

Imperial College London
Department of Chemical Engineering

Industrial fault detection and diagnosis using alarms and process measurements

Matthieu Lucke

Academic supervisor: Prof. Nina F. Thornhill
Industrial supervisor: Dr. Moncef Chioua

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Process Systems Engineering of
Imperial College London, November 2019

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. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Implementing data-driven fault detection and diagnosis methods on process plants can be a challenge. Constraints due to the availability and the variability of the process measurements as well as constraints due to the characteristics of the industrial systems impact the reliability of fault detection and diagnosis methods. This thesis aims at increasing the reliability of data-driven fault detection and diagnosis methods on process plants to extend their use in industry.

The core idea of the thesis is to bring together the disciplines of alarm management and fault detection and diagnosis. The first part of the thesis suggests a fault detection and diagnosis approach to the problem of classification of ongoing abnormal situations based on alarm data alone. The second part of the thesis investigates the integration of alarms, alarm settings, and alarm management practices into traditional fault detection and diagnosis methods based on process measurements. Both parts emphasize the robustness of the proposed methods with regard to the variability in the input data, as well as the industrial applicability of the methods. The results are validated on an oil and gas separation plant and on a multiphase flow facility.

The last part of the thesis focuses on root cause analysis of process disturbances. Many data-driven root cause analysis methods have been proposed in process literature in the past twenty years, but their reliability depends on the properties of the industrial system and on the properties of the disturbance. This thesis provides a comparative review of data-driven root cause analysis methods clarifying the scope of application of each method. The objective is to guide practitioners during the root cause analysis and facilitate the use of data-driven root cause analysis methods in industry. The comparative review also highlights the gap of knowledge in root cause analysis of transient disturbances and suggests a new approach based on transient disturbance detection methods to fill this gap.

Acknowledgements

Professor Nina Thornhill had the largest influence on this work. She set the standards for the thesis and showed impressive commitment until they have been reached.

Moncef Chioua has been a great source of inspiration. Moncef, I really appreciated the conversations we had and the hours you did not report.

Reinhard Bauer, Barbara Beck, Cornelia Glock, Iiro Harjunoski, Thorsten Schindler, and Marco Ulrich ensured this project could be led smoothly until the end.

I would like to mention the people from the Optimization and Operational Analytics group at ABB Corporate Research Germany, the early stage researchers of the PRONTO ITN, and the students of the Centre for Process Systems Engineering at Imperial College for the friendly atmospheres. Special thanks to Abdolreza, Andrew, Anna, Brian, Chriss, Francesco, Frederik, Giancarlo, James, Marta, Martin, Mei, Nuo, and Ruomu for the fruitful technical (and non-technical) discussions. I am also adding a thought for Margret Bauer who brought me into this.

And thanks to family and friends on both sides of the Rhine.

Financial support is gratefully acknowledged from the Marie Skłodowska Curie Horizon 2020 EID-ITN project "PROcess NeTWork Optimization for efficient and sustainable operation of Europe's process industries taking machinery condition and process performance into account PRONTO", Grant agreement No 675215.

Contents

Abstract	1
Acknowledgements	2
1 Introduction	18
1.1 Description of the chapter	18
1.2 Introduction to the project: PRONTO	19
1.3 Introduction to fault detection and diagnosis in industrial processes	20
1.3.1 Introduction to fault detection and diagnosis	20
1.3.2 Fault detection and diagnosis for safe and profitable process operation	21
1.3.3 Alarm systems and alarm management	23
1.4 Motivation of the thesis	24
1.4.1 Increasing reliability of data-driven fault detection and diagnosis . . .	24
1.4.2 Integrating alarm systems in fault detection and diagnosis	25
1.4.3 Clarifying the field of application of data-driven root cause analysis methods	25
1.5 Introduction to the thesis	26
1.5.1 Objectives of the thesis	26
1.5.2 Contributions of the thesis	27
1.5.3 Outline of the thesis	29
1.6 Publications	30
2 Background	31
2.1 Description of the chapter	31
2.2 Data sources in the process industry	32
2.2.1 Process measurements	32
2.2.2 Alarms and events	32
2.2.3 Plant connectivity	33
2.3 Alarm systems in the process industry	34
2.3.1 Introduction to alarm systems	34
2.3.2 Alarm design	40

2.3.3	Alarm processing	42
2.3.4	Alarm data representations	43
2.4	Fault detection and diagnosis in industrial processes	45
2.4.1	Faults and disturbances	45
2.4.2	Fault detection and diagnosis	47
2.4.3	Root cause analysis	50
2.5	Case studies	52
2.5.1	Oil and gas separation plant	52
2.5.2	Multiphase flow facility	54
2.6	Overview of the assumptions of the thesis	56
2.6.1	Definitions and concepts	56
2.6.2	Assumptions on the data	56
2.7	Summary of the chapter	57
3	State of the art	58
3.1	Description of the chapter	58
3.2	Alarm data analysis	59
3.2.1	State of the art in alarm data analysis	59
3.2.2	Sequence mining methods	60
3.2.3	Time series analysis methods	65
3.2.4	State of the art in alarm flood classification	67
3.2.5	Discussion on the review	69
3.3	Online fault detection and diagnosis in industrial processes	70
3.3.1	State of the art in online fault detection and diagnosis	70
3.3.2	Preprocessing	73
3.3.3	Variable selection	74
3.3.4	Fault detection	75
3.3.5	Fault isolation	77
3.3.6	Fault identification	77
3.4	Root cause analysis in industrial processes	78
3.4.1	State of the art in root cause analysis	80
3.4.2	Univariate methods	82
3.4.3	Time-delay estimation methods	83
3.4.4	Prediction-based methods in the time-domain	85
3.4.5	Prediction-based methods in the frequency-domain	91
3.4.6	Comparative analyses of data-driven root cause analysis methods	93
3.5	Summary of the chapter	94

4	A binary series approach to online alarm flood classification	96
4.1	Description of the chapter	96
4.2	Background	97
4.2.1	A gap in online alarm flood classification for the industry	97
4.2.2	A case study in alarm flood classification	98
4.2.3	Alarm binary series	99
4.3	Method development	99
4.3.1	Alarm coactivations	102
4.3.2	Classification of the alarm floods	103
4.3.3	Online implementation	107
4.4	Application to industrial case study	109
4.4.1	Assessment of performance of the detection stage	110
4.4.2	Assessment of performance of the classification stage	110
4.4.3	Assessment of performance of the online implementation	111
4.5	Results	111
4.5.1	Results of the detection stage	111
4.5.2	Results of the classification stage	112
4.5.3	Results of the online implementation	112
4.5.4	Analysis of the parameters	118
4.6	Summary of the chapter	119
5	Integrating alarm design in online fault detection and diagnosis through alarm-range normalization	120
5.1	Description of the chapter	120
5.2	Background	121
5.2.1	Integrating alarm systems in online fault detection and diagnosis . . .	121
5.2.2	Motivation for the method	122
5.3	Method development	123
5.3.1	Alarm-range normalization	123
5.3.2	Fault detection and isolation	124
5.3.3	Fault identification	125
5.4	Application to industrial case study	126
5.4.1	Description of the industrial case study	126
5.4.2	Design of experiment	128
5.5	Results	130
5.5.1	PCA modeling	130
5.5.2	Fault detection	132
5.5.3	Fault isolation	135
5.5.4	Fault identification	136
5.6	Summary of the chapter	141

6	Fault detection and identification combining process measurements and statistical alarms	142
6.1	Description of the chapter	142
6.2	Background	143
6.2.1	Variable selection with mRMR criterion for process measurements . .	143
6.2.2	Motivation for the method	146
6.3	Method development	147
6.3.1	Variable selection with mRMR on level-2 alarms	147
6.3.2	Fault detection and identification with Bayesian fusion of process measurements and level-1 alarms	148
6.4	Tests of the method in simulation	150
6.4.1	Simulation example	150
6.4.2	Results	152
6.4.3	Comparison of the methods	152
6.5	Application to industrial case study	153
6.5.1	The multiphase flow facility	153
6.5.2	Case study	154
6.6	Results of the case study	160
6.6.1	Variable selection	160
6.6.2	Fault detection and identification	162
6.6.3	Analysis of the parameters	167
6.7	Summary of the chapter	169
7	Review of existing methods for root cause analysis of oscillatory and non-oscillatory disturbances	171
7.1	Description of the chapter	171
7.2	Criteria for the comparative review	172
7.2.1	Properties of the methods	172
7.2.2	Impact of the systems' properties on the results of the methods	174
7.2.3	Types of disturbances	176
7.3	Comparative review of existing data-driven root cause analysis methods . . .	179
7.3.1	Review of the properties of the methods	179
7.3.2	Review of impact of the systems' properties on the results of the methods	182
7.3.3	Review of the types of disturbances covered by the methods	186
7.4	Root cause analysis of persistent disturbances	190
7.4.1	Root cause analysis of oscillatory disturbances	190
7.4.2	Root cause analysis of non-oscillatory persistent disturbances	191
7.5	Root cause analysis of transient disturbances	192
7.5.1	Root cause analysis of transient disturbances with existing methods .	192

7.5.2	Limitations of the time-delay estimation approach	193
7.5.3	Root cause analysis using methods for detection of transient disturbances	195
7.6	Summary of the chapter	197
8	Critical evaluation and suggestions for future work	199
8.1	Description of the chapter	199
8.2	Evaluation of the achievements of the objectives	199
8.2.1	Objective 1: online FDD approach to alarm data analysis	199
8.2.2	Objective 2: integration of alarm systems into online FDD	202
8.2.3	Objective 3: RCA, from oscillatory to non-oscillatory disturbances . .	204
8.2.4	Contribution to the main motivation of the thesis	206
8.3	Future work and future research directions	207
8.3.1	Future work for Objective 1	207
8.3.2	Future work for Objective 2	207
8.3.3	Future work for Objective 3	208
8.3.4	Future research directions in data-driven fault detection and diagnosis	209
8.4	Summary of the chapter	210
9	Conclusion	211
9.1	Conclusion on Objective 1	212
9.2	Conclusion on Objective 2	212
9.3	Conclusion on Objective 3	212
9.4	Final comments	213
A	Variable selection in the multiphase flow facility	214
B	Review of existing root cause analysis methods	219
B.1	Review of the properties of the methods	219
B.1.1	Univariate methods	219
B.1.2	Time-delay estimation methods	220
B.1.3	Prediction-based methods in the time-domain	221
B.1.4	Prediction-based methods in the frequency-domain	224
B.2	Review of the impact of the systems' properties on the results of the methods	225
B.2.1	Univariate methods	225
B.2.2	Time-delay estimation methods	225
B.2.3	Prediction-based methods in the time-domain	226
B.2.4	Prediction-based methods in the frequency-domain	228
B.3	Review of the types of disturbances covered by the methods	229
B.3.1	Univariate methods	229
B.3.2	Time-delay estimation methods	230

B.3.3	Prediction-based methods in the time-domain	231
B.3.4	Prediction-based methods in the frequency-domain	233
Bibliography		234

List of Tables

2.1	Example of alarm record.	35
2.2	Recommended alarm performance metrics.	39
2.3	Process variables in the PWRI section.	53
2.4	Process variables in the multiphase flow facility.	55
2.5	Data sources considered in the technical chapters.	56
3.1	Approaches to alarm data analysis.	61
4.1	Description of the selected types of alarm floods.	99
4.2	Time-stamped sequences of alarm floods 70 and 78.	101
4.3	Results of FDR and MDR tests.	112
4.4	Average accuracy of classification (offline analysis).	112
4.5	Online results for selected classes.	113
4.6	Online results for leftover classes.	114
5.1	Selected faults in the PWRI section.	126
5.2	Process variables, alarm thresholds, and standard deviations in the PWRI section.	127
5.3	False detection rates on the test set Ω_{N0} with N-standardization and AR-normalization.	133
6.1	Classification accuracies on the simulation example.	153
6.2	Classification accuracies on the test dataset for scenarios AB and BA with the selected set of variables $\mathcal{X}_2$	162
6.3	Confusion matrices on the test dataset for scenarios AB and BA with the selected set of variables $\mathcal{X}_2$	166
7.1	Review of the properties of the methods.	178
7.2	Review of impact of the systems through which the disturbance propagates on the results of the methods.	184
7.3	Review of types of disturbances covered by the methods.	189
7.4	Estimated time-delays with cross-correlation function on the transient weld disturbance.	195

7.5	Maximum time-delayed cross-correlation values with cross-correlation function on the transient weld disturbance.	196
7.6	Index of the first abnormal embedded vector for each variable affected by the transient weld disturbance.	196
A.1	Sets of selected variables $\mathcal{X}_M$ with mRMR on level-2 alarms and with mRMR on process measurements in scenario AB.	217
A.2	Sets of selected variables $\mathcal{X}_M$ with mRMR on level-2 alarms and with mRMR on process measurements in scenario BA.	218

List of Figures

1.1	Online process monitoring loop.	20
1.2	Layers of protection.	22
1.3	Feedback model of operator interaction.	23
1.4	Graphical summary of the thesis.	27
2.1	Alarm management lifecycle.	37
2.2	Alarm response timeline.	41
2.3	Fault detection and diagnosis workflow	48
2.4	Signed directed graph for causal modeling between four variables.	51
2.5	Process overview of the separation plant.	52
2.6	Process diagram of the PWRI section.	53
2.7	Process diagram of the multiphase flow facility.	54
3.1	Family tree of methods for alarm data analysis.	62
3.2	Family tree of methods for data-driven root cause analysis.	79
4.1	Binary alarm series for alarm flood 70 and alarm flood 78.	100
4.2	Alarm coactivation matrix for alarm flood 70.	105
4.3	Alarm coactivation matrix for alarm flood 78.	106
4.4	Flowchart of the proposed online implementation.	108
4.5	Classification outcomes for ongoing alarm floods from selected classes and leftover classes in the low-MDR configuration.	115
4.6	Classification outcomes for ongoing alarm floods from selected classes and leftover classes in the trade-off configuration.	116
4.7	Classification outcomes for ongoing alarm floods from selected classes and leftover classes in the low-FDR configuration.	117
5.1	Process measurements in the training set $\tilde{\Omega}_{N0}$ normalized with N-standardization and AR-normalization.	128
5.2	Process measurements in the training set $\tilde{\Omega}_F$ normalized with F-standardization and AR-normalization.	129
5.3	Projection of the process measurements of $\tilde{\Omega}_{N0}$ on the first ten loading vectors with N-standardization and AR-normalization.	130

