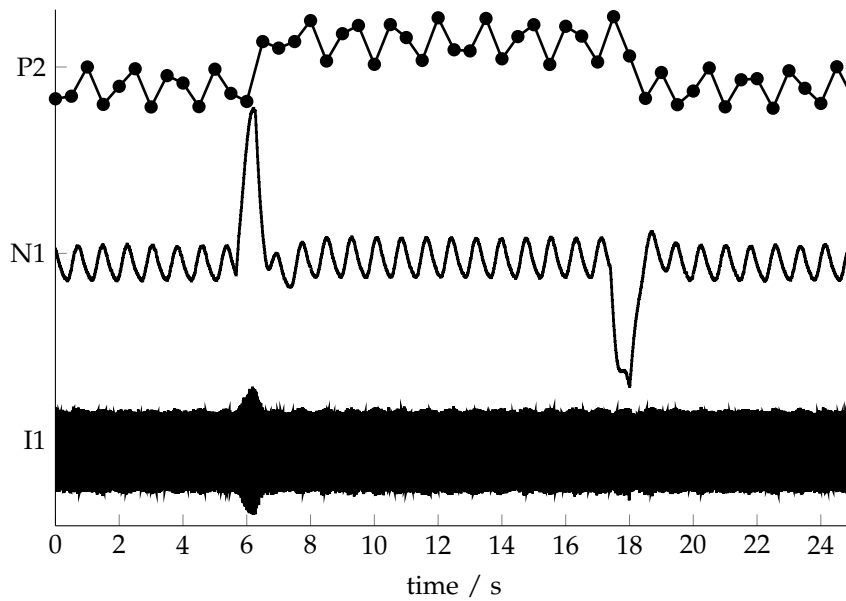


Figure 6.1: Measurements from the reference example used to demonstrate the multi-rate detection method. Measurement P1 has been resampled with a sampling interval of 0.5 s to create a multi-rate data set.

while leaving the torque and current measurements, N1 and I1, with a sampling interval of 1 ms.

The interesting events in this case study are the two transient disturbances occurring around 6 and 18 s. The objective of the proposed method is to detect those transients. Additionally, the three measurements are affected by oscillations, which are due to unstable operation of the compressor close to the surge limit.

6.2.2 Algorithm

To handle multi-rate systems, the core modification to the method in Chapter 5 is in the way the anomaly index vectors \mathbf{ai} are constructed. The anomaly index vectors \mathbf{ai}_r of measurement r is the sequence of anomaly indices $ai_{r,i}$ for each embedded vector \mathbf{x}_i of that measurement.

Embedding dimensions m^f and m^s

As with the uni-rate method, embedded vectors are defined as segments of m samples from a measurement, which are ordered in time and τ instants apart. However, in the multi-rate method the number of samples is different for

fast-sampled and slow-sampled measurements. The requirement is that the embedded vectors of both fast- and slow-sampled measurements must cover the same period of time. The period of time covered by an embedded vector is $(m - 1)\tau\Delta t$, where Δt is the sampling interval. The embedding dimension of the fast-sampled measurement is denoted as m^f , while the embedding dimension of the slow-sampled measurement is denoted as m^s . The relation between m^f and m^s must follow equation (6.1), where Δt^f and Δt^s denote the sampling intervals of fast and slow-sampled measurements, respectively.

$$\frac{(m^f - 1)\tau^f}{(m^s - 1)\tau^s} = \frac{\Delta t^s}{\Delta t^f} \quad (6.1)$$

The optimum period of time covered by an embedded vector follows the recommendations in Table 5.1 in section 5.3.

Numbers of embedded vectors N_E^f and N_E^s

The number N_E of embedded vectors generated from a measurement is given by relation (6.2), where n indicates the number of samples in the measurement. The incomplete brackets indicate a floor function, which maps a real number to the largest previous integer.

$$1 + (N_E - 1)\delta + (m - 1)\tau \leq n \quad \Leftrightarrow \quad N_E = \left\lfloor \frac{n - (m - 1)\tau - 1}{\delta} + 1 \right\rfloor \quad (6.2)$$

In a fast-sampled measurement, the number of samples is denoted as n^f , while the number of samples in the slow-sampled measurements is denoted as n^s . If all measurements cover the same period of time, then n^f and n^s are related by equation (6.3).

$$\frac{n^f - 1}{n^s - 1} = \frac{\Delta t^s}{\Delta t^f} \quad (6.3)$$

This equation and equation 6.1 imply the relationship

$$\frac{(N_E^f - 1)\delta^f}{(N_E^s - 1)\delta^s} = \frac{\Delta t^s}{\Delta t^f} \quad (6.4)$$

where N_E^f is the number of embedded vectors in the fast-sampled measurements, and N_E^s is the number of embedded vectors in the slow-sampled measurements.

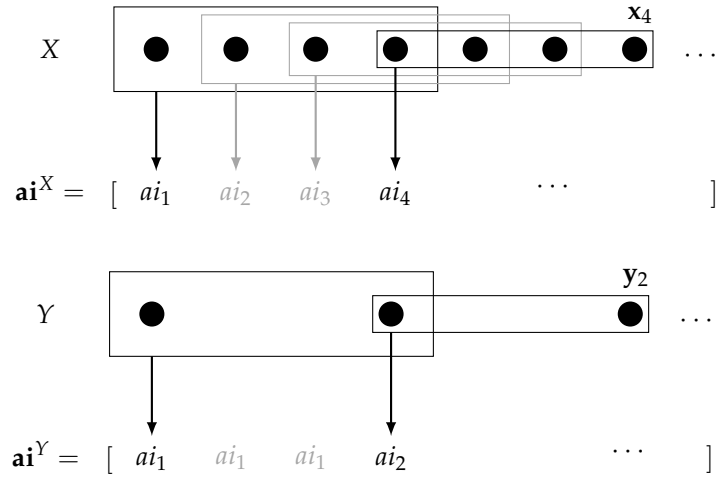


Figure 6.2: Construction of the anomaly index vectors \mathbf{ai} for a fast-, X , and a slow-sampled measurement, Y . The black dots represent samples in the time series of the measurements. The rectangles show the samples included in the embedded vectors \mathbf{x}_i and \mathbf{y}_i . ai are the corresponding anomaly indices. The sampling intervals are related as $\Delta t^s / \Delta t^f = 3$, and the embedding dimensions are $m^f = 4$ and $m^s = 2$.

Parameters δ^f and δ^s in equation (6.4) are the embedding steps for fast and slow-sampled measurements, respectively. Their recommended values are the same as for δ , shown in Table 5.1.

Expanded anomaly index vector for slow-sampled measurements

An anomaly index ai_i is computed from each embedded vector \mathbf{x}_i as in the uni-rate method of Chapter 4, that is, as the k^{th} smallest distance between embedded vector \mathbf{x}_i and all other embedded vectors. For the fast-sampled measurement, such anomaly indices are assembled in the anomaly index vector \mathbf{ai} sequentially, as in Chapter 4 and Chapter 5. This is illustrated in the top part of Figure 6.2. The black dots represent the initial samples of the fast-sampled measurement X , and each rectangle indicates the samples included in an embedded vector, for which $m^f = 4$ in the illustration. The anomaly index of embedded vector \mathbf{x}_i is, thus, assigned to position i in the anomaly index vector.

The construction of the anomaly index vector for a slow-sampled measurement differs from the uni-rate method. The bottom part of Figure 6.2 illustrates this new procedure, where Y is the slow-sampled measurement and its

embedded vectors have $m^s = 2$ samples. The anomaly index of embedded vector \mathbf{y}_i is now assigned to position $1 + (i - 1)\Delta t^s / \Delta t^f$. This means that the anomaly indices of embedded vectors such as \mathbf{x}_4 and \mathbf{y}_2 in Figure 6.2 are assigned to the same position in the anomaly index vectors. Embedded vectors \mathbf{x}_4 and \mathbf{y}_2 cover the same period of time in X and Y .

The empty positions in the anomaly index vector of the slow-sampled measurement are populated with the previous anomaly index, as illustrated in the bottom part of Figure 6.2 by the grey-colored anomaly indices. This is equivalent to assuming that the trend of the measurement stayed equally anomalous during those periods.

If measurements X and Y originate from different sources, it may happen that their sampling instants are not synchronized, that is, each sample from the slow-sampled measurement falls in between two samples of the fast sampled measurement. In such cases, the fast-sampled measurement X can be resampled by interpolation. Lehmann et al. [1999] discusses methods for interpolation.

Figure 6.3 shows the anomaly index vectors \mathbf{a}_i for the development case study. These are univariate anomaly index vectors, because they were built for each measurement separately. The first anomaly index vector corresponds to the slow-sampled measurement, and the step-wise shape is the result of expanding the anomaly indices as proposed.

Multivariate analysis

The anomaly index vectors constructed as above have the same number of elements, N_E^f . Therefore, it is now possible to build the anomaly index matrix \mathbf{A} as in Chapter 5. The subsequent multivariate analysis and threshold-based detection follows as in section 5.2. In brief, the anomaly index matrix \mathbf{A} is decomposed with SVD, and two thresholds, α and β , are used to eliminate terms of the decomposition. The final anomaly index vectors of each measurement are denoted as $\tilde{\mathbf{a}}_i$.

Figure 6.4 shows the final anomaly index vectors $\tilde{\mathbf{a}}_i$ for the development case study. The black lines result from the multi-rate method, and the black dashed lines indicate the detection thresholds. In order to demonstrate the improvement obtained with the multi-rate method, Figure 6.4 also shows, with grey lines, the results with the uni-rate method of Chapter 5. To use the uni-

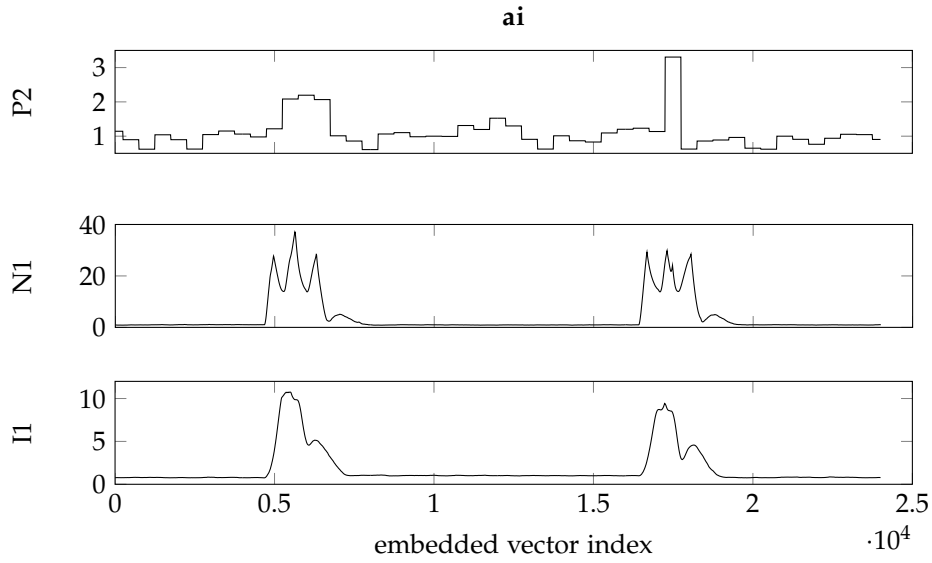


Figure 6.3: Univariate anomaly index vectors for multi-rate measurements of the development case study. The first \mathbf{ai} is an expanded anomaly index vector, for a slow-sampled measurement.

rate method, the fast-sampled measurements had to be downsampled to the slow rate. As desired, the multi-rate method correctly identifies the transient disturbances in all measurements. On the other hand, the uni-rate method at slow rate misses the transients in the measurement I1, and estimates shorter durations for the transients in the other measurements, specially P2.

6.2.3 Comparison of the performance of the multi-rate and uni-rate methods

The purpose of this section is to quantify the improvement obtained when using the multi-rate method, in comparison to using the uni-rate method with measurements downsampled to the lower rate. This is done using the measurements from the development case study.

The sampling interval Δt^s of the slow-sampled measurements is the limiting factor to the performance of both methods, since it affects the characterization of the trends of the slow-sampled measurements. Therefore, performance is assessed for different values of Δt^s .

In the multi-rate method, measurement P2 has the sampling interval Δt^s . Measurements N1 and I1 have a smaller sampling interval, Δt^f . The Δt^f used

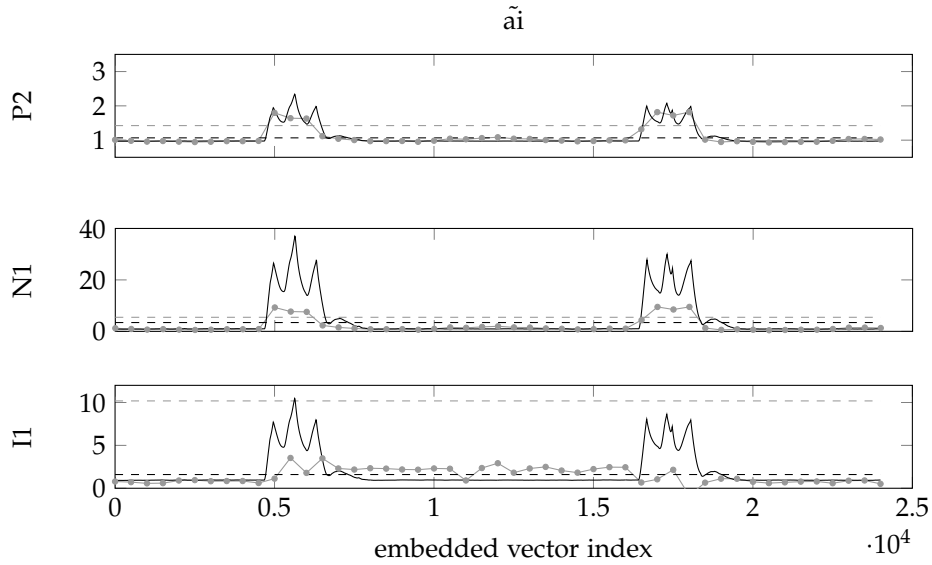


Figure 6.4: Final anomaly index vectors for the reference example. Black lines refer to the multi-rate method. Grey lines refer to the uni-rate method with the fast-sampled measurements downsampled to the slow rate. Dashed lines represent the corresponding detection thresholds.

is that leading to $N_t = 40$ samples in the transients, which is the minimum recommended number of sample as shown in section 5.3. The improvement obtained with the multi-rate method as determined in this section is, therefore, associated to $N_t = 40$. Any number of samples N_t greater than 40 should lead to improvements in performance equal to, or above, the improvements determined in this section.

Performance is measured by applying the metrics FP and FN presented in section 4.3.2 to the detection results of each measurement. The metrics are then averaged over the three measurements. Equation (6.5) recalls the definition of these metrics.

$$FN = \frac{N_{TP}}{N_E^{trans}} \quad (6.5a)$$

$$FP = \frac{N_{TN}}{N_E^{norm}} \quad (6.5b)$$

Metric FN (false negatives) assesses whether embedded vectors that cover a transient are incorrectly classified as normal. It compares the total number N_E^{trans} of embedded vectors that include a transient and the number N_{TP} of

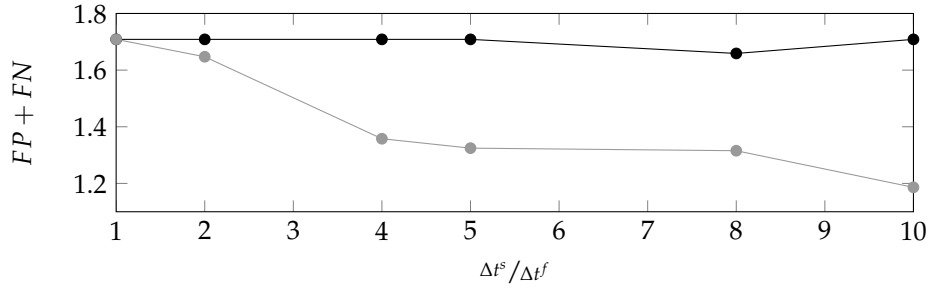


Figure 6.5: Performance of the multi-rate (black line) and uni-rate methods (grey line) as a function of the sampling interval of the slow-sampled measurements Δt^s . The sampling interval of the fast-sampled measurements Δt^f is fixed. The performance index can vary between zero and two.

embedded vectors correctly classified as anomalous. Metric FN can vary between zero and one. Metric FP (false positives) assesses whether embedded vectors not covering a transient were incorrectly classified as anomalous. It compares the total number N_E^{norm} of embedded vectors that do not include a transient and the number N_{TN} of embedded vectors correctly classified as normal. Metric FP can vary between zero and one.

Figure 6.5 shows the influence of the sampling interval Δt^s on a performance index consisting of the sum of the global FP and FN metrics. The black line corresponds to the multi-rate method and the grey line to the uni-rate method. It is clear that the multi-rate method always performs better. At $\Delta t^s / \Delta t^f = 1$, the methods are equivalent. As expected, the relative improvement achieved by the multi-rate method increases as the sampling interval of the slow-sampled measurements increases.

Nonetheless, it should be noted that the time to compute the univariate anomaly index vectors \mathbf{a}_i is different for the two methods. Specifically, each fast-sampled measurement requires approximately $(\Delta t^s / \Delta t^f)^2$ times longer to process than the slow-sampled measurements.

6.3 Application to test case study

The *Turbocharger case* is the case study used to test the multivariate multi-rate detection method. The case study derives from performance tests to a commercial turbocharger compressor, as explained in section 2.5. The compressor is driven by an asynchronous motor, and is controlled by an a.c. measurement

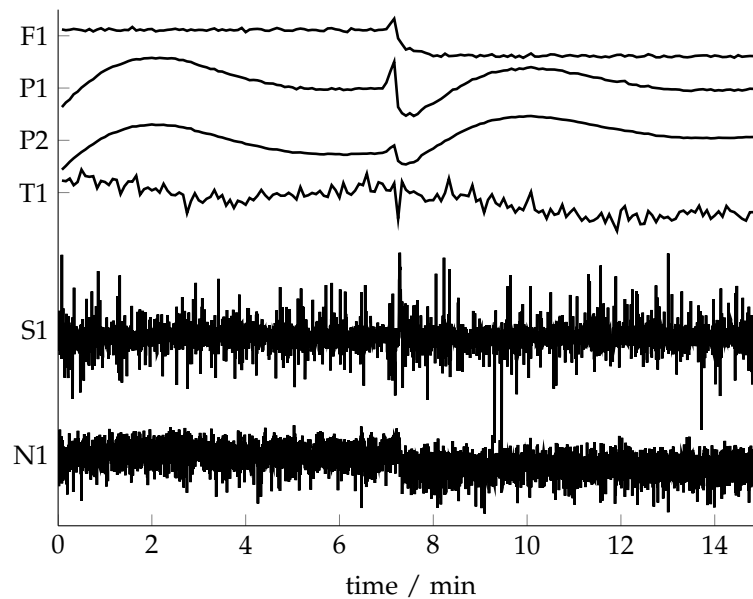


Figure 6.6: Time series of the measurements in the test case study *Turbocharger case*.

speed drive. The tests subjected the compressor to varying mass flow rates, at different rotational speeds. Figure 6.6 shows the time series from six measurements monitored during these tests. The process measurements of flow (F1), inlet pressure (P1), outlet pressure (P2) and inlet temperature (T1) were collected by a process control module, and are given with a sampling interval of 5 s. The speed (S1) and torque (N1) of the motor shaft are measured by a monitoring module in the drive, and are given with a sampling interval of 0.1 s.

All measurements are affected by a sharp spike happening around minute 7. The transient is not very distinct in measurements N1 and T1, and the trends of most measurements are not constant. Specifically, measurements P1 and P2 show slow oscillations, and measurements T1, S1 and N1 are noisy. The expectation for this case study is to detect the transient disturbance in all measurements. As in Chapter 5, the multivariate approach is used to improve the detection results in measurements in which the transient is less clear. In contrast to Chapter 5, this is a multi-rate system so the multivariate multi-rate method is used.

The time scale of the transient disturbance is approximately 35 to 40 s. The transient in the measurements collected by the drive is characterized by

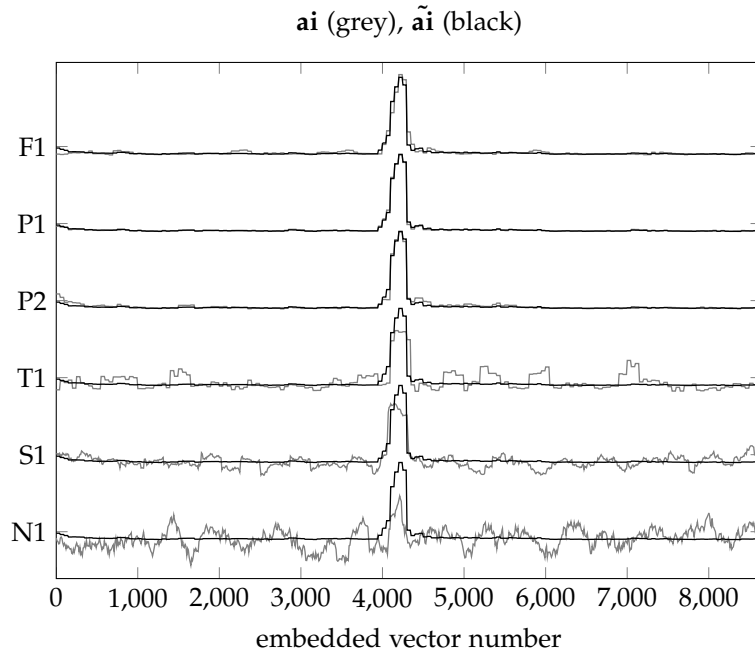


Figure 6.7: Anomaly index vectors for the measurements in the compressor tests case study. Grey lines are the univariate anomaly index vectors \mathbf{a}_i . Black lines are the final anomaly index vectors $\tilde{\mathbf{a}}_i$ after the multivariate step.

400 samples. This is above 40 samples, as recommended. The transient in the process measurements is characterized by 8 samples. As shown in section 6.2.3, the performance of the multi-rate method is similar whether the slow-sampled measurements have 40 samples or less.

Considering the sampling intervals, the recommended values for the embedding dimensions are $m^f = 250$, for the drive measurements, and $m^s = 5$, for the process measurements. The other parameters are $\tau = 1$, $\delta = 1$, $k = 3$, $\alpha = 0.05$ and $\beta = 0.15$.

Figure 6.7 shows, in black lines, the reconstructed anomaly index vectors for each measurement. The grey lines represent the univariate anomaly index vectors. Those from process measurements are expanded as described in section 6.2.2. This case study shows that the multi-rate method enables the multivariate analysis and, as desired, it retrieves the common features of the anomaly index vectors. The anomaly indices of embedded vectors covering the transient become more distinct particularly for measurements T1, S1 and N1, leading to the correct detection results.

The outcome of the analysis is that a transient disturbance was detected starting at 7.1 min and ending at 7.5 min.

6.4 Chapter summary and discussion

This chapter has developed a new method for the detection of transient disturbances in a multivariate data set comprising process, electrical, and mechanical measurements. This method extends the multivariate detection method of Chapter 5 to the case of a multi-rate system. This is justified because, in industry, the measurements from process, electrical, and mechanical systems are commonly available with different sampling rates. In a multi-rate system, the uni-rate method would require the fast-sampled measurements to be downsampled to line up with the slow-sampled measurements. This chapter has demonstrated that the multi-rate method achieves improved detection results in comparison to that scenario.

The multi-rate method proposed in section 6.2.2 is based on expanding the anomaly index vectors of slow-sampled measurements to match those of fast-sampled measurements. Section 6.2.3 quantified the improvement achieved by this method in a multi-rate system, in comparison to using the uni-rate method with the fast-sampled measurements downsampled to the lower rate. Even when the slow sampling rate is only four times lower than the fast sampling rate, the multi-rate method improves on the uni-rate method by approximately 30%.

The multi-rate detection method is a new contribution to the field of process monitoring and diagnosis, and its concepts have application beyond the detection of transient disturbances. The reason is the formulation of embedded vectors which imposes the same time span for the embedded vectors of all measurements, instead of the same number m of samples. This formulation can be used in any analyses which construct embedded vectors from measurements with different sampling rates, or from a measurement whose sampling rate is not regular. Section 9.1 and section 9.2 will discuss future research directions which use the time-based formulation to develop other disturbance analysis methods for measurements with different sampling rates and with irregular sampling rates.

Chapter 7

Removal of transient disturbances from oscillating measurements

The previous chapters proposed methods to detect the presence of transient disturbances. The main motivation to detect those disturbances is that they disrupt the operation of process, mechanical and electrical systems. Another motivation is that transient disturbances can interfere with other methods of signal analysis. Specifically, the book on valve stiction by Jelali and Scali [2010] gives examples of methods to analyse oscillatory disturbances whose reliability is affected by transient disturbances. The reason is that the transients modify time and frequency properties of the time series on which the methods rely.

This chapter proposes a method to remove transient disturbances from the time series of a measurement. The segment replacing the removed transient agrees with the underlying dynamics of the original time series. In other words, the replacing segment is an estimate of what the measurement would have been had the transient not been present. This method is part of the task of data treatment described in section 3.1.1.

The method is based on a nearest neighbours imputation technique. It is applied to the time series of one measurement at a time. As a result, the challenge of measurements with different sampling rates does not arise.

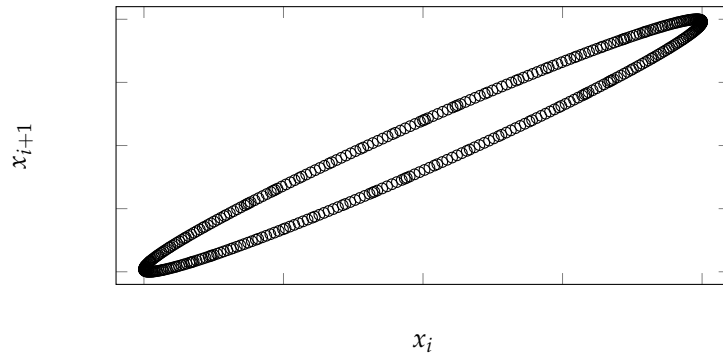


Figure 7.1: State space of dimension $M = 2$ for a sinusoidal time series.

7.1 Background

This section introduces the concept of non-linear deterministic systems, and explains how nearest neighbours techniques can determine unknown segments in time series generated from those systems.

7.1.1 Non-linear deterministic systems

A deterministic system is defined by the fact that its present state unambiguously determines its future state. Mathematically, such systems are described by ordinary differential equations, or their equivalent discrete-time maps,

$$\mathbf{x}_{i+1} = \mathbf{F}(\mathbf{x}_i), \quad i \in \mathbb{Z}. \quad (7.1)$$

The state of a system is observed through a measurement X of that system. The state at instant i is defined by the ordered sequence of M samples of X

$$\mathbf{x}_i = [x_i \quad x_{i+1} \quad \cdots \quad x_{i+M-1}]. \quad (7.2)$$

The number M of samples which defines the state is such that the condition (7.1) of determinism is verified.

The group of all possible states of a deterministic system defines its state space. Figure 7.1 shows the state space for a system which generates sinusoidal time series. In such a system, a dimension $M = 2$ is enough to observe its determinism. This is visible in Figure 7.1 by the well defined structure in the two-dimensional plot.

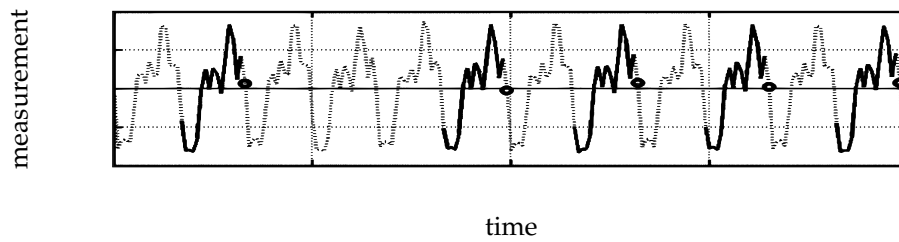


Figure 7.2: A non-linear deterministic time series with five nearest neighbour cycles followed by five similar points (adapted from Thornhill [2005]).

In non-linear systems, the mapping F is non-linear. A linear system depends on a linear combination of past values, while a non-linear system has terms such as x_{i-1}^2 or $x_i \times x_{i-1}$. Kantz and Schreiber [2003] present a comprehensive explanation of non-linear systems, and methods to analyse their time series. Time series generated by non-linear systems often show repeated states, that is, they are often oscillatory. However, the oscillations may not be purely periodic because the cycles may have different shapes and frequencies.

7.1.2 Predictability in non-linear deterministic systems

If the mapping F were known, then future states of the system could be predicted from past states. Even if this is not the case, predictability can be guaranteed if the mapping F is assumed to be continuous. If this is true, then two states which are close in the state space will lead to future states which are also close. If one of these future states were unknown, the other would be a good predictor. The best predictor is that with the most similar past state.

A simple example of predictability in deterministic systems is that resulting in purely periodic time series. Once one full cycle is observed, the rest of the time series is known to be a repetition of that same cycle. Most real time series are not purely periodic, but in non-linear deterministic systems they have some form of pattern which can be exploited. Figure 7.2 illustrates this property in a time series from a real system. The time series is not purely periodic because the cycles are not all equal. Nonetheless, five cycles with similar sequences of samples can be identified (highlighted in thick lines). Due to determinism, these cycles are followed by five similar samples (marked with circles).

7.1.3 Nearest neighbour imputation

Real systems may not be purely deterministic. Even if they are, the measurements through which they are observed are often affected by random inputs, such as instrument noise. In such cases, the prediction described before is not necessarily the most accurate. Before, only the most similar state to the past of the unknown state was used. This may not be accurate due to the existence of some uncertainty in the time evolution of the system. Therefore, another less similar past state could in theory lead to a better prediction of the unknown future state.

The method of nearest neighbour imputation [Kantz and Schreiber, 2003] searches for the k states which are the most similar to the past of the unknown state. The k most similar states are known as the k nearest neighbours. Similarity in time series data can be assessed using any of the measures discussed in Chapter 4. The predictor of the unknown state is then the average of the k states adjacent to the nearest neighbours.

Nearest neighbour imputation has been used to test the determinism [Kennel and Isabelle, 1992] and the non-linearity of a time series [Thornhill, 2005], to replace missing values in survey data [Chen and Shao, 2000], to determine the precedence relationship between two measurements [Bauer et al., 2007a], and to determine the time lag between two measurements [Stockmann et al., 2012]. In particular, Thornhill [2005], Bauer et al. [2007a] and Stockmann et al. [2012] reported applications in process systems. The measurements in these cases originated from processes under the influence of non-linear oscillatory disturbances, such as non-linear hydrodynamic instabilities and limit cycles generated by control valves with excessive static friction. This chapter focuses on measurements whose underlying dynamic trends are oscillatory, hence the motivation to use nearest neighbour imputation.

7.2 Method development

This section explains the method proposed to remove transient disturbances from the time series of a measurement. The development case study is first presented and then used to illustrate the explanation. Finally, this section discusses criteria to evaluate the adequacy of the segment which replaces the removed transient.

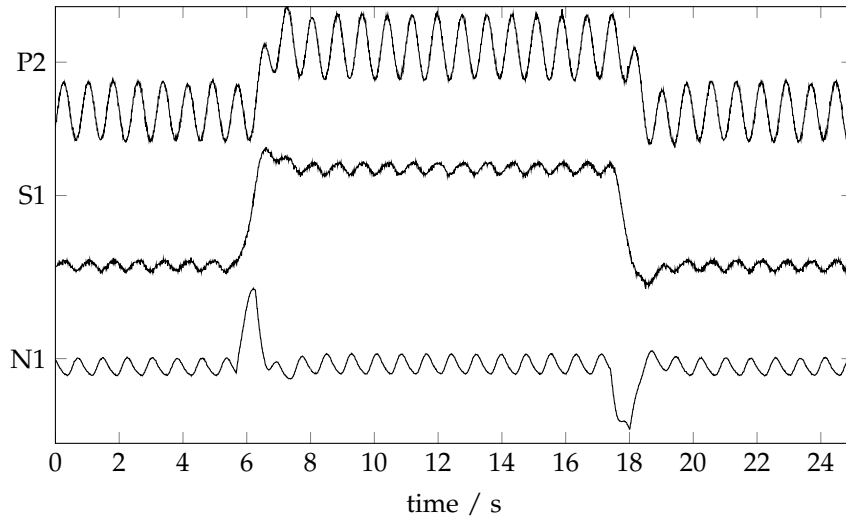


Figure 7.3: Development case study: time series of measurements from the *Compressor rig case 2* case study.

7.2.1 Development case study

Figure 7.3 shows the time trends of three measurements from the case study *Compressor rig case 2*. Section 2.5.1 explained the origin of the data and the set-up of the system.

The measurements are affected by an oscillatory disturbance which is due to compressor surge. Furthermore, the three measurements clearly show two transient disturbances, around 6 s and 18 s. Chapters 5 and 6 used this case study to develop methods to detect those transients. Here, the objective is to replace the segments affected by the transients with estimates that agree with the underlying dynamics of the original time series.

7.2.2 Algorithm

The method handles a measurement as a time series X . Therefore, the transient disturbance can be associated with the segment of the time series that starts at time instant t_i and ends at time instant t_f , that is $[x_{t_i} \cdots x_{t_f}]$. Instants t_i and t_f can be determined with the detection methods developed earlier in this thesis. Figure 7.4 shows a close-up of measurement S1. The light grey segment corresponds to the transient disturbance.