5.4	Cumulative variance explained by the loading vectors on $\tilde{\Omega}_{N0}$ with N-standardization and AR-normalization.	131
5.5	Process measurements in the test set Ω_{N0} normalized with N-standardization and AR-normalization.	132
5.6	T^2 statistic for the test set Ω_{N0} with N-standardization and AR-normalization.	134
5.7	Q statistic for the test set Ω_{N0} with N-standardization and AR-normalization.	134
5.8	T^2 and Q statistics with N-standardization and with AR-normalization for one occurrence of each type of fault in $\tilde{\Omega}_F$	135
5.9	T^2 reconstruction-based contribution plots with N-standardization and with AR-normalization for one occurrence of each type of fault in $\tilde{\Omega}_F$	137
5.10	Q reconstruction-based contribution plots with N-standardization and with AR-normalization for one occurrence of each type of fault in $\tilde{\Omega}_F$	137
5.11	Process measurements in the test set Ω_F normalized with F-standardization and AR-normalization.	138
5.12	Classification outcomes with 1NN and F-standardization.	139
5.13	Classification outcomes with 1NN and AR-normalization.	139
5.14	Normalized plant vectors for the training occurrence of Fault 2 and Fault 3, and for Fault 3-7.	140
6.1	System of the simulation example.	150
6.2	Standardized measurements, level-1 alarms, and classification outcomes on the simulation example.	151
6.3	Standardized measurements captured during each fault introduced in operating point A.	156
6.4	Standardized measurements captured during each fault introduced in operating point B.	157
6.5	Standardized process measurements of the first ten ranked variables with mRMR on level-2 alarms in the training set for scenario AB.	158
6.6	Standardized process measurements of the first ten ranked variables with mRMR on process measurements in the training set for scenario AB.	159
6.7	Classification accuracy on training and test data for scenario AB and scenario BA for the first 10 variables with mRMR applied to process measurements and level-2 alarms.	160
6.8	Standardized measurements, level-1 alarms, and classification outcomes in scenario AB.	163
6.9	Standardized measurements, level-1 alarms, and classification outcomes in scenario BA.	164
6.10	Kernel density estimates of the probability density functions of the standardised variables for normal operation N0 and faults F1, F2, and F3 in operating point A.	168
7.1	Examples of persistent disturbances.	175
7.2	Examples of transient disturbances.	177
7.3	Transient weld disturbance in a glass forming process.	194

7.4	Anomaly indices for the transient weld disturbance.	194
7.5	Response to step changes in the shaft speed setpoint with overlapping oscillatory disturbance on a compressor rig.	197
A.1	Standardized process measurements of the first ten ranked variables with mRMR on level-2 alarms in the training set for scenario BA.	215
A.2	Standardized process measurements of the first ten ranked variables with mRMR on process measurements in the training set for scenario BA.	216

Nomenclature

Abbreviations

1NN	1 Nearest Neighbour
AAFT	Adjusted Amplitude Fourier Transform
ACT	Activation
AF	Alarm Flood
CCF	Cross-Correlation Function
CCM	Convergent Cross Mapping
CFT	Causal Fourier Transform
CRRA	Causal Relationship based on Residual Analysis
CVA	Canonical Variate Analysis
DCC	Alarm floods Detected and Correctly Classified
DCI	Dynamic time warping Causality Index
DCS	Distributed Control System
DIC	Alarm floods Detected and Incorrectly Classified
DTF	Directed Transfer Function
DTW	Dynamic Time Warping
FDR	False Detection Rate
GC	Granger Causality
HMI	Human-Machine Interface
ICA	Independent Component Analysis
kNN	k Nearest Neighbours
LDA	Linear Discriminant Analysis
MDR	Missed Detection Rate
MSPM	Multivariate Statistical Process Monitoring
MSW	Modified Smith Waterman
NLI	Nonlinearity Index
NN	Nearest Neighbours
NPMR	Nonparametric Multiplicative Regression
P&ID	Piping and Instrumentation Diagram
PCA	Principal Component Analysis
PDC	Partial Directed Coherence
PFD	Process Flow Diagram
PLS	Partial Least Squares
PWRI	Produced Water Reinjection
QDA	Quadratic Discriminant Analysis
RCA	Root Cause Analysis
RTN	Return To Normal
SVM	Support Vector Machine
TE	Transfer Entropy

Constants

$\tilde{N}$	number of vectors in the training set
M	number of variables X_m selected for the FDD
$M_{\mathcal{E}}$	number of unique alarm identifiers e_m
$M_{\mathcal{X}}$	total number of variables X_m
N	number of samples of X and Y
N_C	number of classes c (of alarm floods or faults)
N_v	number of embedded vectors $\vec{x}$ included in X

Indices

e_m	alarm identifier
I_n	time interval for computing the ACM
t_n	sampling times
$t_{\tilde{n}}$	sampling times corresponding to the $\tilde{N}$ vectors in the training set

Parameters

$\bar{k}$	multiplying constant for the level-1 alarm threshold
$\bar{\bar{k}}$	multiplying constant for the level-2 alarm threshold
κ	time delay
$\mathbf{b}$	kernel tolerance vector (for NPMR)
ν_{α_1}	alarm flood threshold
ν_{α_2}	intermediate alarm threshold
τ	sampling constant of the embedded vectors
h	prediction horizon
K	number of nearest neighbours
L	embedding dimension
l	lag for computing the ACM
M_a	number of principal components for PCA
P_{α}	probability detection threshold
Q_{α}	Q detection threshold
R	model order for the regression
T_{α}^2	T^2 detection threshold
t_e	end time of the alarm flood episode
t_s	start time of the alarm flood episode
T_w	size of the time window
T_{off}	time delay for the off-delay timer
T_{on}	time delay for the on-delay timer

Sets

$\tilde{\Omega}_F$	training set of process measurements in faulty operation
$\tilde{\Omega}_{N0}$	training set of process measurements in normal operation
$\mathcal{E}$	set of (unique) alarm identifiers
$\mathcal{L}$	set of lags
$\mathcal{X}$	set of all variables X_m
$\mathcal{X}_M$	set of M selected variables X_m
Ω_F	test set of process measurements in faulty operation
Ω_{N0}	test set of process measurements in normal operation

Observations

$\dot{x}_m(t_n)$	raw observation of X_m at time t_n
$\hat{c}(I_n)$	estimated class on interval I_n

$\hat{c}(t_n)$	estimated class at time t_n
$\tilde{\mathbf{x}}(t_n)$	normalized plant vector at time t_n
$\mathbf{r}(t_n)$	residual vector at time t_n
$\mathbf{S}(I_n)$	alarm coactivation matrix on interval I_n
$\mathbf{s}(I_n)$	input vector for the online alarm flood classification on interval I_n
$\mathbf{x}(t_n)$	plant vector at time t_n
$\nu(t_n, T_w)$	alarm rate at t_n over the last T_w window
$\tilde{x}_m(t_n)$	normalized observation of X_m at time t_n
$\tilde{\mathbf{x}}(t_n)$	embedded vector of $\mathbf{X}$ at time t_n
$c(I_n)$	class on interval I_n
$c(t_n)$	value of the class variable at time t_n
$f_{Q,m}(t_n)$	Q contribution of X_m at time t_n
$f_{T^2,m}(t_n)$	T^2 contribution of X_m at time t_n
$Q(t_n)$	value of the Q statistic at time t_n
$T^2(t_n)$	value of the T^2 statistic at time t_n
$x_m(t_n)$	mean-centered observation of X_m at time t_n

Random variables

$\bar{A}_m$	level-1 statistical alarm generated from X_m
$\bar{\bar{A}}_m$	level-2 statistical alarm generated from X_m
$\bar{C}$	estimated class
$\bar{\mathbf{A}}$	plant vector of level-1 statistical alarms
$\bar{\bar{\mathbf{A}}}$	plant vector of level-2 statistical alarms
$\mathbf{X}$	plant vector
$\mathcal{S}_A$	alarm sequence
A_m	alarm series
C	class
M_X	shadow manifold of X
X_m	process variable

Others

$\eta_{X \rightarrow Y}$	causality measure from X to Y
λ_m	m th singular value
$\mathbf{1}_N$	full one vector of size N
Σ_{M_a}	matrix containing the first M_a rows and columns of Σ
Σ	matrix with singular values for PCA
$\mathbf{I}_N$	identity matrix of size N
$\mathbf{P}$	matrix of the loading vectors for PCA
$\mathbf{U}, \mathbf{V}$	orthogonal matrices for PCA
$\mathcal{K}'(\cdot)$	kernel function for the kernel density estimation
$\mathcal{K}(\cdot, \cdot)$	kernel function for the support vector machine
$\phi(\cdot)$	mapping function corresponding to the kernel function $\mathcal{K}(\cdot, \cdot)$
σ_m	standard deviation of X_m
D_X	self-predictability of X
h_m	high absolute threshold of X_m
l_m	low absolute threshold of X_m
$s_{A_{m_1}, A_{m_2}}$	similarity between A_{m_1} and A_{m_2} based on the Jaccard similarity index

Chapter 1

Introduction

1.1 Description of the chapter

This thesis advocates the integration of alarm systems in industrial Fault Detection and Diagnosis (FDD). FDD is traditionally based on process measurements and the integration of new types of data from new data sources in FDD is a key research direction identified by the European project PRONTO that hosted this PhD. The PRONTO project is introduced in Section 1.2.

Section 1.3 gives an introduction to FDD in industrial processes, defining basic terminology and clarifying the objectives of FDD. Alarms, alarm systems, and the related discipline of alarm management are also introduced.

Section 1.4 discusses the main motivation for the thesis, which is increasing the reliability of data-driven FDD methods to improve their industrial applicability. Bringing information from the alarm systems into FDD is a means to increase the reliability of FDD methods, and this strategy is pursued in the first three technical chapters of this thesis. The last technical chapter aims at clarifying the field of applicability of data-driven root cause analysis methods based on process measurements, in order to tell the industrial practitioners in which scenarios they can confidently trust the results of the methods. The objectives of the thesis are formulated in Section 1.5 along with the corresponding contributions.

Finally, Section 1.6 gives an overview of the publications that resulted from this thesis.

1.2 Introduction to the project: PRONTO

The work presented in this thesis was carried out as part of a European Union Horizon 2020 innovation training network called PRONTO (for PROcess NeTwork Optimization). PRONTO focuses on process network optimization for efficient and sustainable operation of Europe’s process industries taking machinery condition and process performance into account [1]. PRONTO is a European Industrial Doctorate funded by the Marie Skłodowska-Curie Actions scheme under Horizon 2020, coordinated by Imperial College London.

PRONTO offered industrial PhD training to 14 Marie Skłodowska-Curie early stage researchers to tackle industrial challenges in the process industries. The training is offered under the European Industrial Doctorate scheme involving the non-academic sector in joint supervision of the doctoral training. The scheme emphasizes PhD projects with high industrial relevance and practical demonstrations.

The PRONTO consortium includes six universities, three research centers of an industrial technology company, and research groups of a chemicals manufacturer and of a stainless steel manufacturer. Three additional industrial partners provide application cases from oil and gas, chemicals, and refining. The research topics covered by the consortium are defined in [2] as:

- Data analytics to extract information from large, heterogeneous assemblies of data to determine machinery condition and process performance.
- Optimization of operations, materials, energy and wastes taking machinery condition and process performance into account.

PRONTO is divided into three Work Packages (WP) defined according to [2]:

- *WP1: Managing information.* WP1 addresses data analytics for assessment of the condition and performance of networks of equipment used for production in the process industries. Modern automation systems treat the process, low-pass filters, and electrical hardware of a site as an integrated system but there is a need for new methods able to merge information from disparate sources, both quantitative and qualitative.
- *WP2: Managing resources.* WP2 considers optimization of use of resources in process networks taking account of real-time information about the condition and performance of the process equipment.

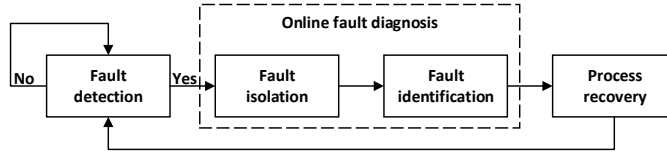


Figure 1.1: Online process monitoring loop adapted from [6].

- *WP3: Operating processes.* WP3 demonstrates new concepts for process operation based on real-time knowledge of the condition of interconnected equipment to achieve the best possible management of equipment and resources.

This thesis fits into WP1. The work presented in this thesis was conducted with the Optimization and Operational Analytics group at ABB Corporate Research Germany. The academic partner was Imperial College London.

1.3 Introduction to fault detection and diagnosis in industrial processes

1.3.1 Introduction to fault detection and diagnosis

Fault detection and diagnosis is usually integrated in the online process monitoring loop shown in Figure 1.1. The main steps of the online process monitoring loop are fault detection, fault identification, and fault isolation:

- *Fault detection* determines if a fault is present in a measurement and the time when the fault occurred [3].
- *Fault isolation* determines the location of the fault, i.e. the variables affected by the fault [4].
- *Fault identification* determines the type and magnitude of the fault [5]. The types of faults can be defined based on historical faults encountered on the plant, in which case the fault identification is equivalent to a classification problem that matches the new fault with one of the classes of historical faults.

Once the fault has been detected, isolated and identified, *process recovery* reverses the effects of the fault [4]. Process recovery can involve fault-tolerant control algorithms that

automatically perform process recovery, or can involve (real-time) manual intervention [7]. Process recovery is not in the scope of this thesis, however.

The *online fault diagnosis stage* covers fault isolation and fault identification, as indicated in Figure 1.1. When the results of the online fault diagnosis are not satisfying, i.e. that the type of fault could not be identified or the affected variables selected could not provide enough information on the fault, an offline fault diagnosis stage is performed [8, 9]. The *offline fault diagnosis stage* includes:

- *Root cause analysis* where the root cause or root causes of the fault are investigated.
- *Propagation path analysis* where the propagation of the fault is retraced.

In this thesis, the following terminology is adopted:

- *Online Fault Detection and Diagnosis* (online FDD) refers to fault detection, fault isolation, and fault identification.
- *Offline fault diagnosis* refers to Root Cause Analysis (RCA) and propagation path analysis.
- *Fault Detection and Diagnosis* (FDD) refers to both online fault detection and diagnosis and offline fault diagnosis.

FDD methods can be methods based on first-principle models, which means that they rely on a model built from *a priori* knowledge, or data-driven methods, which means that they learn directly from the available data [10]. Data-driven FDD methods have become more popular than model-based methods in the industry since they are cheaper and faster to implement. This thesis examines data-driven FDD methods, both for online FDD and offline fault diagnosis.

1.3.2 Fault detection and diagnosis for safe and profitable process operation

Fault detection and diagnosis plays an important role in the safe, profitable, and environmentally responsible operation of the plant [12]. Process safety is a major concern for all companies in the process industry and the fields of process safety and abnormal situation management receive a lot of attention from the academic community [13, 14, 15]. Process safety is extensively regulated with formal methods to design and maintain safety critical systems and promote functional safety [16]. Figure 1.2 summarizes the layers of protection designed to ensure a safe operation of the plant. Process design, process control, operator in-

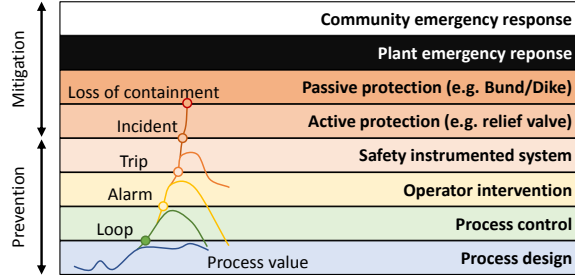


Figure 1.2: Layers of protection adapted from [11].

tervention, and the safety instrumented system constitute the prevention layers. The safety dimension in those four layers, including process design and process control, is critical. If an incident occurs, active and passive protection constitute the first part of the mitigation layers. If the incident cannot be contained, emergency response is required, both internal and external to the plant.

Fault detection and diagnosis belongs to the process control layer, at the border with the operator intervention layer. FDD is intended as a decision-support tool for operators to eventually trigger operator intervention. However, from a process safety perspective, FDD does not have the same value as the alarm system, which should constitute the primary source of information for operators during abnormal situations. Alarm systems are introduced in the next sub-section.

Nevertheless, the amount of data collected on industrial plants is transforming the field of abnormal situation management and data-driven FDD methods are expected to take an increasing role in supporting process safety [17, 18]:

- For the design and maintenance of safety systems (including alarm systems).
- For online decision-support systems during the operation of the plant.