The first step of the method is to remove the segment $[x_{t_i} \cdots x_{t_f}]$ from time series X . The objective is to replace segment $[x_{t_i} \cdots x_{t_f}]$ by a different

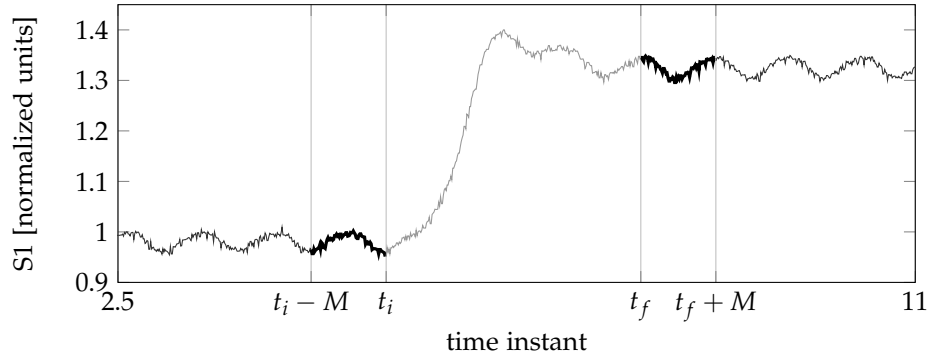


Figure 7.4: Close-up of measurement S1 highlighting the transient (grey line) and the pre- and post-transient references (black thick lines).

segment, $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$, which agrees with the overall dynamics of time series X . The replacing segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ will be computed in the next subsections using the property of predictability of deterministic systems (section 7.1).

States and embedding vectors

A state is represented by M samples. This is the dimension for which the system is deterministic. States are approximated by embedded vectors which span those M samples. An embedded vector \mathbf{x}_i is formed by m samples, τ instants apart, that is,

$$\mathbf{x}_i = \begin{bmatrix} x_i & x_{i+\tau} & \cdots & x_{i+(m-1)\tau} \end{bmatrix}. \quad (7.3)$$

The dimension M is related to m and τ according to $M = (m-1)\tau + 1$. The reason to consider τ greater than one is to minimize the computational effort of the algorithm. This is discussed in section 7.3.

The embedded vectors generated from time series X can be arranged in the rows of an embedding matrix, as shown in equation (7.4).

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{N_E} \end{bmatrix} = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_{1+\delta} & x_{1+\delta+\tau} & \cdots & x_{1+\delta+(m-1)\tau} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1+(N_E-1)\delta} & x_{1+(N_E-1)\delta+\tau} & \cdots & x_{1+(N_E-1)\delta+(m-1)\tau} \end{bmatrix} \quad (7.4)$$

Each embedded vector in the matrix lags the previous by δ samples. The number N_E of embedded vectors depends on parameters m , τ , and δ , as well as on the total number of samples n . The use of δ greater than one corresponds to omitting states of the time series. As with τ , the reason may be to minimize the computational effort of the algorithm.

Before further processing, each embedded vector in the matrix is centered around its own mean. As in previous chapters, this is done to recognize that embedded vectors have the same dynamic behaviour, even if they are centered around a different mean value. Examples of such embedded vectors are those before t_i and after t_f in Figure 7.4.

Pre- and post-transient references

The embedded vectors \mathbf{x}_{t_i-M} and \mathbf{x}_{t_f} are those immediately before and immediately after the removed segment $[x_{t_i} \cdots x_{t_f}]$. For the development case study, Figure 7.4 highlights \mathbf{x}_{t_i-M} and \mathbf{x}_{t_f} with thick black lines. Embedded vectors \mathbf{x}_{t_i-M} and \mathbf{x}_{t_f} are denoted as *pre-transient reference* and *post-transient reference*, and will be used to predict the empty samples between t_i and t_f .

If the system generating time series X is purely deterministic, then the prediction based on the pre-transient reference \mathbf{x}_{t_i-M} will be as correct as the prediction based on the post-transient reference \mathbf{x}_{t_f} . The reason is that in purely deterministic systems, the mapping $\mathbf{x}_{i+1} = \mathbf{F}(\mathbf{x}_i)$ is continuous.

However, it is unlikely that real systems are purely deterministic. Therefore, the samples $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ predicted by the pre-transient reference and by the post-transient reference may be different. The prediction which fits better to the time series is not known in advance, hence the method computes the two predictions $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ in parallel, and assesses the adequacy of each at the end. The predictions based on the pre-transient reference and on the post-transient reference have similar algorithms. Therefore, the following description only highlights the necessary differences.

Similarity

For both references \mathbf{x}_{t_i-M} and \mathbf{x}_{t_f} , the similarity between the reference and all other embedded vectors is assessed with the Euclidean distance metric. With the embedded vectors arranged in the embedding matrix, the implementation of this similarity assessment is straightforward.

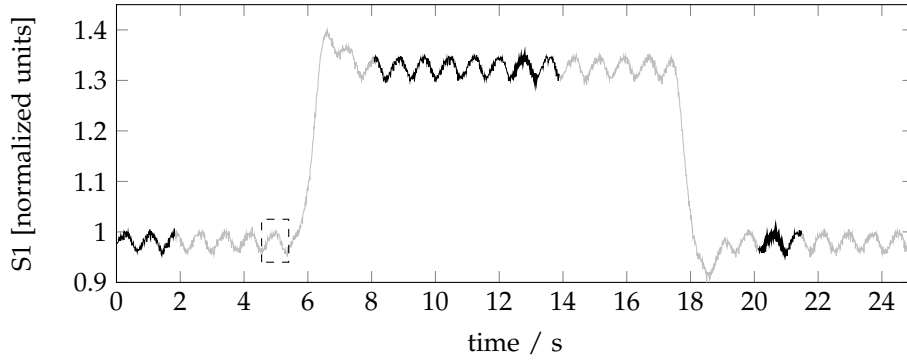


Figure 7.5: Measurement S1 highlighting the $k = 2$ nearest neighbours (thick black lines) and the embedded vectors excluded from the neighbourhood assessment (grey lines), for the pre-transient reference of the first transient.

Nearest neighbours of the pre- and post-transient references

Nearest neighbours of a reference denote the embedded vectors which have the smallest distances to that reference. The k nearest neighbours of the pre-transient reference \mathbf{x}_{t_i-M} are represented by $\mathbf{x}_{t_j}^{t_i-M}$, with $j = 1 \dots k$. The k nearest neighbours of the post-transient reference \mathbf{x}_{t_f} are represented by $\mathbf{x}_{t_j}^{t_f}$, with $j = 1 \dots k$.

Figure 7.5 shows the complete measurement S1, and highlights the $k = 2$ nearest neighbours of the pre-transient reference \mathbf{x}_{t_i-M} for the first transient. The pre-transient reference is enclosed by a dashed rectangle, while its $k = 2$ nearest neighbours are highlighted with thick black lines.

Some embedded vectors are excluded from being considered near neighbours of the pre- and post-transient references. The next subsection will clarify the reasons for this. Table 7.1 indicates that the excluded embedded vectors are at one of the extreme ends of the time series, as well as next to and including a transient disturbance.

Figure 7.5 highlights with grey lines the embedded vectors which were excluded from the neighbourhood assessment for the pre-transient reference of the first transient.

Replacing segment

For each nearest neighbour $\mathbf{x}_{t_j}^{t_i-M} = \left[x_{t_j}^{t_i-M} \dots x_{t_j+M-1}^{t_i-M} \right]$ of the pre-transient reference \mathbf{x}_{t_i-M} , its subsequent segment $\left[x_{t_j+M}^{t_i-M} \dots x_{t_j+M+T}^{t_i-M} \right]$ is considered, where $T = t_f - t_i$.

Table 7.1: Embedded vectors excluded from the neighbourhood assessment of the pre- and post-transient references. Excluded embedded vectors are at one of the extreme ends of the time series X , and next to and including a transient disturbance. $T = t_f - t_i$, and n is the total number of samples of X .

Location of embedded vectors		Pre-transient	Post-transient
Extreme end of time series	start	-	$\mathbf{x}_1 \cdots \mathbf{x}_{1+T}$
	end	$\mathbf{x}_{n-M-T} \cdots \mathbf{x}_{n-M}$	-
Next to and including transient $[x_{t_i} \cdots x_{t_f}]$	after	-	$\mathbf{x}_{t_i} \cdots \mathbf{x}_{t_f+T}$
	before	$\mathbf{x}_{t_i-M-T} \cdots \mathbf{x}_{t_f-M}$	-

Each segment $[x_{t_j+M}^{t_i-M} \cdots x_{t_j+M+T}^{t_i-M}]$ is mean centred, and the mean-centred segments are denoted as $[x_{0,t_j+M}^{t_i-M} \cdots x_{0,t_j+M+T}^{t_i-M}]$. The k mean-centred segments are averaged according to

$$\left[\frac{1}{k} \sum_{j=1}^k x_{0,t_j+M}^{t_i-M} \quad \cdots \quad \frac{1}{k} \sum_{j=1}^k x_{0,t_j+M+T}^{t_i-M} \right]. \quad (7.5)$$

The mean-centring step is done because of segments $[x_{t_j+M}^{t_i-M} \cdots x_{t_j+M+T}^{t_i-M}]$ such as those after the two nearest neighbours in Figure 7.5. These segments have the same dynamic behaviour, but are centered around a different mean value.

Equation (7.5) results in the segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$, which replaces the empty samples between t_i and t_f . Segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ should agree with the overall dynamics of time series X .

Segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ can also derive from the post-transient reference \mathbf{x}_{t_f} . In this case, the previous segment $[x_{t_j-T}^{t_f} \cdots x_{t_j-1}^{t_f}]$ of each nearest neighbour $\mathbf{x}_{t_j}^{t_f} = [x_{t_j}^{t_f} \cdots x_{t_j+M-1}^{t_f}]$ is considered. The mean-centring and averaging steps are applied on the k segments $[x_{t_j-T}^{t_f} \cdots x_{t_j-1}^{t_f}]$.

The previous subsection indicated that some embedded vectors are excluded from being considered near neighbours of the pre- and post-transient references. The reason is that the replacing segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ is defined by the T -sample segments after the nearest neighbours, in the case of the pre-transient reference, or before the nearest neighbours, in the case of the post-transient reference. Therefore, embedded vectors at one of the ends of the time series, as well as next to and including a transient disturbance need to be excluded.

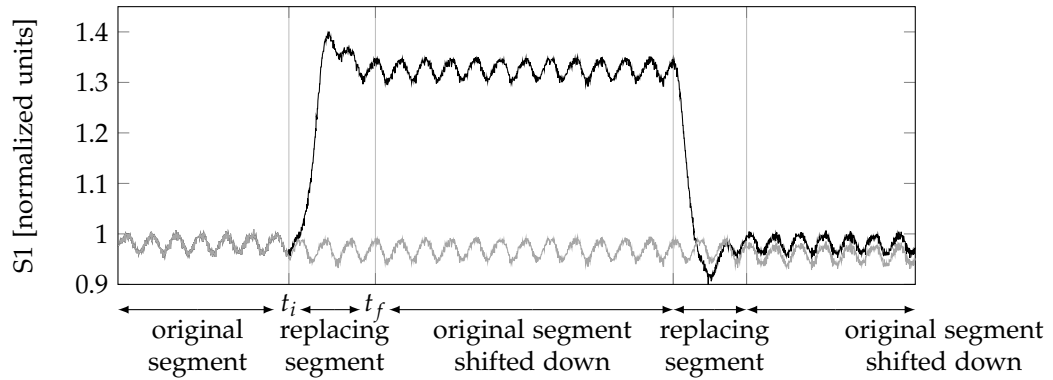


Figure 7.6: Generation of the reconstructed time series (grey line) by concatenating the replacing segments, and the transient-free segments of the original time series (black line).

Concatenation of segments and reconstructed time series Y

A reconstructed time series Y without transients is generated by concatenating the replacing segments $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ and the transient-free segments of the original time series.

The concern when adjoining the ends of two segments is to avoid introducing an artificial discontinuity. Discontinuities happen if adjoining segments are centered around a different mean value, such as the segments before t_i and after t_f in Figure 7.4.

Figure 7.6 illustrates the concatenation procedure based on the pre-transient references x_{t_i-M} , and the list below describes the steps.

1. The first segment of the reconstructed time series Y is the first transient-free segment of the original time series X , that is,

$$\begin{bmatrix} y_1 & y_2 & \cdots & y_{t_i-1} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{t_i-1} \end{bmatrix} \quad (7.6)$$

2. The numerical values of the replacing segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ are shifted to match the point y_{t_i-1} according to

$$\begin{bmatrix} y_{t_i} & \cdots & y_{t_f} \end{bmatrix} = \begin{bmatrix} \hat{x}_{t_i} & \cdots & \hat{x}_{t_f} \end{bmatrix} - (\hat{x}_{t_i} - (y_{t_i-1} + \Delta y_{t_i-1})) \quad (7.7)$$

where Δy_{t_i-1} is the average of the differences at the end of each nearest neighbour $\mathbf{x}_{t_j}^{t_i-M}$, that is,

$$\Delta y_{t_i-1} = \frac{1}{k} \sum_{j=1}^k \left(x_{t_j+M}^{t_i-M} - x_{t_j+M-1}^{t_i-M} \right) \quad (7.8)$$

The purpose of shifting is to avoid discontinuities between the two adjoining segments. The difference Δy_{t_i-1} approximates what the difference $\Delta x_{t_i-1} = x_{t_i} - x_{t_i-1}$ would have been if not for the transient.

3. The next transient-free segment of the original time series is the one which starts at x_{t_f+1} and ends before the next transient disturbance. The numerical values of this transient-free segment are shifted to match the point y_{t_f} according to

$$\left[y_{t_f+1} \quad \cdots \quad y_{t_{i,2}} \right] = \left[x_{t_f+1} \quad \cdots \quad x_{t_{i,2}} \right] - (x_{t_f+1} - (y_{t_f} + \Delta y_{t_f})) \quad (7.9)$$

where the time instant $t_{i,2}$ refers to the start of the second transient disturbance. The value Δy_{t_f} is the average of the differences at the start of each nearest neighbour $\mathbf{x}_{t_j}^{t_f}$. Therefore, Δy_{t_f} approximates what the difference $\Delta x_{t_f} = x_{t_f+1} - x_{t_f}$ would have been if not for the transient.

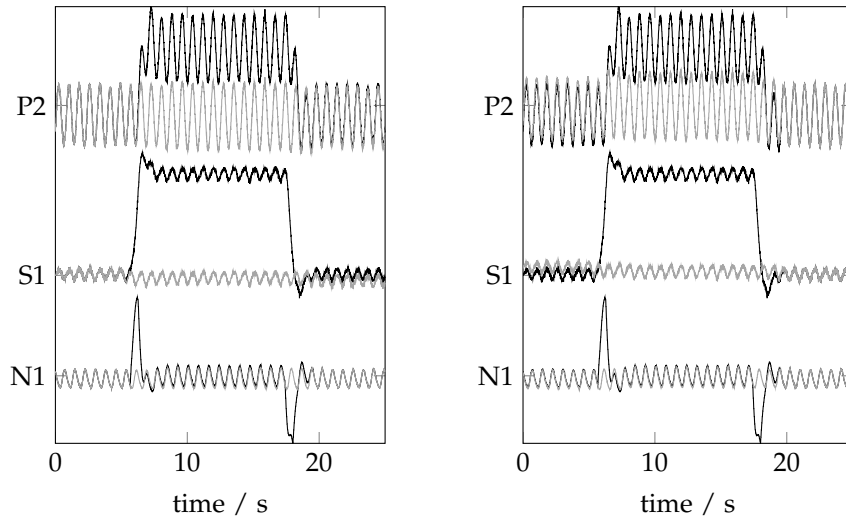
These shifting steps are repeated for the following replacing segments and transient-free segments of the original time series X .

The concatenation procedure based on the post-transient reference \mathbf{x}_{t_f} is done in the opposite direction, that is, from the end of the reconstructed time series Y to its start.

Both Figure 7.7a and Figure 7.7b show the original time series X of the measurements in the development case study in black lines. Superimposed on these and in grey lines, Figure 7.7a shows the reconstructed time series based on the pre-transient references \mathbf{x}_{t_i-M} , and Figure 7.7b shows the reconstructed time series based on the post-transient reference \mathbf{x}_{t_f} .

7.2.3 Adequacy of the replacing segment

The concern with the removal of transient disturbances is that the replacing segment should agree with the underlying dynamics of the original time se-



(a) Based on the pre-transient references. (b) Based on the post-transient references.

Figure 7.7: Reconstructed time series (grey lines) after removal of transients, superimposed on the original time series (black lines).

ries. To assess possible mismatches, this section proposes the two following indices:

- Index D_k measures how similar the reference embedded vector is to its k nearest neighbours.
- Index s_k measures how similar the replacing segment is to the k mean-centred segments which are averaged into it (equation 7.5).

Indices D_k and s_k should be minimized because they assess inconsistencies in the prediction of the replacing segments. In particular, if the system generating the time series is purely deterministic, and the choice of dimension M is correct, then D_k and s_k should be zero. In this case, the replacing segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ will exactly match the underlying dynamics of the original time series.

For both the pre- and post-transient references, index D_k is the average of the squared distances between the reference embedded vector and its k nearest neighbours. This criterion is normalized by the variance of the reference embedded vector. Equation (7.10) describes the computation of D_k from the pre-transient reference $\mathbf{x}_{t_i-M} = [x_{t_i-M} \cdots x_{t_i-1}]$ and its nearest neighbours $\mathbf{x}_{t_j}^{t_i-M} = [x_{t_j}^{t_i-M} \cdots x_{t_j+M-1}^{t_i-M}]$.

$$D_k = \frac{1}{k} \sum_{j=1}^k \left(\frac{\sum_{i=1}^M (x_{t_i-M+(i-1)} - x_{t_j+(i-1)}^{t_i-M})^2}{\sum_{i=1}^M (x_{t_i-M+(i-1)} - \bar{x}_{t_i-M})^2} \right) \quad (7.10)$$

The variance of the reference embedded vector, in the denominator of equation (7.10), is equivalent to the square of the distance between the reference embedded vector and its mean, respectively \mathbf{x}_{t_i-M} and \bar{x}_{t_i-M} in the case of the pre-transient reference. As a result, the normalization proposed compares the similarity of the k nearest neighbours to the similarity of the mean of the reference embedded vector.

Index s_k is the average of the squared distances between the replacing segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ and the k mean-centred segments which are averaged into it. This criterion is normalized by the variance of the replacing segment. Equation (7.11) describes the computation of s_k based on the pre-transient reference \mathbf{x}_{t_i-M} . In this case, the k mean-centred segments are denoted by $[x_{0,t_j+M}^{t_i-M} \cdots x_{0,t_j+M+T}^{t_i-M}]$.

$$s_k = \frac{1}{k} \sum_{j=1}^k \left(\frac{\sum_{i=1}^{T+1} (\hat{x}_{t_i+(i-1)} - x_{0,t_j+M+(i-1)}^{t_i-M})^2}{\sum_{i=1}^{T+1} (\hat{x}_{t_i+(i-1)} - \bar{\hat{x}})^2} \right), \quad (7.11)$$

where $T = t_f - t_i$. The variance of the replacing segment, in the denominator of equation (7.11), is equivalent to the square of the distance between the replacing segment and its mean $\bar{\hat{x}}$. As a result, the normalization proposed compares the similarity of the k mean-centred segments to the similarity of the mean of the replacing segment.

For the development case study, table 7.2 indicates the values of D_k and s_k for the replacing segments based on the pre- and post-transient references. The values are averaged over the two transient disturbances.

Indices D_k and s_k are able to distinguish between the adequacy of the replacing segments which were calculated based on the pre-transient reference and on the post-transient reference. For practical applications, the reconstructed time series Y of a measurement is the one that derives from the replacing segment with lower indices.

For the development case study, Figure 7.7 suggests that the reconstructed time series (grey lines) are equally adequate whether based on the pre- or post-

Table 7.2: Inconsistency indices D_k and s_k based on the pre- and post-transient references for the development case study.

Tag	D_k		s_k	
	Pre-transient	Post-transient	Pre-transient	Post-transient
P2	0.0067	0.0050	0.0017	0.0007
S1	0.1491	0.1225	0.0029	0.0032
N1	0.0049	0.0065	0.0011	0.0011

transient references. For measurements S1 and N1, indices D_k and s_k in table 7.2 agree with this observation. In fact, the values of indices D_k and s_k are not consistently lower for either the pre-transient reference or the post-transient reference. On the other hand, for measurement P2 the values of indices D_k and s_k are both lower based on the post-transient references. This result indicates that, for measurement P2, the replacing segments based on the post-transient references are the ones which best agree with the underlying dynamics of the original time series.

7.3 Parameter settings and sensitivity

The removal of transient disturbances involves the following parameters:

- embedding dimension m ,
- embedding granularity τ ,
- embedding step δ , and
- number of nearest neighbours k .

The objective of this section is to find optimal values for these parameters, given the dynamics of a particular system. Before this is done, the section analyses how each parameter influences the removal task.

7.3.1 Influence of parameters in the removal task

The embedding dimension m , granularity τ , and step δ determine which embedded vectors are identified as nearest neighbours.

Embedding granularity τ and embedding step δ

Parameters τ and δ reduce the computational time, without having to down-sample the time series. The embedding granularity τ increases the number of sampling intervals between each sample included in an embedded vector. Similarly, the embedding step δ increases the number of sampling intervals between consecutive embedded vectors. As a result, the algorithm is $\mathcal{O}(1/\tau\delta)$. The reason to use τ and δ instead of downsampling is that the reconstructed time series Y should retain the same samples as the original time series. This is important because the reconstructed time series will be used in other analyses.

On the other hand, parameters τ and δ affect the accuracy of finding the true nearest neighbours of the reference embedded vectors. If τ is large, the algorithm may assess wrongly the similarity between the trends of embedded vectors because, with fewer samples, outliers to the trend have a larger influence. If δ is large, the algorithm may skip the embedded vectors which are the true nearest neighbours.

Embedding dimension m

For a given τ , the embedding dimension m approximates the number of samples of a state of the system, M , according to $(m - 1)\tau + 1$. States are segments of the time series with a length M such that the adjacent segments can be known unambiguously. Therefore, $(m - 1)\tau + 1$ should span the length of a state so that the method can exploit the determinism of the system.

Number of nearest neighbours k

Parameter k determines the number of nearest neighbours of the reference embedded vector and, by extension, the number of mean-centred segments which are averaged into the replacing segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$. As discussed in section 7.1.3, a real system is not purely deterministic. As a result, the first nearest neighbour to the reference embedded vector may be adjacent to a segment which is not the best predictor for $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$. Therefore, k mean-centred segments should be considered.

7.3.2 Recommendations for parameters and analysis of sensitivity

The optimal values for the parameters are determined using the development case study. To that end, different parameter values are used in order to generate the reconstructed time series Y for each measurement. This is done based on the pre-transient reference and the post-transient reference of each transient disturbance. As defined in section 7.2.2, *reference* means the embedded vector immediately before (pre-transient reference) or after (post-transient reference) the removed segment $[x_{t_i} \cdots x_{t_f}]$. The adequacy of the replacing segments is evaluated based on the indices D_k and s_k introduced in section 7.2.3. These indices measure inconsistencies in the estimation of the replacing segment, and thus should be minimized.

Embedding granularity τ and embedding step δ

Setting $\tau = \delta$ is equivalent to downsampling the time series by a factor of τ . The optimal factor will depend on the original sampling rate, therefore the optimisation results are plotted as function of T_P/τ . T_P is the number of samples in one cycle of the oscillatory time series. T_P/τ corresponds to the number of samples in one cycle if the time series were actually downsampled by a factor of τ .

Figure 7.8 shows the influence of T_P/τ on the indices D_k (panel 7.8a) and s_k (panel 7.8b). In this case study, T_P is approximately 800 samples. Each plot in a panel shows the optimisation results for each variable of the development case study. Each plot combines the results based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).

In general, all plots show lower indices D_k and s_k at higher T_P/τ , indicating that τ and δ should be equal to one. This is expected, because increasing τ and δ means discarding more information.

However, the computational time of the algorithm increases with T_P/τ , as shown in Figure 7.9. Thus, if the computational time is prohibitive, it is important to note is that above $T_P/\tau = 40$ the improvement in performance is moderate. Specifically, the reduction in D_k and s_k is usually below 20%.

Embedding dimension m

The purpose of m is to ensure that $(m - 1)\tau + 1$ spans the number of samples of a state of the system, as explained in section 7.3.1. As a result, the value of parameter m depends on the chosen τ , and on the original sampling rate. With regards to τ , this optimisation uses $\tau = 1$, which is the optimal value. If higher values of τ are needed, then the value of m must be varied accordingly. To take into account the original sampling rate, the results are plotted as a function of m/T_p .

Figure 7.10 shows the influence of m on the indices D_k (panel 7.10a) and s_k (panel 7.10b). Each plot in a panel shows the optimisation results for each variable of the development case study. Each plot combines the results based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).

In general, all figures show a steep decrease in the indices D_k and s_k when m increases until $m = T_p$. Above $m = T_p$, index D_k increases, while s_k stays approximately at the same level. These results suggest that, for $\tau = 1$, m should be set equal to the number of samples in one cycle, T_p . Similar conclusions are reported elsewhere [Thornhill, 2005].

Number of nearest neighbours k

Figure 7.11 shows the influence of k on the indices D_k (panel 7.11a) and s_k (panel 7.11b). Each plot in a panel shows the optimisation results for each variable of the development case study. Each plot combines the results based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).

All figures show a consistent increase in D_k and s_k when k increases, particularly above $k = 3$. These results suggest an optimal value of $k = 2$.

It should be recalled that indices D_k and s_k are averages over the k nearest neighbours. This means that if all neighbour embedded vectors were equal to each other, then D_k would be constant. Likewise, if all segments averaged into the replacing segment $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ were equal to each other, then s_k would be constant. However, a real system such as the one in the development case study seldom generates identical states. Therefore, neighbours become increasingly different from the reference embedded vector, and the segments averaged into $[\hat{x}_{t_i} \cdots \hat{x}_{t_f}]$ become increasingly different from each other.

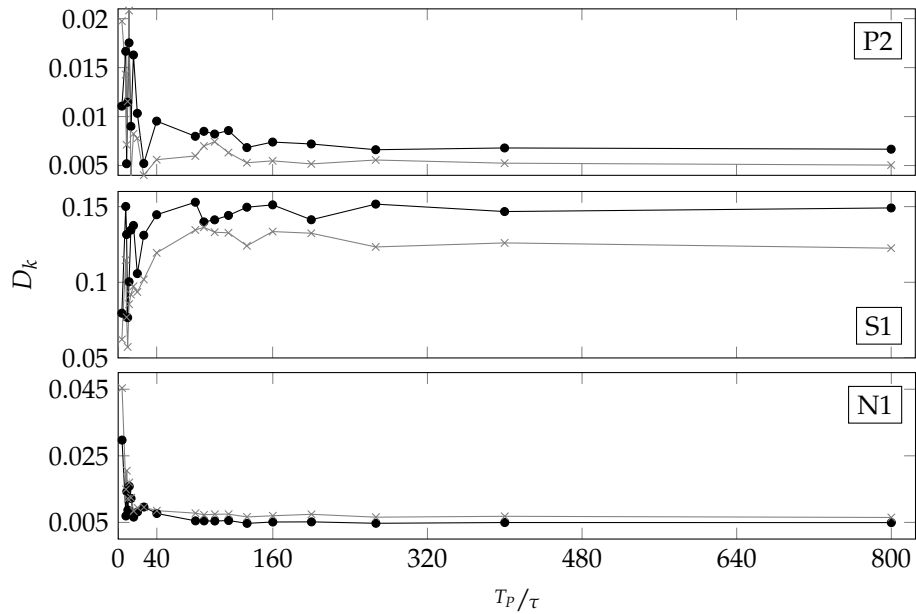
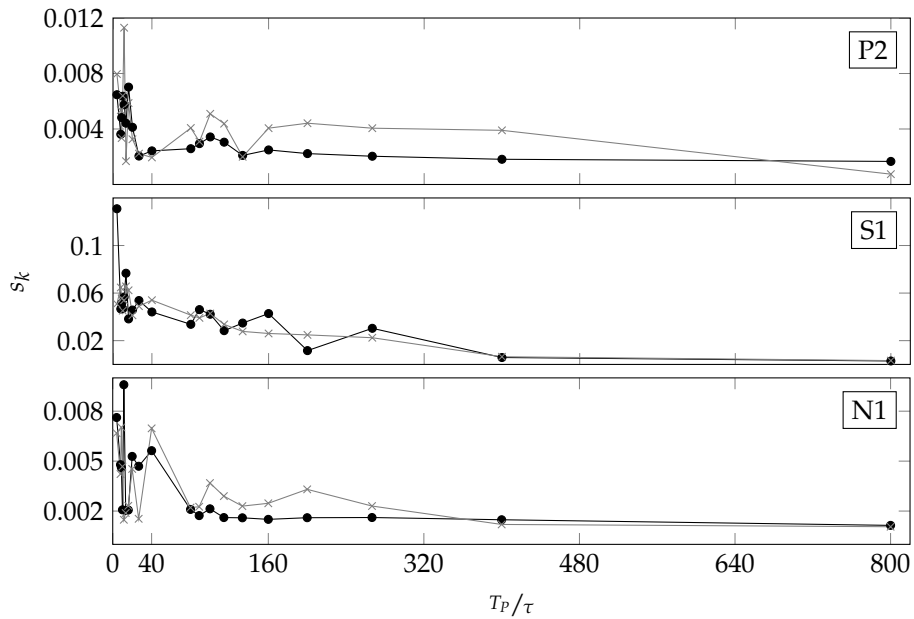
(a) Performance measure D_k .(b) Performance measure s_k .

Figure 7.8: Performance of the removal method as a function of τ and δ , when $\tau = \delta$. Parameters fixed in the analyses were $m = 800$ and $k = 2$. The number of samples per cycle is $T_p = 800$. Results are shown for each variable, and for the estimations based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).

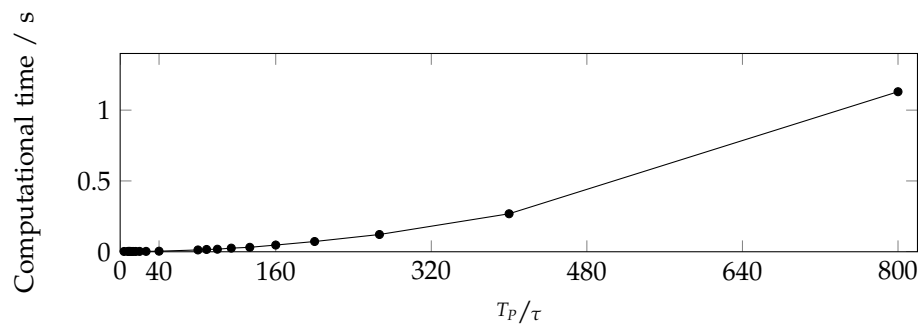
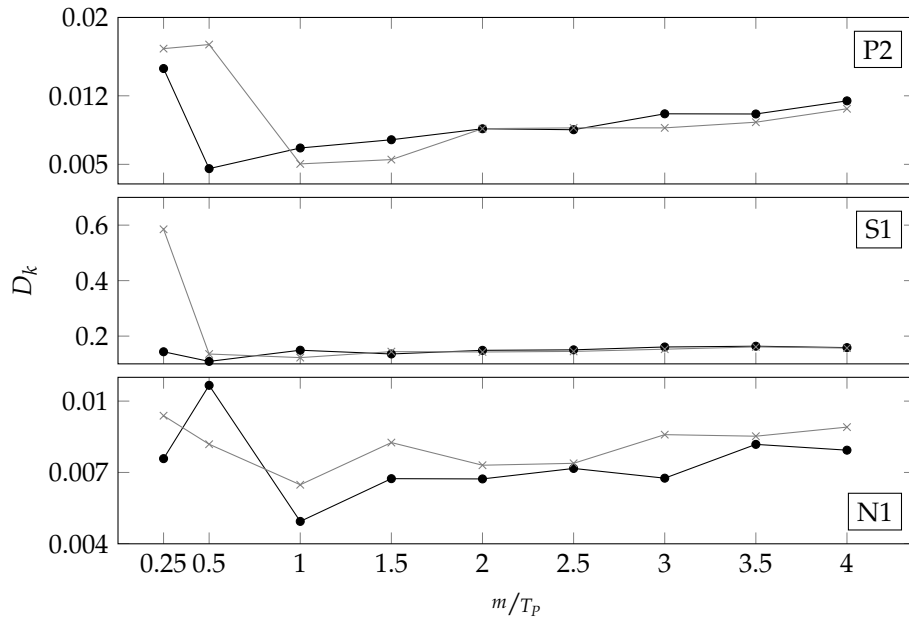
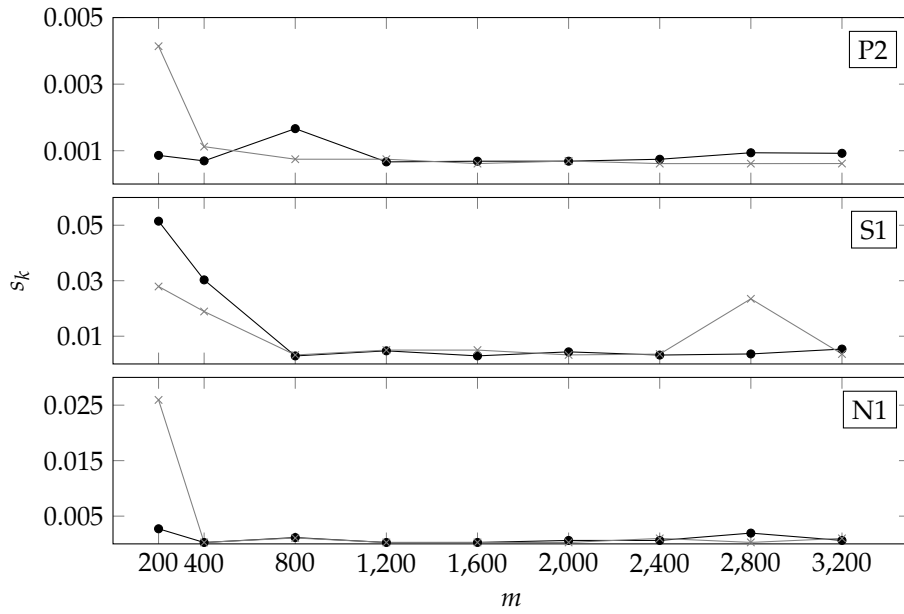


Figure 7.9: Computational time as a function of T_P/τ . The computational time is an average over the three variables, and includes the removal of the two transients.



(a) Performance measure D_k .



(b) Performance measure s_k .

Figure 7.10: Performance of the removal method as a function of m/T_p . Parameters fixed in the analyses were $\tau = 1$, $\delta = 1$ and $k = 2$. Results are shown for each variable, and for the estimations based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).

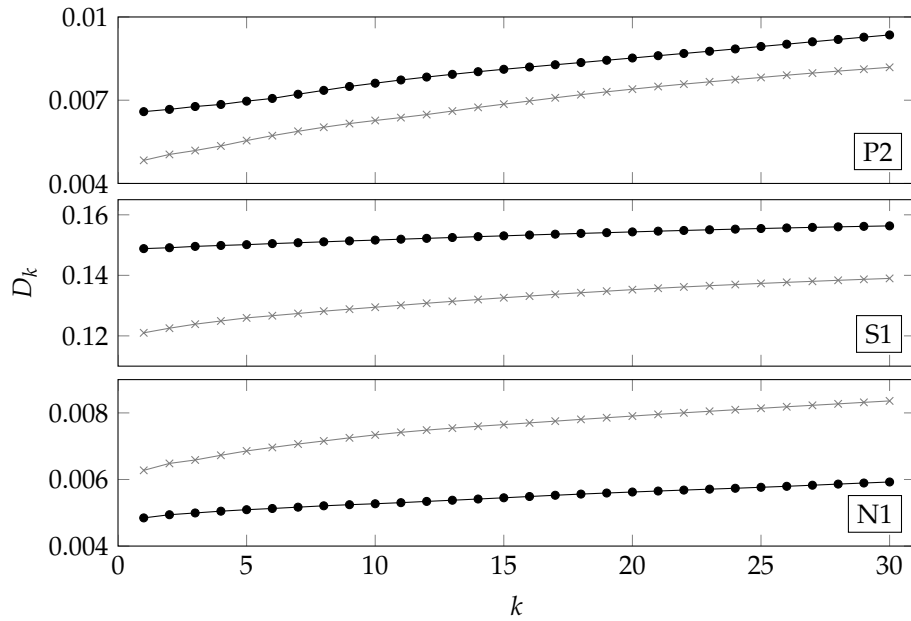
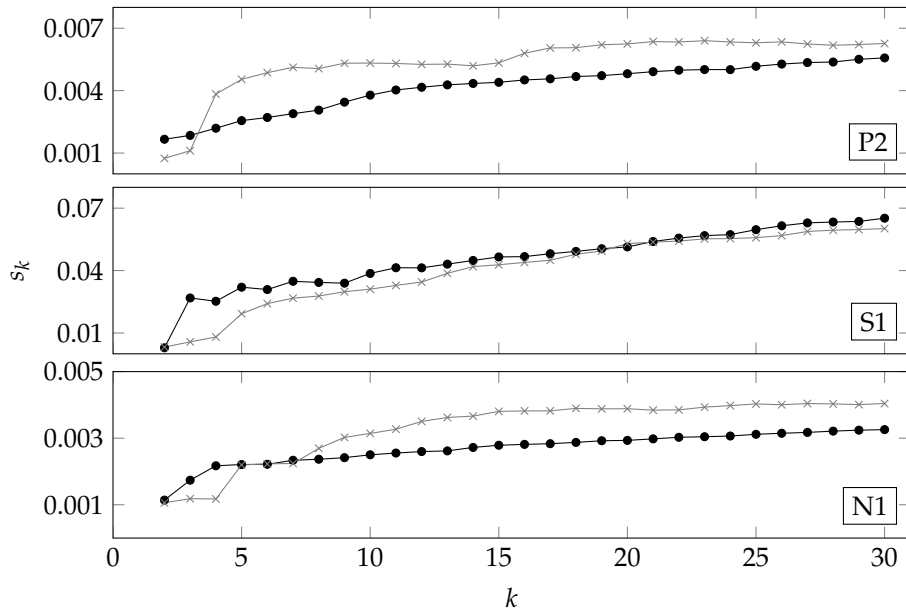
(a) Performance measure D_k .(b) Performance measure s_k .

Figure 7.11: Performance of the removal method as a function of k . Parameters fixed in the analyses were $\tau = 1$, $\delta = 1$ and $m = T_p = 800$. Results are shown for each variable, and for the estimations based on the pre-transient reference (black line with round marker) and the post-transient reference (grey line with cross marker).

Table 7.3: Summary of the roles and recommended values for the parameters.

Parameters	Influence	Recommended value
τ	reduces accuracy of similarity measure between the trends of embedded vectors, and reduces computational time	$\tau < T_p/40$ preferably $\tau = 1$
δ	omits possible embedded vectors from the similarity assessment, and reduces computational time	$\delta < T_p/40$ preferably $\delta = 1$
m	approximates the dimension of a state of the system	$m = T_p/\tau$
k	defines the number of segments averaged into the replacing segment	$k = 2$

Summary

Table 7.3 summarizes the values recommended for all the parameters, as well as their influence in the removal task. Some of these parameters depend on the original sampling rate of the time series. This is taken into account by referring those parameters to T_p , the number of samples in one cycle of the oscillatory time series.

7.4 Application to test case study

The *Gas plant case 2* is the case study used to test the removal method. The case study derives from routine operation of an industrial gas processing plant, as explained in section 2.5. It is the same case study used in Chapter 5, in which the objective was to detect the transient disturbance.

The top panel of Figure 7.12 shows the measurements in the case study. The transient affecting these measurements occurs around hour 4 in the figure. Each time series has 1200 samples taken with a sampling interval of 30 s. The expectation for this test case study is that the nearest neighbours imputation method correctly removes the transients. The reconstructed time series should be an estimate of what the measurement would have been had the transient not been present.

Following the recommendations in section 7.3, parameters τ and δ are set to one. Parameter m is set to 100, which is approximately the number of samples in one cycle of the oscillations. This length is the same in all measurements. Parameter k is set to two.

The middle and bottom panels of Figure 7.12 show the reconstructed time series, which derive from the pre- and post-transient references, respectively. The comparison shows that the removal method and the parameters proposed are able to generate time series which are free of transients and maintain the original oscillatory dynamics.

The reconstructed time series are different for the results based on the pre- and post-transient references. Visual analysis suggests that the time series based on pre-transient references present less distortion of the original time series. This is more evident for measurements P1, P2, S1 and P_B. These observations are confirmed by the inconsistency indices D_k and s_k , which are indicated in table 7.4.

It is generally not possible to know in advance which reference embedded vectors will lead to the most adequate reconstructed time series. In the test case study, the distortions of the reconstructed time series relative to the original time series are a result of the trends of the original time series. As the top panel of Figure 7.12 shows, the two oscillation cycles after the transient disturbance are different from the other cycles in the measurement. Therefore, the replacing segment will join two sections of the measurement which have

Table 7.4: Inconsistency indices D_k and s_k for the new times series of the measurement in the test case study.

Measurement	D_k		s_k	
	Previous	Subsequent	Previous	Subsequent
T1	0.007	0.018	0.001	0.005
T2	0.009	0.039	0.009	0.005
T3	0.041	0.271	0.009	0.007
P1	0.065	0.215	0.009	0.016
P2	0.095	0.226	0.022	0.035
S1	0.071	0.192	0.017	0.298
P _B	0.094	0.229	0.012	0.015

different trends. In this case study, the joining is smoother with the replacing segment based on the pre-transient reference.

The conclusions from these observations are that it is important to calculate reconstructed time series based on both pre- and post-transient references, and to have indices such as D_k and s_k which can decide on the best reconstructed time series on a case-to-case basis.

7.5 Chapter summary

This chapter has presented a method to remove transient disturbances from otherwise oscillating time series. Transient disturbances affect the reliability of several oscillation analysis methods.

The method is based on a nearest neighbours imputation technique. This technique works based on the assumption that the system generating the time series is, to some extent, deterministic. The system can be linear or nonlinear. Time series with repeating patterns may be generated by nonlinear deterministic systems.

Section 7.1 presented the background on nonlinear deterministic systems, and on how these systems have been exploited with nearest neighbours techniques for the purpose of imputation.

Section 7.2 presented the algorithm for the method. The idea is to consider a segment of the time series adjacent to the removed transient disturbance, and look for its k most similar segments in the time series. These are known as the k nearest neighbours. The segments adjacent to the k nearest neighbours are then used as replacements for the removed transient disturbance.

This section also presented two indices which assess inconsistencies associated with the replacing segment. The purpose of the indices is to measure the agreement of the replacing segment to the underlying dynamics of the original time series.

Section 7.3 analysed the sensitivity of the method to its parameters and optimised the values of those parameters. It used the inconsistency indices as measures to compare the different results. Some of the parameters depend on the original sampling rate of the time series, thus optimal values were given in relation to the number of samples in one cycle of the oscillatory time series.

The method proposed, the inconsistency indices, and the optimal parameters were demonstrated in a development case study as well as in a test case study which derived from routine operation of an industrial plant (section 7.4). The results showed that the method and parameters are able to generate time series which are free of transients and agree with the original oscillatory dynamics. Furthermore, the inconsistency indices agree with visual analyses of the adequacy of the reconstructed time series. Therefore, these indices are useful for automated applications of the method.

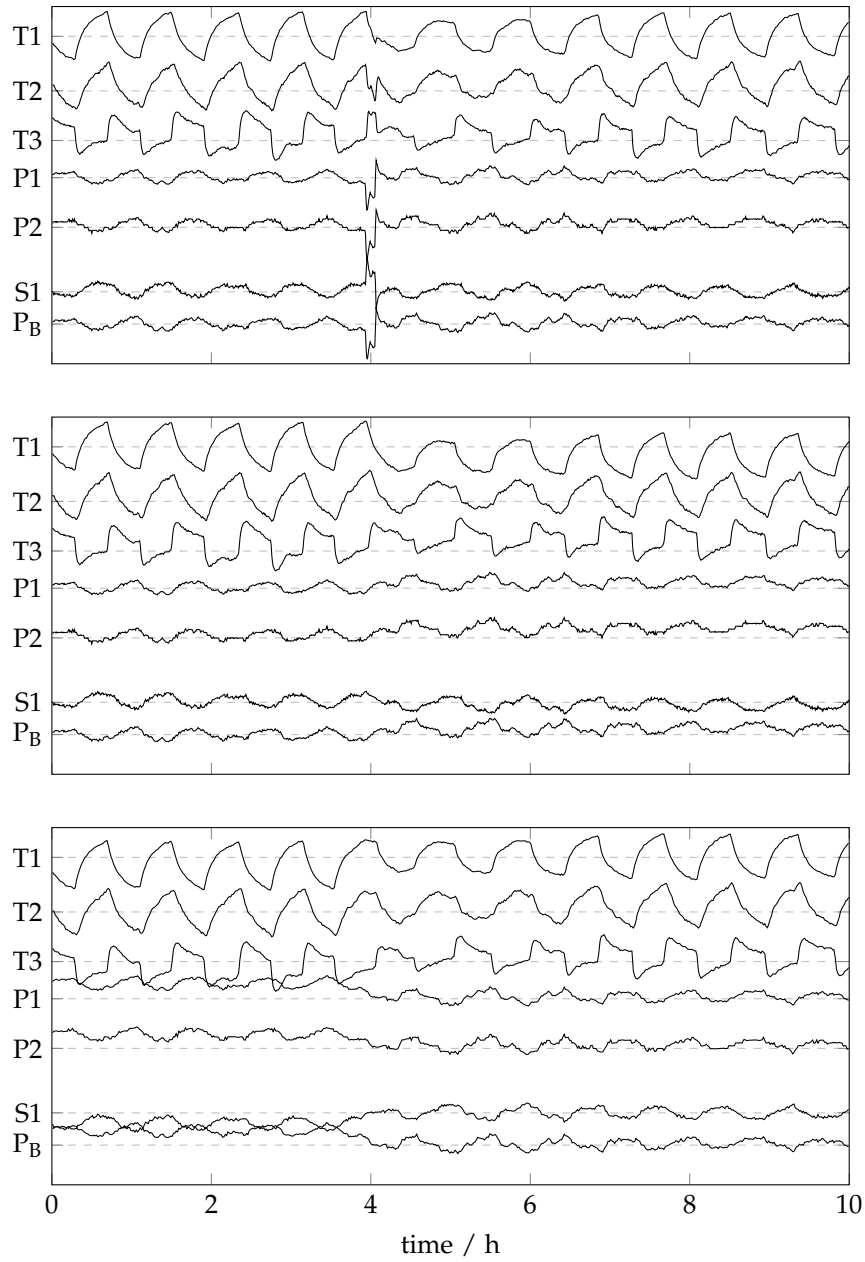


Figure 7.12: Original time series (top panel) and reconstructed time series based on the pre-transient reference (middle panel) and post-transient reference (bottom panel).

Part III

Conclusions

Chapter 8

Summary and discussion

The aim of this chapter is to recall the achievements of this thesis, and to compare them with the initial research questions. Additionally, section 8.2 presents summaries of each of the methods developed, and gives a critical discussion of their merits and limitations.

8.1 Answer to the research questions

This thesis addressed two main research questions: (i) whether it is possible to integrate the electrical and mechanical systems in the analysis of process disturbances, and (ii) how to do it. The main conclusions are (i) that the state-of-the-art offered limited solutions to the integrated analysis, and (ii) that the four new methods developed in this thesis have contributed to solving two of the main gaps in the state-of-the-art. As a result, the extension of process monitoring and diagnosis to the electrical and mechanical systems has become more feasible. However, there are still gaps in the state-of-the-art which require further contributions.

The motivation to extend process monitoring and diagnosis to the electrical and mechanical systems is that the reliability, safety and energy efficiency of processes increasingly depend on the condition of the electrical supply and electromechanical equipment used in the process. This dependence is a result of the interaction between the process, electrical and mechanical systems, the increasing use of electromechanical equipment in industrial processes, and the increase in electrical disturbances.

Table 8.1: Contributions in this thesis framed by the conditions addressed.

Sampling rate	Time scale of disturbance	
	Persistent	Transient
Measurements with equal sampling rates	State-of-the-art	Chapters 4, 5, 7 [Cecílio et al., 2013] [Cecílio et al., 2014]
Measurements with different sampling rates	Future work	Chapter 6 [Cecílio et al., 2014]

The integration of electrical and mechanical measurements in the analysis of process disturbances implies challenges not found in previous work. These challenges include, but are not limited to, the following:

- analysis of transient disturbances,
- access to process and electromechanical measurements,
- analysis of process and electromechanical measurements with irregularities such as
 - different sampling rates for each system,
 - time-misalignment between systems, and
 - irregular sampling rate within a system, and
- application of the methods to large-scale systems.

Except for the access to process and electromechanical measurements, the challenges listed above should be handled through appropriate methods of process monitoring and diagnosis. However, this thesis showed that the current state-of-the-art in process monitoring and diagnosis is particularly limited in the analysis of transient disturbances, of measurements sampled with irregular rate, and of measurements with different sampling rates.

In order to extend the state-of-the-art and address an industrial problem, this thesis developed four new methods which contribute to the analysis of transient disturbances, and of measurements with different sampling rates. Table 8.1 recalls a table which was presented in section 1.4. This table frames the methods developed in this thesis, and indicates the chapters and publications where the methods have been discussed.

All four methods contribute to the analysis of transient disturbances. Two of the methods apply to a single measurement at a time, and hence are applicable whether the process, electrical and mechanical systems have equal or different sampling rates. The other two methods apply to groups of measurements. In one of them, all measurements must have the same sampling rate, while in the other method the measurements can have different sampling rates. The next section presents an extended summary of the four methods, and evaluates the merits and limitations of each.

All methods developed are of advanced signal analysis. The distinguishing feature of advanced signal analysis compared to other methods of process monitoring and diagnosis is that it handles measurement data as time series. This thesis argued that methods of advanced signal analysis are the most suitable for the extension of process monitoring and diagnosis to the electrical and mechanical utilities. In brief, the reasons are that the time series approach takes into account the dynamic behaviour of the system, that advanced signal analysis can exploit complex relations between the trends of process, electrical and mechanical measurements, and that advanced signal analysis methods are not developed for a specific system or operating condition.

8.2 Summary and discussion of the methods developed

This section summarizes the four methods developed in this thesis, and evaluates the merits and limitations of each.

8.2.1 Univariate detection of transient disturbances

Chapter 4 developed a method to detect transient disturbances in the measurement of a single variable. This method was needed because the existing signal analysis methods focused on persistent disturbances, and hence relied on the repetition of the abnormal deviation episode. Transient disturbances, on the other hand, are characterized by being infrequent deviations from normal operation.

Following the advanced signal analysis approach, the transients detection method handles the measurement of a variable as a time series. As a result, the transient disturbance is seen as an unusual segment. The problem of de-

detecting unusual segments is framed as an anomaly detection problem, and solved with a nearest neighbours technique. The basic idea is to use a similarity measure to evaluate the similarity between each segment in the time series and all other segments. Segments that are similar are known as near neighbours. An anomalous segment is detected because it does not have any near neighbours.

Chapter 4 also analysed the statistical significance of the detection threshold, and provided clear recommendations for the parameters of the method. Furthermore, Chapter 4 proposed a colour plot to visualize the detection results in a compact manner which suggests the propagation of the disturbance through the system.

Merits of the method

A significant merit of the transients detection method is that it can detect any deviation of a measurement from its underlying trend. This means that the method is not limited to transient disturbances which have a magnitude or wavelet coefficients different from the underlying trend of the measurement. It also means that the underlying trend of the measurement can consist of oscillations, ramps, noise, or changes in operating level. Furthermore, the method does not require the development of data-based models, and hence is not specific to a particular system or mode of operation.

The colour plot has the merit of conveying the detection results in an industrially relevant manner. The reason is that it is a compact visualisation tool that conveys the plant-wide impact of a certain transient, the variables which are most often affected, and delays between the onsets of the transient, which suggest a propagation path.

Another merit of the method is its utility as a preceding step for other analyses, such as root-cause diagnosis and the removal of transients to enhance the analysis of oscillating measurements.

Limitations of the method

The application of the transients detection method may be limited if the measurement to analyse has a large number n of samples. For instance, if the measurement has 20,000 samples, then the method takes approximately one and a half hours. The reason is that the traditional algorithm for nearest

neighbours is $\mathcal{O}(n^2)$. The number n of samples can be large if one aims to analyse a long data history, or if the sampling rate is fast. In the case of a long data history, the measurement can be divided into shorter segments with no loss for the method. Caution is needed only when the underlying trend of the measurement is periodic. In that case, the number of samples n_{short} of the shortened measurement should be such that there are more than k cycles of the periodic trend, or such that $n_{\text{short}} > 10k$. Parameter k is the number of nearest neighbours, and recommended values were given in section 4.3.2. In the case of a fast sampling rate, section 4.3.2 also concluded that the measurement may be downsampled until the transients have 30 samples. Any extra samples above this number contribute marginally to improving the performance of the method.

Two other limitations of the transients detection method are the need for measurements with regular sampling rate, and the univariate approach. The univariate approach, in particular, means that the presence of the same disturbance in different measurements is not exploited. A multivariate analysis would do this, and could improve the detection of transients in measurements with strong oscillatory trends or noise.

8.2.2 Multivariate detection of transient disturbances in univariate systems

Chapter 5 developed a method to detect transient disturbances in a multivariate context. The multivariate method was needed in order to improve the univariate method of Chapter 4 in the detection of transients in measurements with strong oscillatory trends or noise. The multivariate method successfully detects transients in those measurements because it can recognize if the transient is also present in other measurements. The identification of a common transient is done regardless of the transient having different shapes in the various measurements.

The multivariate detection method uses the univariate anomaly index vectors of several measurements, as constructed in Chapter 4. The multivariate set of anomaly index vectors is analysed through a singular value decomposition, which identifies representative features across the various anomaly index vectors. The representative features correspond to transient disturbances which are particularly strong or are present in several of the measurements.

The anomaly index vector of each measurement is reconstructed as to retain only the features which are both representative in the multivariate set and similar to the behaviour of the univariate anomaly index vector.

Chapter 5 also analysed the statistical significance of the detection threshold, and provided clear recommendations for the parameters of the method. Furthermore, Chapter 5 proposed a plant-wide anomaly index vector which combines the anomaly index vectors of each measurement, and hence provides a global characterization of the presence of transient disturbances in the group of measurements.

Merits of the method

The most notable merit of the multivariate detection method is that it exploits the presence of the same transient disturbance in different measurements, regardless of the transient having different shapes in the various measurements. This means that transient disturbances can also be detected in measurements in which the transients occur but are not evident. It also means that the method continues to apply to transients of any shape.

The plant-wide anomaly index vector has the merit of conveying a global picture of the state of the plant with regards to transient disturbances. Transients which are present in most measurements or are particularly strong will correspond to anomaly indices of larger magnitude, while transients which are present in only a few measurements will correspond to smaller anomaly indices.

The multivariate detection method also maintains the merits of the univariate method, namely being independent of data-based models and being useful as a preceding step for other analyses.

Limitations of the method

The application of the multivariate detection method is limited to groups of measurements of equal sampling rate. However, process, electrical, and mechanical measurements are commonly available with different sampling rates. In this situation, the multivariate method can only be applied if the fast-sampled measurements are downsampled to align with the slow-sampled measurements. Downsampling may lead to loss of information and, potentially, incorrect detection results.

As with the univariate version, the multivariate method is also limited to measurements with regular sampling rate.

8.2.3 Multivariate detection of transient disturbances in multi-rate systems

Chapter 6 extended the multivariate method of transients detection to multi-rate systems. The multi-rate method was needed in order to apply the multivariate analysis when the measurements are available at different sampling rates, without having to downsample the fast-sampled measurements. As shown in the chapter, downsampling leads to loss of information and incorrect detection results. The multi-rate method, on the other hand, maintains correct detection results in the fast-sampled measurements and successfully detects the transients in the slow-sampled measurements.

In the uni-rate methods of Chapter 4 and Chapter 5, embedded vectors are defined as segments from a time series with the same number m of samples. The multi-rate method uses an alternative approach to the construction of embedded vectors which is based on imposing instead the same time span for all embedded vectors. As a result, embedded vectors of fast-sampled measurements have more samples than embedded vectors of slow-sampled measurements. Slow-sampled measurements also have less embedded vectors. Therefore, the multi-rate method expands the anomaly index vectors of slow-sampled measurements to match those of fast-sampled measurements. In brief, the anomaly indices of embedded vectors which cover the same period of time in the fast- and slow-sampled measurements are assigned to the same positions in the corresponding anomaly index vectors. Then, the empty positions in the anomaly index vector of the slow-sampled measurement are populated with the previous anomaly index. This is equivalent to assuming that the trend of the slow-sampled measurement stayed equally anomalous during the periods with no samples.

Chapter 6 also quantified the improvement in the detection of transients when using the multi-rate method in a multi-rate data set compared to the alternative approach of downsampling the fast-sampled measurements to the lower sampling rate. Even when the slow sampling rate was only four times lower than the fast sampling rate, the multi-rate method improved on the uni-rate method by approximately 30%.

Merits of the method

The multi-rate multivariate method has the merits of performing a multivariate analysis and of integrating fast-sampled with slow-sampled measurements. The multivariate approach, as discussed previously, has the advantage of exploiting the presence of the same transient disturbance in different measurements for an improved outcome. Avoiding to downsample fast-sampled measurements to the lower sampling rate prevents the loss of information and incorrect detection results.

An additional merit of the multi-rate method is the new formulation of embedded vectors which imposes the same time span for the embedded vectors of all measurements, instead of the same number m of samples. This formulation has application beyond the detection of transient disturbances. It can be used in any analyses which construct embedded vectors from measurements with different sampling rates, or from a measurement whose sampling rate is not regular. Section 9.1 and section 9.2 will discuss future research directions which use the time-based formulation to develop other disturbance analysis methods for measurements with different sampling rates and with irregular sampling rates.

Limitations of the method

As with the methods developed in Chapter 4 and Chapter 5, the multi-rate method is also limited to measurements with regular sampling rate. The reason is that the embedded vectors constructed from an irregularly sampled measurement have different numbers of samples, and the spacing between samples is not regular. However, the similarity measure used is only defined for embedded vectors which have the same number of samples and whose samples are synchronized.

8.2.4 Removal of transient disturbances from oscillating measurements

Chapter 7 developed a method to remove transient disturbances from the time series of an oscillating measurement. Such a method is needed because transient disturbances affect the reliability of several oscillation analysis methods.

Oscillating measurements have repeated patterns but may not be purely periodic because the repeating patterns may have different shapes and frequencies. Chapter 7 exploited the repeating patterns with a nearest neighbours imputation technique. The idea is to consider a segment of the time series which is adjacent to the removed transient disturbance, and look for its k most similar segments in the time series. These are known as the k nearest neighbours. The segments adjacent to the k nearest neighbours are then used as replacements for the removed transient disturbance.

Chapter 7 also presented two indices to assess inconsistencies associated with the replacing segment. The purpose of the indices is to measure the agreement of the replacing segment to the underlying dynamics of the original time series. The inconsistency indices were used in analysing the sensitivity of the removal method to its parameters and in optimising the values of those parameters.

Merits of the method

A merit of the transients removal method is that it can reconstruct a time series without the transient disturbance while at the same time maintaining the original dynamics of the time series. Furthermore, the method is capable of doing this in any measurement that has some repeating patterns, even if there are patterns of different shapes and frequencies. The method is also applicable to measurements which are mostly noise. Although these measurements do not have repeating patterns, any replacement segment with a noisy trend will easily agree with the original noisy dynamics of the time series.

The inconsistency indices are an additional merit of the transients removal method because they agree with visual analyses of the adequacy of the reconstructed time series. This agreement to the visual evaluation means that the inconsistency indices enable the automation of the method.

Limitations of the method

The nearest neighbours imputation technique relies on the existence of repeating patterns so that future samples can be predicted. As a result, the transients removal method is inherently unsuitable for measurements which have some trend, but this neither repetitive nor noise. Figure A.1a in the appendices shows an example of such a measurement.

Chapter 7 also showed that reconstructed time series may show distortion relative to the original time series in the case that the original time series has different trends before and after the transient.

Another limitation of the transients removal method is the need for measurements with regular sampling rate. One reason is that the embedded vectors from one measurement are all constructed with the same number of samples. Another reason is that conventional similarity measures are defined for embedded vectors which have the same number of samples and whose samples are synchronized.

Chapter 9

Future research directions

The previous chapter recalled the challenges which result from the extension of process monitoring and diagnosis to the electrical and mechanical utilities. That chapter also summarized the contributions of the thesis to solve some of those challenges. The current chapter discusses three ideas for future research which aim to address the following unsolved challenges:

- identification of the propagation path of a persistent disturbance when process and electromechanical measurements have different sampling rates (section 9.1),
- detection and diagnosis of disturbances in measurements with irregular sampling rates (section 9.2), and
- detection and diagnosis of disturbances in large-scale processes (section 9.3).

9.1 Propagation path of a persistent disturbance in multi-rate systems

A disturbance originates at the root cause, and then propagates away from it, affecting other measurements. The propagation path of the disturbance is a qualitative model of the affected system, and shows the affected measurements in a directed succession according to the order of propagation of the disturbance. The relevance of the propagation path for process monitoring

and diagnosis is that the measurement closest to the root cause can be inferred by tracking the disturbance up the propagation path.

The survey in section 3.4 reviewed methods to determine qualitative models of process systems, and section 3.4.2 focused in particular on the methods that determine the propagation path. Methods that determine the propagation path of a disturbance search for features in measurement data such as time delays and attenuation. The reason is that the dynamic characteristics of the system change the shape of the disturbance when it propagates to subsequent measurements. The methods that have been developed so far to determine the propagation path require that the disturbance be persistent.

Methods that determine the propagation path quantify, for example, the nonlinearity of time series [Thornhill, 2005], the transfer entropy between two time series [Bauer et al., 2007b, Duan et al., 2012, Naghoosi et al., 2013], or the non-linear mutual prediction between two time series [Bauer et al., 2007a, Stockmann et al., 2012]. These methods have been successfully used in the analysis of persistent disturbances in process systems, and are available in commercial tools. These methods are applicable only to uni-rate systems, that is, systems whose measurements are all available with the same sampling rate. However, systems with process and electromechanical measurements can be multi-rate because process measurements are usually sampled approximately 1000 times slower than electromechanical measurements. In this situation, the electromechanical measurements have to be downsampled to the process rate in order to apply the existing methods. However, downsampling may compromise the accuracy of the results. For instance, if the disturbance lasts shorter in the electromechanical measurements than in the process measurements, the slow process sampling rate may be enough to capture the disturbance in the process measurements but not in the electromechanical measurements.

This section presents an idea to adapt the non-linear mutual prediction method by Bauer et al. [2007a] in order to extend it to multi-rate systems. The objectives of this section are to present the idea and the current algorithm for the multi-rate method, and to use examples to compare the multi-rate method and the uni-rate method by Bauer et al. [2007a]. This section will show that, in the selected case study, the method by Bauer et al. [2007a] is able to determine the correct propagation path even with the measurements significantly downsampled. Nonetheless, this section will also speculate about other data sets in which downsampling the fast-sampled measurements may

Table 9.1: Common changes in a propagating disturbance due to dynamic characteristics of the system.

Change	Underlying dynamic characteristic
Time lag between the disturbance in the measurements of two variables	Dead time
Low pass filtering, <i>i.e.</i> smoothing of the disturbance trend	Time constant
Decrease in the disturbance magnitude	Gain smaller than one
Addition of noise	Measurement noise or outside influences

compromise the analysis results. For those data sets, the idea proposed in this section could be a relevant direction for research.

9.1.1 Background

Changes in a propagating disturbance

As discussed in section 2.4.1, a disturbance changes when it propagates along a system due to the dynamic characteristics of that system. Table 9.1 indicates four changes which are commonly observed.

Some of these changes can be observed in Figure 9.1. The time series in this figure belong to the case study which will be used in sections 9.1.3 to 9.1.5. The sequence of the measurements reflects, from top to bottom, the propagation path of a disturbance, whose start is seen after the 30 s time instant. The effect of additional time constants is best observed from measurement S1.sp to N1, and from N1 to S1, whereas the effect of dead time is best observed from measurement S1 to P1, and from P1 to P2. Methods which are sensitive to the changes indicated in table 9.1 can be used to determine which measurements precede others in the propagation path of the disturbance.

Non-linear mutual prediction

The top panel in Figure 9.2 shows a time series, X , which has repeating patterns. As discussed earlier, repeating patterns can be exploited to predict values of the time series. For instance, the three highlighted segments are all similar, and the figure shows that their future samples, which are marked

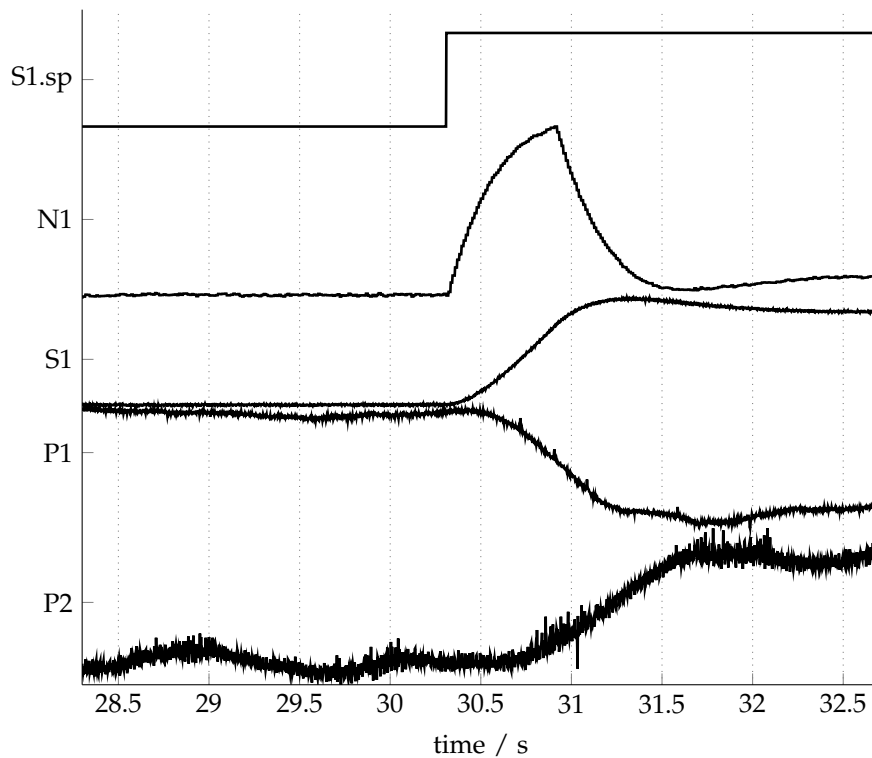


Figure 9.1: Close-up on the start of a disturbance induced in measurement $S1.sp$. The sequence of plots reflects, from top to bottom, the propagation path of the disturbance.

by crosses, are also similar. This means that if the crossed sample x_{i+h} were unknown, it could be predicted from the other two crossed samples. The prediction of a time series from its own past is known as self-predictability [Kantz and Schreiber, 2003]. This property was explored in Chapter 7 in order to replace transient disturbances by predictions that agreed with the underlying dynamics of the time series.

Another way to predict sample x_{i+h} is if there is another time series, Y , which is related to X . A relationship between X and Y may occur if one of the time series is an input to a system and the other time series is an output, if both time series are outputs of the same system, or if the systems that generate the time series are connected by a common driver [Schiff et al., 1996].