Figure 1.3 shows the feedback model of the operator interaction [19]. FDD methods are increasingly integrated in the operator sub-system as support for decision-making for detecting, diagnosing and responding to deviations from the reference operation.

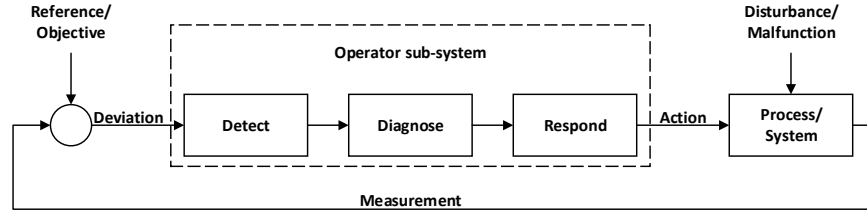


Figure 1.3: Feedback model of operator interaction reproduced from [19].

As highlighted in [12], FDD also has an important role in ensuring profitable operation by predicting and diagnosing abnormal situations to minimize plant shut down. According to [17], abnormal situations cost 27 billion US dollars per year in the United-Kingdom (estimation from 2005).

1.3.3 Alarm systems and alarm management

Alarms belong to the operator intervention layer of Figure 1.2 and play a major role in process safety since they are guiding human intervention in case of an abnormal situation. Alarms are generated by alarm systems that obey strict rules defined by interface standards [20] and alarm management standards [19, 21], including the IEC 62682 standard on management of alarms systems for the process industries [21]. According to the IEC 62682 standard, the primary function of the alarm system is to notify operators of abnormal process conditions or equipment malfunctions and support the response. The IEC 62682 standard suggests the following definitions:

- An *alarm* is an audible or visible means of indicating to the operator an equipment malfunction, process deviation, or abnormal condition requiring a timely response.
- An *alarm system* refers to the operator support system for generating and handling alarms for managing abnormal situations.
- An *alarm record* refers to a set of information which documents an alarm state change. The alarm record can be extracted from the alarm log or from the alarm historian. *Alarm log* refers to the short term repository for alarm records and *alarm historian* refers to the long term repository for alarm records.

Alarms and alarm systems are organized according to the principles of alarm management. The IEC 62682 standard defines *alarm management* as a collection of processes and practices for determining, documenting, designing, operating, monitoring, and maintaining

alarm systems. Alarm management is related to process safety and follows formal methods and procedures such as the ones that can be found in process safety.

1.4 Motivation of the thesis

1.4.1 Increasing reliability of data-driven fault detection and diagnosis

Much research has been published on data-driven FDD methods in the past years as highlighted by bibliometric reviews [22, 23]. Nevertheless, the limited number of FDD publications with industrial applications indicates that the presence of FDD methods in the industry is still limited compared to other process control methods [17]. The authors of [24] list the main challenges for the implementation of data-driven methods in the chemical industry, which can be applicable to the process industry as a whole:

- *Challenges due to diversity of data.* The diversity of data covers the heterogeneity of the recorded data (measurements such as flow and temperatures, alarm and events, analysers) and the variability in the operating conditions during the recording of the data (e.g. equipment added, operating mode changed). The integration of *a priori* knowledge such as information about the structure of the process (e.g. process connectivity and causality) is another challenge.
- *Challenges due to the quality of data.* The quality of data covers measurement uncertainty and variability of the measurements due to e.g. disturbances, sample randomness, measurement noise, or machine drifting.
- *Challenges due to the speed of collection of data.* Challenges due to the speed of collection of data include dealing with various sampling rates and developing adaptive methods that can learn online.

FDD methods are traditionally based on process measurements but the variability in the measurements and in the operating conditions have an impact on the reliability of data-driven FDD methods. The integration of new types of data (alarm and events, analysers) and *a priori* knowledge about the plant (plant connectivity) in FDD methods is a recurrent research direction to increase the reliability of the methods. This is one of the key research directions of the PRONTO innovative training network that hosted this project. This thesis advocates the integration of a particular data type, alarm data, into the FDD framework as

well as the integration of *a priori* knowledge included in the corresponding alarm systems. The next sub-section justifies this choice.

1.4.2 Integrating alarm systems in fault detection and diagnosis

Alarm management and fault detection and diagnosis are two disciplines that are closely related but still separated in the industry. Alarm management belongs to the sphere of process safety with established practices [25] and standards [21]. Alarms constitute the primary source of information for operators when abnormal situations occur, since they have been designed to guide operator intervention. In contrast, FDD has emerged from the process control community. FDD is also intended as a decision-support tool for the operators during abnormal situations, but the credit given to FDD method depends on the reliability of the methods in industrial environments. Most data-driven FDD methods have a limited transparency and explainability: for this reason, reliability of the FDD outcome plays an important role in the acceptance of the technology. While alarms will continue to remain the first and foremost decision-support system during abnormal situations for safety reasons, FDD methods can bring a precious complement of information to the operator. The motivation of this thesis is to increase the reliability of data-driven FDD methods by integrating data and *a priori* knowledge from the alarm system.

FDD is traditionally based on process measurements that correspond to the measurements of the physical quantities (i.e. the process variables) from the industrial process. Process measurements are key elements in alarm management, in particular for applications such as alarm design [26], but FDD applications taking alarm systems into account are limited. Recent works [27, 28] demonstrated the benefits of a preliminary analysis of alarm records before carrying the root cause analysis on the process measurements. The authors of [29], also part of PRONTO, recently proposed a Bayesian fusion framework to improve online fault detection and identification by combining process measurements and alarm records. Nevertheless, the integration of alarm systems in FDD has not been treated as full research question and remains a gap in the literature.

1.4.3 Clarifying the field of application of data-driven root cause analysis methods

Data-driven root cause analysis methods are another example of the importance of the reliability of data-driven FDD methods. RCA methods do not impact process safety since

they are applied offline to investigate the cause of the disturbance, but they provide a basis for the resolution of the issue that led to the fault. Manual root cause analysis can be complex in industrial processes that are highly interconnected, and data-driven methods are used to speed up the analysis. However, the acceptance of data-driven RCA methods is conditioned to the reliability of the results.

A first step to increase the reliability of data-driven RCA methods is to clarify the scope of each method. Many methods have been proposed in the literature on specific case studies but the field of application of each method has not been discussed in a systematic manner. For this reason, it is currently difficult for industrial practitioners to know when a given method should be trusted. In particular, most methods have been designed for RCA of oscillatory disturbances, and RCA of non-oscillatory disturbances is considered as an unsolved problem in the industry. The motivation of this thesis is to increase the reliability of data-driven RCA methods by determining the scenarios where each method is applicable and by providing recommendations on the RCA procedure to adopt for each scenario, including RCA of non-oscillatory disturbances.

1.5 Introduction to the thesis

1.5.1 Objectives of the thesis

The main motivation of this thesis is increasing the reliability of data-driven FDD methods to improve the applicability and the acceptance of FDD methods in the industry. This thesis investigates one strategy in particular, which is taking alarm systems into consideration in FDD methods. This strategy is illustrated through the first two objectives of this thesis. The third objective relates to the applicability of data-driven RCA methods. The three objectives of the thesis are:

- *Objective 1:* to evaluate an online FDD approach to online alarm data analysis in order to provide support to the operators during abnormal situations, e.g. during alarm flood episodes where the number of alarms is high. Online FDD approach means online detection and diagnosis of abnormal situations when they occur.
- *Objective 2:* to bring alarm systems and alarm management into online FDD to improve the robustness of data-driven online FDD methods with regard to the variability in the process measurements. This includes bringing data (e.g. alarm records), contextual information (e.g. alarm settings), and practices (e.g. alarm processing tech-

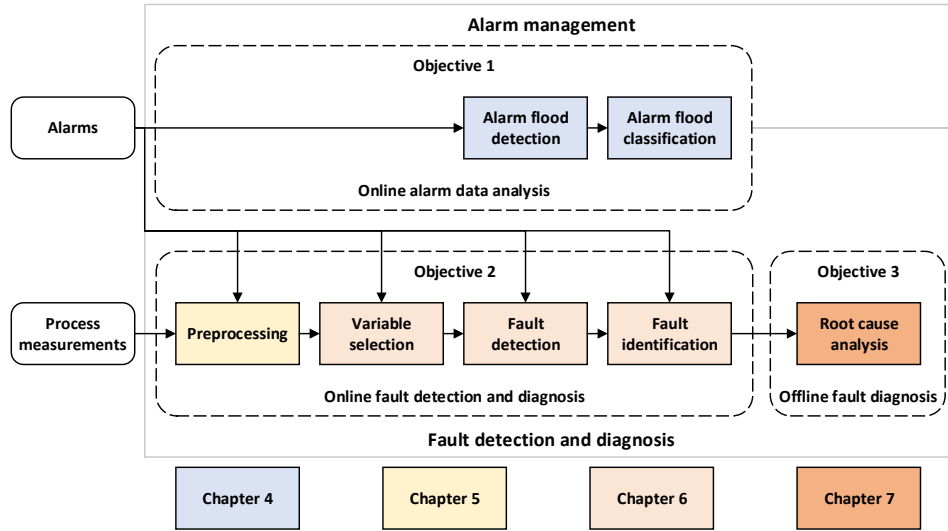


Figure 1.4: Graphical summary of the thesis. The colour codes show in which chapter the topics are discussed.

niques) from alarm management and alarm systems into online FDD that was, until now, mostly based on process measurements.

- *Objective 3:* to determine the scope and applicability of existing data-driven root cause analysis methods for offline fault diagnosis to guide industrial practitioners and facilitate the application of those methods. Since most methods have been designed for oscillatory disturbances, this includes clarifying the procedure for RCA of oscillatory disturbances, and proposing a procedure for RCA of non-oscillatory disturbances.

1.5.2 Contributions of the thesis

The contributions of the thesis can be linked to the objectives of the previous sub-section. Figure 1.4 gives a graphical summary of the contributions of the thesis and how they match the research objectives.

Contribution to Objective 1

As indicated in Figure 1.4, Objective 1 deals with online alarm data analysis for alarm flood detection and diagnosis. The state of the art presented in Chapter 3 highlights that alarm data analysis methods for online decision-support have all been treated as a sequence

mining problem, i.e. that the alarm record is considered as an alarm sequence (defined in Chapter 2). This thesis argues that those sequence mining methods have not initially been designed for online applications and that their online implementation in the plant can be challenging. Chapter 4 suggests a new time series analysis approach to the problem of online alarm flood classification that is similar to the traditional online FDD approach, except that it deals with alarm binary series (defined in Chapter 2) instead of continuous process measurements. The simplicity of the online implementation of the method makes it a good candidate for industrial applications.

Contribution to Objective 2

Objective 2 deals with the integration of data, contextual information, and practices from alarm systems and alarm management into online FDD methods. The contribution of the thesis to this objective has been separated into two technical chapters:

- Chapter 5 demonstrates how contextual information from the alarm systems, namely the alarm settings, can be used to improve the preprocessing of the process measurements for online FDD. Chapter 5 suggests a new normalization of the process measurements based on the corresponding alarm thresholds defined based on engineering and safety considerations. The proposed normalization constitutes an improvement compared to traditional normalization practices since it places the variations of the process measurements in the context of their safe range of operation, which reduces the impact of noise in the FDD.
- Chapter 6 demonstrates how statistical alarms along with delay timers can be used to improve the robustness of online FDD methods to noise. A statistical alarm is a specific type of alarm generated from statistical processing of the process measurements and delay timers are a common alarm processing technique that delays the alarm transition until the process measurement has been continuously crossing the alarm threshold for a given period of time. The statistical alarms help with the selection of the most relevant variables for online FDD during the modeling stage, and the statistical alarms are used in combination with the process measurements for fault detection and identification.

Contribution to Objective 3

Objective 3 covers offline root cause analysis. Chapter 3 introduces twelve data-driven RCA methods proposed in the literature in the last years. Those methods are reviewed in Chapter

7 according to three criteria: the characteristics of the method, the effect of the disturbance propagation on the method, and the types of disturbances covered by the method. This review derives guidelines for RCA of oscillatory and non-oscillatory disturbances in the industry. Chapter 7 discusses the case of RCA of non-oscillatory disturbances, which has not been addressed in the literature.

1.5.3 Outline of the thesis

This chapter set the frame of this thesis and introduced the motivations and research objectives that guided the work.

Chapter 2 provides necessary background material, including a review of data sources in the process industry and a technical introduction to alarm systems and alarm management. The notations and terminology relating to both online FDD and offline fault diagnosis are also introduced, along with the case studies used in Chapter 4 to 6.

Chapter 3 gives an overview of the state of the art in alarm data analysis and FDD. The state of the art in alarm data analysis distinguishes sequence mining methods and time series analysis methods to justify the binary series approach to online alarm flood classification proposed in Chapter 4. The state of the art in online FDD covers the stages of the FDD workflow that are of interest for this thesis, i.e. preprocessing, variable selection, fault detection, fault isolation, and fault identification. The last part of Chapter 3 introduces the twelve data-driven RCA methods reviewed in Chapter 7, grouped according to the RCA strategy adopted: univariate methods, time-delay estimation methods, prediction-based methods in the time-domain, and prediction-based methods in the frequency-domain.

Chapter 4 to 7 correspond to the technical chapters of this thesis as indicated in Figure 1.4. Chapter 4 presents a binary alarm series approach to online alarm flood classification. Chapter 5 introduces the alarm-range normalization based on alarm thresholds for online FDD using process measurements. Chapter 6 suggests the use of statistical alarms in online FDD, combining the statistical alarms and the original process measurements. Chapter 7 provides a review and discussion of existing data-driven RCA methods and presents some thoughts to open the debate on RCA of non-oscillatory disturbances.

Chapter 8 provides a critical evaluation of the contributions of the thesis compared to the research objectives and identifies opportunities for further work. Concluding remarks on the project are given in Chapter 9.

1.6 Publications

Some material presented in this thesis has been published in conferences or journals. The corresponding references are given at the beginning of each chapter, when relevant. The conference proceedings include:

- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “Online alarm flood classification using alarm coactivations,” in *Proceedings of 10th ADCHEM Symposium, Shenyang, China, July 25-27, IFAC-PapersOnLine*, vol. 51, pp. 345–350, 2018
- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “On improving fault detection and diagnosis using alarm-range normalisation,” in *Proceedings of 10th SAFEPROCESS Symposium, Warsaw, Poland, August 29-31, IFAC-PapersOnLine*, vol. 51, pp. 1227–1232, 2018
- M. Lucke, X. Mei, A. Stief, M. Chioua, and N. F. Thornhill, “Variable selection for fault detection and identification based on mutual information of alarm series,” in *Proceedings of 12th DYCOPS Symposium, Florianópolis, Brazil, April 23 - 26, IFAC-PapersOnLine*, vol. 52, pp. 673–678, 2019

The journal publications include:

- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “Advances in alarm data analysis with a practical application to online alarm flood classification,” *Journal of Process Control*, vol. 79, pp. 56–71, 2019
- M. Lucke, A. Stief, M. Chioua, J. R. Ottewill, and N. F. Thornhill, “Fault detection and identification combining process measurements and statistical alarms,” *Control Engineering Practice*, vol. 94, p. 104195, 2020
- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “Integration of alarm design in fault detection and diagnosis through alarm-range normalization,” *Submitted to Control Engineering Practice*

Chapter 2

Background

2.1 Description of the chapter

The integration of heterogeneous data sources in fault detection and diagnosis is an active research direction to improve the robustness of FDD methods to variability in the input data and in the operating conditions, and to improve the accuracy of the methods. Section 2.2 introduces three major types of data that will be considered in this thesis: process measurements, alarm and events, and plant connectivity.