If time series X and Y are related, then the repetition of a pattern in one of the time series should imply also the repetition of a pattern in the other time series. This is illustrated in Figure 9.2. The figure shows that the similar segments highlighted in X correspond to segments in Y which are also similar.

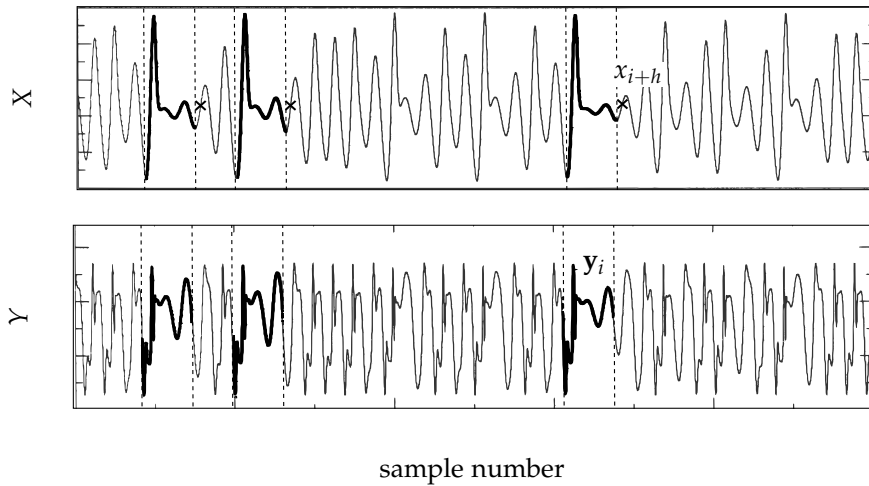


Figure 9.2: Illustration of predictability: three identical X samples, marked with crosses, can be indicated by three past identical segments in X (self-predictability) or in Y (non-linear mutual predictability). Adapted from Le Van Quyen et al. [1998].

This means that the samples to predict x_{i+h} can also be indicated with the aid of time series Y . The highlighted Y segment y_i is the segment which occurs before sample x_{i+h} . The other two highlighted Y segments are significantly similar to y_i . Therefore, the X samples which occur h samples after those two highlighted Y segments can be used as predictors for x_{i+h} . The prediction of a time series from the past of another time series as described here is known as non-linear mutual predictability [Schiff et al., 1996, Le Van Quyen et al., 1998].

Predictability improvement and directionality

The prediction of X may be more accurate from the past of Y than from its own past. The same comments can be made about the predictability of time series Y from its own past compared to using the past of time series X . The predictability improvements of X and Y can be compared, and this comparison allows to ascertain whether there is directionality in the relationship between the two time series [Feldmann and Bhattacharya, 2004]. If X improves the prediction of Y more than Y improves the prediction of X , then one can say that the relationship between the two time series has a direction from X to Y .

There has been some debate about the nature of this directionality. Some authors have interpreted the directionality as a cause-and-effect, or drive-and-response, relationship [Schiff et al., 1996, Le Van Quyen et al., 1998, Feldmann

and Bhattacharya, 2004]. Other authors have stressed that such interpretation can only be made in some circumstances [Arnhold et al., 1999].

Regardless of this debate, what is important for the determination of the propagation path is to understand how non-linear mutual predictability is affected by the presence of the propagation features indicated in table 9.1. As discussed in Bauer et al. [2007a], if time series Y has time lag, attenuation, or added noise in relation to X , then these features make Y easier to predict from the past of X than from its own past. Therefore, the comparison of predictability improvements can be used to determine the propagation path of a disturbance, as shown by Bauer et al. [2007a]. The comparison of predictability improvements will be used in the current section with the same purpose. The implementation uses nearest neighbours of embedded vectors, as done by the other authors referred in this review.

9.1.2 Illustrative example

Figure 9.3 shows the time series of five measurements from the *Compressor rig* data set. All measurements are available with the sampling rate of 1 kHz. The schematic of the compressor rig was presented in section 2.5.1, and shows the location of these measurements.

The time series show a train of pulses induced in the set-point S1.sp of the shaft speed. The deviations in the time series of the other measurements result from the propagation of the set-point disturbance. The sequence of measurements reflects, from top to bottom, the propagation path of the disturbance. The changes in the propagating disturbance due to the dynamic characteristics of the system were discussed in Figure 9.1.

The reason to use the *Compressor rig* data set is because the methods discussed in the current section aim to infer the propagation path of the disturbance from the measurement data. To evaluate the performance of those methods, the real propagation path must be known. The experimental system which generated the *Compressor rig* data set is understood, and its models were presented in section 2.3. The root cause of the disturbance is known, and its expected propagation path can be derived from the models of the system. From the models in section 2.3, the changes induced in S1.sp should propagate to the other measurements according to the order of the measurements in Figure 9.3.

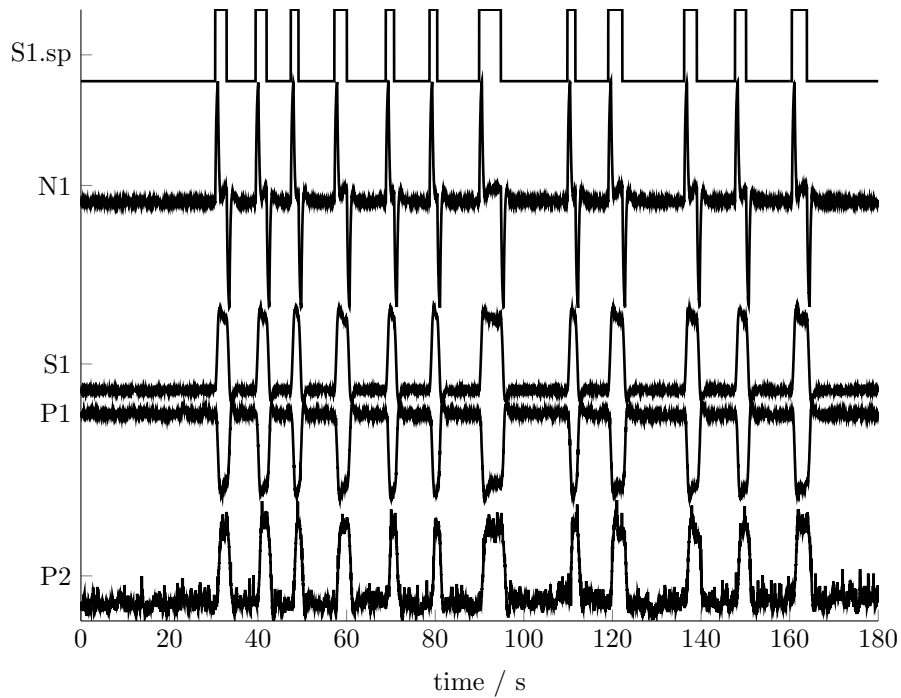


Figure 9.3: Time series of measurements from the *Compressor rig* data set.

It should be noted that measurement P1 in Figure 9.3 pertains to the inlet pressure of the compressor. In the models of section 2.3, this pressure is constant because the models assume that the inlet pipe has unrestricted flow and is open to the air. In the experimental rig, the flow in the pipe is restricted by an inlet valve, as shown in the schematic of Figure 2.12. Therefore, the pressure at the point of measurement is affected by the pressure drop in the inlet valve, and hence is related to the flow of air that enters the compressor. The models in section 2.3 indicate that, in the propagation path of the disturbance, the flow entering the compressor is located between the speed S1 and the outlet pressure P2. This reason justifies the order of P1 in Figure 9.3, and in the expected propagation path.

9.1.3 Algorithm for uni-rate method

This subsection presents an algorithm to implement the uni-rate method by Bauer et al. [2007a]. Showing this algorithm is necessary to highlight the differences of the new idea for a multi-rate method.

Embedding matrices and predicted samples

The uni-rate method considers two measurements, $X = \{x_1 \ x_2 \ x_3 \ \cdots \ x_N\}$ and $Y = \{y_1 \ y_2 \ y_3 \ \cdots \ y_N\}$, which are sampled synchronously with the same sampling interval. Measurements X and Y are arranged in embedding matrices \mathbf{X} and \mathbf{Y} , according to

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_m \\ \mathbf{x}_{m+1} \\ \vdots \\ \mathbf{x}_{n-h} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \vdots & \vdots & & \vdots \\ x_{N_E} & x_{N_E+1} & \cdots & x_{n-h} \end{bmatrix} \quad (9.1a)$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_m \\ \mathbf{y}_{m+1} \\ \vdots \\ \mathbf{y}_{n-h} \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \\ y_2 & y_3 & \cdots & y_{m+1} \\ \vdots & \vdots & & \vdots \\ y_{N_E} & y_{N_E+1} & \cdots & y_{n-h} \end{bmatrix} \quad (9.1b)$$

The number of N_E rows in the embedding matrices derives from n , m and h according to

$$N_E = n - h - m + 1 \quad (9.2)$$

Each row in the matrices is an embedded vector \mathbf{x}_i or \mathbf{y}_i with m samples.

To assess the predictability of X , each embedded vector \mathbf{x}_i and \mathbf{y}_i (on the left of equation (9.3)) is used to predict the sample of X that occurs h sampling intervals after the end of the embedded vector (on the right of equation (9.3)). Array \mathbf{X}_h in equation (9.3) aligns the predicted samples of X with the corresponding embedded vectors of \mathbf{X} and \mathbf{Y} .

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_m \\ \mathbf{y}_{m+1} \\ \vdots \\ \mathbf{y}_{n-h} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_m \\ \mathbf{x}_{m+1} \\ \vdots \\ \mathbf{x}_{n-h} \end{bmatrix} \quad \mathbf{X}_h = \begin{bmatrix} x_{m+h} \\ x_{m+h+1} \\ \vdots \\ x_{n-h} \end{bmatrix} \quad (9.3)$$

The prediction of the samples in \mathbf{X}_h from the embedded vectors in \mathbf{X} will yield the self-predictability of X , while the prediction of the samples in \mathbf{X}_h from the embedded vectors in \mathbf{Y} will yield the mutual predictability of X by Y .

The self and mutual predictability of Y are formulated analogously. In the first case, the samples in an array \mathbf{Y}_h are predicted from the embedded vectors in \mathbf{Y} , and in the second case from the embedded vectors in \mathbf{X} .

Similarity measure and nearest neighbours

The Euclidean distance metric is used to assess the similarity between each pair of embedded vectors in one embedding matrix. The purpose is to retrieve, for each embedded vector, the indices of its k^{th} most similar embedded vectors, which are known as the k nearest neighbours.

For each embedded vector \mathbf{x}_i , the indices of its k nearest neighbours are denoted $r_{i,j}$, where $j = 1 \cdots k$. For each embedded vector \mathbf{y}_i , the indices of its k nearest neighbours are denoted $s_{i,j}$, where $j = 1 \cdots k$.

Self-predictability

To assess the self-predictability of X , each sample x_{i+h} in array \mathbf{X}_h is compared to its k predictors $x_{r_{i,j}+h}$. Each of these predictors is the sample which occurs h -steps ahead of the embedded vector $\mathbf{x}_{r_{i,j}}$. The comparisons of x_{i+h} with its predictors are averaged according to

$$e_i(x_{i+h}|X) = \frac{1}{k} \sum_{j=1}^k |x_{i+h} - x_{r_{i,j}+h}| \quad (9.4)$$

Therefore, quantity $e_i(x_{i+h}|X)$ gives the prediction error of sample x_{i+h} given the past of measurement X . Vector $\mathbf{e}(X|X)$ arranges the prediction errors $e_i(x_{i+h}|X)$ sequentially for all predicted samples in X .

To assess the self-predictability of Y , the quantities $e_i(y_{i+h}|Y)$ and the vector $\mathbf{e}(Y|Y)$ are defined from samples y_{i+h} and its k predictors $y_{s_{i,j}+h}$. Each of these predictors is the sample which occurs h -steps ahead of the embedded vector $\mathbf{y}_{s_{i,j}}$.

The top panels of Figure 9.4 show the vectors of self-predictability errors for measurements S1.sp and S1 of the reference case study.

The sample-wise errors in vector $\mathbf{e}(X|X)$ are averaged over all samples to give the average self-predictability error $\bar{e}(X|X)$. The self-predictability error $\bar{e}(Y|Y)$ for Y is computed in the same manner. The top panels of Figure 9.4 also indicate the average self-predictability errors for measurements S1.sp and S1 in the reference case study.

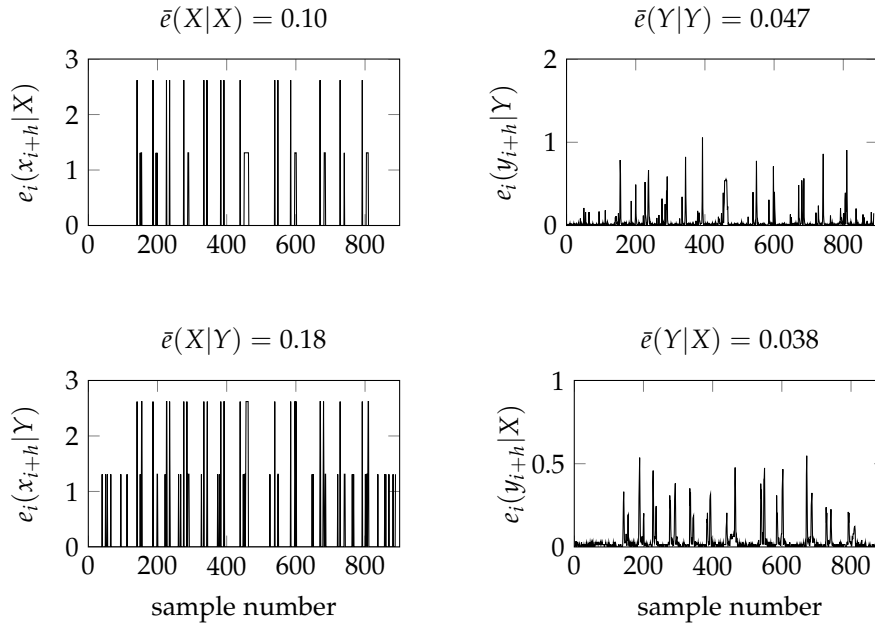


Figure 9.4: Predictability errors for measurements S1.sp (represented by X) and S1 (represented by Y) of the reference case study. Top panels: self-predictability. Bottom panels: mutual predictability. Values on top of plots indicate the average predictability error.

Mutual predictability

To assess the mutual predictability of X by Y , each sample x_{i+h} in array \mathbf{X}_h is compared to its k predictors $x_{s_{i,j}+h}$. Each of these predictors is based on the embedded vector $\mathbf{y}_{s_{i,j}}$. The comparisons of x_{i+h} with its predictors are averaged according to

$$e_i(x_{i+h}|Y) = \frac{1}{k} \sum_{j=1}^k |x_{i+h} - x_{s_{i,j}+h}| \quad (9.5)$$

Therefore, quantity $e_i(x_{i+h}|Y)$ gives the prediction error for sample x_{i+h} given the past of measurement Y . Vector $\mathbf{e}(X|Y)$ arranges the prediction errors $e_i(x_{i+h}|Y)$ sequentially for all predicted samples in X .

To assess the mutual predictability of Y by X , the quantities $e_i(y_{i+h}|X)$ and the vector $\mathbf{e}(Y|X)$ are defined from samples y_{i+h} and its k predictors $y_{r_{i,j}+h}$. Each of these predictors is now based on the embedded vector $\mathbf{x}_{r_{i,j}}$.

The bottom panels of Figure 9.4 show the vectors of mutual predictability errors for measurements S1.sp and S1 in the reference case study.

As with self-predictability, the sample-wise errors in vectors $\mathbf{e}(X|Y)$ and $\mathbf{e}(Y|X)$ are averaged over all samples to give the average self-predictability errors $\bar{e}(X|Y)$ and $\bar{e}(Y|X)$. The bottom panels of Figure 9.4 also indicate the average mutual predictability errors for measurements S1.sp and S1 in the reference case study.

Comparing the values in the top left panel of Figure 9.4 with the values in the bottom left panel shows that the past of S1 does not improve the prediction of S1.sp because the prediction error increases. Conversely, comparing the values in the top right panel with the values in the bottom right panel shows that the prediction of S1 is more accurate from the past of S1.sp than from its own past. These observations are quantified numerically with a predictability improvement measure.

Predictability improvement and directionality

Equation (9.6) defines the predictability improvement measure $H(X|Y)$, which compares the prediction of X from the past of Y with the prediction of X from its own past. The complementary measure $H(Y|X)$ is defined analogously.

$$H(X|Y) = \frac{\bar{e}(X|Y)}{\bar{e}(X|X)} \quad (9.6)$$

The predictability improvements of X and Y are compared using equation (9.7). This comparison yields $H_{X \rightarrow Y}$, a measure of the directionality of the influence between X and Y . If $H_{X \rightarrow Y}$ is positive the influence is directed from X to Y , if $H_{X \rightarrow Y}$ is negative the influence is directed from Y to X .

$$H_{X \rightarrow Y} = H(X|Y) - H(Y|X) \quad (9.7)$$

The directionality measure for measurements S1.sp and S1 of the reference case study is $H_{S1.sp \rightarrow S1} = 0.93$. This value indicates that the disturbance affecting these two measurements propagated from S1.sp to S1. This result agrees with the knowledge about the system, that is, the disturbance was initiated by inducing changes in the compressor speed set-point S1.sp, and as a result it affected the speed of the shaft S1.

9.1.4 Idea for a multi-rate method

This subsection proposes an adaptation of the uni-rate method, so that the propagation path may also be determined in multi-rate systems.

Consequences of downsampling in the uni-rate method

The uni-rate method presented before requires that embedding matrices \mathbf{X} and \mathbf{Y} have the same number of embedded vectors N_E . Furthermore, the method assumes that the number of samples m in each embedded vector is the same in the two matrices. Therefore, to apply the uni-rate method to a pair of measurements with different sampling intervals, samples and embedded vectors of the fast-sampled measurement need to be discarded to match the slow-sampled measurement. Equation (9.8) illustrates the discarded samples and embedded vectors of \mathbf{X} for the case of measurement $X = \{x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ \cdots \ x_{N_X}\}$ having half the sampling interval of measurement $Y = \{y_1 \ y_3 \ y_5 \ \cdots \ y_{N_Y}\}$. This is illustrated for $m = 3$ and $h = 1$. The discarded samples and embedded vectors of \mathbf{X} are indicated by the grey colour.

$$\mathbf{Y} = \begin{bmatrix} y_1 & y_3 & y_5 \\ y_3 & y_5 & y_7 \\ \vdots & \vdots & \vdots \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ x_2 & x_3 & x_4 & x_5 & x_6 \\ x_3 & x_4 & x_5 & x_6 & x_7 \\ x_4 & x_5 & x_6 & x_7 & x_8 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad \mathbf{X}_h = \begin{bmatrix} x_7 \\ x_8 \\ x_9 \\ x_{10} \\ \vdots \end{bmatrix} \quad (9.8)$$

Discarding samples of an embedded vector means that the segment of X spanned by that embedded vector is less well characterized, and hence the accuracy of the search for nearest neighbours may be compromised. Discarding embedded vectors means that the samples x_{i+h} which would be predicted from those embedded vectors will not be predicted. The predicted samples which are discarded from \mathbf{X}_h are shown in equation (9.8) in grey colour.

The multi-rate method avoids discarding samples of the embedded vectors, and predicts the samples x_{i+h} which correspond to discarded embedded vectors.

New formulation for embedding matrices

For multi-rate systems, Chapter 6 proposed to construct embedded vectors by imposing the same time span for the embedded vectors of all measurements, instead of the same number m of samples. The same idea is used here.

Measurement Y is considered as the slow-sampled measurement, with a sampling interval of Δt^s , and a number of samples m^s in each embedded vector y_i . The formulation of embedding matrix \mathbf{Y} is the same as in section 9.1.3. The embedding matrix \mathbf{Y} in the left side of equation (9.8) illustrates the case of Δt^s being two time units, and $m^s = 3$.

Measurement X is considered as the fast-sampled measurement, with a sampling interval of Δt^f . Its embedded vectors must span the same time interval as the embedded vectors of Y , hence the number of samples m^f in each embedded vector x_i is given by

$$\frac{(m^f - 1)}{(m^s - 1)} = \frac{\Delta t^s}{\Delta t^f} \quad (9.9)$$

Furthermore, each embedded vector x_i in the embedding matrix \mathbf{X} must lag the previous by δ^f samples, where δ^f is given by

$$\delta^f = \frac{\Delta t^s}{\Delta t^f} \quad (9.10)$$

The purpose of δ^f is that embedding matrices \mathbf{X} and \mathbf{Y} have the same number of embedded vectors, and that these span the same time intervals. The left side of equation (9.11) illustrates the resulting embedding matrix \mathbf{X} for the case of Δt^f being one time unit. As a result, $m^f = 5$ and $\delta^f = 2$.

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ x_3 & x_4 & x_5 & x_6 & x_7 \\ \vdots & & \vdots & & \vdots \end{bmatrix} \quad \mathbf{X}_h = \begin{bmatrix} x_6 & x_7 \\ x_8 & x_9 \\ \vdots & \end{bmatrix} \quad (9.11)$$

This formulation for the embedding matrix \mathbf{X} assures that the embedded vectors in \mathbf{X} and \mathbf{Y} span the same time intervals, without discarding samples from X .

New formulation for predicted samples

The formulation of the array \mathbf{Y}_h is the same as in section 9.1.3, that is, each row has the sample y_{i+h} which occurs h sampling intervals after the end of the embedded vector \mathbf{y}_i .

On the other hand, array \mathbf{X}_h is expanded to include the samples which are discarded with the uni-rate method. These are the samples shown in light grey colour in equation (9.8). The expansion is illustrated in equation (9.11). As a result, embedded vectors \mathbf{x}_i and \mathbf{y}_i are used to predict not only sample x_{i+h} , but also the $\delta^f - 1$ samples before that.

Predictability improvement and directionality

The rest of the algorithm follows as in section 9.1.3. The objective again is to determine the predictability improvement measures $H(X|Y)$ and $H(Y|X)$, and the directionality measure $H_{X \rightarrow Y}$.

To illustrate the multi-rate method with measurements S1.sp and S1, measurement S1 was downsampled by a factor of 5000. The reason is that the reference case study is not a multi-rate data set because all the measurements, including the process measurements, were sampled at 1 kHz. This situation is unusual because process measurements are normally sampled at a much lower rate than electromechanical measurements. Therefore, the slow-sampled measurement had to be obtained by downsampling.

The top panels of Figure 9.5 plot the vectors of self-predictability errors for measurements S1.sp and S1 of the reference case study, and indicate the corresponding average self-predictability errors. The bottom panels plot the vectors of mutual predictability errors, and also indicate the corresponding average self-predictability errors.

Comparing the top left panel with the bottom left panel shows that the past of S1 does not improve the prediction of S1.sp. Comparing the top right panel with the bottom right panel shows that the prediction of S1 is more accurate from the past of S1.sp than from its own past. These results are the same as obtained with the uni-rate method when both measurements had the fast sampling rate.

The directionality measure for this example is $H_{S1.sp \rightarrow S1} = 0.71$. This value indicates the same directionality between S1.sp and S1 as obtained with the uni-rate method when both measurements had the fast sampling rate.

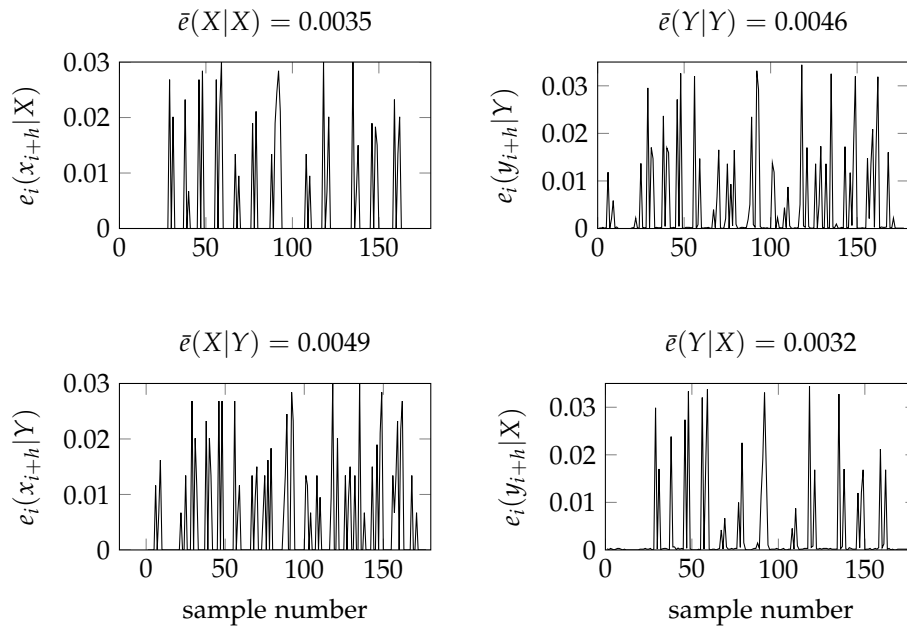


Figure 9.5: Predictability errors for measurements S1.sp (represented by X) and S1 (represented by Y) of the reference case study determined with the multi-rate method. Top panels: self-predictability. Bottom panels: mutual predictability. Values on top of plots indicate the average predictability error.

9.1.5 Application of the uni-rate and multi-rate methods

This section presents results of applying the uni-rate and multi-rate methods to the reference case study in four different tests:

- the uni-rate method is applied to pairs of measurements with the original fast-sampling rate,
- the uni-rate method is applied to the same pairs of measurements after downsampling,
- the multi-rate method is applied to the same pairs of measurements in which the first measurement has the fast sampling rate and the second measurement has the slow sampling rate, and
- the multi-rate method applied to the same pairs of measurements in which the first measurement has the slow sampling rate and the second measurement has the fast sampling rate.

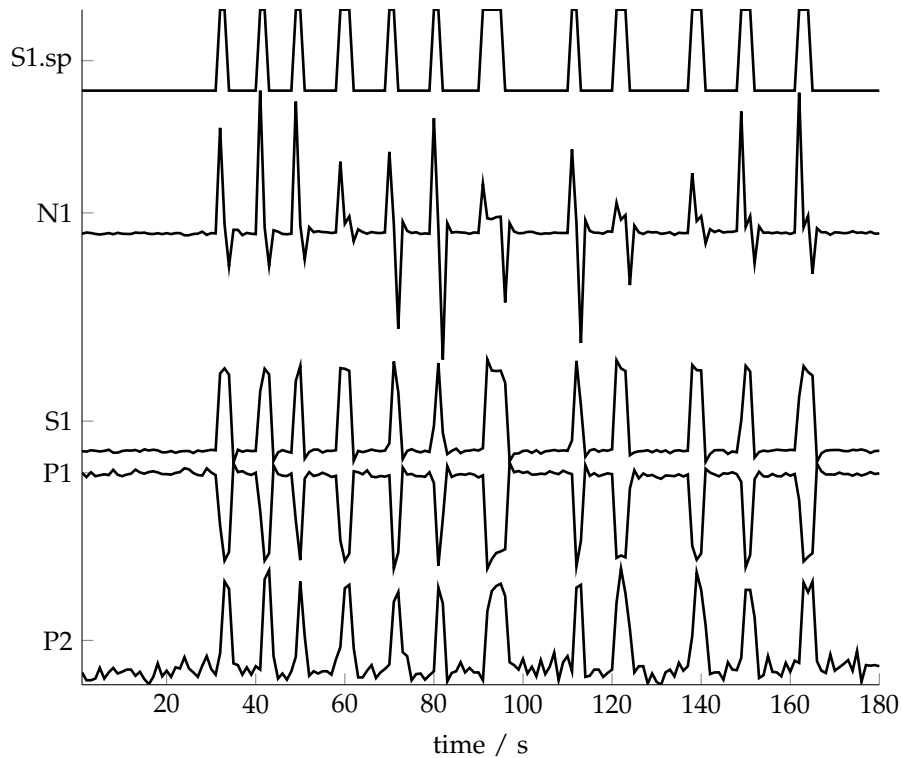


Figure 9.6: Measurements from the reference case study after downsampling by a factor of 5000.

As mentioned, the reference case study was not a multi-rate data set because all the measurements, including the process measurements, were sampled at 1 kHz. Therefore, slow-sampled measurements had to be obtained by downsampling the measurements from the reference case study. The downsampling factor was 5000. The effect of the downsampling in the trends of the measurements is shown in Figure 9.6.

Table 9.2 summarizes the directionality measures $H_{X \rightarrow Y}$ in the four tests for selected pairs of measurements. Not all possible pairs of measurements are shown because the objective of the current chapter is to present ideas for future research and not to analyse the ideas extensively. The expected result is that all selected pairs have positive $H_{X \rightarrow Y}$ measures. The reason is that the models presented in section 2.3 were consulted so that the measurements chosen as X precede in the propagation path the measurements chosen as Y .

The results in table 9.2 show that, in this case study, the uni-rate method by Bauer et al. [2007a] is able to determine the correct direction of propagation

Table 9.2: Directionality measures $H_{X \rightarrow Y}$ for selected pairs of measurements in the reference case study calculated in four different tests. F and S refer to the fast and the slow-sampled measurement, respectively.

	Uni-rate method		Multi-rate method	
	$F \rightarrow F$	$S \rightarrow S$	$F \rightarrow S$	$S \rightarrow F$
$H_{S1.sp \rightarrow N1}$	0.82	0.68	0.66	0.50
$H_{N1 \rightarrow S1}$	0.52	0.03	0.25	0.35
$H_{S1 \rightarrow P1}$	0.96	0.46	0.71	0.62
$H_{P1 \rightarrow P2}$	1.90	0.81	0.73	0.81

of the disturbance, even with the measurements significantly downsampled. The results also show that a multi-rate method as suggested in this section is on a par with the established method.

9.1.6 Comments on the proposed research direction

Section 9.1 discussed as a future research direction the identification of the propagation path of a persistent disturbance when process and electromechanical measurements have different sampling rates. The challenge with different sampling rates is that downsampling the fast-sampled measurement to the slower rate may cause loss of information, while interpolating the slow-sampled measurements to the faster rate may create false information. The purpose of the multi-rate method is to use the information available.

A possible direction towards the goal of a multi-rate method for determining propagation is to adapt the method by Bauer et al. [2007a], which is only applicable to measurements with the same sampling rate. This section suggested an algorithm to adapt that uni-rate method to a multi-rate data set, and compared results of the two methods using one case study.

The uni-rate method applied to measurements with a fast sampling rate is able to determine the correct propagation path of the disturbance. The initial analysis suggests that the new algorithm achieves comparable results in a multi-rate data set.

In a multi-rate data set, the uni-rate method requires the fast-sampled measurements to be downsampled to the slower rate. In the case study, the uni-rate method was able to determine the correct propagation path when the measurements were significantly downsampled. However, in other data sets downsampling the fast-sampled measurements may compromise the results.

For instance, if the disturbance lasts shorter in the fast-sampled electromechanical measurements than in the slow-sampled process measurements, the slower sampling rate may be enough to capture the disturbance in the process measurements but not in the electromechanical measurements. For those data sets, the idea in this section could be a relevant direction for research.

9.2 Monitoring and diagnosis in irregularly sampled time series

As discussed in section 1.2, the analysis of process, electrical and mechanical measurements may be hindered by practices adopted in the storage of those measurements. To save memory, measurements are sometimes compressed, either by eliminating samples or by substituting the values of the samples by a constant value, for example, the average over a period. As a result, the interval between samples in a measurement is not constant. If a measurement with varying intervals between samples is handled as a time series, then it is known as an irregularly sampled time series [Rehfeld et al., 2011].

A time series X is a finite sequence of n sample values $x(t_i)$ (equation (9.12b)). The values $x(t_i)$ are ordered according to the sampling instants t_i , which are strictly increasing and form a sequence T (equation (9.12a)).

$$T = \{t_1, t_2, \dots, t_n\} : t_1 < t_2 < \dots < t_n \quad (9.12a)$$

$$X = \{x(t_1), x(t_2), \dots, x(t_n)\} : t_1 < t_2 < \dots < t_n \quad (9.12b)$$

When the interval between each sample Δt is constant, the time instants t_i are strictly increasing multiples of that interval, that is $t_i = i \cdot \Delta t$. In this case, replacing t_i with its index i becomes a valid representation of time, and the time series can be simply represented by its sequence of samples in the form $X = \{x_1, x_2, \dots, x_n\}$. On the other hand, irregularly sampled time series have varying intervals between samples, and hence both sequences T and X in equation (9.12) are required to describe the time series.

In process monitoring and diagnosis, the methods which analyse process measurements as time series have been reviewed in section 3.6.2. All the methods discussed in that section were developed for time series with constant

sampling intervals Δt and are not applicable to measurements such as those affected by data compression. For instance, several methods obtain the spectral information of the measurements from their Fourier transforms [Thornhill et al., 2002, Choudhury et al., 2004, Tangirala et al., 2007, Zang and Howell, 2007, Babji and Tangirala, 2010]. The Fourier transform assumes that the measurement samples are taken at regular intervals. Another example is the use of cross-correlation to determine the time lag between two measurements [Bauer and Thornhill, 2008].

The methods developed in this thesis also analyse measurements as time series. All the methods are based on nearest neighbours of embedded vectors. Embedded vectors are segments of a time series which should span a constant interval of time. Therefore, in irregularly sampled time series those segments will have a varying number of samples and varying intervals between samples. The nearest neighbours approach implies measuring the similarity between those segments. The conventional similarity measures are defined between two ordered sequences \mathbf{p} and \mathbf{q} which have the same number n of samples and whose samples are synchronized. For example, the Euclidean distance metric which is used in the methods of this thesis is defined as

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}. \quad (9.13)$$

As a result, the methods developed in this thesis also are not directly applicable to measurements such as those affected by data compression. This section presents an idea to reformulate the construction of embedded vectors and the computation of similarity for the case of irregularly sampled time series. The new formulation should allow to adapt all the methods developed in this thesis. The objectives of this section are to present the idea and the current algorithm, and to use examples to demonstrate the potential of this research direction.

9.2.1 Background

Analysis of irregularly sampled time series

Research in irregularly sampled time series is commonly found in domains such as astronomy [Scargle, 1989, Bos et al., 2002], finance [Zumbach and Müller, 2001], and geophysics [Rehfeld et al., 2011].

The methods to analyse irregularly sampled time series can be grouped into four categories [Rehfeld et al., 2011]: (i) reconstruction methods, (ii) spectral transforms, (iii) ARMA model fitting, and (iv) weighing methods.

Reconstruction methods resample the time series into a regular time grid and then apply existing methods developed for regularly sampled time series. Common techniques of resampling include linear and spline interpolation, regression, and approximation by the value of the sample closest in time [Lall and Sharma, 1996].

A common spectral transform for irregularly sampled time series is the Lomb-Scargle Fourier transform [Scargle, 1989]. It determines the spectrum of a measurement from a least squares fit of sine curves to the time series of the measurement. It is suitable for measurements with periodic components and no outliers [Stoica et al., 2009]. The wavelet transform can also be computed for irregularly sampled time series if implemented through the lifting scheme [Sweldens, 1998].

Fitting autoregressive-moving-average (ARMA) models to a time series involves determining the coefficients of the ARMA model. To determine the coefficients from irregularly sampled time series, research focused on adapting estimation algorithms such as maximum-likelihood estimation [Isaksson, 1993] and the Burg algorithm [Bos et al., 2002].

Weighing methods generalize measures, such as distance and correlation, which are conventionally computed between pairs of aligned samples [Rehfeld et al., 2011]. The conventional implementation of these measures is illustrated in Figure 9.7a. For the case of the distance measure, the arrows in the figure indicate that only differences between aligned samples are considered. Instead, the weighing method calculate differences between all possible pairs of samples, and weight each difference according to the time misalignment between the pair. This idea is illustrated in Figure 9.7b for sample p_{10} and sample q_6 . Larger weights are represented in the figure by darker tones on the arrows. The weighing function is such that the more aligned samples are, the more their difference counts towards the distance metric.

The weighted version of the Euclidean distance metric is defined as

$$d(\mathbf{p}, \mathbf{q}, w) = \sqrt{\sum_{i=1}^{n_p} \sum_{j=1}^{n_q} w_{i,j} (p_i - q_j)^2} \quad (9.14)$$

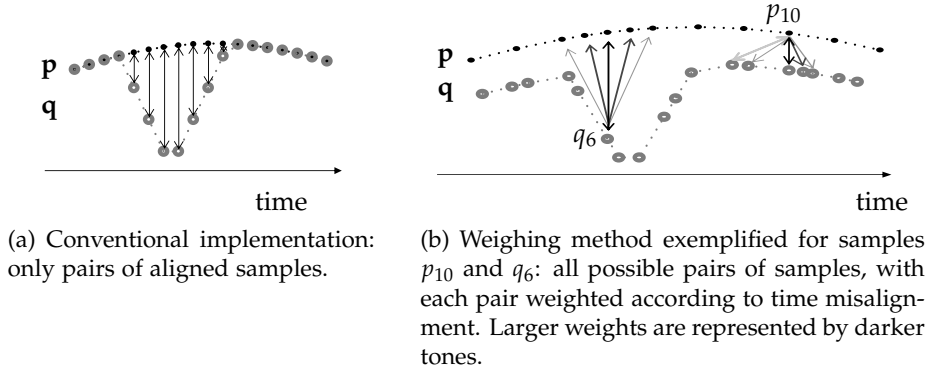


Figure 9.7: Pairs of differences (represented by arrows) used in assessing the distance between segments \mathbf{p} (black line and markers) and \mathbf{q} (grey line and markers).

where $w_{i,j}$ is the weight attributed to the difference between sample p_i of time series \mathbf{p} and sample q_j of time series \mathbf{q} . Examples of weight functions w found in the literature are sinc and Gaussian functions [Rehfeld et al., 2011]. In particular, the Gaussian function (equation (9.15)) is a positive function which decays smoothly to zero, and is symmetric with relation to the time misalignment $(t_i - t_j)$ between samples p_i and q_j .

$$w_{i,j} = w(t_i, t_j) = \frac{1}{\sqrt{2\pi}L} \exp\left(-\frac{(t_i - t_j)^2}{2L^2}\right) \quad (9.15)$$

Since a distance metric should be non-negative and symmetric, the Gaussian function is a relevant alternative for a weighing function. The Gaussian weighing function has a width parameter L which determines the rate of decay of the weight values $w_{i,j}$ with the time misalignment between the two samples.

One of the challenges discussed in section 9.2 is measuring similarity between segments whose sample are irregularly spaced. The weighted Euclidean distance metric using a Gaussian weighing function is a possible direction towards that goal.

Time-based construction of embedded vectors

Embedded vectors are conventionally defined as segments from a time series with the same number m of samples, with each segment lagging the previous by δ samples [Kantz and Schreiber, 2003]. If the same number m of sam-

ples were imposed to an irregularly sampled time series, then the embedded vectors would not span the same duration of time. If the same step δ were imposed, then embedded vectors of different measurements would not be aligned.

Chapter 6 proposed an alternative approach to the construction of embedded vectors which is based on imposing the same time span M for all embedded vectors. In that chapter, the focus was on groups of measurements which had different sampling rates, although the sampling rate of each measurement was regular. As a result, all embedded vectors of a single measurement had the same number of samples, but this number was different from the embedded vectors of another measurement with a different sampling rate.

For measurements with irregular sampling rates, the implications of using the constant time span M is that the embedded vectors of a single measurement may all have different number of samples. The weighted Euclidean distance metric discussed previously is able to compute distances between embedded vectors with these sampling characteristics.

9.2.2 Idea to adapt embedded vectors and similarity to irregularly sampled time series

The first two steps of all methods developed in this thesis involve the construction of embedded vectors from a measurement, and an assessment of the similarity between all pairs of embedded vectors. This subsection presents a possible algorithm to construct embedded vectors and compute similarity in the case of a measurement with irregular sampling rate.

Embedded vectors

The construction of embedded vectors from the time series of a measurement is represented in Figure 9.8. An embedded vector \mathbf{x}_r is defined as a segment of a time series X which spans M time units, as defined in Chapter 6. Furthermore, embedded vector \mathbf{x}_r lags the previous \mathbf{x}_{r-1} by a constant number Δ of time units. This leads to a total of N_E embedded vectors. It should be noted that the embedded vectors of a measurement cannot be arranged in an embedding matrix, as in the case of regularly sampled measurements. This is due to the different number of samples in each embedded vector, as illustrated in Figure 9.8.

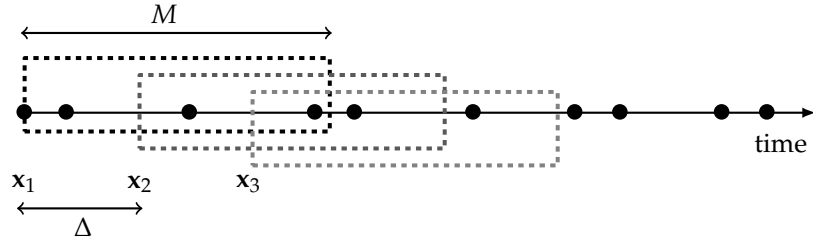


Figure 9.8: Construction of first three embedded vectors from the irregularly sampled time series X . Each embedded vector spans M time units and lags the previous by Δ time units.

Additionally, for each embedded vector \mathbf{x}_r , a time vector \mathbf{t}_r should be created to arrange the time instants of each sample in \mathbf{x}_r .

Similarity

Each pair of embedded vectors \mathbf{x}_r and \mathbf{x}_s is then compared using the weighted Euclidean distance metric

$$d(\mathbf{x}_r, \mathbf{x}_s, w)^s = \frac{\sqrt{\sum_{i=1}^{m_r} \sum_{j=1}^{m_s} w_{i,j} (x_{r,i} - x_{s,j})^2}}{\sqrt{\sum_{i=1}^{m_r} \sum_{j=1}^{m_s} w_{i,j}}} \quad (9.16)$$

where i and j represent the indices of the samples in \mathbf{x}_r and \mathbf{x}_s , respectively. This equation scales the weighted Euclidean distance metric presented in equation (9.14). The aim of scaling is to have a metric $d(\mathbf{x}_r, \mathbf{x}_s, w)^s$ which is independent of the number of samples in \mathbf{x}_r and \mathbf{x}_s .

The weighing function $w = w(t_{r,i} - t_{s,j})$ is defined as in equation (9.15), and depends on the time instants of the samples of \mathbf{x}_r and \mathbf{x}_s . The width parameter L may be optimized or used as suggested in Rehfeld et al. [2011], that is,

$$L = \frac{\bar{\Delta}t}{4} \quad (9.17)$$

where $\bar{\Delta}t$ is the mean value of the sampling intervals in measurement X .

The next subsection will show the alternative embedded vectors and similarity described here in the univariate detection of transient disturbances.

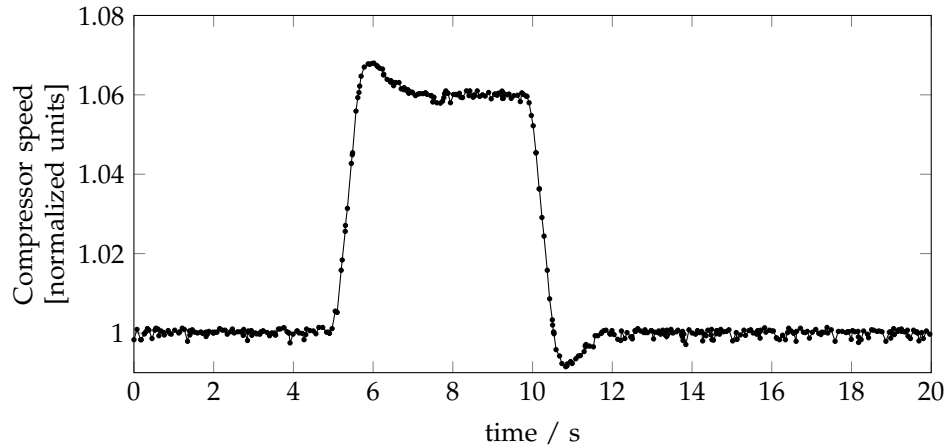


Figure 9.9: Compressor speed measurement from *Compressor rig case 1*. The measurement was manipulated in order to have an irregularly sampled time series. The values are normalized by the initial value.

9.2.3 Application to univariate detection of transient disturbances

This subsection illustrates the alternative construction of embedded vectors and similarity measure on the *Compressor rig case 1* measurement. This is the same measurement as used in Chapter 4 for the univariate detection of transient disturbances. Therefore, the results of the method in Chapter 4, for regularly sampled measurements, can be used as a benchmark for the method for irregularly sampled measurements.

The measurement in the *Compressor rig case 1* represents the shaft speed of the compressor during 20 seconds. The shaft speed was measured at a regular rate of 1 kHz, therefore its measurement had to be manipulated in order to have an irregularly sampled time series for illustration of the method. This was done by randomly eliminating samples from the original measurement, which also resulted in a decrease in the total number of samples from 20,000 samples to approximately 400. The time instants of the retained samples were stored. Figure 9.9 shows the speed measurement after this manipulation. Figure 9.10 shows a close-up to highlight the irregular spacing between samples.

Two step changes, around 6 and 11 s, were imposed in the drive of the compressor by changing its speed set-point, resulting in the two transients seen in the figure. The objective of the proposed method is to detect those transients. Detection is achieved through an anomaly index vector \mathbf{a}_i , whose

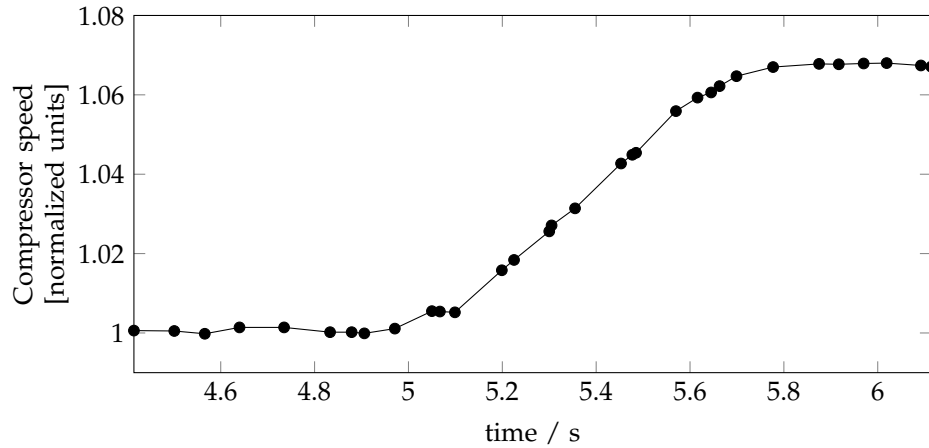


Figure 9.10: Close-up on the ompressor speed measurement to highlight the irregular spacing between samples.

values should be high for embedded vectors including the transients, and low for embedded vectors including periods of normal operation.

Figure 9.11 shows the anomaly index vector \mathbf{ai} computed from the measurement in Figure 9.9. The computation used the idea proposed in section 9.2.3 for the two initial steps, and then followed the method in Chapter 4. As in Chapter 4, \mathbf{ai} was normalized by its median so that $ai = 1$ now approximates the average anomaly index of non-anomalous embedded vectors.

The figure shows that the construction of embedded vectors and similarity measure suggested in this section are able to cope with the sampling irregularity and achieve the detection of the two transients. The positive detection is indicated by the fact that the embedded vectors which correspond to transient disturbances have anomaly indices above the detection threshold, which is represented by the dashed line in Figure 9.11. Comparing the anomaly index vector in this figure with that from Figure 4.6a in Chapter 4 shows the similarity between the two results.

9.2.4 Comments on the proposed research direction

Section 9.2 discussed as a future research direction the adaptation of the methods in this thesis to the case of measurements with irregular sampling rates. The challenge with irregular sampling rates is that the conventional construction of embedded vectors and similarity measure cannot be applied. Therefore, this section suggested an alternative formulation for embedded vectors

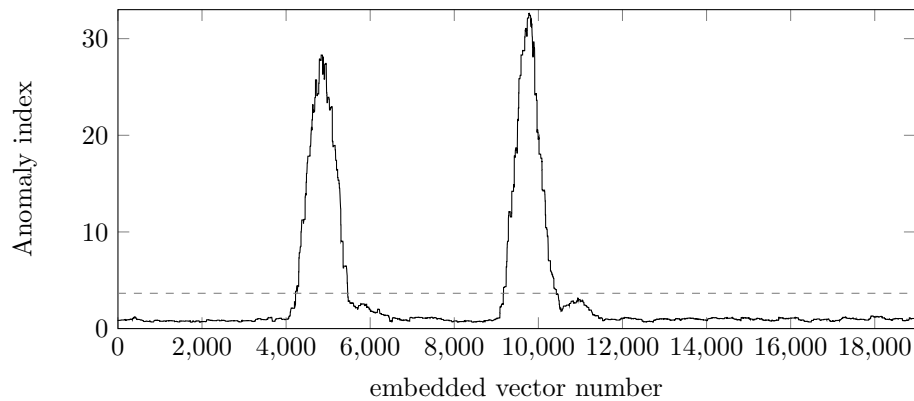


Figure 9.11: Normalized anomaly index vector. The dashed line indicates the detection threshold.

and a weighted distance metric to assess the similarity between these embedded vectors.

Construction of embedded vectors and similarity assessment constitute the first two steps of all methods developed in this thesis. Therefore, with irregularly sampled time series the methods could employ the two steps as formulated in this section, and then follow with the subsequent steps.

The initial analysis of the new formulation was done with the univariate detection of transient disturbances. The univariate detection method had been developed in Chapter 4 for regularly sampled measurements. The results showed that the new formulation achieves results in an irregularly sampled time series comparable to the results in Chapter 4. This indicates that the idea proposed in this section could be a relevant direction for research.

Open questions about the idea proposed include:

- studying if, and in which conditions, the weighted Euclidean metric converges to conventional Euclidean metric,
- determining the statistical behaviour of the anomaly index vectors in order to attribute a confidence level to the selected threshold, in the case of the detection methods,
- optimize the width parameter L , and re-evaluate the parameter optimisation done for methods with regularly sampled measurements, and

- analyse the sensitivity of the methods to the distribution of samples in the measurement.

9.3 Large-scale systems: Systematic approach of integrated analysis and functional specifications for a semi-automated tool

As discussed in section 1.2, one of the challenges in process monitoring and diagnosis is to carry out the analyses in large-scale systems. Large-scale industries such as oil and gas may have several thousands of measured variables from the process system alone. This dimension affects visualisation, requires significant computational memory and speed, and increases the complex dependencies amongst variables due to recycles, energy integration and control paths. As a result, it becomes difficult to extract useful information.

In industry, process monitoring and diagnosis is normally undertaken by contracted services companies. Experts of such companies consult different sources of information to guide them in evaluating the state of the process and the causes of unwanted operation. The sources of information include:

- measurements from process and control variables, usually collected continuously by process control systems,
- measurements from the utility systems, machinery and electrical drives, usually collected by condition monitoring systems,
- event logs, recording occurrences such as interventions of operators and alarms, and
- process schematics, which provide qualitative topological information and help to give meaning to data.

The reason why all these sources are informative is that the behaviour of a process results from a combination of factors, namely the process conditions, control tuning and structure, the condition of equipment and utility systems, and the actions of operators.

Existing commercial tools can automate the analysis of some of these sources, although separately. For instance, the Plant Disturbance Analyser tool by ABB automates the detection of oscillations and root-cause diagnosis in process and control measurements [Horch et al., 2007]. Another example

is the DriveMonitor tool, also by ABB, which collects and analyses measurements from electrical drives [Wnek et al., 2006]. However, the integration of information from all the sources is mostly done manually, hence being a time-consuming task particularly in large-scale systems.

As reviewed in section 3.4, there has been progress recently in integrating some of the sources of information listed above. Yim et al. [2006] and Thambirajah et al. [2009] used electrical process schematics to automatically extract process connectivity, and then combine this information with results from analyses of process measurements. Schleburg et al. [2013] have combined the process connectivity information, with rules and alarm logs for alarm management. This thesis has also contributed with advanced signal analysis methods which overcome some of the challenges in analysing together process, electrical and mechanical measurements.

Based on these contributions, a possible research direction is to develop a tool which integrates different sources of information in a systematic manner which is compatible with large-scale system. The value of the integrated analysis has been recognized by the ARC Advisory Group in a report for a process analytics client [Fiske, 2009]. The systematisation and automation of the integrated analysis should help the experts responsible for analysing the plant operation. This section suggests a systematic approach to integrate the analysis of different sources of information in large-scale systems, and defines functional specifications for a semi-automated tool which implements the proposed approach.

9.3.1 Background

Process monitoring and diagnosis in large-scale systems

Large-scale systems represent a challenge in any domain dealing with data analysis and visualisation. In process monitoring and diagnosis, the contributions in the literature can be divided into reducing the large set of data to a smaller set, and separating the measurements into conceptually meaningful blocks [Qin, 2012]. These two approaches were introduced in section 1.3 and are now further discussed.

The reduction of large sets of data to smaller sets uses the dependencies between measurements. In chemical processes, these dependencies result from the physical and chemical principles governing the process operation, such

as mass and energy balances. Some of the most popular techniques for dimensionality reduction are Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Partial Least Squares (PLS), which were discussed in section 3.6.1. Using this approach, Maurya et al. [2005] reduced by 40% the computational complexity in qualitative trend analysis by exploiting with PCA the redundancy of the measurements. Thornhill et al. [2002] used the scores of spectral PCA to extract smaller clusters of measurements having the same disturbance.

Wold et al. [1996] introduced the approach of separating the measurements into conceptually meaningful blocks. These authors proposed algorithms such as consensus PCA, hierarchical PCA, multiblock PLS, and hierarchical PLS. Westerhuis et al. [1998] and Qin et al. [2001] demonstrated that multiblock data-based models can be calculated directly using the standard PCA and PLS techniques on each block. The success of this approach depends on the appropriate separation the measurements, which requires knowledge of the process or access to the process diagrams. Qin et al. [2001] separated the measurements according to process units or a single process operation.

The systematic approach suggested in this section differs from the data reduction and data blocking approaches. The approach suggested proceeds by steps, and at each step a different reduced set of measurements is analysed. With each step the focus of analysis moves along the process towards the root-cause.

Differences between sources of information

Each of the sources of information used in process monitoring and diagnosis has unique characteristics. The dissimilarities between sources of information need to be understood in order to outline the technical challenges of their integrated analysis. Section 1.2 discussed specifically the dissimilarities between process, electrical and mechanical measurements, which include differences in dynamics, sampling rate and time synchronisation. When integrating additionally event logs and process schematics, differences in format, databases, and diagrams should also be considered. These are detailed below.

- **Format:** measurements from process, utility, mechanical and electrical systems are numerical and continuous if pertaining to physical variables; for on-off variables, such as valve openings, measurements are generally

numerical but binary. Alarms are held in text format, and diagrams are most commonly supported in paper or pdf files.

- Dynamics: process and utility systems are mostly dependent on mass and heat phenomena with time constants in the order of minutes. On the contrary, the dynamics of rotating machinery are determined by the rotational speed, in the order of 1 rotation per second or faster; similarly the duration of events in electrical systems is typically below 1 second.
- Sampling rate: consistent with the systems dynamics, process and utility measurements are sampled at a slower rate than mechanical and electrical measurements. Mechanical and electrical measurements may be available at the same sampling rate as process measurements but, in such cases, the dynamic events are usually not captured.
- Databases: it is common that different engineering domains in the plant have technology supplied by different vendors or different business units in the same company, which leads to data being stored in different databases.
- Time synchronisation: different databases may have different clocks, raising a problem of time synchronisation between different sources of data and alarms.
- Diagrams: tags from different systems and equipment are generally figured in different diagrams, which poses practical problems of localising and handling several diagrams.

9.3.2 Idea for a systematic approach to integrated analysis

This subsection explains the systematic approach to integrated analysis using a case study from the *Gas plant* set of data and schematics. The *Gas plant* system can be classified as large-scale since it is formed by 35 areas, described by more than 300 schematics, and monitored by measurements from more than 7000 variables. The motivation for the integrated analysis is an unplanned process shut-down resulting from the propagation of disturbances through a series of areas of the system. The aim is to find the closest possible root-cause.

The systematic approach to the root-cause analysis follows the structure of Table 9.3. The flow of analysis proceeds by steps, where each step represents

an envelope of analysis. The envelope refers to the reduced set of measurements which is being analysed. Each step is represented by one row in Table 9.3. With each step the envelope of analysis moves along the system towards the root-cause.

The columns in Table 9.3 distinguish the sources of information used within each envelope. The flow of analysis proceeds from the event log, to the process schematics, to the measurement data. In Table 9.3, this flow of analysis is structured from the left column to the right column.

As an example, the first row of the table should be read as follows. To start investigating the unplanned shut-down at 11:19, the alarm log is searched around that time and a high pressure alarm in system IS is found to be responsible for triggering the shut-down. The safety instrument associated with this alarm is the first envelope of analysis. Therefore, the instrument is localised in a diagram to understand its location, and the corresponding measurements are analysed around that time. The analysis and conclusions drawn from this envelope are indicated in the table. The conclusions lead to a new envelope, in the second row, and the description of the integrated analysis goes forth in this manner.

A group of tasks involved in this systematic approach is searching the information sources for a particular alarm, schematic or measurement. Another group of tasks is the use of connectivity information and process knowledge to analyse the measurements and alarms. The approach is still manual but the case study shows that these tasks lend themselves to automation. Therefore, the next subsection presents functional requirements to this end.

Table 9.3: Root cause diagnosis with integrated analysis applied to an industrial case study.

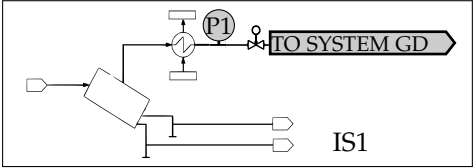
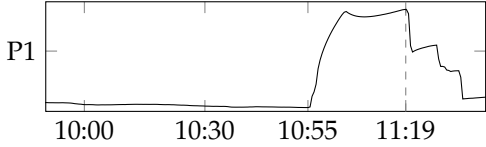
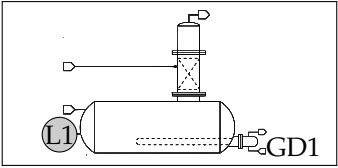
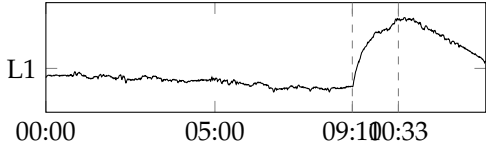
Alarms	Diagrams	Measurements
<p>Step 1 At 11:19 the alarm log reports a high-pressure shut-down order by instrument P1-IS</p>	 <p style="text-align: center;">IS1</p>	 <p style="text-align: center;">time</p>
<p>P1-IS is localised in drawing IS1; it is situated at the outlet of system IS, immediately before system GD.</p>		
<p>Step 2 At 10:55 the alarm log reports a high-level trip of system GD (L1-GD). As seen in IS1, GD is immediately downstream of the previous envelope; when GD trips, the valve at its inlet (outlet of IS1) is closed.</p>	 <p style="text-align: center;">GD1</p>	 <p style="text-align: center;">time</p>
<p>The envelope of analysis is now around L1-GD. A second drawing shows that L1-GD indicates the level of a reboiler.</p>		

Table 9.3: Root cause diagnosis with integrated analysis applied to an industrial case study.

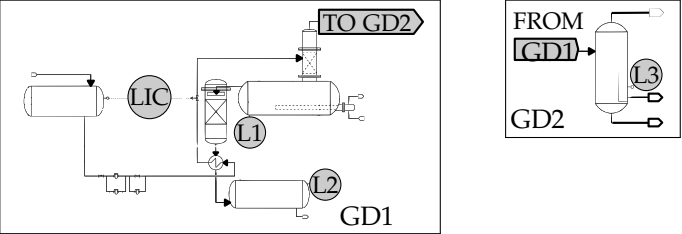
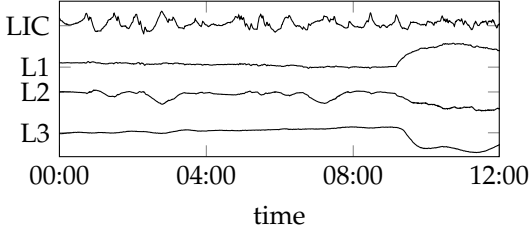
Alarms	Diagrams	Measurements
<p>Step 3 No records of alarms before this point because main alarm log is overwritten periodically.</p>	 <p>The envelope of analysis is expanded upstream and downstream of the reboiler, along the process streams. A third diagram, GD2, is needed to visualise one of the downstream directions. The level tags in this new envelope are selected.</p>	 <p>The level disturbance was also present in the equipment downstream of the reboiler. However, while in the reboiler the level increased (L1-GD), downstream it decreased (L2-GD, L3-GD). This suggests a reduction of the flow out of the reboiler.</p>

Table 9.3: Root cause diagnosis with integrated analysis applied to an industrial case study.

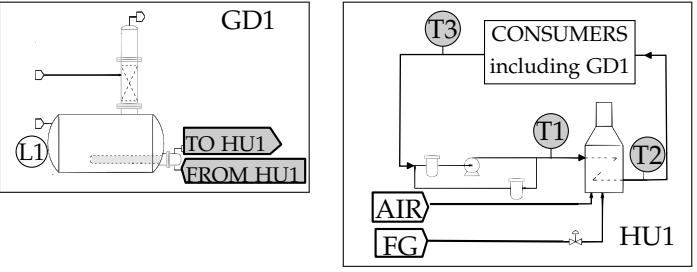
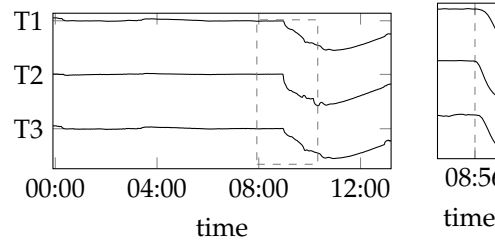
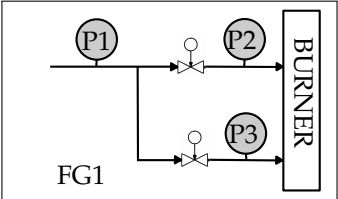
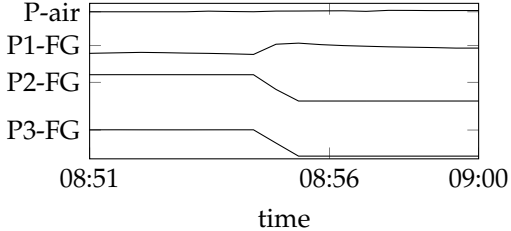
Alarms	Diagrams	Measurements
Step 4	 <p data-bbox="638 734 1355 893">Diagram GD1 is searched for the elements that can influence the flow between the reboiler and the equipment downstream. The heating utility is identified, and the relevant temperature tags are selected from the corresponding diagram, HU1.</p>	 <p data-bbox="1377 710 1886 933">The temperature of the heating utility suffered a decrease throughout, which can explain the reduction of gas flow from the reboiler. The time delays between the start of the disturbance suggest that it entered this system between T1-HU and T2-HU, in the burner.</p>

Table 9.3: Root cause diagnosis with integrated analysis applied to an industrial case study.

Alarms	Diagrams	Measurements
Step 5		
Step 6	<p>Diagram HU1 (row above) reveals the burner is fed by a fuel gas stream (FG) and by air. Thus, the envelope of analysis is expanded to these two systems. Pressure tags upstream of the burner in both systems are selected from the corresponding diagrams (above, only a part of the fuel gas system is shown).</p> <p>Diagram FG 1 (row above) is searched for the elements that can influence the flow between P1 and the next sensors, P2 and P3. Two valves are identified and the actuator tags are selected.</p>	<p>The pressure of air (P_{air}) shows no disturbance, whereas the inlet pressures of fuel gas (P2 and P3) drop. The reduced supply of gas can explain the decrease of the burner outlet temperature. Moreover, it can be seen that the behaviour of the fuel gas pressure is different along the supply line; up to P1 the pressure increases and beyond that it decreases. This indicates the flow of gas was stopped after P1.</p> <p>The data for the actuator tags is not present in this dataset. The conclusion of the analysis is that attention should be focused on why valves were closed.</p>

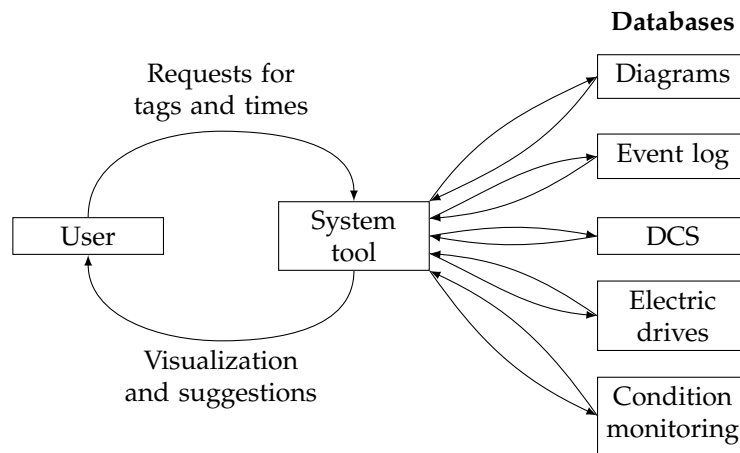


Figure 9.12: Context diagram of the proposed tool representing its interactions with the main external entities.

9.3.3 Functional specifications for semi-automated tool

One of the reasons to automate the systematic analysis described before is to liberate the expert from the tasks related to searching the information sources. These tasks are time-consuming and do not add value to the analysis. The other reason is to implement the structured approach of navigating the system towards the root-cause.

In brief, a semi-automated tool should perform certain operations on the selected envelope of measurements and time. The outputs of these operations should be the efficient visualisation of measurement trends, alarms and diagrams, and suggestions for the next envelope of analysis. The decision for the next envelope should remain with the user because it involves process knowledge. This way the users can use their insights to evaluate the suggestions given, and be confident with the final outcomes.

Figure 9.12 shows a context diagram of the tool. A context diagram represents the relationships between the tool, in the centre, and the main external entities. As shown, the tool should interact with the various information sources to search for particular tags. It should also interact with the user, who decides the envelope of analysis, and receives visualisation and suggestions.

The tool should be appropriate for users both with and without experience with the plant and the flow of analysis. The experienced user may only use level 1 functions, while the inexperienced user should also require level 2 functions. Level 1 functions mainly concern the interaction with the databases

and visualisation. Level 2 functions analyse together the different sources of information in order to provide suggestions.

Level 1 functional requirements

Level 1 functions should facilitate the interaction between the user and the databases, by attaining and presenting the data and diagrams corresponding to the measurements and times selected by the user. The efficient visualisation of measurement data may require data processing functions. Specifically, the tool should:

- produce the diagram from the tag name and highlight the queried tag in the diagram,
- produce the diagram from part of its name,
- allow the selection of one or more tags in diagrams, to bring out the corresponding data, alarms or diagrams,
- produce time trends from the tags name and time instant,
- allow simultaneous visualisation of multiple time histories,
- offer different techniques of visualising time histories,
- produce a scrollable alarm list centred around a queried time instant, and
- allow the selection of an alarm tag, to bring out the corresponding data or diagram.