This thesis deals specifically with the integration of alarm systems within the traditional FDD framework. Section 2.3 provides necessary background information about alarm systems in the process industry and details the alarm design and alarm processing techniques relevant for Chapter 5 and Chapter 6. The various approaches to represent alarm data from the alarm records are also presented.

The discipline of fault detection and diagnosis in industrial processes is introduced in Section 2.4. Faults and disturbances are defined, and the traditional workflow for industrial FDD is presented, covering online FDD and offline fault diagnosis. Special attention is also given to root cause analysis that will be the topic of Chapter 7.

Section 2.5 introduces the case studies used in the rest of the thesis: the offshore oil and gas separation plant used in Chapter 4 and Chapter 5, and the multiphase flow facility used in Chapter 6. Section 2.6 summarizes the assumptions holding in the thesis.

2.2 Data sources in the process industry

2.2.1 Process measurements

Control systems require measurements which are generated by sensors. While some of the sensors are local sensors (i.e. only available on site for the operator to read directly), most of the measurements are linked to a Distributed Control System (DCS) where they are first temporarily stored for a few months, and then stored in a process historian for long-term archiving.

The measurements include:

- *Process measurements*, such as temperature, level, flow rate and pressure. The standard sampling rate in a control loop is 10 Hz, but the signals are down-sampled to the minute when stored in the process historian.
- *Mechanical measurements*, such as vibrations. The sampling rate is usually 1 kHz.
- *Electrical measurements*, such as voltage and current. They are usually sampled at several kHz (e.g. 5 kHz).

Other available measurements are measurements from analysers, which can feature even longer sampling periods (from 10 to 15 minutes for online analysers if they are used in the control loop to a few hours for offline measurements done manually in the laboratory).

2.2.2 Alarms and events

The term *alarms and events* covers all types of event data available on a plant. *Event data* refer to event-triggered data that are recorded when an event occurs, as opposed for example to process measurements that are time-triggered with a given sampling time. Event data can be automatically generated by a system or can be entered manually.

Alarms are generated by alarm systems according to the rules defined in interface standards [20] and alarm management standards [19, 21]. The alarms are generated from logic and from measurements of process conditions based on field devices (i.e. sensors, logic solvers, and final elements) of a safety instrumented system and basic process control system. The safety instrumented system and the basic process control system are both included in the alarm system. According to [21], the alarm system also contains:

- A *Human-Machine Interface* (HMI) to communicate the alarms to the operator.
- An *alarm and event log* and an *alarm historian* as defined in Section 1.3.3.

- A *master alarm database* that stores the list of alarms and the associated alarm attributes defined in Section 2.3.1.

Events generally refer to event data types other than alarms. An example of an event is manually-entered events by the operators to record e.g. manual interventions. This thesis examines alarms and alarm systems, and Section 2.3 gives a detailed overview of alarm systems in the process industry.

2.2.3 Plant connectivity

Plant connectivity describes physical or information flows between process units. Process flow diagrams (PFD) and piping and instrumentation diagrams (P&ID) are the main resources for specifying connectivity. PFDs and P&IDs are described in many books [36] and follow the ANSI/ISA-5.1 standard.

Work has been done to convert PFDs and P&IDs into computer-friendly standard formats in order to facilitate automatic processing of the information. Those standard formats include [37]:

- *Adjacency matrix.* The adjacency matrix is a matrix representation of the direct connectivity between process units. Direct connections are indicated by ones, the rest by zeros [38]. Another binary matrix, the reachability matrix can be derived from the adjacency to indicate both direct and indirect connections.
- *Digraph.* The digraph is a graphical representation of the adjacency matrix where process units are represented by nodes and ones in the adjacency matrix are represented by arcs [39].
- *Semantic web model.* Semantic web models such as the ontology framework represent conceptual models of process units and their links defined using computer understandable language. Computer Aided Engineering Exchange (CAEX) description and eXtensible Markup Language (XML) files are generally used to construct the process topology, following the work of [40] and [41].

This thesis does not directly deal with plant connectivity information but several methods in the literature exploit plant connectivity for alarm data analysis or FDD. In particular, plant connectivity information can be very helpful for root cause analysis and it has become a common practice in the industry to combine data-driven methods based on process measurements for causality analysis with plant connectivity from PFDs and P&IDs [41].

Additional details about the link between connectivity analysis and causality analysis are given in Section 2.4.3, and methods for RCA are reviewed in Section 3.4.

2.3 Alarm systems in the process industry

2.3.1 Introduction to alarm systems

Alarm systems are central to this thesis. This sub-section presents the key characteristics of alarms and alarm systems. The alarm attributes define the alarm settings and the format of alarm data. The alarm types describe the alarm generation mechanism. The alarm management lifecycle explains how alarms are designed, monitored, and maintained. Alarm performance metrics give indications on what a good alarm system is, and on the problems that need to be addressed.

Alarm attributes

The *alarm attributes* provide information about the alarm settings of each alarm configured in the alarm system [25]. Alarm settings are conceived during the engineering stage of the plant based on safety and technical considerations [21]. The alarm settings are stored in the master alarm database. The IEC 62682 standard [21] lists and defines the alarm attributes that should be available for each alarm:

- *Alarm description*: all alarms shall have an informative text provided as a tag description, or alarm description, or both. The alarm description usually follows a structured format.
- *Alarm threshold or logical conditions*: threshold value of a process variable or discrete state that triggers the alarm indication.
- *Alarm type*: alarm attribute which gives a distinction of the alarm condition. The various alarm types are reviewed in the next sub-section.
- *Alarm priority*: relative importance assigned to an alarm within the alarm system to indicate the urgency of response (e.g., seriousness of consequences and allowable response time).
- *Alarm group*: set of alarms with common association (e.g., process unit, process area, equipment set, or service).

Table 2.1: Example of alarm record captured during an alarm flood 70 on the oil and gas separation plant presented in Section 4.2.3.

Tag name	Description	State change	Priority	Type	Time
AFICA130	Alarm L	ACT	850	ABS	06:26:47
AFZAL115	$\leq 340.0 \text{ Sm}^3 \text{ h}^{-1}$ Alarm LL	ACT	760	ABS	06:26:47
AFICA116	Alarm L	ACT	850	ABS	06:26:48
AFZT115	$\leq 240.0 \text{ Sm}^3 \text{ h}^{-1}$ Alarm LL	ACT	760	ABS	06:26:48
AFZT114	$\leq 470.0 \text{ Sm}^3 \text{ h}^{-1}$ Alarm L	ACT	860	ABS	06:26:48
AFZT127	Alarm LL	ACT	750	ABS	06:26:48
APT121	Alarm L	ACT	850	ABS	06:26:49
AFZT114	$\leq 470.0 \text{ Sm}^3 \text{ h}^{-1}$ Alarm L	ACT	860	ABS	06:26:49
AFZT127	Alarm LL off	RTN	750	ABS	06:26:51
AFICA130	Alarm L off	RTN	850	ABS	06:26:53
AFZT115	$\leq 240.0 \text{ Sm}^3 \text{ h}^{-1}$ Alarm LL	ACT	760	ABS	06:26:54
AFZAL115	$\leq 340.0 \text{ Sm}^3 \text{ h}^{-1}$ Alarm LL	ACT	760	ABS	06:26:55
AFICA116	Alarm L off	RTN	850	ABS	06:26:56
AFZAL115	$\leq 340.0 \text{ Sm}^3 \text{ h}^{-1}$ Alarm LL	ACT	760	ABS	06:26:57
AFICA116	Alarm L	ACT	850	ABS	06:26:58

- *Alarm on-delay*: time a process measurement remains in the alarm state before the alarm is annunciated. Alarm on-delay timers are detailed in Section 2.3.3.
- *Alarm off-delay*: time a process measurement remains in the normal state before the alarm becomes inactive. Alarm off-delay timers are detailed in Section 2.3.3.
- *Alarm deadband*: a change in signal from the alarm threshold necessary for the alarm to return to normal. Alarm deadbands are detailed in Section 2.3.3.
- *Alarm message*: text string displayed with the alarm indication that provides additional information to the operator.

According to the IEC 62682 standard [21], an alarm record must contain at least six alarm record attributes. An example of alarm record taken from the alarm log of the oil and gas separation plant during an alarm flood is presented in Table 2.1 to illustrate the structure of the alarm record. The six required alarm record attributes are [21]:

- *Alarm tag name*. The tag name refers to the unique identifier assigned to a process measurement, calculation, or device within the control system [21]. The tag name

follows a defined structure. In Table 2.1, the structure indicates respectively the process area, the measurement type, and the measurement number. In AFICA130 for example, A refers to process area A, FICA to Flow Indicator Control, and 130 to the measurement number 130.

- *Alarm description.* The alarm description complements the alarm tag. In Table 2.1, it can indicate the alarm level (i.e. L for low level, LL for very low level, H for high level, and HH for very high level) and the alarm threshold.
- *Alarm state change.* The alarm state can be active or inactive. The transition from an inactive state to an active state is called ACTivation (ACT) and the transition from an active state to an inactive state is called Return To Normal (RTN). Other alarm states exist but, in this thesis, the alarm state change record attribute will be limited to ACT and RTN (such as in Table 2.1).
- *Alarm priority.* The alarm priority attributes of Table 2.1 correspond to positive numbers: the higher the number, the higher the priority. It is common practice in industry to define high-priority alarms as alarms with a priority greater than 750 [42]. Table 2.1 only shows high-priority alarms.
- *Alarm type.* The alarm type attributes of Table 2.1 only show ABS which stands for ABSolute alarms, defined in the next sub-section. Absolute alarms are the most common type of alarms in the industry.
- *Date and time of occurrence of the alarm state change.* The date and time of the alarm state change are usually shown but Table 2.1 only shows the time for confidentiality reasons.

Alarm types

The alarm type is an important alarm attribute. Various types of alarms can be defined during the alarm design stage corresponding to various alarm generation mechanisms [21]:

- *Absolute alarm:* alarm generated when the alarm threshold is exceeded.
- *Adaptive alarm:* alarm for which the threshold is changed by an algorithm (e.g., rate based).
- *Deviation alarm:* alarm generated when the difference between two values exceeds a limit (e.g. deviation between primary and redundant instruments or a deviation between process measurement and threshold).

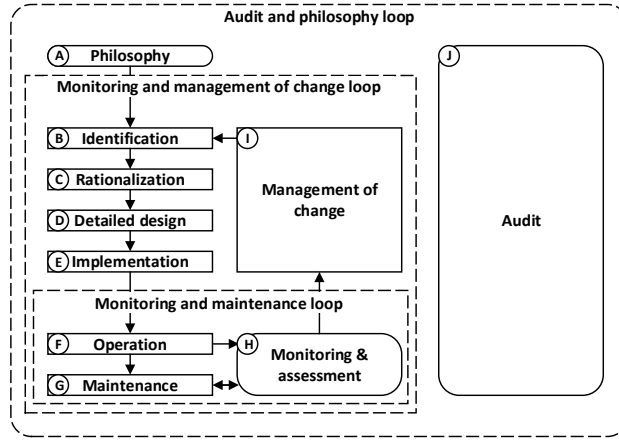


Figure 2.1: Alarm management lifecycle reproduced from [19].

- *Rate-of-change alarm*: alarm generated when the change in process measurement per unit time exceeds a defined threshold.
- *Statistical alarm*: alarm generated based on statistical processing of a process measurement or measurements.

Two types of alarms are investigated in this thesis: absolute alarms in Chapter 5 and statistical alarms in Chapter 6. Sub-section 2.3.2 gives more details about those two types of alarms.

Alarm management lifecycle

Alarm systems are defined and maintained according to the rules of alarm management. The stages of the alarm management lifecycle are summarized in Figure 2.1 and defined in the IEC 62682 standard [21]:

1. *Philosophy*. The alarm philosophy stage serves as the basis for the specifications of the alarm system. The philosophy provides operational definitions and criteria (e.g. alarm prioritization guidelines or performance metrics) used to design the alarm system and specifies processes for the alarm management lifecycle stages [21].
2. *Identification*. The identification stage identifies the potential alarms needed on the plant based on a set of methods. Those methods include process hazards analysis, safety requirements specifications, operating procedure reviews, P&ID development,

and environmental permits. Process modifications and monitoring of the alarm system can also create the need for new alarms or alarm changes [21].

3. *Rationalization.* The rationalization stage aligns the need for a potential alarm with the alarm philosophy. The requirements serving as the basis for the design (e.g. alarm threshold, consequence threshold and the corresponding consequence, corrective action that can be taken by the operator) are generated during this stage and stored in the master alarm database [21].
4. *Detailed design.* The design stage corresponds to the specification of the alarm attributes following the requirements from the rationalization stage. The design stage includes the basic alarm design where each alarm is designed according to guidelines depending on the alarm type, the HMI design, and the design of advanced alarming techniques [21].
5. *Implementation.* The implementation stage corresponds to the installation of the alarm which includes physical and logical installation, functional verification of the system, as well as operator training [21].
6. *Operation.* In the operation stage, the alarm is in service [21].
7. *Maintenance.* In the maintenance stage, the alarm is being tested or repaired [21].
8. *Monitoring and assessment.* The monitoring and assessment stage compares the performance of the alarm system and individual alarms with the performance objectives from the alarm philosophy. Monitoring can highlight need for changes in the alarm system or in the operating procedures [21].
9. *Management of change.* The management of change stage suggests modifications to the alarm system and each new change shall follow the alarm management lifecycle stages from identification to implementation [21].
10. *Audit.* The audit stage corresponds to a review of the alarm system and alarm management processes to highlight gaps that could not be detected in the monitoring stage [21].

Alarm management is safety-critical and must obey established guidelines derived from the fields of process safety and functional safety. Still, a growing interest of the academia in alarm management has been observed [26].

Table 2.2: Recommended alarm performance metrics adapted from the IEC 62682 standard [21].

Metric	Target value
Annunciated alarms per operator console per day	Acceptable value: 144 Maximum value: 288
Annunciated alarms per operator console per hour	Acceptable value: 6 (average) Maximum value: 12 (average)
Annunciated alarms per operator console per minute	Acceptable value: 1 (average) Maximum value: 2 (average)
% of hours containing more than 30 alarms	< 1%
% of 10-minute intervals containing more than 10 alarms	< 1%
Maximum number of alarms in a 10-minute interval	≤ 10
% of time the alarm system is in a flood condition	< 1%
% contribution of the top ten most frequent alarms to the total alarms	< 1% to 5% maximum
Quantity of chattering alarms	Zero
Annunciated priority distribution	80% low, 15% medium, 5% high

Alarm performance metrics

Table 2.2 lists recommended alarm performance metrics from the IEC 62682 standard [21]. It highlights several performance indicators:

- The total number of alarms (or the alarm rate, i.e. number of annunciated alarms per operator in a specific time interval [21]). The accepted value should correspond to an average of one alarm per operator console per minute, i.e. 144 alarms per operator console per day.
- The frequency of *alarm flood* episodes, defined as condition during which the alarm rate is greater than the operator can effectively manage [21]. In the standard [21], an episode is considered as an alarm flood if it contains more than 10 alarms per 10 minutes. According to Table 2.2, less than 1% of the 10-minute intervals should be alarm floods.
- The frequency of occurrence of the most frequent alarms (referred to as “bad actors”). Table 2.2 indicates that the number of occurrences of the top ten most frequent alarms should not constitute more than 5% of the total number of alarm occurrences.