Level 2 functional requirements

Level 2 functions should combine the analysis of measurements and alarms with connectivity information and some rules based on first principles in order to suggest the most likely envelope to approach the root-cause. Specifically, the tool should:

- extract process connectivity from the diagrams, distinguishing the different connections, namely mass, heat, momentum and information,
- find the alarm tag(s) closest to a requested time instant and a requested tag, using the connectivity information and information about time,

- within the time window being analysed, detect the time instant t when a measurement suffered a transient disturbance, and use the output $\{t, \text{tag}\}$ to search the alarm list,
- extract precedence relationships between the measurements in the envelope using advanced signal analysis to determine the propagation path,
- compare the disturbance propagation path with the connectivity information to determine if the disturbance entered the envelope of analysis between two measurements in the envelope, named boundary measurements,
- if so, use the connectivity information to find the external influences (mass, heat, momentum and information) between the two boundary measurements, and extract the name of the first measurement in those other systems,
- if not, or if an alarm search gives no results, expand the set of analysed measurements upstream and downstream of the current set along mass streams.

9.3.4 Comments on the proposed research direction

Section 9.3 discussed as a future research direction the development of a systematic approach to analyse large-scale system, and the implementation of that systematic approach with a semi-automated tool. Large-scale systems are more difficult to visualize, require significant computational memory and speed, and have a higher number of complex dependencies amongst measurements due to recycles, energy integration and control paths. Therefore, this section suggested a flow of analysis which proceeds by steps, and at each step a different reduced set of measurements is analysed. With each step the focus of analysis moves along the system towards the root-cause.

This systematic approach should integrate the complementary information captured by different sources, such as schematics, process and electromechanical measurements. The reason why all these sources are informative is that the behaviour of a process results from a combination of factors, namely the process conditions, control tuning and structure, the condition of equipment and utility systems, and the actions of operators.

The systematic approach to integrated analysis was illustrated in an industrial plant which has over 7000 measurement tags. The outcome of the analysis was that maintenance attention could be focused on a particular area. The example also illustrated the advantage of integrated analysis when necessary data is not available or cannot be found. From step 3 onwards, there were no records of alarms because the main alarm log is overwritten periodically. Nonetheless, the analysis could proceed thanks to the process measurements and the schematics. Therefore, the integration of different sources can help by providing redundancy and robustness to the decision-making process.

The illustration of the systematic approach to integrated analysis suggests that this could be a relevant direction for research. The approach is still manual, therefore this section also defined functional requirements for a semi-automated tool. Some of the required functions for the tool belong to the domain of informatics and are already common in commercial software, for example, clicking on a name on a figure and producing some information related to that entity.

Some of the functions related to the analysis of information are also available. This thesis contributed to some of these functions, namely the detection of transient disturbances in process, electrical and mechanical measurements (Chapter 4 to Chapter 6). Section 9.1 also presented ideas towards determining precedence relationships between measurements with different sampling rates. Furthermore, section 3.4 reviewed other contributions to the determination of precedence relationships, and to the extraction of connectivity from process schematics. A strength of the proposed approach and tool is that each function can be achieved with more than one method.

9.4 Chapter summary

The current chapter discussed three directions for future research, namely:

- the identification of the propagation path of a persistent disturbance when process and electromechanical measurements have different sampling rates,
- the detection and diagnosis of disturbances in measurements with irregular sampling rates, and
- the detection and diagnosis of disturbances in large-scale processes.

These directions had been identified in Chapter 1 as challenges which result from the extension of process monitoring and diagnosis to the electrical and mechanical utilities. Therefore, research along these directions will contribute to that extension, but also to monitoring and diagnosis in general. The reason is that the challenges of different sampling rates, irregular sampling rates, and large-scale systems also affect systems other than the process, electrical and mechanical systems.

Section 9.1 suggested a multi-rate method to identify the propagation path of a persistent disturbance when process and electromechanical measurements have different sampling rates. The suggested method is an adaptation of the uni-rate method developed in Bauer et al. [2007a]. Using a case study, this section showed that the multi-rate method can achieve results in a multi-rate data set which are comparable to the results of the uni-rate method applied to process and electromechanical measurements all with the fast sampling rate. Both methods achieved the identification of the correct propagation path of the disturbance.

Section 9.2 suggested an alternative formulation for embedded vectors and a weighted distance metric to overcome the fact that, in measurements with irregular sampling rates, the conventional construction of embedded vectors and similarity measure cannot be applied. Construction of embedded vectors and similarity assessment constitute the first two steps of all methods developed in this thesis. Using a case study, this section demonstrated the new formulation with the method of univariate detection of transient disturbances. The initial analysis showed that the new formulation can achieve results in an irregularly sampled time series comparable to the results in regularly sampled time series.

Section 9.3 suggested a systematic approach to analyse large-scale system, and the implementation of that systematic approach with a semi-automated tool. This systematic approach integrates the complementary information captured by different sources, such as schematics, process and electromechanical measurements. The approach was illustrated in a case study from an industrial plant with more than 7000 measurement tags. The outcome of the analysis was that maintenance attention could be focused on a particular area. This section also defined functional requirements for the semi-automated tool that implements the systematic approach. This thesis contributed to some of the functions which the tool should perform.

Chapter 10

Conclusion

This thesis set out to extend process monitoring and diagnosis to measurements from the electrical and mechanical utilities, and has documented the work done to achieve that goal.

The thesis supported its proposition with convincing cases, and clearly defined the new developments required to extend process monitoring and diagnosis to the electrical and mechanical utilities. It presented comprehensive surveys of methods in process monitoring and diagnosis as well as other fields such as computer science and statistics, and analysed the suitability of each method for the proposed extension.

Having defined its roadmap and identified the most suitable tools, the thesis developed four new methods to analyse transient disturbances and measurements with different sampling rates. The performances of the methods were explored in depth and optimised for a wide range of cases. Additionally, three ideas for future research were suggested in order to address challenges not solved in the thesis, such as measurements with irregular sampling rates and large-scale systems. The future research ideas were supported by surveys on their background and context, as well as demonstrations with promising preliminary results. The fully-developed methods and the ideas for future research were all successfully demonstrated in experimental and industrial case studies.

The work in the thesis has led to two submissions for journal papers, two published conference papers, three technical reports documenting industrial secondments, as well as a conference paper and a software toolbox from work done in parallel to the thesis.

Analyses of transient disturbances, measurements with different or irregular sampling rates, and large-scale systems are relevant far beyond process, electrical and mechanical systems. Therefore, this thesis has successfully contributed to extending the field of process monitoring and diagnosis to the electrical and mechanical utilities, and also to extending the whole field of monitoring and diagnosis.

References

- ABB. System 800xA AC 800M Control and I/O Overview. [http://www05.abb.com/global/scot/scot343.nsf/veritydisplay/f0ee41263daf4cbfc12579df0054d998/\\$file/3BSE047351_F_en_System_800xA_5.1_AC_800M_Control_and_I_O_Overview.pdf](http://www05.abb.com/global/scot/scot343.nsf/veritydisplay/f0ee41263daf4cbfc12579df0054d998/$file/3BSE047351_F_en_System_800xA_5.1_AC_800M_Control_and_I_O_Overview.pdf), 2012. [Online; accessed 09/01/14].
- Chrissanthi Angeli. Online expert systems for fault diagnosis in technical processes. *Expert Systems*, 25(2):115–132, 2008.
- Fabrizio Angiulli and Clara Pizzuti. Fast outlier detection in high dimensional spaces. In *Proceedings of the 6th European Conference on Principles of Data Mining and Knowledge Discovery*, pages 15–27, Helsinki, 2002. Springer Berlin Heidelberg.
- Jochen Arnhold, Peter Grassberger, Klaus Lehnertz, and Christian E. Elger. A robust method for detecting interdependences: application to intracranially recorded EEG. *Physica D: Nonlinear Phenomena*, 134(4):419–430, 1999.
- Srinivasan Babji and Raghunathan Rengaswamy. Automatic oscillation detection and characterization in closed-loop systems. *Control Engineering Practice*, 20(8):733–746, 2012.
- Srinivasan Babji and Arun K. Tangirala. Source separation in systems with correlated sources using NMF. *Digital Signal Processing*, 20(2):417–432, 2010.
- Bhavik R. Bakshi. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal*, 44(7):1596–1610, 1998.
- Bhavik R. Bakshi and George Stephanopoulos. Wave-net: a multiresolution, hierarchical neural network with localized learning. *AIChE Journal*, 39(1): 57–81, 1993.

- Bhavik R. Bakshi and George Stephanopoulos. Representation of process trends – III. Multiscale extraction of trends from process data. *Computers & Chemical Engineering*, 18(4):267–302, 1994.
- P. P. Bansal and Alan J. Ardell. Average nearest-neighbor distances between uniformly distributed finite particles. *Metallography*, 5(2):97 – 111, 1972.
- Mike Barth and Alexander Fay. Automated generation of simulation models for control code tests. *Control Engineering Practice*, 21(2):218 – 230, 2013.
- Michèle Basseville. Detecting changes in signals and systems – A survey. *Automatica*, 24(3):309–326, 1988.
- Margret Bauer and Nina F. Thornhill. A practical method for identifying the propagation path of plant-wide disturbances. *Journal of Process Control*, 18(7–8):707–719, 2008.
- Margret Bauer, John W. Cox, Michelle H. Caveness, James J. Downs, and Nina F. Thornhill. Nearest neighbors methods for root cause analysis of plantwide disturbances. *Industrial & Engineering Chemistry Research*, 46(18):5977–5984, 2007a.
- Margret Bauer, John W. Cox, Michelle H. Caveness, James J. Downs, and Nina F. Thornhill. Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Transactions on Control Systems Technology*, 15(1):12–21, 2007b.
- Ashish Bendre, Deepak Divan, William Kranz, and William Brumsickle. Equipment failures caused by power quality disturbances. In *Conference Record of the 2004 IEEE Industry Applications Conference*, volume 1. IEEE, 2004.
- Hassan Bevrani. *Robust Power System Frequency Control*. Springer, 2009.
- Mani Bhushan and Raghunathan Rengaswamy. Design of sensor network based on the signed directed graph of the process for efficient fault diagnosis. *Industrial & Engineering Chemistry Research*, 39(4):999–1019, 2000.
- L.A. Biber, Yu Ya Lubarskii, V.M. Nadtochii, and A.S. Tseitlin. Expert system for vibration control of a hydroelectric generator. *Electrical Technology*, 3:45–52, 1990.

- Jeff A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 4(510):126, 1998.
- Math H. J. Bollen. *Understanding Power Quality Problems: Voltage Sags and Interruptions*. Wiley-IEEE Press, 2000.
- Robert Bos, Stijn de Waele, and Piet M. T. Broersen. Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data. *IEEE Transactions on Instrumentation and Measurement*, 51(6):1289–1294, 2002.
- Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. LOF: Identifying density-based local outliers. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 93–104, Dallas, 2000. ACM.
- Murray Bullis. *Book of SEMI Standards*. Semiconductor Equipment and Materials International, Mountain View, CA, 1996.
- Inês M. Cecílio. Electrical interactions with process systems. WP8.1 – Use of causal signal analysis – Part I. Technical report, Project REAL-SMART Secondment Report, 2011.
- Inês M. Cecílio. Electrical interactions with process systems. WP8.1 – Use of causal signal analysis – Part II. Technical report, Project REAL-SMART Secondment Report, 2012.
- Inês M. Cecílio. Detecting and diagnosing disturbances in natural gas processes with signal analysis. In *SET for Britain national competition, Parliamentary and Scientific Committee, London UK*, March 2013a.
- Inês M. Cecílio. Why is the heating off? – Finding the causes of disturbances in natural gas processes. In *Chemical Engineering PhD Research Symposium*, July 2013b.
- Inês M. Cecílio. Electrical interactions with process systems. WP8.1 – Use of causal signal analysis – Part II.2. Technical report, Project REAL-SMART Secondment Report, 2013.
- Inês M. Cecílio. FRECOL project. <http://www.frecol.co.uk>, 2013. [Online; accessed June 2013].

- Inês M. Cecílio and Moncef Chioua. Informatics in process operations – Academic algorithms, practical applications. In *CPSE Industrial Consortium Meeting*, June 2012.
- Inês M. Cecílio and Nina F. Thornhill. Data fusion in support of operational decision-making. In *CPSE Industrial Consortium Meeting*, April 2010.
- Inês M. Cecílio and Nina F. Thornhill. Improving process performance looking at the whole picture. In *CPSE Industrial Consortium Meeting*, December 2011.
- Inês M. Cecílio, Su-Liang Chen, and Nina F. Thornhill. Importance of auxiliary systems for process fault detection and diagnosis. In *Proceedings of the 19th Mediterranean Conference on Control & Automation (MED)*, pages 952–957, Corfu, 2011. IEEE.
- Inês M. Cecílio, Knut Rapp, and Nina F. Thornhill. Process performance analysis in large-scale systems integrating different sources of information. In *Proceedings of the 8th IFAC International Symposium on Advanced Control of Chemical Processes*, pages 45–50, Singapore, 2012.
- Inês M. Cecílio, Anne Mai Ersdal, Davide Fabozzi, and Nina F. Thornhill. An open-source educational toolbox for power system frequency control tuning and optimization. In *Proceedings of the 4th IEEE European Innovative Smart Grid Technologies*, Lyngby, 2013.
- Inês M. Cecílio, James R. Ottewill, John Pretlove, and Nina F. Thornhill. Nearest neighbors method for detecting transient disturbances in process and electromechanical systems. Submitted to. *Journal of Process Control*, 2013. [Preprint].
- Inês M. Cecílio, James R. Ottewill, Harald Fretheim, and Nina F. Thornhill. Multivariate detection of transient disturbances for uni- and multi-rate systems combining nearest neighbors methods and SVD. Submitted to. *IEEE Transactions on Control System Technology*, 2014. [Preprint].
- Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):1–58, 2009.
- Subrahmanyan Chandrasekhar. Stochastic problems in physics and astronomy. *Reviews of Modern Physics*, 15(1):1, 1943.

- Sylvie Charbonnier and François Portet. A self-tuning adaptive trend extraction method for process monitoring and diagnosis. *Journal of Process Control*, 22(6):1127–1138, 2012.
- Jiahua Chen and Jun Shao. Nearest neighbor imputation for survey data. *Journal of Official Statistics*, 16(2):113–132, 2000.
- Jie Chen, Guozhu Zhang, and Zhiping Li. On-line parameter estimation for a class of time-varying continuous systems with bounded disturbances. *International Journal of Adaptive Control and Signal Processing*, 25(1):18–32, 2011.
- Tao Chen and Jie Zhang. On-line multivariate statistical monitoring of batch processes using Gaussian mixture model. *Computers & Chemical Engineering*, 34(4):500–507, 2010.
- Jarvis T. Cheung and George Stephanopoulos. Representation of process trends – Part I. A formal representation framework. *Computers & Chemical Engineering*, 14(4–5):495–510, 1990.
- Saneej B. Chitralkha, Sirish L. Shah, and Jagadeesan Prakash. Detection and quantification of valve stiction by the method of unknown input estimation. *Journal of Process Control*, 20(2):206–216, 2010.
- M. A. A. Shoukat Choudhury, Sirish L. Shah, and Nina F. Thornhill. Diagnosis of poor control-loop performance using higher-order statistics. *Automatica*, 40(10):1719–1728, 2004.
- M. A. A. Shoukat Choudhury, Sirish L. Shah, Nina F. Thornhill, and David S. Shook. Automatic detection and quantification of stiction in control valves. *Control Engineering Practice*, 14(12):1395–1412, 2006.
- MooiChoo Chuah and Fen Fu. ECG anomaly detection via time series analysis. In *Proceedings of the ISPA International Workshops on Frontiers of High Performance Computing and Networking*, pages 123–135, Niagara Falls, 2007. Springer Berlin Heidelberg.
- James R. Couper, W. Roy Penney, James R. Fair, and Stanley M. Walas. *Chemical Process Equipment, Second Edition: Selection and Design*. Gulf Professional Publishing, 2005.
- Johan De Kleer and John Seely Brown. A qualitative physics based on confluences. *Artificial Intelligence*, 24(1):7–83, 1984.

- Lane Desborough and Randy Miller. Increasing customer value of industrial control performance monitoring – Honeywell’s experience. In *AICHE symposium series*, pages 169–189, New York, 2002. AIChE.
- Anjali P. Deshpande and Sachin C. Patwardhan. Online fault diagnosis in nonlinear systems using the multiple operating regime approach. *Industrial & Engineering Chemistry Research*, 47(17):6711–6726, 2008.
- Håvard Devold. Oil and gas production handbook-an introduction to oil and gas production, May 2009.
- Feng Ding and Tongwen Chen. Hierarchical identification of lifted state-space models for general dual-rate systems. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 52(6):1179–1187, 2005.
- Feng Ding and Jie Ding. Least-squares parameter estimation for systems with irregularly missing data. *International Journal of Adaptive Control and Signal Processing*, 24(7):540–553, 2010.
- Dong Dong and Thomas J. McAvoy. Nonlinear principal component analysis – based on principal curves and neural networks. *Computers & Chemical Engineering*, 20(1):65–78, 1996.
- Richard C. Dorf and James A. Svoboda. *Introduction to electric circuits*. Wiley, 2010.
- Ping Duan, Fan Yang, Tongwen Chen, and Sirish L. Shah. Detection of direct causality based on process data. In *American Control Conference (ACC), 2012*, pages 3522–3527. IEEE, 2012.
- Norbert Edomah. Effects of voltage sags, swell and other disturbances on electrical equipment and their economic implications. In *20th International Conference and Exhibition on Electricity Distribution-Part 1*, pages 1–4. IET, 2009.
- Sami El-Ferik, Mohammed Nadeemullah Shareef, and Lahoucine Ettaleb. Detection and diagnosis of plant-wide oscillations using ga based factorization. *Journal of Process Control*, 22(1):321–329, 2012.
- ENTSO-E. Technical features of automatic FRR in interim period. In *Workshop Nordic Automatic FRR*, 2012a. [Online; accessed May 2013].

- ENTSO-E. Operational Reserve Ad Hoc Team Report. https://www.entsoe.eu/fileadmin/user_upload/_library/resources/LCFR/2012-06-14_S0C-AhT-OR_Report_final_V9-3.pdf, 2012b. [Online; accessed May 2013].
- Anne Mai Ersdal, Inês M. Cecílio, Davide Fabozzi, Lars Imsland, and Nina F. Thornhill. Applying model predictive control to power system frequency control. In *Proceedings of the 4th IEEE European Innovative Smart Grid Technologies*, Lyngby, 2013.
- Reza Eslamloueyan. Designing a hierarchical neural network based on fuzzy clustering for fault diagnosis of the tennessee-eastman process. *Applied Soft Computing*, 11(1):1407–1415, 2011.
- European Commission. EN Standard 50160. *Voltage characteristics of electricity supplied by public distribution networks*, 2002.
- Dafydd Evans, Antonia J. Jones, and Wolfgang M. Schmidt. Asymptotic moments of near-neighbour distance distributions. *Proceedings: Mathematical, Physical and Engineering Sciences*, 458(2028):2839–2849, 2002.
- Davide Fabozzi and Thierry Van Cutsem. Assessing the proximity of time evolutions through dynamic time warping. *IET Generation, Transmission & Distribution*, 5(12):1268–1276, 2011.
- Davide Fabozzi, Nina F. Thornhill, and Bikash C. Pal. Frequency restoration reserve control scheme with participation of industrial loads. In *IEEE PowerTech Conference Proceedings*, pages 1–6, June 2013.
- Ute Feldmann and Joydeep Bhattacharya. Predictability improvement as an asymmetrical measure of interdependence in bivariate time series. *International Journal of Bifurcation and Chaos*, 14(2):505–514, 2004.
- David Alan Fink, Nicholas A Cumpsty, and Edward M Greitzer. Surge dynamics in a free-spool centrifugal compressor system. *Journal of Turbomachinery*, 114(2):321–331, 1992.
- Tom Fiske. Aegis analytical delivers process intelligence from disparate data sources. *ARC Advisory Group*, <http://www.arcweb.com/strategy-reports/2009-03-13/>

- aegis-analytical-delivers-process-intelligence-from-disparate-data-sources-1.aspx:[Online; accessed 8th December 2011], March 2009.
- Paul M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. *Automatica*, 26(3):459–474, 1990.
- Dong Gao, Chongguang Wu, Beike Zhang, and Xin Ma. Signed directed graph and qualitative trend analysis based fault diagnosis in chemical industry. *Chinese Journal of Chemical Engineering*, 18(2):265–276, 2010.
- Zhiqiang Ge, Furong Gao, and Zhihuan Song. Batch process monitoring based on support vector data description method. *Journal of Process Control*, 21(6): 949–959, 2011.
- Zhiqiang Ge, Zhihuan Song, and Furong Gao. Review of recent research on data-based process monitoring. *Industrial & Engineering Chemistry Research*, 52(10):3543–3562, 2013.
- Janos Gertler. Analytical redundancy methods in fault detection and isolation. In *Proceedings of IFAC/IAMCS Symposium on Safe Processes*, volume 1, pages 9–21, 1991.
- Janos Gertler, Weihua Li, Yunbing Huang, and Thomas J. McAvoy. Isolation enhanced principal component analysis. *AIChE Journal*, 45(2):323–334, 1999.
- Sakti Prasad Ghoshal. Optimizations of PID gains by particle swarm optimizations in fuzzy based automatic generation control. *Electric Power Systems Research*, 72(3):203–212, 2004.
- Sebastian Gigi and Arun K. Tangirala. Quantitative analysis of directional strengths in jointly stationary linear multivariate processes. *Biological Cybernetics*, 103(2):119–133, 2010.
- Gene H. Golub and Charles F. Van Loan. *Matrix Computations*, volume 3 of *Johns Hopkins Studies in the Mathematical Sciences*. Johns Hopkins University Press, 2012.
- Ruben Gonzalez, Biao Huang, Fangwei Xu, and Aris Espejo. Dynamic Bayesian approach to gross error detection and compensation with application toward an oil sands process. *Chemical Engineering Science*, 67(1):44–56, 2012.

- Jose Maria González-Martínez, Johan A. Westerhuis, and Alberto Ferrer. Using warping information for batch process monitoring and fault classification. *Chemometrics and Intelligent Laboratory Systems*, 127:210 – 217, 2013.
- Jan Tommy Gravdahl and Olav Egeland. *Compressor surge and rotating stall: Modelling and control*. Springer-Verlag, 1999.
- Jan Tommy Gravdahl, Olav Egeland, and Svein Ove Vatland. Drive torque actuation in active surge control of centrifugal compressors. *Automatica*, 38 (11):1881–1893, 2002.
- Edward M Greitzer. Surge and rotating stall in axial flow compressors – Part I: Theoretical compression system model. *Journal of Engineering for Power*, 98:190, 1976.
- Allan R. Hambley. *Electrical Engineering: Principles and Applications (5th Edition)*. Prentice Hall, 2010.
- Paul C. Hanlon. *Compressor Handbook*. McGraw-Hill, 2001.
- Ville Hautamäki, Ismo Kärkkäinen, and Pasi Fränti. Outlier detection using k -nearest neighbour graph. In *Proceedings of the 17th International Conference on Pattern Recognition*, pages 430–433, Cambridge, 2004. IEEE.
- Q. Peter He, Jin Wang, Martin Pottmann, and S. Joe Qin. A curve fitting method for detecting valve stiction in oscillating control loops. *Industrial & Engineering Chemistry Research*, 46(13):4549–4560, 2007a.
- Qingbo He, Fanrang Kong, and Ruqiang Yan. Subspace-based gearbox condition monitoring by kernel principal component analysis. *Mechanical Systems and Signal Processing*, 21(4):1755–1772, 2007b.
- David M. Himmelblau. *Fault Detection and Diagnosis in Chemical and Petrochemical Processes*. Elsevier Scientific Pub. Co., Amsterdam, 1978.
- Hannele Holttinen. *The impact of large scale wind power production on the nordic electricity system*. PhD thesis, Helsinki University of Technology, 2004.
- Joachim Holtz. Sensorless control of induction motor drives. *Proceedings of the IEEE*, 90(8):1359–1394, 2002.
- Alexander Horch. A simple method for detection of stiction in control valves. *Control Engineering Practice*, 7(10):1221–1231, 1999.

- Alexander Horch. *Process Control Performance Assessment*, chapter Benchmarking control loops with oscillations and stiction, pages 227–257. In Ordys et al. [2007], 2007.
- Alexander Horch, John Cox, and Nunzio Bonavita. Peak performance – Root cause analysis of plant-wide disturbances. *ABB Review*, 1:24–29, 2007.
- Ye Hu and Nael H. El-Farra. Robust fault detection and monitoring of hybrid process systems with uncertain mode transitions. *AIChE Journal*, 57(10):2783–2794, 2011.
- Biao Huang. Bayesian methods for control loop monitoring and diagnosis. *Journal of Process Control*, 18(9):829–838, 2008.
- Biao Huang and Sirish L. Shah. *Performance assessment of control loops: Theory and Applications*. Advances in Industrial Control. Springer-Verlag, London, 1999.
- Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics, Second Edition*. John Wiley & Sons Inc., 2009.
- Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430, 2000.
- IEEE. IEEE Standard 1159. *IEEE Recommended Practice for Monitoring Electric Power Quality*, 1995.
- Masao Iri, Katsuaki Aoki, Eiji O’Shima, and H. Matsuyama. An algorithm for diagnosis of system failures in the chemical process. *Computers & Chemical Engineering*, 3(1-4):489–493, 1979.
- Alf J. Isaksson. Identification of ARX – models subject to missing data. *IEEE Transactions on Automatic Control*, 38(5):813–819, 1993.
- Rolf Isermann. Process fault detection based on modeling and estimation methods – A survey. *Automatica*, 20(4):387–404, 1984.
- Rolf Isermann. Fault diagnosis of machines via parameter estimation and knowledge processing – Tutorial paper. *Automatica*, 29(4):815–835, July 1993.
- Rolf Isermann. Model-based fault-detection and diagnosis – Status and applications. *Annual Reviews in Control*, 29(1):71–85, 2005.

- Margaret E. Janusz and Venkat Venkatasubramanian. Automatic generation of qualitative descriptions of process trends for fault detection and diagnosis. *Engineering Applications of Artificial Intelligence*, 4(5):329–339, 1991.
- Andrew K. S. Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.
- Mohieddine Jelali and Biao Huang, editors. *Detection and Diagnosis of Stiction in Control Loops: State of the Art and Advanced Methods*. Springer London, 2010.
- Mohieddine Jelali and Claudio Scali. *Detection and Diagnosis of Stiction in Control Loops: State of the Art and Advanced Methods*, chapter Comparative Study of Valve-stiction-detection Methods, pages 295–358. In Jelali and Huang [2010], 2010.
- F Jia, Elaine B. Martin, and Julian Morris. Non-linear principal components analysis with application to process fault detection. *International Journal of Systems Science*, 31(11):1473–1487, 2000.
- Hailei Jiang, M. A. A. Shoukat Choudhury, and Sirish L. Shah. Detection and diagnosis of plant-wide oscillations from industrial data using the spectral envelope method. *Journal of Process Control*, 17(2):143–155, 2007.
- Hailei Jiang, Rohit Patwardhan, and Sirish L. Shah. Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix. *Journal of Process Control*, 19(8):1347–1354, 2009.
- Taiwen Jiang, Bingzhen Chen, Xiaorong He, and Paul Stuart. Application of steady-state detection method based on wavelet transform. *Computers & Chemical Engineering*, 27(4):569–578, 2003.
- Ian Jolliffe. *Principal Component Analysis*. John Wiley & Sons, Ltd, 2005.
- Petr Kadlec, Bogdan Gabrys, and Sibylle Strandt. Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, 33(4):795–814, 2009.
- Henry F. Kaiser. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20:141–151, 1960.

- Manabu Kano, Shinji Hasebe, Iori Hashimoto, and Hiromu Ohno. Evolution of multivariate statistical process control: application of independent component analysis and external analysis. *Computers & Chemical Engineering*, 28(6-7):1157 – 1166, 2004.
- Holger Kantz and Thomas Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 2003.
- Srinivas Karra and M Nazmul Karim. Comprehensive methodology for detection and diagnosis of oscillatory control loops. *Control Engineering Practice*, 17(8):939–956, 2009.
- Matthew B. Kennel and Steven Isabelle. Method to distinguish possible chaos from colored noise and to determine embedding parameters. *Physical Review A*, 46(6):3111, 1992.
- Eamonn Keogh, Jessica Lin, and Ada Fu. HOT SAX: Efficiently finding the most unusual time series subsequence. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 226–233, Houston, 2005. IEEE.
- Shima Khatibisepehr, Biao Huang, and Swanand Khare. Design of inferential sensors in the process industry: A review of Bayesian methods. *Journal of Process Control*, 23(10):1575 – 1596, 2013.
- Edwin M. Knorr, Raymond T. Ng, and Vladimir Tucakov. Distance-based outliers: algorithms and applications. *The VLDB Journal*, 8(3–4):237–253, 2000.
- Teuvo Kohonen. *Self-organizing Maps*, volume 30. Springer, 2001.
- Theodora Kourti and John F. MacGregor. Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, 28(4), 1996.
- Mark A. Kramer and B. L. Palowitch. A rule-based approach to fault diagnosis using the signed directed graph. *AIChE Journal*, 33(7):1067–1078, 1987.
- James V. Kresta, John F. Macgregor, and Thomas E. Marlin. Multivariate statistical monitoring of process operating performance. *The Canadian Journal of Chemical Engineering*, 69(1):35–47, 1991.

- Wenfu Ku, Robert H. Storer, and Christos Georgakis. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30(1):179 – 196, 1995.
- Benjamin Kuipers. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge (Artificial Intelligence)*. The MIT Press, 1994.
- Upmanu Lall and Ashish Sharma. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32(3):679–693, 1996.
- Michel Le Van Quyen, Claude Adam, Michel Baulac, Jacques Martinerie, and Francisco J. Varela. Nonlinear interdependencies of EEG signals in human intracranially recorded temporal lobe seizures. *Brain Research*, 792(1):24–40, 1998.
- Jong-Min Lee, ChangKyoo Yoo, Sang Wook Choi, Peter A Vanrolleghem, and In-Beum Lee. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 59(1):223–234, 2004.
- Jong-Min Lee, S. Joe Qin, and In-Beum Lee. Fault detection and diagnosis based on modified independent component analysis. *AIChE Journal*, 52(10): 3501–3514, 2006.
- Thomas Martin Lehmann, Claudia Gonner, and Klaus Spitzer. Survey: Interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging*, 18(11):1049–1075, 1999.
- Stefaan Lhermitte, Jan Verbesselt, Willem W . Verstraeten, and Pol Coppin. A pixel based regeneration index using time series similarity and spatial context. *Photogrammetric Engineering and Remote Sensing*, 76(6):673–682, 2010.
- Dongguang Li, Sirish L. Shah, Tongwen Chen, and Kent Z. Qi. Application of dual-rate modeling to CCR octane quality inferential control. *IEEE Transactions on Control Systems Technology*, 11(1):43–51, 2003.
- Xinong Li, Jiandong Wang, Biao Huang, and Sien Lu. The DCT-based oscillation detection method for a single time series. *Journal of Process Control*, 20(5):609–617, 2010.
- Shu-Hsien Liao. Expert system methodologies and applications – A decade review from 1995 to 2004. *Expert Systems with Applications*, 28(1):93–103, 2005.

- Patrick L. Love and Marwan Simaan. Automatic recognition of primitive changes in manufacturing process signals. *Pattern Recognition*, 21(4):333–342, 1988.
- Jan Machowski, Janusz Bialek, and Jim Bumby. *Power system dynamics: stability and control*. Wiley, 2011.
- Richard S. H. Mah. *Chemical Process Structures and Information Flows*. Butterworth-Heinemann, 1990.
- Sankar Mahadevan and Sirish L. Shah. Fault detection and diagnosis in process data using one-class support vector machines. *Journal of Process Control*, 19(10):1627–1639, 2009.
- K. F. Martin. A review by discussion of condition monitoring and fault diagnosis in machine tools. *International Journal of Machine Tools and Manufacture*, 34(4):527–551, 1994.
- Toru Matsuo, Isao Tadakuma, and Nina F. Thornhill. Diagnosis of a unit-wide disturbance caused by saturation in a manipulated variable. In *IEEE Advanced Process Control Applications for Industry Workshop*, Vancouver, April 26–28 2004.
- Mano Ram Maurya, Raghunathan Rengaswamy, and Venkat Venkatasubramanian. A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis. *Industrial & Engineering Chemistry Research*, 42(20):4789–4810, 2003.
- Mano Ram Maurya, Raghunathan Rengaswamy, and Venkat Venkatasubramanian. Fault diagnosis by qualitative trend analysis of the principal components. *Chemical Engineering Research and Design*, 83(9):1122–1132, 2005.
- Mano Ram Maurya, Praveen K Paritosh, Raghunathan Rengaswamy, and Venkat Venkatasubramanian. A framework for on-line trend extraction and fault diagnosis. *Engineering Applications of Artificial Intelligence*, 23(6):950–960, 2010.
- Raghunathan Rengaswamy Maurya, Mano Ram and Venkat Venkatasubramanian. A signed directed graph and qualitative trend analysis-based framework for incipient fault diagnosis. *Chemical Engineering Research and Design*, 85(10):1407–1422, 2007.