- The quantity of *chattering alarms*, i.e. alarms with repeated transitions between the alarm state and the normal state in a short period of time [21]. There should be no chattering alarms according to Table [21]. Methods to address alarm chattering such as delay timers and alarm deadbands are detailed in Section 2.3.3.
- The alarm priority distribution. High priority alarms should constitute no more than 5% of the total alarms.

Those performance indicators give an overview of the objectives pursued by the various alarm data analysis methods reviewed in Section 3.2 such as reducing the total number of alarms, reducing the frequency of alarm flood intervals or supporting the operators during the alarm flood intervals, and reducing the quantity of alarm chattering.

Chapter 4 tackles the problem of online classification of alarm floods to provide guidance to the operators when alarm flood intervals occur. Chapter 6 uses common methods from alarm management to deal with alarm chattering, namely on-delay and off-delay timers, to improve the robustness of the online fault detection and identification to noise.

2.3.2 Alarm design

This section gives additional details about absolute alarms used in Chapter 5 and statistical alarms used in Chapter 6.

Absolute alarms

Absolute alarms are activated when the alarm threshold is exceeded. The choice of the alarm threshold is a major component of the alarm design stage. Figure 2.2 illustrates the alarm design procedure for absolute alarms and their alarm thresholds. The *consequence threshold* is first defined as the limit after which a consequence begins to occur (e.g. for the level of a tank, it could be the maximum level the tank can admit before overflow). The alarm threshold is defined from the consequence threshold taking into account the process dead-time, the operator response delay and the acknowledgement delay for the corresponding process measurement.

A high and a low alarm threshold are usually defined for each process variable with an absolute alarm. For each absolute alarm A_m associated with a process variable X_m , the high alarm threshold is h_m and the low alarm threshold is l_m .

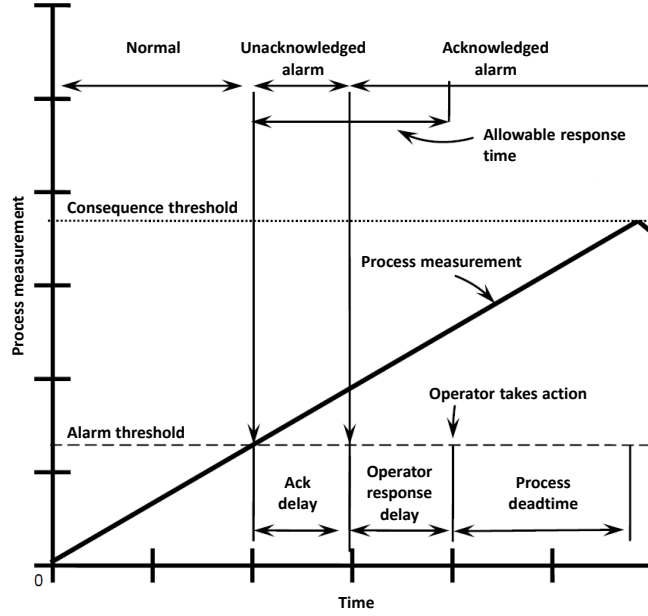


Figure 2.2: Alarm response timeline adapted from the IEC 62682 standard [21].

Chapter 5 examines absolute alarms and their alarm thresholds for use in preprocessing measurements for FDD.

Statistical alarms

Statistical alarms are a specific type of alarm generated based on statistical processing of a process measurement [21]. A common statistical indicator used to generate statistical alarms is the standard deviation σ_m of each process variable X_m during normal operation in the historical data. A statistical alarm $\bar{A}_m$ is generated from X_m as:

$$\bar{a}_m(t_n) = \begin{cases} -1, & \text{if } x_m(t_n) \leq -\bar{k}\sigma_m \\ 0, & \text{if } -\bar{k}\sigma_m < x_m(t_n) < \bar{k}\sigma_m \\ 1, & \text{if } x_m(t_n) \geq \bar{k}\sigma_m \end{cases} \quad (2.1)$$

where $x_m(t_n)$ is the mean-centered observation of X_m at time t_n (defined in Sub-section 2.4.2), $\bar{k}$ is a constant ($\bar{k} > 0$) that can be tuned for each alarm, and $\bar{k}\sigma_m$ is called the alarm threshold.

In the following, the alarm $\bar{A}_m$ generated from X_m according to Eq. 2.1 with the constant $\bar{k}$ is described as level-1 alarm. A level-2 alarm $\bar{\bar{A}}_m$ is also generated from X_m according to Eq. 2.1 with the constant $\bar{\bar{k}}$ ($\bar{\bar{k}} > \bar{k} > 0$). $\bar{k}\sigma_m$ indicates a low severity threshold and $\bar{\bar{k}}\sigma_m$ indicates a high severity threshold. Commonly used values for $\bar{k}$ and $\bar{\bar{k}}$ are respectively 3 and 5, following statistical process control rules.

Chapter 6 investigates the use of statistical alarms in fault detection and identification.

2.3.3 Alarm processing

Alarm chattering is usually removed from alarms to limit nuisance to the operators [43]. Several methods have been introduced to limit alarm chattering while maintaining a trade-off between false alarms and missed alarms, such as delay timers and deadbands [44].

Alarm delay timers

The on-delay and the off-delay are two alarm attributes as indicated in Sub-section 2.3.1. The IEC 62682 standard [21] suggests the following definitions:

- The *on-delay* is used to avoid unnecessary alarms when a measurement temporarily overshoots its threshold, thus preventing the alarm from being triggered until the signal remains in the alarm state continuously for a specified length of time T_{on} .
- The *off-delay* is used to reduce chattering alarms by locking in the alarm indication for a certain holding period T_{off} after the process condition has returned to normal.

In Chapter 6, on-delay and off-delay are applied to statistical alarms $\bar{A}_m$ and $\bar{\bar{A}}_m$, i.e. the alarm should cross the threshold for T_{on} before being activated (on-delay timer) and the alarm should go below the threshold for T_{off} before being deactivated (off-delay timer). T_{on} and T_{off} are tuned according to the dynamics of the corresponding variable.

Alarm deadband

The alarm deadband is another alarm attribute listed in Sub-section 2.3.1. According to the IEC 62682 standard [21], the alarm deadband requires the process variable to cross the alarm threshold into the normal operating range by some defined increment or percentage of the range. Deadbands are typically set based on the normal operating range of the process variable, measurement noise, and the type of process variable.

2.3.4 Alarm data representations

There are two main alternatives to represent alarm data from an alarm record: alarm sequences and alarm series. An alarm sequence represents alarm data as a sequence of events (in this case, events are alarm activations and possibly alarm deactivations), that might or might not be supplemented by the timestamp at which the event occurred. By contrast, an alarm series is a time series that indicates the status of alarm variables at regular sampling times. The alarm data representation has an incidence on the type of method used for analysing the alarm record, as explained in the state of the art for alarm data analysis in Section 3.2. This section presents alarm sequences and alarm series that have been used in the alarm data analysis literature.

Alarm sequences

An alarm sequence $\mathcal{S}_{\mathcal{A}}$ consists of a sequence of alarm tuples $\mathcal{A}_n$ ($n = 1 \dots N_{\mathcal{A}}$) ordered by their time of appearance, where n indicates here the position of $\mathcal{A}_n$ in the sequence and $N_{\mathcal{A}}$ is here the number of alarm tuples in the sequence, i.e.:

$$\mathcal{S}_{\mathcal{A}} = (\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_{N_{\mathcal{A}}}) \quad (2.2)$$

Each alarm tuple contains at least an alarm identifier e which is the unique identifier of the alarm (e.g. FC101), and can also contain the activation time t of the alarm (e.g. 06:12:02) and the deactivation time t' of the alarm (e.g. 06:14:41). The activation of an alarm is referred to as ACT, and the deactivation of an alarm is referred to as RTN. In the following, the alarm identifier corresponding to the alarm tuple $\mathcal{A}_n$ is indicated as $e_{\mathcal{A}_n}$, the activation time of $\mathcal{A}_n$ is indicated as $t_{\mathcal{A}_n}$, and the deactivation time of $\mathcal{A}_n$ is indicated as $t'_{\mathcal{A}_n}$.

According to the definition of the alarm tuple $\mathcal{A}_n$, the alarm sequence can have various names:

- *Alarm sequence.* The alarm tuple contains only the alarm identifier, i.e. $\mathcal{A}_n = (e_{\mathcal{A}_n})$, and therefore the alarm sequence is defined as:

$$\mathcal{S}_{\mathcal{A}} = ((e_{\mathcal{A}_1}), (e_{\mathcal{A}_2}), \dots, (e_{\mathcal{A}_{N_{\mathcal{A}}}})) \quad (2.3)$$

- *Time-stamped sequence.* The alarm tuple contains the alarm identifier and the activation time of the alarm, i.e. $\mathcal{A}_n = (e_{\mathcal{A}_n}, t_{\mathcal{A}_n})$, and therefore the alarm sequence is

defined as:

$$\mathcal{S}_{\mathcal{A}} = ((e_{\mathcal{A}_1}, t_{\mathcal{A}_1}), (e_{\mathcal{A}_2}, t_{\mathcal{A}_2}), \dots, (e_{\mathcal{A}_{N_{\mathcal{A}}}}, t_{\mathcal{A}_{N_{\mathcal{A}}}})) \quad (2.4)$$

- *RTN time-stamped sequence.* The alarm tuple contains the alarm identifier, the activation time of the alarm and the deactivation time of the alarm, i.e. $\mathcal{A}_n = (e_{\mathcal{A}_n}, t_{\mathcal{A}_n}, t'_{\mathcal{A}_n})$, and therefore the alarm sequence is defined as:

$$\mathcal{S}_{\mathcal{A}} = ((e_{\mathcal{A}_1}, t_{\mathcal{A}_1}, t'_{\mathcal{A}_1}), (e_{\mathcal{A}_2}, t_{\mathcal{A}_2}, t'_{\mathcal{A}_2}), \dots, (e_{\mathcal{A}_{N_{\mathcal{A}}}}, t_{\mathcal{A}_{N_{\mathcal{A}}}}, t'_{\mathcal{A}_{N_{\mathcal{A}}}})) \quad (2.5)$$

An alternative to the alarm sequence $\mathcal{S}_{\mathcal{A}}$ is the *alarm set* $\mathcal{S}'_{\mathcal{A}}$, where the set of the $N'_{\mathcal{A}}$ unique alarm identifiers e appearing in the alarm sequence $\mathcal{S}_{\mathcal{A}}$ is represented:

$$\mathcal{S}'_{\mathcal{A}} = \{e_1, e_2, \dots, e_{N'_{\mathcal{A}}}\} \quad (2.6)$$

Alarm series

Alarm data can also be represented as a time series. An alarm series A_m can be defined for each unique alarm identifier $e_m \in \mathcal{E}$, where $\mathcal{E}$ is the set of all the $M_{\mathcal{E}}$ alarm identifiers defined on the plant (e.g. $\mathcal{E} = \{\text{FC101}, \text{FC102}, \text{FZT101}, \dots\}$) and $m = 1 \dots M_{\mathcal{E}}$.

Let $\mathcal{T}_m^1$ be the set of times when the alarm with identifier e_m gets activated, $\mathcal{T}_m^0$ be the set of times when the alarm with identifier e_m gets deactivated, and $\mathcal{T}_m$ be the set of times when the alarm with identifier e_m is active (i.e. all the times between the activation and the deactivation times).

The alarm series A_m associated with the alarm identifier e_m can take various forms:

- *ACT binary series.* A_m is set to 1 only on the activation times of the corresponding alarm:

$$a_m(t_n) = \begin{cases} 1 & \text{if } t_n \in \mathcal{T}_m^1 \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

- *Binary series.* A_m is set to 1 on the times when the corresponding alarm is active:

$$a_m(t_n) = \begin{cases} 1 & \text{if } t_n \in \mathcal{T}_m \\ 0 & \text{otherwise} \end{cases} \quad (2.8)$$

- *Multi-valued series.* A_m takes a value based on the alarm level that is active for the alarm with alarm identifier e_m corresponding alarm. For example, $e_{m_{LL}}$, e_{m_L} , e_{m_H} and $e_{m_{HH}}$ correspond respectively to the alarm identifier e_m with Low Low (LL) level, Low (L) level, High (H) level and High High (HH) level. A_m can then take the following form:

$$a_m(t_n) = \begin{cases} -2 & \text{if } t_n \in \mathcal{T}_{m_{LL}} \\ -1 & \text{if } t_n \in \mathcal{T}_{m_L} \text{ and } t_n \notin \mathcal{T}_{m_{LL}} \\ 1 & \text{if } t_n \in \mathcal{T}_{m_H} \text{ and } t_n \notin \mathcal{T}_{m_{HH}} \\ 2 & \text{if } t_n \in \mathcal{T}_{m_{HH}} \\ 0 & \text{otherwise} \end{cases} \quad (2.9)$$

- *Continuous-valued series.* A continuous-valued time series can be generated from the binary alarm series. An example proposed in [45] is:

$$a_m(t_n) = \sum_{t_{n'} \in \mathcal{T}_m} \mathcal{K}(t_n - t_{n'}) \text{ with } \mathcal{K}(t_n - t_{n'}) = \frac{1}{\sqrt{2\pi}\beta} e^{-\frac{(t_n - t_{n'})^2}{2\beta^2}} \quad (2.10)$$

where β is a smoothing constant here. This example from [45] has the advantage of being continuous and smooth, and when the value of the original binary alarm series is 1 on a certain number of consecutive samples, the value of $a_m(t_n)$ is also 1.

Chapter 4 suggests a method for online alarm flood classification based on the alarm series of Eq. 2.8. While the state of the art in alarm flood classification only contains methods based on alarm sequences, Chapter 4 demonstrates the benefits of using a time series analysis method on alarm series for online alarm flood classification. The method of Chapter 4 uses the alarm coactivations (that indicate the times where two alarms are active at the same time) as features for the online alarm flood classification, and the alarm coactivations can be computed directly from the alarm series of Eq. 2.8 as shown in Subsection 4.3.1.

2.4 Fault detection and diagnosis in industrial processes

2.4.1 Faults and disturbances

Some relevant definitions are:

- A *disturbance* refers to a deviation of the variables of the system from their desired values or trends [46].
- A *fault* can be defined as an unpermitted deviations in the variables of the system [5].
- A *failure* is a permanent interruption of a system’s ability to perform a required function under specified operating conditions [6].

In the process literature in general, it is common to talk about online detection and diagnosis of a fault, and about root cause analysis of a disturbance. The same terminology is adopted for this thesis. The rest of this sub-section describes the nature of disturbances encountered in the process industry, which has an impact on the RCA strategy adopted as discussed in Chapter 7.

Disturbances can be distinguished according to their time scales according to the definitions suggested in [8]:

- *Incipient (or slowly developing) disturbances*, i.e. long-term trend drifting from the desired values.
- *Persistent (or repeating) disturbances*, i.e. disturbances in which a deviation episode repeats tens to hundreds of times. Oscillatory disturbances are a specific type of persistent disturbances common in the process industry.
- *Transient disturbances*, i.e. short-lasting deviations of a measurement from its previous and subsequent trend. In addition, the deviation seldom repeats within the time horizon of analysis.

Oscillatory disturbances play an important role in the process industry and in the field of root cause analysis. According to [47], oscillatory disturbances are usually due to one of these reasons: poor process and control system design, aggressive controller tuning, nonlinearities in control-loop hardware or external oscillatory disturbances. The authors of [48] identify three major causes for oscillatory disturbances that match the ones of [47]: process or constraint problem, controller tuning problems, and valve problems.