- Manish Misra, H Henry Yue, S. Joe Qin, and Cheng Ling. Multivariate process monitoring and fault diagnosis by multi-scale PCA. *Computers & Chemical Engineering*, 26(9):1281–1293, 2002.
- Elham Naghoosi, Biao Huang, Elom Domlan, and Ramesh Kadali. Information transfer methods in causality analysis of process variables with an industrial application. *Journal of Process Control*, 23(9):1296–1305, 2013.
- Subhasis Nandi, Hamid A. Toliyat, and Xiaodong Li. Condition monitoring and fault diagnosis of electrical motors – A review. *IEEE Transactions on Energy Conversion*, 20(4):719–729, 2005.
- Abolfath Nikranjbar, Kambiz M. Ebrahimi, and Alastair S. Wood. Model-based fault diagnosis of induction motor eccentricity using particle swarm optimization. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 223(3):607–615, 2009.
- Paul Nomikos and John F. MacGregor. Monitoring batch processes using multiway principal component analysis. *AIChE Journal*, 40(8):1361–1375, 1994.
- Andrzej W. Ordys, Damien Uduehi, and Michael A. Johnson, editors. *Process Control Performance Assessment*. Springer, 2007.
- James R. Ottewill and Michal Orkisz. Condition monitoring of gearboxes using synchronously averaged electric motor signals. *Mechanical Systems and Signal Processing*, 38(2):482–498, 2013.
- Olayiwola O. Oyeleye and Mark A. Kramer. Qualitative simulation of chemical process systems: Steady-state analysis. *AIChE Journal*, 34(9):1441–1454, 1988.
- Girish Keshav Palshikar. Distance-based outliers in sequences. In *Proceedings of the Second International Conference on Distributed Computing and Internet Technology*, pages 547–552, Bhubaneswar, 2005. Springer Berlin Heidelberg.
- Sachin C. Patwardhan, Shankar Narasimhan, Prakash Jagadeesan, Bhushan Gopaluni, and Sirish L. Shah. Nonlinear Bayesian state estimation: A review of recent developments. *Control Engineering Practice*, 20(10):933 – 953, 2012.
- Perceptive Engineering. MonitorMV. <http://www.perceptiveapc.com/products/monitormv/>, 2014. [Online; accessed 25/01/14].

- Anna Pernestal. *A Bayesian Approach to Fault Isolation with Application To Diesel Engine Diagnosis*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 2007.
- Robert H. Perry and Don W. Green. *Perry's Chemical Engineers' Handbook*. McGraw-Hill Professional, 1997.
- Sylvia Pfeifer. Weather and low stocks add to gas worries. <http://www.ft.com>, October 2013a. [Online; accessed 09/01/14].
- Sylvia Pfeifer. Gas price spike underlines UK supply fears. <http://www.ft.com>, October 2013b. [Online; accessed 09/01/14].
- Dragoljub Pokrajac, Aleksandar Lazarevic, and Longin Jan Latecki. Incremental local outlier detection for data streams. In *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining*, pages 504–515, Honolulu, 2007. IEEE.
- Vinay Prasad, Matthias Schley, Louis P. Russo, and B. Wayne Bequette. Product property and production rate control of styrene polymerization. *Journal of Process Control*, 12(3):353–372, 2002.
- Fei Qi and Biao Huang. Bayesian methods for control loop diagnosis in the presence of temporal dependent evidences. *Automatica*, 47(7):1349–1356, 2011.
- Fei Qi, Biao Huang, and Edgar C. Tamayo. A Bayesian approach for control loop diagnosis with missing data. *AIChE Journal*, 56(1):179–195, 2010.
- S. Joe Qin. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2):220 – 234, 2012.
- S. Joe Qin and Jie Yu. Recent developments in multivariable controller performance monitoring. *Journal of Process Control*, 17(3):221–227, 2007.
- S. Joe Qin, Sergio Valle, and Michael J. Piovoso. On unifying multiblock analysis with application to decentralized process monitoring. *Journal of Chemometrics*, 15(9):715–742, 2001.
- Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. In *Proceedings of the ACM SIGMOD*

- International Conference on Management of Data*, pages 427–438, Dallas, 2000. ACM.
- REAL-SMART. Marie Curie FP7-IAPP project. <http://www3.imperial.ac.uk/realsmart>, 2011. [Online; accessed 20/10/13].
- Yann G. Rebours, Daniel S. Kirschen, Marc Trotignon, and Sébastien Rossignol. A survey of frequency and voltage control ancillary services-Part I: Technical features. *IEEE Transactions on Power Systems*, 22(1):350–357, 2007.
- Todd Reeves. Optimising process equipment performance. *Petroleum Technology Quarterly*, 10(4):93–97, 2005.
- Kira Rehfeld, Norbert Marwan, Jobst Heitzig, and Juergen Kurths. Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Processes in Geophysics*, 18(3):389–404, 2011.
- Raghunathan Rengaswamy, Tore Hägglund, and Venkat Venkatasubramanian. A qualitative shape analysis formalism for monitoring control loop performance. *Engineering Applications of Artificial Intelligence*, 14(1):23–33, 2001.
- Dulpichet Rerkpreedapong, Amer Hasanovic, and Ali Feliachi. Robust load frequency control using genetic algorithms and linear matrix inequalities. *IEEE Transactions on Power Systems*, 18(2):855–861, 2003.
- Carlos Rojas-Guzman and Mark A. Kramer. Comparison of belief networks and rule-based expert systems for fault diagnosis of chemical processes. *Engineering Applications of Artificial Intelligence*, 6(3):191–202, 1993.
- Peter J Rousseeuw and Annick M Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, Inc, 2003.
- Evan L. Russell, Leo H. Chiang, and Richard D. Braatz. *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes*. Springer, 1st edition, 2000.
- S Saksena, Baozhuang Shi, and George Karady. Effects of voltage sags on household loads. In *Power Engineering Society General Meeting*, pages 2456–2461. IEEE, 2005.

- Biswanath Samanta and Chandrasekhar Nataraj. Use of particle swarm optimization for machinery fault detection. *Engineering Applications of Artificial Intelligence*, 22(2):308–316, 2009.
- N Saravanan, V N S Siddabattuni, and K I Ramachandran. Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM). *Applied Soft Computing*, 10(1):344–360, 2010.
- Claudio Scali and Claudio Ghelardoni. An improved qualitative shape analysis technique for automatic detection of valve stiction in flow control loops. *Control Engineering Practice*, 16(12):1501–1508, 2008.
- Jeffrey D. Scargle. Studies in astronomical time series analysis. III. Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. *Astrophysical Journal*, 343:874–887, 1989.
- Steven J. Schiff, Paul So, Taeun Chang, Robert E. Burke, and Tim Sauer. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Physical Review E*, 54(6):6708, 1996.
- Richard Schiltz. Condition monitoring under control. *Plant Services*, 29(9):39 – 42, 2008.
- Markus Schlegel, Lars Christiansen, Nina F. Thornhill, and Alexander Fay. A combined analysis of plant connectivity and alarm logs to reduce the number of alerts in an automation system. *Journal of Process Control*, 23(6): 839 – 851, 2013.
- Till Schmidberger, Alexander Horch, Alexander Fay, and Rainer Drath. Rule based engineering of asset management system functionality. In *5th Vienna Symposium on Mathematical Modelling*, volume 8, 2006.
- Dale E. Seborg, Duncan A. Mellichamp, Thomas F. Edgar, and Francis J. Doyle III. *Process Dynamics and Control*. Wiley, 2010.
- Warren D. Seider. *Product and Process Design Principles: Synthesis, Analysis and Design, 3rd Edition*. Wiley, 2011.
- Xinguang Shao, Biao Huang, Jong Min Lee, Fangwei Xu, and Aris Espejo. Bayesian method for multirate data synthesis and model calibration. *AIChE Journal*, 57(6):1514–1525, 2011.

- Yuri A. W. Shardt, Yu Zhao, Kwan Ho Lee, Xinyi Yu, Biao Huang, and Sirish L. Shah. Determining the state of a process control system: Current trends and future challenges. *The Canadian Journal of Chemical Engineering*, 90(2): 217–245, 2012.
- Jie Sheng, Tongwen Chen, and Sirish L. Shah. Generalized predictive control for non-uniformly sampled systems. *Journal of Process Control*, 12(8):875–885, 2002.
- Siemens Energy & Automation, Inc. DCS or PLC? Seven Questions to Help You Select the Best Solution. http://www.automation.siemens.com/w2/efiles/pcs7/support/marktstudien/PLC_or_DCS.pdf, 2007. [Online; accessed 09/01/14].
- Harry Silla. *Chemical Process Engineering: Design And Economics (Chemical Industries)*. CRC Press, 2003.
- Silvio Simani and Ron J. Patton. Fault diagnosis of an industrial gas turbine prototype using a system identification approach. *Control Engineering Practice*, 16(7):769–786, 2008.
- Hua Song and Hongyue Zhang. An approach to sensor fault diagnosis based on fully-decoupled parity equation and parameter estimation. In *Proceedings of the 4th World Congress on Intelligent Control and Automation, Shanghai*, pages 2750–2754, 2002.
- Ranganathan Srinivasan, Raghunathan Rengaswamy, and Randy Miller. A modified empirical mode decomposition (EMD) process for oscillation characterization in control loops. *Control Engineering Practice*, 15(9):1135–1148, 2007.
- Statnett. Systemdrifts- og markedsutviklingsplan 2012. http://www.statnett.no/Documents/Kraftsystemet/Systemansvaret/Statnett_SMUP_24.05_lnk_Low.pdf, 2012. [Online; accessed 14/04/13].
- George Stephanopoulos. *Chemical Process Control: An Introduction to Theory and Practice*. PTR Prentice Hall, 1984.
- Markus Stockmann, Robert Haber, and Ulrich Schmitz. Source identification of plant-wide faults based on k nearest neighbor time delay estimation. *Journal of Process Control*, 22(3):583–598, 2012.

- Petre Stoica, Jian Li, and Hao He. Spectral analysis of nonuniformly sampled data: a new approach versus the periodogram. *IEEE Transactions on Signal Processing*, 57(3):843–858, 2009.
- Wim Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis*, 29(2):511–546, 1998.
- Arun K. Tangirala, Jitendra Kanodia, and Sirish L. Shah. Non-Negative matrix factorization for detection and diagnosis of plantwide oscillations. *Industrial & Engineering Chemistry Research*, 46(3):801–817, 2007.
- Jegatheeswaran Thambirajah, Lamia Benabbas, Margret Bauer, and Nina F. Thornhill. Cause-and-effect analysis in chemical processes utilizing XML, plant connectivity and quantitative process history. *Computers & Chemical Engineering*, 33(2):503–512, 2009.
- Nina F. Thornhill. Finding the source of nonlinearity in a process with plant-wide oscillation. *IEEE Transactions on Control Systems Technology*, 13(3):434–443, 2005.
- Nina F. Thornhill. *Process Control Performance Assessment*, chapter Locating the source of a disturbance, pages 199–225. In Ordys et al. [2007], 2007.
- Nina F. Thornhill and Tore Hägglund. Detection and diagnosis of oscillation in control loops. *Control Engineering Practice*, 5(10):1343–1354, 1997.
- Nina F. Thornhill and Alexander Horch. Advances and new directions in plant-wide disturbance detection and diagnosis. *Control Engineering Practice*, 15(10):1196–1206, 2007.
- Nina F. Thornhill, Sirish L. Shah, Biao Huang, and Anand Vishnubhotla. Spectral principal component analysis of dynamic process data. *Control Engineering Practice*, 10(8):833–846, 2002.
- Nina F. Thornhill, Biao Huang, and H. Zhang. Detection of multiple oscillations in control loops. *Journal of Process Control*, 13(1):91–100, 2003.
- Nina F. Thornhill, M. A. A. Shoukat Choudhury, and Sirish L. Shah. The impact of compression on data-driven process analyses. *Journal of Process Control*, 14(4):389–398, 2004.

- Nina F. Thornhill, Hallgeir Melbø, and Jan Wiik. Multidimensional visualization and clustering of historical process data. *Industrial & Engineering Chemistry Research*, 45(17):5971–5985, 2006.
- Martin Todd, Stephen D. J. McArthur, Jim R. McDonald, and Sally J. Shaw. A semiautomatic approach to deriving turbine generator diagnostic knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(5):979–992, 2007.
- Lloyd N. Trefethen and David Bau III. *Numerical Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, 1997.
- Umetrics. SIMCA – the standard in multivariate data analysis. <http://www.umetrics.com/products/simca>, 2014. [Online; accessed 25/01/14].
- Varanon Uraikul, Christine W. Chan, and Paitoon Tontiwachwuthikul. Artificial intelligence for monitoring and supervisory control of process systems. *Engineering Applications of Artificial Intelligence*, 20(2):115–131, 2007.
- Sergio Valle, Weihua Li, and S. Joe Qin. Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods. *Industrial & Engineering Chemistry Research*, 38(11):4389–4401, 1999.
- Vladimir Vapnik. *Statistical learning theory*. Wiley, New York, 1998.
- Peter Vas. *Sensorless vector and direct torque control*. Oxford University Press, Oxford New York, 1998.
- Venkat Venkatasubramanian. Inexact reasoning in expert systems: a stochastic parallel network approach. In *The Second Conference on Artificial Intelligence Applications: the engineering of knowledge-based systems*, page 13. IEEE Computer Society Press, 1985.
- Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, and Surya N. Kavuri. A review of process fault detection and diagnosis: Part I: quantitative model-based methods. *Computers & Chemical Engineering*, 27(3):293–311, 2003a.
- Venkat Venkatasubramanian, Raghunathan Rengaswamy, and Surya N. Kavuri. A review of process fault detection and diagnosis: Part II: qual-

- itative models and search strategies. *Computers & Chemical Engineering*, 27(3):313–326, 2003b.
- Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, and Kewen Yin. A review of process fault detection and diagnosis: Part III: process history based methods. *Computers & Chemical Engineering*, 27(3):327–346, 2003c.
- Kris Villez, Venkat Venkatasubramanian, and Raghunathan Rengaswamy. Generalized shape constrained spline fitting for qualitative analysis of trends. *Computers & Chemical Engineering*, 58:116–134, 2013.
- Paul Waide and Conrad U. Brunner. Energy-efficiency policy opportunities for electric motor-driven systems. Technical report, International Energy Agency, 2011.
- Jiandong Wang, Biao Huang, and Sien Lu. Improved DCT-based method for online detection of oscillations in univariate time series. *Control Engineering Practice*, 21(5):622–630, 2013.
- Eric W. Weisstein. *CRC Concise Encyclopedia of Mathematics*. CRC press, 2010.
- Johan A. Westerhuis, Theodora Kourti, and John F. MacGregor. Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics*, 12(5):301–321, 1998.
- G. Barrie Wetherill and Don W. Brown. *Statistical Process Control: Theory and Practice*. CRC Texts in Statistical Science. Chapman & Hall, third edition, 1991.
- David Whitley and Ole Gjerde. LFC/AGC - Nordic and european perspective. *Exchange of balancing services between the Nordic and the Central European synchronous systems*, pages 1–13, 2011.
- Achmad Widodo and Bo-Suk Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21(6):2560–2574, 2007.
- Achmad Widodo and Bo-Suk Yang. Wavelet support vector machine for induction machine fault diagnosis based on transient current signal. *Expert Systems with Applications*, 35(1):307–316, 2008.

- Maciej Wnek, Michal Orkisz, Jaroslaw Nowak, and Stefano Legnani. Drive-Monitor – Embedded product intelligence that enhances lifecycle management and performance in drive systems. *ABB Review*, 2:35–38, 2006.
- Svante Wold, Nouna Kettaneh, and Kjell Tjessem. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *Journal of Chemometrics*, 10(5-6):463–482, 1996.
- Chunming Xia, John Howell, and Nina F. Thornhill. Detecting and isolating multiple plant-wide oscillations via spectral independent component analysis. *Automatica*, 41(12):2067–2075, 2005.
- Yoshiyuki Yamashita. An automatic method for detection of valve stiction in process control loops. *Control Engineering Practice*, 14(5):503–510, 2006.
- Fan Yang and Deyun Xiao. Model and fault inference with the framework of probabilistic SDG. In *9th International Conference on Control, Automation, Robotics and Vision*, pages 1–6, 2006.
- Fan Yang and Deyun Xiao. Sensor location problem for large-scale complex systems in the framework of the dynamic signed directed graph. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 2843–2847. IEEE, 2007.
- Fan Yang and Deyun Xiao. Progress in root cause and fault propagation analysis of large-scale industrial processes. *Journal of Control Science and Engineering*, 2012:10, 2012.
- Fan Yang, Sirish L. Shah, Deyun Xiao, and Tongwen Chen. Improved correlation analysis and visualization of industrial alarm data. *ISA Transactions*, 51(4):499–506, 2012.
- Fan Yang, Deyun Xiao, and Sirish L. Shah. Signed directed graph-based hierarchical modelling and fault propagation analysis for large-scale systems. *IET Control Theory Applications*, 7(4):537–550, 2013.
- Yuan Yao and Furong Gao. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annual Reviews in Control*, 33(2):172–183, 2009.

- Sok Yee Yim, Hari Govind Ananthakumar, Lamia Benabbas, Alexander Horch, Rainer Drath, and Nina F. Thornhill. Using process topology in plant-wide control loop performance assessment. *Computers & Chemical Engineering*, 31(2):86–99, 2006.
- Jie Yu and S. Joe Qin. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE Journal*, 54(7):1811–1829, 2008.
- Alexey Zakharov, Elena Zattoni, Lei Xie, Octavio Pozo Garcia, and Sirkka-Liisa Jämsä-Jounela. An autonomous valve stiction detection system based on data characterization. *Control Engineering Practice*, 21(11):1507–1518, 2013.
- Xiaoyun Zang and John Howell. Isolating the source of whole-plant oscillations through bi-amplitude ratio analysis. *Control Engineering Practice*, 15(1):69–76, 2007.
- Ji Zhang and Hai Wang. Detecting outlying subspaces for high-dimensional data: the new task, algorithms and performance. *Knowledge and Information Systems*, 10(3):333–355, 2006.
- Yingwei Zhang and Chi Ma. Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS. *Chemical Engineering Science*, 66(1):64–72, 2011.
- Yingwei Zhang and Yang Zhang. Fault detection of non-Gaussian processes based on modified independent component analysis. *Chemical Engineering Science*, 65(16):4630–4639, 2010.
- Gilles Zumbach and Ulrich Müller. Operators on inhomogeneous time series. *International Journal of Theoretical and Applied Finance*, 4(01):147–177, 2001.

Part IV

Appendices

Appendix A

Appendices for Chapter 4

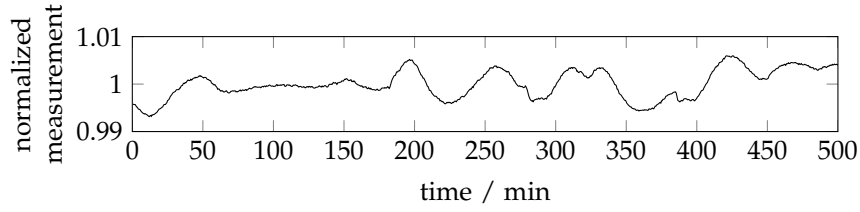
The appendices in group A present additional information relevant to Chapter 4.

A.1 Fitting of anomaly index vector to gamma distribution with skewness smaller than 0.77

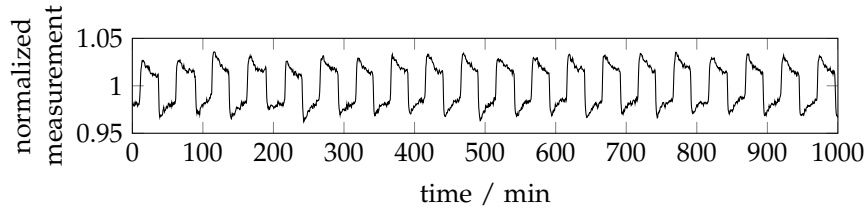
This appendix confirms the validity of assuming that the anomaly index vector generated from a time series without anomalous embedded vectors approximately fits to a gamma distribution with a skewness smaller than 0.77. This is done using Monte-Carlo simulations, which generate time series without transient disturbances, and observing that a significant fraction of these time series lead to anomaly index vectors with an acceptable fit to the assumption.

To be representative of different operation scenarios, this demonstration uses three cases: (i) steady state operation, (ii) operation with non-random variability, and (iii) oscillatory operation. The time series used for case (i) is synthetic while the time series for cases (ii) and (iii) belong to physical measurements monitored in a real gas processing plant and are shown in Figure A.1.

To begin with, three sets of N_{MC} random time series, representing Gaussian noise, are generated and added to the three base time series, thus making up $3N_{MC}$ time series. The anomaly index vector is then computed for each of the $3N_{MC}$ time series by the method developed in Chapter 4, with $m = 50$,



(a) Case (ii): operation with non-random variability, but no transients.



(b) Case (iii): oscillatory operation.

Figure A.1: Base time series for generation of cases representing normal operation scenarios.

$\tau = \delta = 1$ and $k = [1 : 10]$. The choices of these parameters follow the rules proposed in section 4.3.

The $3N_{MC}$ anomaly index vectors are assumed to follow gamma distributions. This distribution has two parameters, a and b , respectively describing the shape and scale of the distribution. Therefore, this section derives maximum likelihood estimators for the $3N_{MC}$ pairs of parameters a and b . The maximum likelihood method finds the parameters for a given statistical distribution that maximize the likelihood of observing the existing data.

Statistics for the estimated shape parameter \hat{a} and derived skewness (A.1) are then calculated for each set of N_{MC} parameters \hat{a} . Figure A.2 shows these statistics for each set. The lines follow the mean of the estimated properties and the error bars represent one standard deviation. The plots indicate skewness values significantly below the assumption limit of 0.77, as we wanted to show.

$$\text{skewness}_{\Gamma} = \frac{2}{\sqrt{a}} \quad (\text{A.1})$$

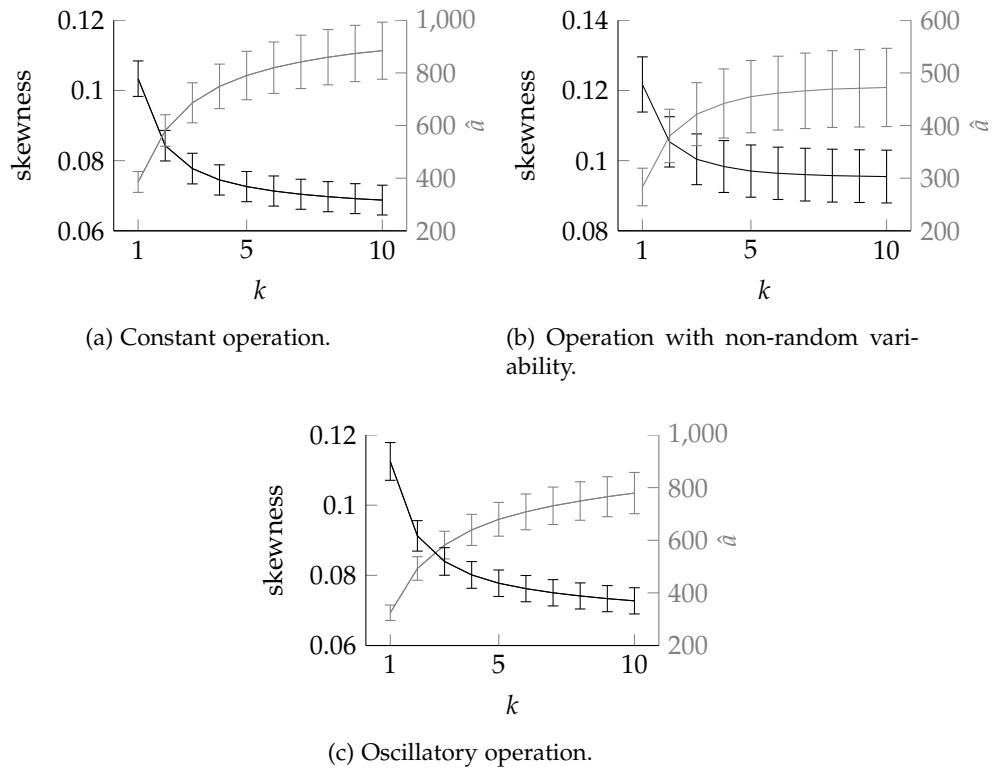


Figure A.2: Estimated gamma shape parameter (grey line) and derived skewness (black line) for three sets of clean signals.

A.2 Probability of false detection in gamma distributions with skewness smaller than 0.77

This appendix proves numerically that, in gamma distributions, a threshold of $Q_{2,\Gamma} + 6 \times \text{IQR}_{\Gamma}$ leads to a detection rate of less than one in a million if that distribution has a skewness of less than 0.77.

The gamma distribution has two parameters, a and b , respectively describing the shape and scale of the distribution. The gamma quantile function $F_{\Gamma}^{-1}(p)$ for the probability of detection $p = 0.999999$ and the statistics $Q_{2,\Gamma} + 6 \times \text{IQR}_{\Gamma}$ can then be calculated for gamma distributions with different pairs of parameters $\{a, b\}$. Because there is no analytical solution to the gamma quantile function, the $F_{\Gamma}^{-1}(p)$ values were found with the MATLAB function `gaminv`, which uses Newton's method to converge to the solution.

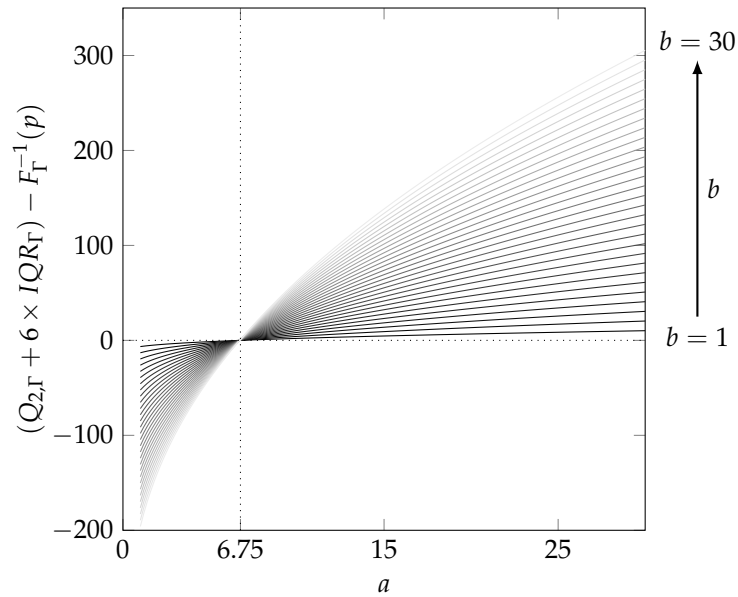


Figure A.3: Influence of gamma parameters a and b on the relation between the proposed threshold and the gamma quantile function, with $p = 0.999999$.

Figure A.3 plots $(Q_{2,\Gamma} + 6 \times IQR_{\Gamma}) - F_{\Gamma}^{-1}(p)$ for each $\{a, b\}$. With the proposed threshold, a detection rate equal or lower than one in a million happens if $F_{\Gamma}^{-1}(p) \leq Q_{2,\Gamma} + 6 \times IQR_{\Gamma}$. Figure A.3 shows that, in gamma distributions, this happens for a shape parameter $a \geq 6.75$ and is independent of the scale parameter b . A gamma distribution with shape parameter $a = 6.75$ has, by definition (A.1), a skewness of 0.77, as had to be proved.

Appendix B

Appendices for Chapter 5

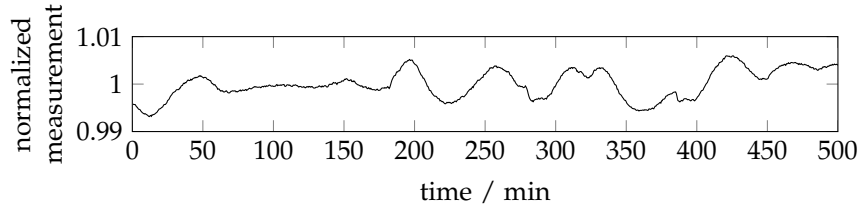
The appendices in group B present additional information relevant to Chapter 5.

B.1 Confidence level of the detection threshold

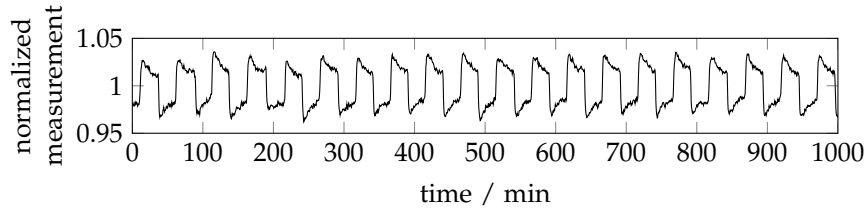
This appendix analyzes the behaviour of the detection threshold for the multivariate method, under the null hypothesis of a time series with no anomalies. Monte-Carlo simulations are used to generate three groups of N_{MC} time series with no anomalies. Each group is based on one time series, which is representative of an operation scenario with no transient disturbances. The three representative time series are: (i) steady state operation, (ii) operation with non-random variability, and (iii) oscillatory operation. The time series used for case (i) is synthetic while the time series for cases (ii) and (iii) belong to physical measurements monitored in a real gas processing plant and are shown in Figure B.1. Each of the N_{MC} time series in a group is given by the representative time series added to one random time series, which represents Gaussian noise.

The univariate detection method, described in Chapter 4, generates $3N_{MC}$ univariate anomaly index vectors \mathbf{ai} from the time series. The parameters used are $m = 50$, $\tau = \delta = 1$ and $k = 3$, as proposed in section 5.3.2.

An anomaly index matrix \mathbf{A} is formed for each univariate anomaly index vector, which is arranged in the first row of the matrix and hence is denoted \mathbf{ai}_1 . A second row stores an additional anomaly index vector \mathbf{ai}_2 , which is generated synthetically and is shown in Figure B.2. This vector resembles the anomaly index vector from a time series with a clear transient disturbance.

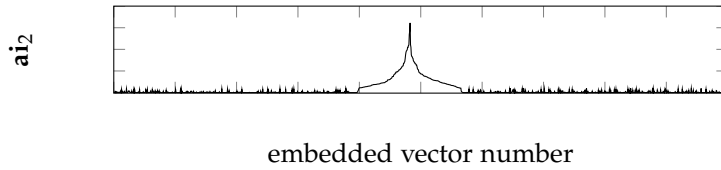


(a) Case (ii): operation with non-random variability, but no transients.



(b) Case (iii): oscillatory operation.

Figure B.1: Base time series for generation of cases representing normal operation scenarios.

Figure B.2: Univariate anomaly index vector \mathbf{ai}_2 used in the statistical study of the detection threshold.

The purpose of the anomaly index vector \mathbf{ai}_2 is two-fold. First, it allows a multivariate analysis. Second, it leads to a SVD basis function \mathbf{v}_j^\top with a clear protusion. Such a basis function can potentially induce false positives in the anomaly index vectors of the time series with no anomalies.

With an anomaly index matrix \mathbf{A} formed by two rows, the final anomaly index vector $\tilde{\mathbf{ai}}_1$ reconstructed after the multivariate step is formed by, at most, two terms. False positive detections on $\tilde{\mathbf{ai}}_1$ happen only if an included term leads itself to positive detections. Therefore, this section determines the probability that a term leading to positive detection is included in $\tilde{\mathbf{ai}}_1$. This gives the probability of $\tilde{\mathbf{ai}}_1$ leading to false positive detections.

A term $u_{1,j}s_j\mathbf{v}_j^\top$ leading to positive detections is included in $\tilde{\mathbf{ai}}_1$ if the ratio

$$\frac{(u_{1,j}s_j)^2}{(N_E - 1)\text{var}(\mathbf{ai}_1)} \quad (\text{B.1})$$

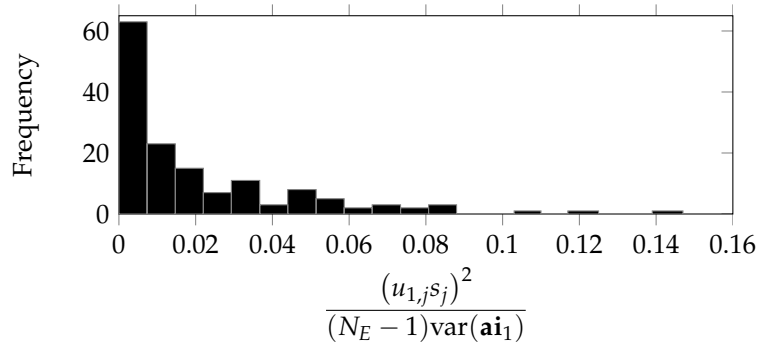


Figure B.3: Distribution of the values of relative variance of the SVD terms with positive detections.

is greater than parameter β (section 5.2.3). Figure B.3 shows the frequency distribution of ratio (B.1) for terms $u_{1,j}s_j\mathbf{v}_j^\top$ leading to positive detections. Positive detections are determined based on the threshold proposed in section 5.2.4.

The exponential distribution fits approximately to the distribution in Figure B.3. The exponential distribution has a single parameter, estimated in this case as $\lambda = 48$. An exponential distribution with $\lambda = 48$ has a probability lower than one in a thousand that the ratio (B.1) is greater than 0.15. This means that, with $\beta > 0.15$, a term leading to positive detections has a probability lower than one in a thousand to be included in the anomaly index vector $\tilde{\mathbf{a}}\mathbf{i}$ of a time series with no anomalies. As a result, with $\beta > 0.15$, the probability of false positives with the proposed threshold is less than one in a thousand.

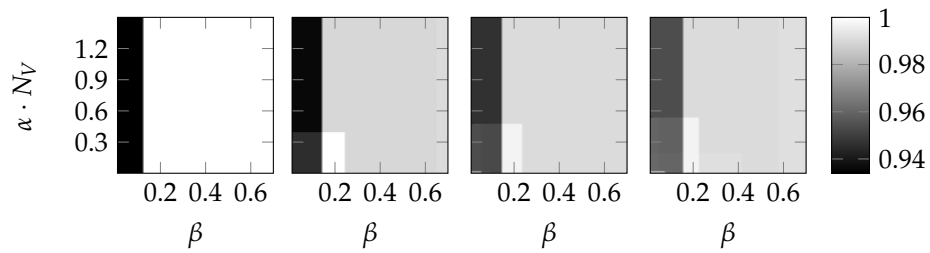
B.2 Optimization of α and β for additional groups of measurements

This appendix presents the optimisation results for five sub-groups of measurements from the development case study *Compressor rig case 2*. Table B.1 indicates the measurements included in each sub-group.