Linear and nonlinear oscillations should be distinguished. A nonlinear oscillation means that the corresponding time series is nonlinear, i.e. that the dependence of the current value on previous values of the time series is nonlinear. Linear oscillations can result from controller tuning, poor process and control system design, or linear external disturbances. [48] tests RCA methods on a linear oscillation that is typically due to poor process and control system design, since it results from the interaction between the pressure control and recycle of intermediate chemical species. Nonlinear oscillations can result from valve

nonlinearity, poor process and control system design, or nonlinear external disturbances. Nonlinear oscillations contain harmonics, linear oscillations do not.

2.4.2 Fault detection and diagnosis

As introduced in Section 1.3, online FDD refers to the online process monitoring steps of fault detection, fault isolation and fault identification. Offline fault diagnosis refers to root cause analysis and propagation path analysis.

Online FDD methods have been traditionally divided into three categories following the distinction introduced in [10]:

- *Quantitative model-based methods* that assume that a quantitative model is known *a priori*. Common examples of quantitative models are observers or extended Kalman filters [10].
- *Qualitative model-based methods* that assume that a qualitative model is known *a priori*. Common examples of qualitative models are causal models (e.g. fault trees or digraphs) or models based on abstraction hierarchy [49].
- *Process history based (or data-driven) methods* that do not assume *a priori* knowledge and rely on historical process data alone [50]. A data-driven model is trained based on historical data.

This thesis examines data-driven methods that learn directly from the data available on process plants.

Data-driven fault detection and diagnosis workflow

A complete workflow for data-driven FDD in an industrial process is presented in Figure 2.3. The workflow includes:

- The *offline modeling stage* where data is selected and prepared and the data-driven model or models for online FDD are selected and trained.
- The *online fault detection and diagnosis stage* where the data-driven model or models are used to detect, identify and isolate faults on new samples coming from the real-time database.
- The *offline diagnosis stage* where a *post mortem* analysis of the fault is carried out.

The offline modeling stage can be divided into:

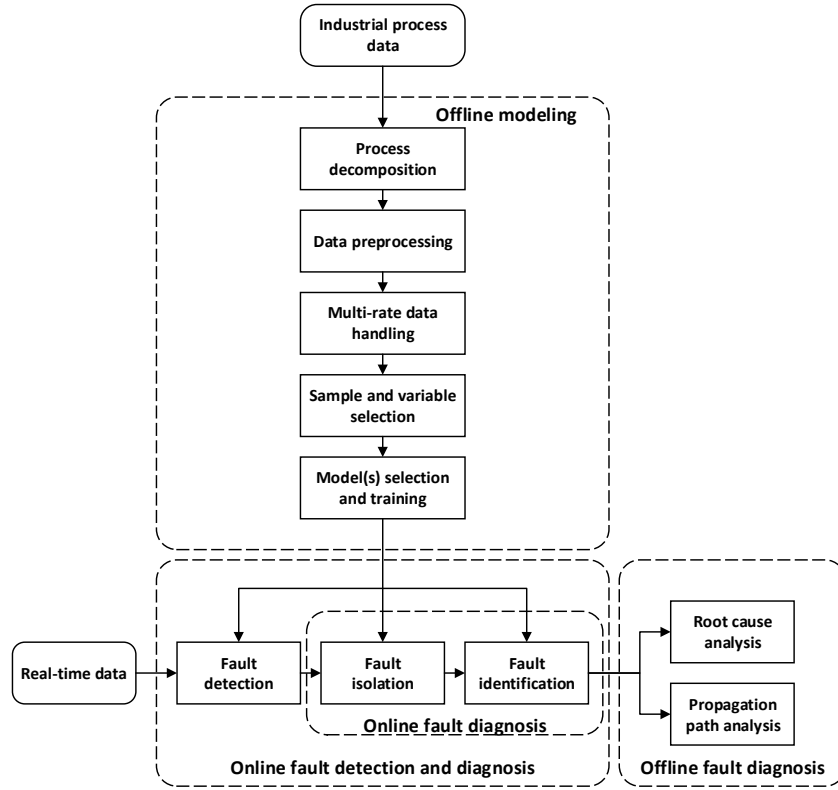


Figure 2.3: Fault detection and diagnosis workflow adapted from [51].

- A *process decomposition stage* where the FDD problem can be divided into several sub-problems on sections of the industrial process called process blocks. An online FDD model is designed for each process block.
- A *data preprocessing stage* where the process data is cleaned and prepared for the modeling. Data preprocessing covers dealing with outliers, dealing with missing data, and normalizing the data. Chapter 5 aims at improving the process measurements normalization stage using the alarm thresholds.
- A *multi-rate data handling stage* that defines a strategy to deal with the various sampling rates of the measurements. A simple strategy is to sub-sample the measurements to the lowest sampling rate of all the measurements but other strategies exist to avoid sub-sampling.
- A *sample and variable selection stage* where the samples (i.e. the sampling times t_n) to be included in the training set are selected, and the variables X_m to be used for the

analysis are selected. Chapter 6 aims at improving the variable selection stage using statistical alarms.

- *A model selection and training stage.* One or several data-driven models are selected and trained using the training set to be used for fault detection, fault isolation, and fault identification.

The data-driven models for online FDD are then used for fault detection, fault isolation, and fault identification. Chapter 5 covers fault detection, fault isolation, and fault identification. Chapter 6 covers fault detection and fault identification.

If the results of the online fault diagnosis are not conclusive, an offline fault diagnosis is performed. Offline fault diagnosis usually includes both root cause analysis and propagation path analysis, although in practice root cause analysis and propagation path analysis are often performed jointly using a similar set of methods [9]. In this thesis, special emphasis is put on root cause analysis since RCA is the subject of Chapter 7. Sub-section 2.4.3 presents additional details about RCA in the process industry.

Notations

FDD on a process plant is traditionally based on the process measurements corresponding to process variables $X_m \in \mathcal{X}$, where $\mathcal{X}$ is the set of the $M_{\mathcal{X}}$ process variables measured on the plant. The objective of the variable selection stage is to select the M process variables X_m that should be included in the FDD method, and $\mathcal{X}_M$ indicates the set of the selected process variables.

The process variables X_m are measured at each sampling time t_n . In this thesis, the following notations are adopted for the observations of X_m :

- $\hat{x}_m(t_n)$ refers to the preprocessed observation of X_m at time t_n before mean-centering and scaling, i.e. the measured values after removing outliers and dealing with missing data.
- $x_m(t_n)$ refers to the preprocessed observation of X_m at time t_n before scaling, i.e. the measured values after removing outliers, dealing with missing data, and mean-centering.
- $\tilde{x}_m(t_n)$ refers to the preprocessed observation of X_m at time t_n after scaling, i.e. the measured values after removing outliers, dealing with missing data, mean-centering, and scaling.

The plant vector $\mathbf{X}$ is defined as the M -dimensional variable taking values $\mathbf{x}(t_n) = [x_1(t_n), x_2(t_n), \dots, x_M(t_n)]^T$ at time t_n . The M -dimensional normalized plant vector at time t_n is $\tilde{\mathbf{X}}(t_n) = [\tilde{x}_1(t_n), \dots, \tilde{x}_M(t_n)]^T$.

Online FDD is usually performed at each new sampling time t_n based on the values $\mathbf{x}(t_n)$ of the plant vector at time t_n (or on the values $\tilde{\mathbf{x}}(t_n)$ of the normalized plant vector, depending on the method). The class label that indicates if $\mathbf{x}(t_n)$ belongs to the class of normal operation or to a specific fault class is $c(t_n)$, and C refers to the corresponding random variable. The values of the plant vector used for training the model are denoted as $\mathbf{x}(t_{\tilde{n}})$, where $t_{\tilde{n}}$ indicates the sampling times in the training set ($\tilde{n} = 1 \dots \tilde{N}$, $\tilde{N}$ being the total number of values used for training). The class label corresponding to $\mathbf{x}(t_{\tilde{n}})$ is $c(t_{\tilde{n}})$. Depending on the method and the application, the training set can be limited to process measurements from normal operation (in this case it is denoted as $\tilde{\Omega}_{N_0}$) or can contain process measurements from faulty operation covering at least one occurrence of each type of fault selected for the study (in this case it is denoted as $\tilde{\Omega}_F$). Ω_{N_0} and Ω_F indicate respectively the test set of process measurements in normal operation and faulty operation that are used to test the FDD methods.

2.4.3 Root cause analysis

Root cause analysis and propagation path analysis both rely on connectivity and causality analysis [37]. In the context of fault detection and diagnosis in the process industry:

- *Connectivity analysis* refers to the determination of the connections between the process variables of the plant in the form of a connectivity model. The connectivity model can be determined from process knowledge or from process data (e.g. process measurements).
- *Causality analysis* refers to the determination of the causal relations between process variables in the form of a causality model. The causality model can be determined from process knowledge or from process data (e.g. process measurements).

Other applications of connectivity and causality analysis in the process industry include topology modeling and closed-loop identification, risk analysis (hazard and operability study), consequential alarm identification, and plant-wide control structure design [37].

Examples of connectivity models have already been given in Sub-section 2.2.3 when presenting the data sources covering the plant connectivity. Examples of causality models used in the process industry are [37]:

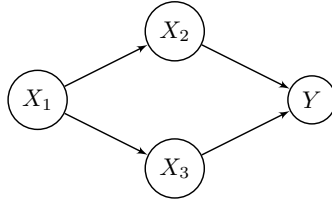


Figure 2.4: Signed directed graph for causal modeling between four variables.

- *Structural equation models.* Structural equation models are a graphical representation of causal links between input and output variables based on statistical techniques [52].
- *Matrices and digraphs.* As in the case of connectivity, matrices and digraphs can be used to describe causality, the only difference being the introduction of positive or negative sign for the connections. In that case, the graph is called Signed Directed Graph (SDG) [53].
- *Matrix layout plots.* Matrix layout plots are matrix representations of the quantitative causality indices derived from quantitative methods between the different measurements.

Causal models can also be determined using process knowledge, using process data, or using a combination of both process knowledge and process data. Causal modeling approaches based on process knowledge include [37]:

- Structural modeling based on first-principle structural models.
- Construction of adjacency and reachability matrices from PFDs and P&IDs.
- Modeling of SDGs using first-principle models.
- Modeling of SDGs using control systems.

This thesis examines causal modeling approaches based on process data. The methods for causal modeling from process data are reviewed in Section 3.4. Signed directed graphs are very common for visualizing the results of causal modeling approaches based on process data. Figure 2.4 pictures a SDG for a causal model between four process variables X_1 , X_2 , X_3 , and the variable considered as the effect variable Y . In this case, X_1 causes X_2 and X_3 , X_2 causes Y , and X_3 causes Y . X_1 has no direct causal link with Y , although it has an indirect causal impact through X_2 and X_3 . This leads to the distinction between *direct causality* which implies a direct causal relation between two variables, and *indirect causality* which implies a causal relation between two variables through one or several intermediate variables.

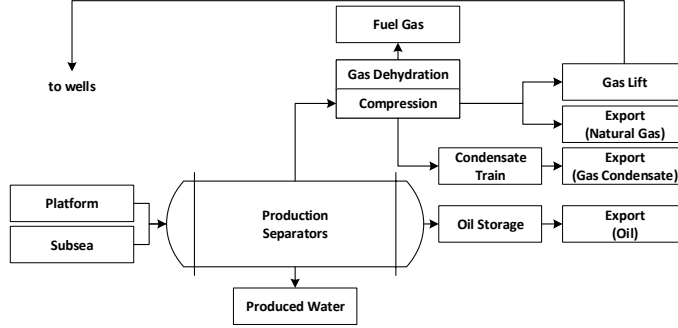


Figure 2.5: Process overview of the separation plant.

2.5 Case studies

2.5.1 Oil and gas separation plant

The oil and gas separation plant separates crude oil, gas and condensates next to the well before export. An overview of the process is given in Figure 2.5. The mixture from the well is going through two stages of separation. The gas lift and export blocks at the output of the compression block in Figure 2.5 correspond to natural gas. The output of the condensate train block corresponds to gas condensate, which is a by-product, also exported. The produced water is going through hydrocyclones and then through a degassing drum. Some of the cleaned water is reinjected into the reservoir via two injection tanks, the water injection tank 1 and the water injection tank 2. The other part of the cleaned water is discharged to the sea. The oil obtained at the output of the hydrocyclones can be recycled to the second stage of separation, or sent directly to storage facilities after going through the degassing drum.

Chapter 4 deals with plant-wide alarm records of the oil and gas separation plant. Alarm floods are extracted from the alarm records and used for training a model to classify new ongoing alarm floods.

Chapter 5 focuses on the Produced Water Reinjection (PWRI) section where the produced water and the produced oil from the hydrocyclones are going through the degassing drum before being redirected respectively to the injection tanks (or to the sea) and to the oil storage. A detailed diagram of the PWRI section is given in Figure 2.6. Four pumps convey the flow of water to the injection tanks. P11 and P12 feed the water injection tank 1, and

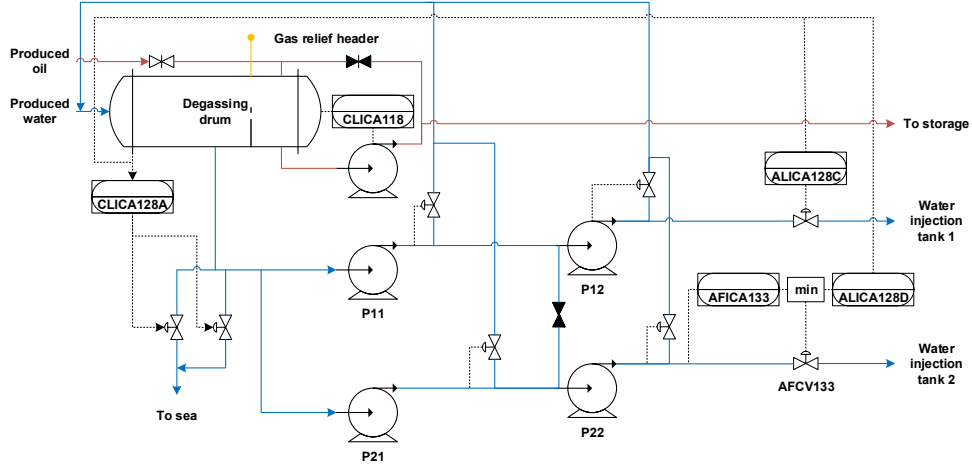


Figure 2.6: Process diagram of the PWRI section dealing with the produced water of Figure 2.5. Blue indicates the water system, brown the oil system, and yellow the gas system.

Table 2.3: Process variables in the PWRI section.

Sensor tag	Description	Unit
CLY128A	Water level degassing drum	%
CPZT151	Suction pressure P11	barg
CPZT112	Suction pressure P21	barg
AFT121	Water flow to water injection tank 2	$\text{m}^3 \text{h}^{-1}$
AFZT114	Inlet flow P22	$\text{Sm}^3 \text{h}^{-1}$
AFZT139	Water flow to water injection tank 2	$\text{m}^3 \text{h}^{-1}$
AFZT115	Discharge flow P21	$\text{m}^3 \text{h}^{-1}$
ALT148	Water level of the injection rundown tank	%
APT123	Suction pressure P12 and P22	barg
APT146	Discharge pressure P22	barg
APZT141	Discharge pressure P11	barg
APZT154	Suction pressure P12	barg
APZT143	Discharge pressure P12	barg
APZT117	Discharge pressure P21	barg
APZT140	Suction pressure P22	barg
APZT142	Discharge pressure P22	barg

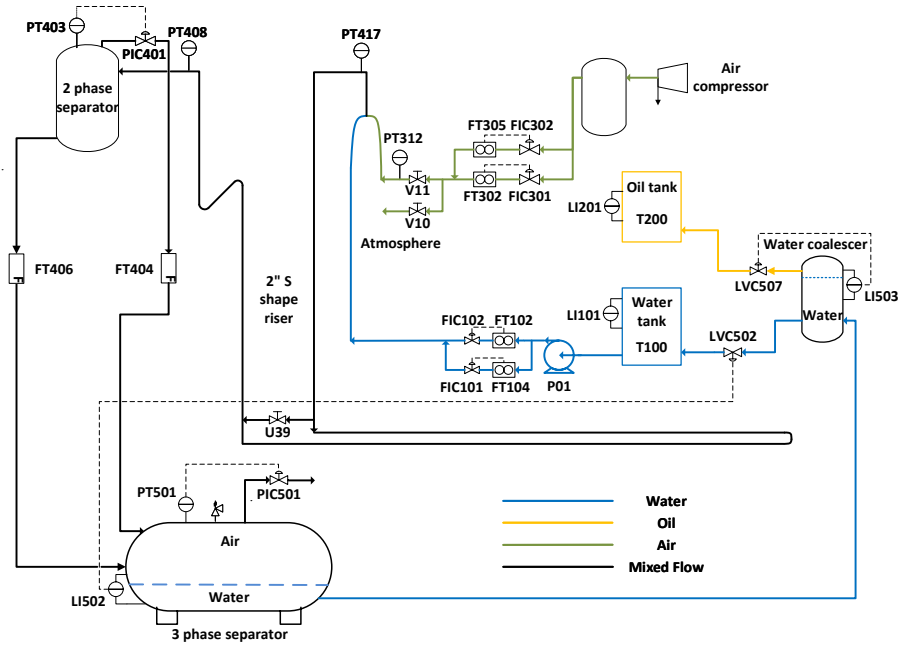


Figure 2.7: Process diagram of the multiphase flow facility reproduced from [54] ¹.

P21 and P22 feed the water injection tank 2. Another pump conveys the flow of oil from the output of the degassing drum to the oil storage. The level of water in the degassing drum is controlled using several valves. The valve to the water injection tank 1 and the valve to the water injection tank 2 can be closed to ensure a minimum flow going into the degassing drum through the recycle. The valves to the sea can also be opened or closed to regulate the water output. The process variables from the PWRI section used in Chapter 5 are described in Table 2.3.

2.5.2 Multiphase flow facility

The multiphase flow case study is located at the Process System Engineering laboratory of Cranfield University described by [54, 55] and depicted in Figure 2.7. Air (the green stream) and water (the blue stream) are mixed just below PT417 in the top part of the plant, and the mixed two-phase flow then goes through the horizontal section in the centre of the figure.

¹©2018 International Federation of Automatic Control. Reproduced with permission from the original publication in IFAC-PapersOnline, 51/18.

Table 2.4: Process variables in the multiphase flow facility.

Sensor tag	Description	Unit
FT305/OUT	Inlet air flow rate 1	$\text{Sm}^3 \text{ h}^{-1}$
FT302/OUT	Inlet air flow rate 2	$\text{Sm}^3 \text{ h}^{-1}$
PT312/OUT	Air delivery pressure	barg
PT417/OUT	Pressure in the mixing zone	barg
PT408/OUT	Pressure at the riser top	barg
PT403/OUT	Pressure in the top separator	barg
FT404/OUT	Top separator output air flow rate	$\text{m}^3 \text{ h}^{-1}$
FT406/OUT	Top separator output water flow rate	kg s^{-1}
PT501/OUT	Pressure in the three-phase separator	barg
PIC501/PID1/OUT	Air outlet valve opening in the three-phase separator	%
LI502/OUT	Water-oil level three-phase separator	%
LI503/OUT	Water coalescer level	%
LVC502/PID1/OUT	Water coalescer outlet valve opening	%
LI101/OUT	Water tank Level	m
FIC302/PID1/OUT	Inlet air flow rate controller 1 valve opening	%
FIC302/PID1/PV	Inlet air flow rate controller 1 process value	$\text{Sm}^3 \text{ h}^{-1}$
FIC301/PID1/PV	Inlet air flow rate controller 2 process value	$\text{Sm}^3 \text{ h}^{-1}$

The three-phase separator at the bottom of the figure separates the mixed flow into water and air, the water is recycled in the system through a water coalescer. Some air is exhausted in the atmosphere though a valve according to the pressure control in the separator.

The multiphase flow facility is used in Chapter 6 to demonstrate how statistical alarms can be integrated in fault detection and identification to improve the robustness of fault detection and identification to noise in the process measurements. The process variables from the multiphase flow facility used in Chapter 6 are described in Table 2.4.

Table 2.5: Data sources considered in the technical chapters.

	Chapter 4	Chapter 5	Chapter 6	Chapter 7
Process measurements		•	•	•
Alarm data	• ²		• ³	
Alarm settings		• ⁴		
Alarm management practices			• ⁵	

2.6 Overview of the assumptions of the thesis

2.6.1 Definitions and concepts

All four technical chapters of the thesis present data-driven FDD methods, but the chapters address various types of abnormal situations:

- Chapter 4 deals with *alarm floods*, defined as condition during which the alarm rate is greater than the operator can effectively manage [21].
- Chapters 5 and 6 deal with *faults*, defined as unpermitted deviations in the variables of the system [5].
- Chapter 7 deals with *disturbances*, defined as deviations of the variables of the system from their desired values or trends [46].

2.6.2 Assumptions on the data

Table 2.5 gives an overview of the data sources considered in the technical chapters of the thesis, which focus specifically on process measurements and data from the alarm systems.

The following assumptions hold throughout the thesis:

- Within each use case, all process measurements have the same sampling time. In particular, the integration of electrical measurements, mechanical measurements, or measurements from analysers with the process measurements is not considered.
- The abnormal situations occur during stationary operation, i.e. the abnormal situation is the only non-stationary phenomena during the considered time intervals. The time

² The full alarm record is considered.

³ Only statistical alarms are considered.

⁴ Alarm thresholds are considered.

⁵ Delay-timers are considered.

intervals can correspond to various operating regimes, but the transitions between operating regimes have been removed.

- The abnormal situations are not overlapping, i.e. only one type of abnormal situation occur during a given time interval.

2.7 Summary of the chapter

This chapter introduced the basic concepts dealing with alarm systems and fault detection and diagnosis in the process industry. The terminology and notations relating to process measurements, alarms, process connectivity and causality have been defined. A large part of the chapter examined alarm systems since the core topic of this thesis is the integration of alarm systems in fault detection and diagnosis. The structure of alarm records as well as alarm management practices have been covered in details to provide a basis for the technical presentation of the contributions in Chapter 4, Chapter 5, and Chapter 6.

The steps of the FDD workflow have been introduced to clarify the context of the contributions in Chapter 5, Chapter 6, and Chapter 7. Basic concepts and definitions relating to root cause analysis of industrial disturbances have been defined to prepare for Chapter 7.

The chapter also introduced the oil and gas separation plant and the multiphase flow facility discussed in Chapter 4, Chapter 5, and Chapter 6.

Chapter 3

State of the art

3.1 Description of the chapter

This chapter presents data-driven methods for alarm data analysis and fault detection and diagnosis that have been increasingly popular in the process industry. Methods for alarm data analysis are reviewed in Section 3.2, covering topics such as alarm rationalization, online operator support, and root cause analysis. The review distinguishes sequence mining methods and time series analysis methods according to the distinction between alarm sequences and alarm series introduced in Sub-section 2.3.4. Sub-section 3.2.4 presents in particular the state of the art in alarm flood classification to support the approach suggested in Chapter 4. The material presented in Section 3.2 has been published in the following journal article:

- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “Advances in alarm data analysis with a practical application to online alarm flood classification,” *Journal of Process Control*, vol. 79, pp. 56–71, 2019

The state of the art in data-driven methods for online fault detection and diagnosis is presented in Section 3.3. Special attention is given to the stages of the FDD workflow that are treated in the this thesis, namely:

- Preprocessing, treated in Chapter 5.
- Variable selection, treated in Chapter 6.
- Fault detection, treated in Chapter 5 and Chapter 6.
- Fault isolation, treated in Chapter 6.
- Fault identification, treated in Chapter 5 and Chapter 6.

Data-driven methods for root cause analysis are introduced in Section 3.4 before the comparative review of Chapter 7. The methods are grouped into four branches corresponding to four different strategies for RCA of industrial disturbances: univariate methods, time-delay estimation methods, prediction-based methods in the time-domain, and prediction-based methods in the frequency-domain.

3.2 Alarm data analysis

There are many open research topics in the area of industrial alarm systems in the process industries [56]. A common goal is the mitigation of the impact of alarm floods during which the alarm rate is greater than the operator can effectively manage [21]. Various approaches exist in practice, such as improving the alarm design, rationalizing the alarm system or providing online operator support. Alarm rationalization involves removing redundant alarms to make sure that the operator receives only alarms that require an operator response [21]. Online operator support aims at providing the operator with contextual support when alarms are appearing. Online operator support is a way to help the operators manage the high rate of alarms during alarm floods that could not be eliminated through alarm rationalization.

3.2.1 State of the art in alarm data analysis

Alarm data analysis methods encompass data-driven methods for the purpose of alarm rationalization, online operator support, or root cause analysis where alarms are used as an alternative to process measurements for diagnosis of process disturbances [57]. This section provides a review of the state of the art in alarm data analysis, and distinguishes sequence mining methods that apply to alarm sequences (defined in Sub-section 2.3.4) and time series analysis methods that apply to alarm series (defined in Sub-section 2.3.4).

Alarm data analysis methods for alarm rationalization include:

- *Alarm similarity analysis methods*, which investigate the similarity between alarms to find and remove redundancy [58].
- *Alarm flood similarity analysis methods*, which identify recurrent alarm floods that need to be addressed [59].

Alarm data analysis methods for online operator support include:

- *Alarm flood template extraction methods*, where a unique alarm flood template is extracted from a group of alarm floods associated with the same physical situation [60].

For example, the template can be an alarm sequence extracted from the alarm sequences of the alarm floods associated with the same physical situation. Reasons to generate an alarm flood template for a given physical situation include eliminating non-relevant alarms and ordering the alarms according to the most frequent order.

- *Online alarm flood classification methods*, where an ongoing alarm flood corresponding to a given abnormal situation is matched to the class of historical alarm floods corresponding to the same abnormal situation [61].
- *Alarm prediction methods*, where the next alarm is predicted based on the ongoing sequence [62].
- *Dynamic alarm suppression methods*, where a set of non-relevant alarms is temporarily hidden from the operator during the alarm flood [63].

Alarm data can also be used as an alternative to process measurements for determining the root cause(s) of process disturbances [64]. The corresponding methods are named *alarm-based root cause analysis methods*.

Table 3.1 summarizes the literature on alarm rationalization, online operator support and root cause analysis. The table distinguishes between approaches using alarm series and those using alarm sequences. Other authors have considered methods combining alarm data with other data sources (such as [82] combining alarm records with process topology).

Figure 3.1 presents a hierarchical tree of the sequence mining methods and time series analysis methods that have been applied to alarm data, with the corresponding references.

3.2.2 Sequence mining methods

Sequence mining methods have been designed for the analysis of data structured as sequences, and have been adapted to the analysis of alarm sequences. Table 3.1 indicates that sequence mining methods have been in use in the alarm data analysis literature since the work of [59], first for alarm flood similarity analysis and only recently for online applications. Figure 3.1 shows the various branches of sequence mining methods.

Table 3.1: Approaches to alarm data analysis.

		Alarm rationalization			Online operator support			Root cause analysis
		Alarm similarity analysis	AF similarity analysis	AF template extraction	Online AF classification	Alarm prediction	Dynamic alarm suppression	Alarm-based RCA
Time series analysis	ACT binary series	[58, 65, 66]						[67]
	Binary series	[65, 68, 69]			[30]			[57]
	Multi-valued series							[70]
	Continuous-valued series	[45, 66]						
Sequence mining	Set		[59, 71]	[61]	[61]		[72]	
	Sequence		[59, 71, 73]	[60]	[60]	[62, 74]		
	Time-stamped sequence	[75]	[76, 77, 78]	[79]	[80]			
	RTN time-stamped sequence				[81]		[63]	

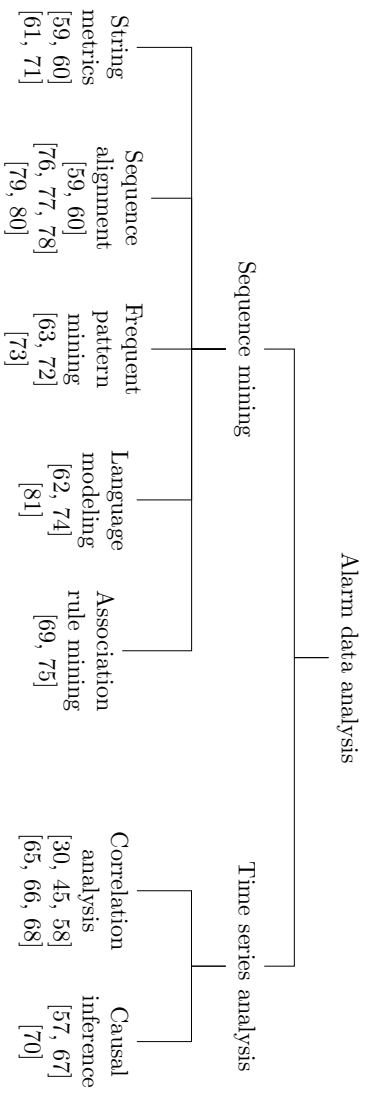


Figure 3.1: Family tree of methods for alarm data analysis.

String metrics

String metrics methods are based on distances that quantify the similarity of two sequences. Some metrics for alarm flood similarity analysis are computed on the alarm sets (alarm sets are defined in Eq. 2.6). The Jaccard distance between two alarm floods [59, 71] indicates the ratio between the number of unique alarms (defined by their unique identifier, e.g. FC101) appearing only in one of the two floods and the number of unique alarms appearing in both alarm floods. Two alarm floods can also be compared using the tf-idf (term frequency inverse document frequency) measure [71] that weights alarms with a certain identifier according to how frequently alarms with the same identifier appear in the complete set of alarm floods.

Some metrics that take into account the order of the alarms in the alarm sequence (alarm sequences are defined in Eq. 2.3) have also been used for alarm flood similarity analysis. A measure of similarity between two alarm floods is the distance (for instance, the normalized Euclidean distance) between the $\mathbf{M}_{\mathcal{A}}$ matrix of each alarm flood, where the $\mathbf{M}_{\mathcal{A}}$ matrix is defined as the matrix of counts of each pair of unique alarms (defined by their identifiers) appearing consecutively in the alarm flood [59, 71]. The Levenshtein distance [71] counts the edits (alarm insertion, alarm deletion or alarm substitution) that are needed to transform one alarm sequence into another one.