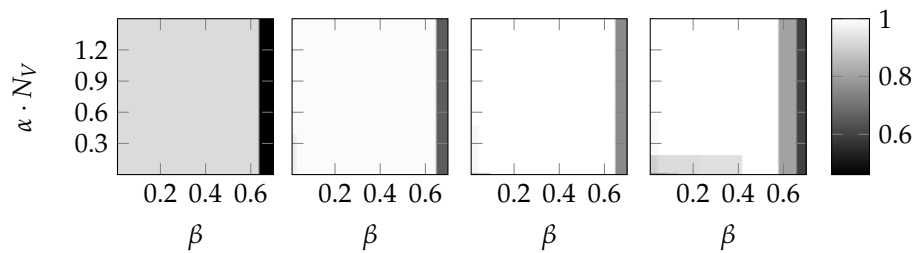
Figure B.4 shows the influence of α and β on the detection results from each of the five sub-groups, as measured by performance metrics FP and FN . The colours represent the magnitude of the metrics according to the scales shown in the figure.

Table B.1: Measurements included in each sub-group used for the optimisation of α and β . N_V refers to the number of measurements in the group.

N_V	Tags					
3	N1	Z1	P1			
4	N1	Z1	P1	S1		
5	N1	Z1	P1	S1	I1	
6	N1	Z1	P1	S1	I1	P2



(a) Metric FP



(b) Metric FN

Figure B.4: Performance of the detection method as a function of parameters α and β for five different sub-groups of measurements. N_V denotes the number of measurements in the group. Lighter tones denote better performance.

Appendix C

Experiments with the gas compression rig

The appendices in group C describe the main characteristics of the experimental facilities and procedures which were used to obtain the *Compressor rig* data set.

C.1 Description of the experimental facilities

The experimental rig is a laboratory-scale gas compression installation located at ABB Corporate Research Center, Kraków, Poland. The rig and its main components are shown schematically in Figure C.1, which also shows variables measured in the process system.

The main components on the process side are the centrifugal compressor, the plenum volume (provided by the tank), the inlet and outlet ducts and valves, and the hot recycle stream with pneumatic control valve. The main electro-mechanical components include the transformer (at the grid supply point), the a.c. variable speed drive and the electric motor which is directly coupled to the compressor. Additional equipment included are the sensors, signal converters, the controller and I/O module, and the computer.

Electrical drive

The ABB ACS800 drive is an a.c. variable speed drive of 11 kW responsible for feeding and controlling the electric motor in the rig. The electrical drive can use scalar or Direct Torque Control strategies, and accepts the motor speed or

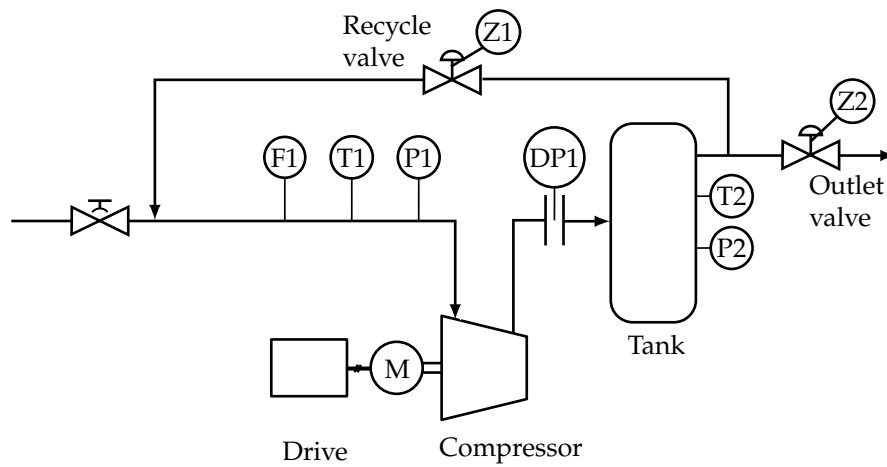


Figure C.1: Simplified schematic of the gas compression experimental rig.

the motor torque as set-points. Measured voltage and current signals in the electrical drive are sent to and recorded by the high-level controller, together with values of motor speed and torque estimated by the electrical drive. These are all 16-bit digital signals.

Electric motor

The ABB M3AA 160 MLB2 motor is a three-phase induction motor with two poles. Its rated power is 15kW and utilized speed range 0 – 6000 rpm. The motor is responsible for driving the compressor, to which it is directly coupled. The speed of the motor is measured by an optical encoder on the shaft. The temperature on the motor windings is also measured. These are all analog signals.

Compressor

The Continental Industrie 020.05 air blower is a five-stage centrifugal compressor with a nominal flow of 1000 m³/h and pressure rise of 300 mbar. Its rated power is 22 kW at maximum speed of 4700 rpm. The temperatures of the bearings on the drive and process sides are measured as analog signals.

Pressure tank

A tank with 0.5 m³ substitutes the volume of the pipeline downstream of the compressor. Together with the pipe length between it and the compressor

outlet, it sets the compressor surge frequency at 1.2 Hz. The tank is also responsible for damping flow disturbances before splitting the gas stream into recycle and outlet. The pressure and temperature in the tank are measured as analog signals.

Process valves

Three valves restrict the gas flow in the system: at the inlet, outlet and recycle streams. The inlet valve is manually operated thus has low position accuracy. The outlet valve is a ball type electric valve with a close-open transition of 15 s and a valve gain of 7000 L/(min.bar). Its control signal is analog with a position feedback signal. The recycle valve is a globe valve equipped with a pneumatic linear actuator and takes less than 5 s in close-open transition. Its control signal is analog with a position feedback signal.

Main controller

The ABB AC800 PEC controller ensures the high level control over the drive and valve controls. Its software is written in MATLAB/Simulink. The controller also gathers and synchronizes the signals from the drive and sensors; the drive (digital) communicates directly and the sensors (analog) through a Combi I/O extension board. All signals can then be passed to the Combi I/O module for recording on to a text file. The controller is assessed via a desktop computer.

Additional equipment, not featured in Figure C.1, was added ad-hoc for the induction of particular electrical disturbances.

C.2 Experimental conditions

Table C.1 indicates the four variable independent conditions and the modes they adopted in the tests.

The choice of electrical disturbances was mainly motivated by industrial relevance, constrained to the viability and safety of implementation. Nonetheless, given the academic motivation of this study, the disturbances were not only limited to those industrially representative and of high-impact. The list

Table C.1: Independent conditions and the modes they assumed in the tests.

Electrical disturbance		Drive control mode	Process control mode	Compressor operating region
Type	frequency			
Change in drive set-point (Torque and Speed controls)	oscillating aperiodic	DTC	Open loop	Stable
Step change in d.c. link voltage	periodic aperiodic	Scalar	Pressure control	Close to surge
One-phase motor current imbalance	periodic			Surge
Motor voltage imbalance	aperiodic			

below indicates the possible causes for each disturbance, and how each disturbance was implemented in the compressor rig.

- The changes in drive set-point may represent communication errors between a main controller and the drive, and were directly implemented in the main controller, using the computer interface.
- The changes in the d.c. link voltage simulate short-term undervoltages coming from the grid, such as dips and sags, which are the most common power quality disturbances affecting industrial plants [Bendre et al., 2004, Saksena et al., 2005]. The frequency and magnitude of these step changes were also implemented in the main controller, using the computer interface.
- The one-phase motor current imbalance may result from a loose terminal connection on one leg of the motor or from a build-up of carbon or dirt on one set of contacts. These two situations force the current to flow through the other phases, which are paths of least resistance. In the compressor rig, the one-phase motor current imbalance was implemented with an extra resistance, placed in parallel to one of stator phase lines. The switching on and off of the resistance was done in the main controller, using the computer interface.
- The voltage imbalance can arise in the power supply, distribution lines or from an unbalanced distribution of single-phase loads. In the compres-

sor rig, this disturbance was created with an auto-transformer placed between the drive and the motor, at a turn increasing or decreasing one of the phase voltages in relation to the other two. This variation was implemented manually.

The compressor operating region was defined by controlling the opening of the outlet valve, and following the trajectory of the operating point in the compressor map.

The dependent variables of the experiments were the physical and control variables of the system. These variables were recorded with a sampling rate of 5 kHz.

Appendix D

Open-Source Educational Toolbox for Power System Frequency Control Tuning and Optimization

The appendices in group D present work developed in parallel to the main contributions of this thesis. The work described is an open-source educational tool for control of the frequency in power systems. The toolbox is available on-line, and is aimed at power and control engineering students to practice frequency control and test tunings and control strategies. This work was presented in the 4th IEEE European conference on Innovative Smart Grid Technologies [Cecílio et al., 2013].

Introduction

The frequency variation of a power system is a direct result of the balance between power generation and consumption. This means that if generation and consumption are not made to match, the excess (deficiency) of energy will be translated into an increase (decrease) of the system frequency, f .

The sensitivity of electrical equipment to frequency deviations is high and the consequences severe - even deviations smaller than 1% can lead to damage, degradation of performance and trips of loads and generators [Bevrani, 2009].

The past decade has seen a significant increase in the number of such frequency incidents which, in part, can be attributed to the increasing penetration of renewable energies. Because the production of renewable generators

cannot be precisely scheduled, the overall system generation loses controllability, and load-generation imbalances are more likely to happen. An obvious example is the Nordic system where roughly a two-fold increase in wind power installed capacity was accompanied by a two-fold increase in the number of frequency incidents [Whitley and Gjerde, 2011], that is, $f < 49.9\text{Hz}$ or $f > 50.1\text{Hz}$.

Commercial power systems software offers detailed representation of electrical equipment but may be unavailable to power and control engineering students.

This appendix presents an educational software toolbox which is available online [Cecilio, 2013]. This long-term dynamic simulation model comprises several features of a realistic frequency control system, conditioned to constant voltage, single control area and all machines connected to the same bus. The constant voltage condition is a valid assumption when effective voltage control is in place. The necessary modifications to reproduce a N -control area power system are discussed in section 6. The control strategy implemented follows the current industrial practice. In addition, the tool is able to generate realistic sequences of power imbalances, provided a worst-case imbalance characteristic which can be defined by the user. This is a complement to the academic step disturbance, against which these systems are mostly tuned [Rerkpreedapong et al., 2003, Ghoshal, 2004].

The tool is aimed at power and control engineering students, to facilitate the learning of frequency control and testing of different control tunings and strategies. An example of using the model proposed by this tool for frequency control design is given in Ersdal et al. [2013] with model predictive control.

Power System Frequency Control

Frequency control is usually provided by generating units which can modulate their generated power. In current industrial practice, these reserves can be used in three different services. The classification adopted by this tool is the one proposed by the European Network of Transmission System Operators for Electricity (ENTSO-E) [ENTSO-E, 2012b], that is, Frequency Containment Reserves (FCR), Frequency Restoration Reserves (FRR) and Replacement Reserves (RR). These services correspond roughly to the terminology of primary, secondary and tertiary services used in continental Europe [Rebours et al.,

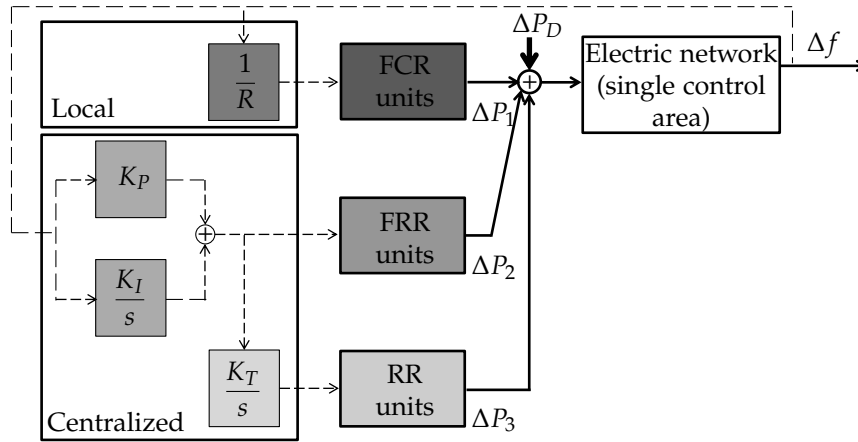


Figure D.1: Single area power system frequency control.

2007]. The commission and decommission of the reserves in each service is controlled by separate control loops, as depicted in Fig.D.1. This simplified block diagram shows the three distinct sets of power reserve units and the three dedicated control loops with their corresponding parameters.

The FCR service is time-critical and must be able to stabilize frequency within tens of seconds. Only generating units that can respond fast enough to frequency variations can participate to this service. Historically in the Nordic network, this service has been almost exclusively provided by hydro power generators. The control of these units is provided by the autonomous proportional action of the speed governor, which is represented by the speed-droop characteristic (R) [Bevrani, 2009].

Because of the proportional law, the FCR service is not capable of restoring the frequency to its nominal value, so the FRR service takes over the former to guarantee frequency restoration. This service is not as time-critical as stabilizing frequency excursions, so units which activate within tens of seconds to minutes can participate. This service is typically provided by thermal power generators and recently there have been suggestions on how to include industrial loads [Fabozzi et al., 2013]. The control of FRR units is provided by a centralized controller, which responds to a regulation signal, which for a single control area is the frequency deviation. Proportional-integral (PI) controller is the widely used mechanism for this service, and the two controller parameters, proportional gain (K_P) and integral gain (K_I), are commonly tuned with heuristics and trial-and-error approaches [Bevrani, 2009].

For longer time horizons, the RR service substitutes the FRR, which has limited energy reserves and is more expensive. Units participating to the RR service are usually asked to respond within several minutes and have to sustain their service for longer periods. This service is provided by thermal generators as well as some controllable loads. The control of these units is often manual; after some time of the secondary reserves being activated, an operator in the system control centre issues an order to activate the replacement reserves.

Representation of power system and frequency control in FRECOL

FRECOL (Frequency REserve Control Open-source Library) is a MATLAB Simulink toolbox which is available online [Cecilio, 2013]. Its purpose is to serve as an open-source educational tool for power and control engineering students to practice frequency control and test different tunings and control strategies against realistic disturbance scenarios.

This section presents the models of the power system and frequency control implemented, using the notation of the tool.

FRECOL represents the power system as a one control area and assumes constant voltage and machines connected to a single bus. Figure D.2 shows the first layer of the Simulink model. The dynamic frequency response of the power system to power imbalances is modelled by the first-order linearized relation shown in the block labelled as Rotating mass and load. Parameters H and D in that equation account for the inertia of the rotating masses and the self regulation of the load, respectively. This is a valid assumption for frequency control in the presence of load imbalance since the dynamics affecting frequency response are relatively slow [Bevrani, 2009].

The three frequency control services are implemented in the blocks labelled FCR, FRR and RR in Figure D.2.

The details of the FCR block are shown in Figure D.3. The gain blocks model the proportional control action of FCR units through the speed-droop characteristic ($1/R$). The power reserve commitment of the generating units should be limited to a predefined amount. This power reserve limitation is implemented in the saturation block in Figure D.3 as the equivalent frequency

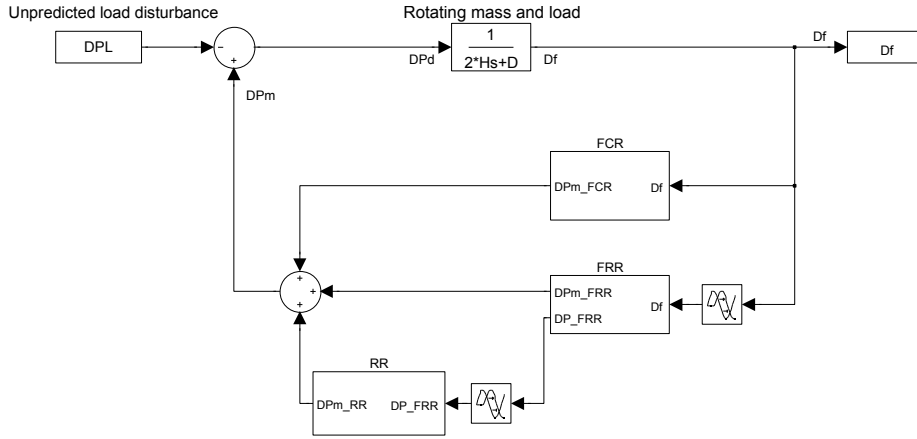


Figure D.2: Simulink model in FRECOL: first layer

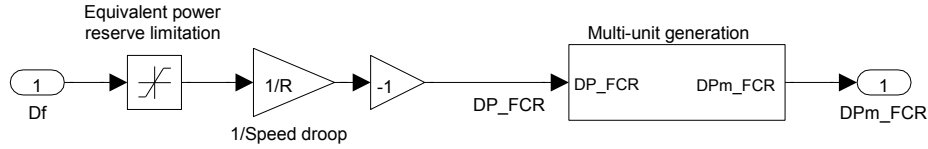


Figure D.3: Under mask of FCR block.

limitation B ; this parameter is specified by grid codes and indicates that at $\pm B$ FCRs should be fully activated.

The generating units in FCR are modelled as hydro units according to Machowski et al. [2011]. In this case, the dynamics of the governor is represented by a one-zero, two-pole equation and the turbine by a one-zero, one-pole equation (D.1).

$$G_{FCR} = \underbrace{\frac{T_3 \cdot s + 1}{(T_2 \cdot s + 1)(T_4 \cdot s + 1)}}_{\text{Governor}} \cdot \underbrace{\frac{-T_w \cdot s + 1}{T_w/2 \cdot s + 1}}_{\text{Turbine}} \quad (\text{D.1})$$

where T_2 , T_3 and T_4 are governor time constants and T_w is the water time constant of hydro turbines.

Frequency control services are usually provided by several units in parallel. As shown in Figure D.4, the model of the FCR service includes multiple generating units, which the user can easily remove or add to. Each unit can have distinct time constants because these are implemented as vectors in FRECOL. The participation of unit i to the FCR service is implemented in the i^{th} gain block in Figure D.4 as $FCR(i)/(B/R)$. This ratio represents the predefined power reserve of unit i relative to the total power reserve of the service.

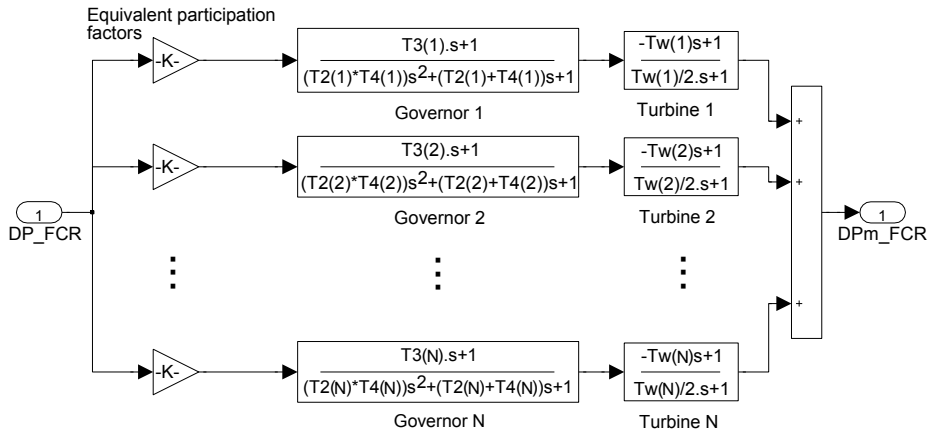
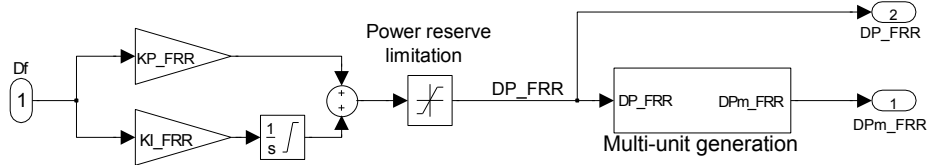
Figure D.4: FCR generation, under *Multi-unit generation* mask.

Figure D.5: Under mask of FRR block.

This implementation is used instead of actual participation factors to allow for the possibility of the units not being capable of generating as much power as demanded by the controller, that is, $\sum FCR(i) < (B/R)$. In this case, the programme returns warning.

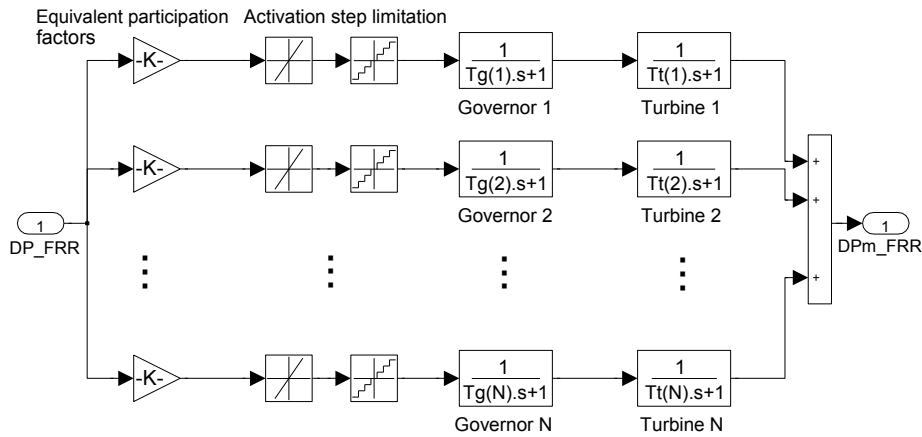
The FRR control loop is implemented with a PI control law as shown in Figure D.5. The proportional and integral gains are denoted as KP_FRR and KI_FRR in the Simulink model.

The total power reserve of this service is limited to $FRRP$ in the saturation block of Figure D.5. Furthermore, to avoid winding-up of the integral control action, the integrator block for the integral action is also limited to $FRRP$.

Generation in FRR is modelled by generic thermal units according to [Bevrani, 2009]. In this case, the dynamics of governor and turbine can be represented by the first-order model (D.2).

$$G_{FRR} = \underbrace{\frac{1}{T_g \cdot s + 1}}_{\text{Governor}} \cdot \underbrace{\frac{1}{T_t \cdot s + 1}}_{\text{Turbine}} \quad (D.2)$$

where T_g and T_t are the governor and turbine time constants.

Figure D.6: FRR generation, under *Multi-unit generation* mask.

As with FCR, the participation of each unit is also implemented in the gain blocks of Figure D.6 as $FRR(i)/FRRP$. A warning message is issued if $\sum FRR(i) < FRRP$.

European guidelines suggest thermal units should be activated with a ramp limited signal [ENTSO-E, 2012a], imposed by the rate limiter blocks in Figure D.6. The following quantizer blocks model the FRR cycle time and the coarseness of the unit activation.

An additional physical constraint of the FRR service featured in the model is the time delay shown in Figure D.2 before the FRR block. This block models the communication delay of the measured frequency between the operating unit and the controller localized in the control center. In a multi-area system this delay also models the time to obtain power imbalances from the SCADA/EMS system.

Block RR in Figure D.2 provides for the decommitment of the FRR units, rather than modelling faithfully the RR service. Therefore, the manual control of the RR units by the operator is mimicked as an integral action on the FRR control signal, with equivalent integral gain KI_RR . The total power reserve limit of this service, RRP , is likewise implemented in the saturation and integrator blocks. This is shown in Figure D.7, where proportional and derivative control actions can also be seen. By default, the tool has the gains in these actions set to zero, but the user may change them to test alternative representations of the manual actions of the operator. The delay before block RR in

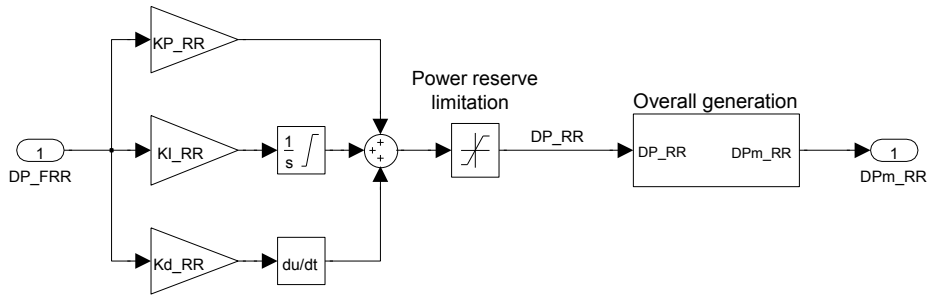


Figure D.7: Under mask of RR block.

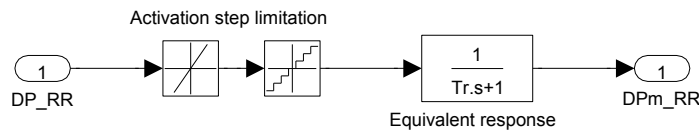


Figure D.8: RR generation, under *Overall generation* mask.

Figure D.2 simulates the overall delay of measurement, communication, state estimation and operator action.

The overall dynamics of RR units is represented as a first-order response with time constant Tr , as shown in Figure D.8. As with the FRR service, limits to activation rate of the units, cycle time and resolution of the activation are modelled by the rate limiter and quantizer blocks.

Power imbalance functions

Block DPL in Figure D.2 injects power imbalances in the system. These can either be given by the user or FRECOL can generate random imbalances constrained to a worst-case area.

With the constrained random disturbance, the imbalance varies continuously with magnitudes of random value but restricted to a worst-case curve, as represented in Figure D.9 (left). At any time n , the future imbalance scenario is constrained to the intersection of similar such curves, centered on the actual imbalances occurred at all the previous time instants. For simplicity, Fig.D.9 (right) illustrates this concept at discrete time instants. The worst-case curve is derived in FRECOL from estimates of worst-case variations from the predicted power output at given time horizons. For instance, in power networks including a certain amount of wind power, such as the Nordic grid, the unpredicted power imbalance can for simplicity be assumed to be dominated

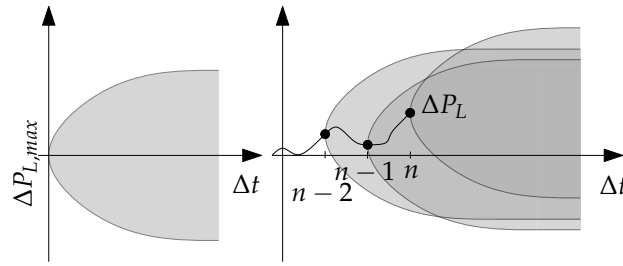


Figure D.9: Area defining the constrained random disturbance $\Delta P_{L,max}$ (left), and a time realization of $\Delta P_{L,max}$ (right).

by the fluctuations in produced wind power. Worst-case variations from the predicted wind power output within several horizons into the future, Δt , were estimated by Holttinen [2004] for a specific wind farm.

The documentation in Cecílio [2013] provides further explanation on the generation of these constrained random disturbances.

Demonstration of the tool

This section presents the simulation results given by FRECOL in three distinct cases. The parameters used in the simulation are the default values suggested in Cecílio [2013]. The nominal frequency is 50 Hz and the FCR, FRR and RR power reserves are 330, 500 and 2500 MW, respectively.

Figure D.10 plots the results of a two-and-a-half-hour simulation of the controlled system. The continuous black line in the lower plot represents the pseudo-random disturbance sequence DPL generated by FRECOL. It presents mostly mild imbalances, as expected in most cases. Under frequency control, and with the mild disturbance scenario, the frequency in the system is kept close to nominal and without incidents, as shown in the upper plot.

The simulated responses of the three generation services to the frequency variations are represented by the patterned blue lines in the lower plot, and show clearly the functions of each service. As expected from the proportional control law and the two-pole model, the response of the FCR service resembles a symmetric and filtered version of the frequency signal, this way providing for the quick stabilization of the frequency. The response of the FRR service, on the other hand, closely follows the power imbalance to restore the energy in the system and keep the frequency at its nominal value. As time progresses

the FRR signal becomes more distanced from the load signal because the RR service slowly activates to replace it.

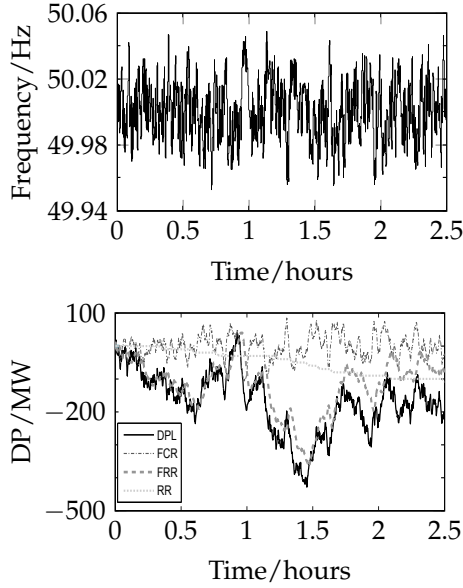


Figure D.10: System response to a mild power imbalance *DPL*

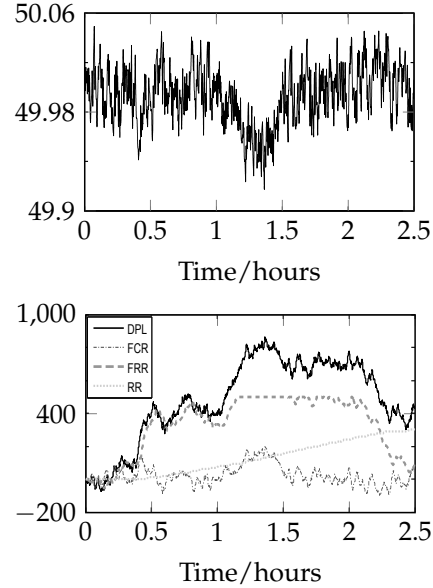


Figure D.11: System response to a more severe power imbalance *DPL*

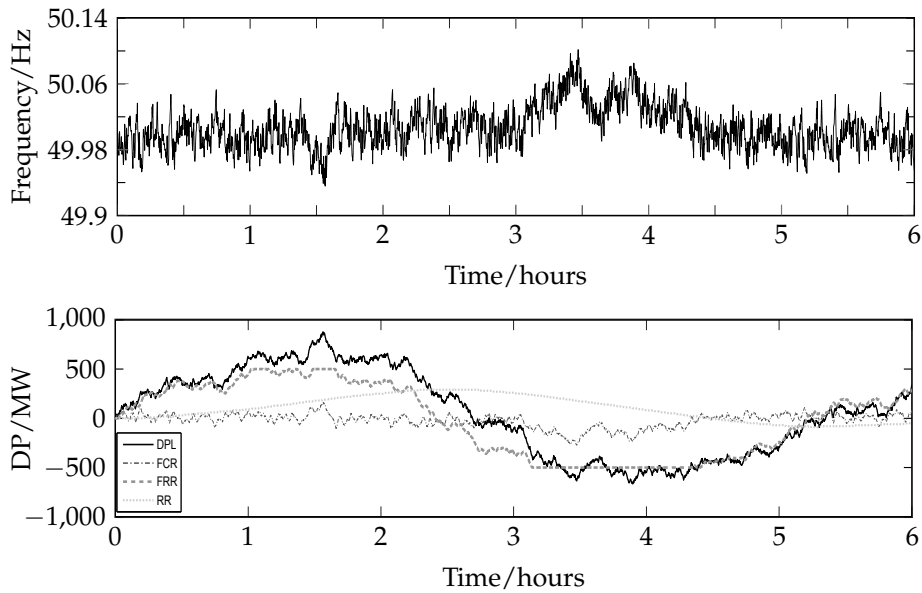


Figure D.12: System response to power imbalance *DPL* with variations close to worst-case scenario.

Figure D.11 plots the evolution of the controlled system for two and a half hours in response to a second pseudo-random disturbance sequence generated by FRECOL. The disturbance is shown in the continuous black line in the lower plot and is significantly more severe than the first, with two consecutive surges in demand of approximately 400 MW in 15 minutes.

The FRR control loop is fast enough to follow the demand but reaches its limit before the RR service fully activates. The consequent power imbalance leads to a sustained frequency deviation, as shown in the upper plot between 1 and 1.5 hours.

The slow activation of the RR service could suggest the need for a more aggressive control tuning. It should be noted, however, that in industrial practice the RR service is manually controlled by an operator. It is thus expected that the operator may vary the aggressiveness of control to the severity of the situation. Possible ways of mimicking this nonlinear reasoning with automatic control include the use of the squared control error as regulation signal or to have different control gains for different operating points of the system (gain scheduling).

Figure D.12 shows the results of a third simulation which exemplify a trying disturbance scenario to the system. In this case the pseudo-random disturbance sequence quickly changes from a significant excess of energy (+600MW) to a deficiency (-600MW), with a variation close to the worst-case scenario: 1200 MW drop in about one hour. The simulation by FRECOL shows that such a situation leads to saturation of the FRRs first in the positive and then in the negative sides of the reserve. Furthermore, this scenario actually forces the RR service to reverse direction. This is a slow change due to the large time constant of this service, so after 2.5 hours the RR service is actually increasing the energy imbalance. This leads to the saturation of FRRs and also the use of the FCRs for energy restoration. The frequency is shown in Figure D.12, which evidences a sustained deviation of the frequency after 2.5 hours and a frequency incident (over 50.1Hz) around 3 hours.

Conclusions

This appendix presented an open-source educational software tool for the long-term simulation of power system frequency control. The tool is aimed at

power and control engineering students and is available online. The parameters of the system can be tuned by the user to resemble the desired network.

The models for the generation of power and the control services take into account the nonlinearities of the system as well as the specifics of each of the three control services, including modelling the manual control by the operator in the RR service. So far, only a single control area is considered. Nonetheless, extension to a multi-area power system can be achieved by replicating the model and including the power flow on the tie-line between control areas in the mismatch power signal and the regulation signal, as explained by Bevrani [2009].

An additional contribution of the tool is the generation of random disturbance sequences which resemble realistic imbalance scenarios. The tool was tested with a common disturbance scenario and with two more severe cases, and the results showed that the responses simulated by the tool agreed with the expected system behaviour. It should be noted that the random disturbance generation feature facilitates extensive statistical analyses of power imbalances, control system performance and sizing of reserves.