String metrics between alarm sets have been used for alarm flood template extraction and online alarm flood classification [61]. The work in [61] defines an alarm set template for a specific class of alarm floods retaining only the alarms with an identifier that appears in more than half of the alarm floods of that class. Weights are applied to enhance the importance of alarms with identifiers that are specific to a class and decrease the importance of alarms with identifiers shared by several classes. A weighted dissimilarity index is then used for online alarm flood classification using a nearest neighbour strategy.

Sequence alignment

Sequence alignment methods operate pattern matching of two sequences through alignment of one sequence with the other. Alarm flood sequence alignment methods have been proposed for alarm flood similarity analysis, based on the alarm sequences (e.g. dynamic time warping [59]) or based on the time-stamped alarm sequences as defined in Eq. 2.4 (e.g. Smith Waterman algorithm [76], BLAST algorithm [77] and the match-based accelerated alignment algorithm [78]). The modified Smith Waterman method [76], derived from bioinformatics

and modified to take into account the time dimension of the alarm floods, has become a benchmark in the literature.

The work in [60] uses the Needleman and Wunsch algorithm to extract an alarm sequence template for a specific class from the alarm sequences of the alarm floods in the same class, keeping only the alarms with identifiers aligned on more than half of the alarm sequences (after the sequences have been aligned). The sequential template is combined with the weighted vectorial template of [61] described in the sub-section on string metrics, and a weighted similarity index is used for online alarm flood classification using a nearest neighbour strategy. The same procedure for alarm flood template extraction and online alarm flood classification is defined for time-stamped alarm sequences in [79] and [80] based on the MSW algorithm. An extension of the MSW method to extract a time-stamped alarm sequence template from time-stamped alarm sequences of the same class is proposed in [79], and a method for online classification of the time-stamped alarm flood sequences using MSW is developed in [80].

Frequent pattern mining

Frequent pattern mining methods find the most frequent combinations of alarms in the available sequences. Frequent pattern mining methods have been used for alarm flood similarity analysis [73] and dynamic alarm suppression [63, 72]. Some of the frequent pattern mining algorithms such as CHARM [72] are applied to the alarm sets. CHARM is an algorithm for closed frequent pattern mining which means that it tracks the subset of frequent patterns to avoid an explosion in the number of patterns. The algorithm in [73] recognizes the most frequent alarm sub-sequences using the alarm sequences. In [63], RTN time-stamped sequences as defined in Eq. 2.5 are transformed into multi-temporal sequences indicating qualitative temporal relations between two alarms (such as equal, before, during and overlap), and frequent patterns are identified using a Bayesian approach.

Statistical language modeling

Statistical language modeling methods are derived from the probabilistic models and word representations used in natural language processing. Statistical language modeling techniques have been applied to alarm prediction [62, 74] and online alarm flood classification [81]. In [62], the most likely n th alarm (here, n is a counter for alarms) following a sequence

of $n - 1$ alarms is computed dynamically using a n -gram model, a probabilistic model following the Markov assumption and based on a $n - 1$ order Markov model.

The works in [74] and [81] use a recurrent neural network for alarm prediction and alarm sequence classification. The inputs of the recurrent neural network are continuous-valued vectorial representations of the multi-temporal sequence representations (indicating qualitative temporal relations between two alarms such as equal, before, during and overlap) derived from the original RTN time-stamped alarm sequences. Each element in the multi-temporal sequence, i.e. each alarm identifier present in the alarm flood and each qualitative temporal relation, is encoded into a binary vector that is then mapped to a continuous-valued vector through an embedding layer, by analogy with word-to-vector transformations in natural language processing.

Association rule mining

Association rule mining discovers relations in data such as temporal dependencies. The conditional probabilities of the temporal dependencies between two alarms (indicating if the two alarms tend to trigger within a given time interval) are used in [75] to identify consequential alarms for the purpose of alarm rationalization. The authors of [69] use a sliding time window in combination with a fuzzy representation of the alarms to extract temporal patterns between alarm series and identify consequential alarms.

3.2.3 Time series analysis methods

Alarm data can also be represented as time series (as defined in Eq. 2.7, Eq. 2.8, Eq. 2.9 and Eq. 2.10), in order to apply time series analysis methods. The two main branches of methods are correlation analysis methods and causal inference methods, as highlighted in Figure 3.1.

Correlation analysis

Several authors use correlation analysis methods for alarm similarity analysis [45, 58, 65, 66, 68]. Various correlation indices for binary data have been applied to ACT binary alarm series as defined in Eq. 2.7, and the recommended indices are the indices that put the weight on the number of ones rather than on the number of zeros in the ACT binary alarm series, since the ones represent the important information. For example, the Jaccard similarity

index between two ACT binary alarm series is suggested in [58], and quantifies the number of coactivations between the two alarm series over the number of activations of each alarm series. The authors define an alarm similarity colour map as the matrix of the $s_{A_{m_1}, A_{m_2}}$, where $s_{A_{m_1}, A_{m_2}}$ is the similarity between the ACT binary series A_{m_1} and A_{m_2} over the whole duration of the available data computed with the Jaccard similarity index defined in Subsection 4.3.1. The authors of [65] review 22 similarity indices for binary data successively on the ACT binary alarm series and on the binary alarm series of Eq. 2.8 and recommends the Sorgenfrei coefficient, which is an expression of the ratio between the number of coactivations between the two alarm series and the number of activations of each alarm series. Correlation indices on the ACT binary alarm series are sometimes used in combination with occurrence delay analysis on the binary series such as in [65, 68] to complete the similarity analysis. The works in [45] and [66] generate continuous-valued alarm series (as defined in Eq. 2.10) from the binary series to apply the Pearson correlation coefficient that measure the linear correlation between two variable, to provide a robust estimation of correlated alarms.

The work published by the author of this thesis in [30] corresponds to Chapter 4 of the thesis. It used the Jaccard similarity indices between each pair alarm binary series on the plant as input of a classifier for online alarm flood classification. To the best knowledge of the author, [30] proposes the first time series analysis method for online classification of alarm floods, and more generally the first time series analysis method for online operator support.

Causal inference

Causal inference methods derive causal relationships between variables and are commonly used for root cause analysis in industrial processes based on process measurements. Causal inference methods in use in the process industry include the transfer entropy method [83, 84], which finds a causal link between one time series and another, and requires $n(n - 1)$ calculations, where here n is the number of time series in the data set. Estimating transfer entropy between two process measurements is computationally demanding, and therefore the method does not scale well when n is large. Estimating the transfer entropy on the alarm series has been proposed as an alternative to speed up the computation and improve the scalability [57, 67, 70], since the estimation of transfer entropy on discrete-valued time series is easier. The transfer entropy [57] and some of its extensions (normalized transfer entropy and normalized direct transfer entropy) [67] are computed on the binary alarm series, while

[70] suggests the computation of the transfer entropy on multi-valued alarm series defined in Eq. 2.9 to improve the accuracy of the estimation.

3.2.4 State of the art in alarm flood classification

Chapter 4 deals specifically with the problem of alarm flood classification. This sub-section presents the state of the art in methods for alarm flood classification. Alarm flood classification matches a new alarm flood that corresponds to a given abnormal situation with the class of historical alarm floods corresponding to the same abnormal situation. Alarm flood classification targets recurrent types of alarm floods since it relies on training of a classifier on past episodes. The alarm flood classes selected by the operators to be monitored are referred to as *selected classes*. The other classes are referred to as *leftover classes*. The classification of the alarm floods is done in two stages:

- A *detection stage*, to determine if the alarm flood to be classified belongs to one of the classes included in the classifier, or if it corresponds to another class of alarm flood (possibly unlabelled), i.e. leftover classes.
- A *classification stage*, where the alarm flood is matched to one of the classes included in the classifier.

The two main approaches for alarm flood classification are similarity-based classification [60, 61, 80] and statistical classification of numerical feature vectors extracted from the alarm sequences [81].

Similarity-based classification

The most common procedure for alarm flood classification in the literature is a similarity-based classification approach [85]. A metric for alarm flood similarity is used to compute the similarity between the new alarm flood and the alarm floods from the training data, and the class of the new alarm flood is defined as the class of the closest alarm flood from the training data. This is called *first nearest neighbour classification*. A alarm flood template can also be extracted beforehand for each class of alarm flood in the training data so that the new alarm flood is only compared to one alarm flood per class.

The first nearest neighbour classification uses the metric to match each new ongoing alarm flood sequence $\mathcal{S}_A$ to the alarm flood that is the closest to $\mathcal{S}_A$. If alarm flood templates are extracted beforehand, $\mathcal{S}_A$ is associated with the closest alarm flood template. When the

distance between $\mathcal{S}_{\mathcal{A}}$ and the closest alarm flood is smaller than a given threshold (the detection threshold), the ongoing alarm flood is considered to belong to one of the selected classes. In this case, the classification stage matches $\mathcal{S}_{\mathcal{A}}$ to the class of the closest alarm flood. Otherwise, $\mathcal{S}_{\mathcal{A}}$ is considered to belong to a leftover class. The works in [60], [61] and [80] follow the above procedure.

In [61], a weighted dissimilarity index is used as string metric between the alarm sets. The weighted dissimilarity index is compared to a standard string metric, the Hamming distance, which indicates the number of positions at which the corresponding strings are different. The authors of [61] show that the weighted dissimilarity index outperforms the Hamming distance for classification of ongoing alarm sets.

The authors of [60] improved the method incorporating the order of the alarms in addition to the set of identifiers, and used a sequence alignment algorithm (the Needle and Wunsch algorithm) for the alarm sequence template extraction and for the definition of the weighted similarity index used in input for the nearest neighbour classification. The work in [79] and [80] extended the approach to time-stamped alarm sequences using a sequence alignment algorithm (the MSW algorithm) for time-stamped alarm sequence template extraction and as input for the nearest neighbour classification of the ongoing time-stamped alarm flood sequences.

Statistical classification of numerical feature vectors

An alternative procedure for alarm flood classification is to transform the alarm floods into numerical feature vectors that can be classified using traditional classification methods such as the methods for fault classification on continuous process data [86]. This is the procedure used in [81] where a recurrent neural network is trained on all the alarm floods in the training set and used for the classification of ongoing alarm floods. In [81], the inputs of the recurrent neural network are continuous-valued vectorial representations of the multi-temporal sequence representations derived from the original RTN time-stamped alarm sequences as described in Sub-section 3.2.2. Each element in the multi-temporal sequence, i.e. each alarm identifier present in the alarm flood and each temporal predicate (indicating qualitative temporal relations between two alarms such as *equal*, *before*, *during*, and *overlap*), is encoded into a binary vector. The binary vector is mapped to a continuous-valued vector (i.e. the numerical feature vector) through an embedding layer. The recurrent neural network is fed with $2N_{\mathcal{A}} - 1$ continuous-valued vectors where $N_{\mathcal{A}}$ here is the number

of alarm tuples in the RTN time-stamped sequence $\mathcal{S}_{\mathcal{A}}$ as defined in Sub-section 2.3.4. Each of the $2N_{\mathcal{A}} - 1$ vectors encodes one of the $N_{\mathcal{A}}$ alarm identifiers or one of the $N_{\mathcal{A}} - 1$ temporal predicate.

The authors of [81] recommended feeding the classifier with only the first three alarm tuples in the alarm floods, and sequences with more than three alarm tuples are truncated. The authors show that incorporating more than three alarm tuples as input of the classifier does not improve the performance of the classification. It is also shown that incorporating the temporal predicates as input of the classifier does not improve the performance of the classification compared to only incorporating the alarm identifiers.

3.2.5 Discussion on the review

This section reviewed the state of the art in methods for alarm data analysis, covering alarm system rationalization, online operator support and root cause analysis. Figure 3.1 and Table 3.1 showed the area is divided into sequence mining methods and time series analysis methods. The methods arise from a variety of sources, including sequence alignment algorithms from bioinformatics as well as natural language processing techniques for the sequence mining methods, and correlation analysis and causal inference for the time series analysis methods. The columns in Table 3.1 show the areas of application. These can be broadly classed as offline and online applications, the former covering alarm system rationalization and root cause analysis, the latter corresponding to online operator support.

Methods for alarm data analysis differ from traditional methods for analysis of continuous process data [86] since they deal with discrete data (sequence mining approach) or address specific problems (e.g. alarm rationalization). Nevertheless, similarities can be found between alarm data analysis problems (e.g. root cause analysis and alarm flood classification) and traditional process data analysis problems (e.g. root cause analysis and fault classification) when a time series approach is adopted. In this case, the main research question is to adapt methods for continuous times series analysis to binary series analysis.

A finding of the review is that there has been little work to date in online applications using a time series approach. The references in Table 3.1 show that problems related to online operator support have been only treated with a sequence mining approach to date. In particular, all the methods for online alarm flood classification [60, 61, 80, 81] are sequence mining methods except [30], which is derived from this thesis.

Sub-section 3.2.4 presented the state of the art on alarm flood classification and highlighted two approaches: similarity-based classification, and statistical classification of numerical feature vectors. Chapter 4 introduces a time series approach for online alarm flood classification using statistical classification of feature vectors computed based on the alarm coactivations.

3.3 Online fault detection and diagnosis in industrial processes

This section gives an overview of the status of online fault detection and diagnosis. The first sub-section discusses two high-level trends for data-driven FDD of relevance to this thesis: the strategies to learn from data and the integration of new data sources. The remaining sub-sections summarize the state of the art in each stage of the online FDD workflow relevant to this thesis, i.e. data preprocessing, variable selection, fault detection, fault isolation, and fault identification.

3.3.1 State of the art in online fault detection and diagnosis

This thesis examines data-driven methods that learn directly from the data available on process plants. This sub-section discusses two important aspects of data-driven methods:

- The learning strategies used for data-driven methods.
- The sources of data used in data-driven methods.

Learning strategies for data-driven methods

There are two major schemes for data-driven online FDD [12]:

- The *unsupervised scheme* where the modeling is based on data from normal operation only. The approach is unsupervised since no additional information about the faults is required.
- The *supervised scheme* that assumes the available process data contain both data from normal operation and data corresponding to faults, each fault being labelled.

Methods in the unsupervised scheme are usually referred to as Multivariate Statistical Process Monitoring (MSPM) methods [87]. MSPM methods are trained on data from normal

operation (i.e. on the training set $\tilde{\Omega}_{N0}$) to provide a simplified model of the normal operating conditions. The space of the model is called the modeling subspace since its dimension is smaller than the dimension of the original space. The fault detection and diagnosis is based on analysis of the deviations from the normal operating conditions in the modeling subspace that are measured by various fault detection indicators. The online FDD workflow follows the structure below:

- The fault detection stage: a new sample t_n is considered faulty if the fault detection indicators computed for the projection of the plant vector $\mathbf{x}(t_n)$ at time t_n in the modeling subspace are above the fault detection indicators thresholds. The fault detection indicators thresholds are determined based on the variations of the fault detection indicators in the training set $\tilde{\Omega}_{N0}$.
- The fault isolation stage: the variables affected by the fault are isolated based on their contribution to the deviation of the fault detection indicators. Contribution plots that highlight visually the contribution of each variable are a common tool for fault isolation.
- The fault identification stage: fault identification is not applicable in an unsupervised framework but the projection of $\mathbf{x}(t_n)$ in the modeling subspace can be used to cluster the faulty samples, i.e. group the faulty samples that present similar characteristics. In this case, historical data from past faults are not used to train the model but it can be used as a reference for visual comparison during the clustering.

In the supervised scheme, the online FDD problem consists of one or several statistical classification problems. In this scheme, both historical data from normal operation (i.e. training set $\tilde{\Omega}_{N0}$) and from past faults (i.e. training set $\tilde{\Omega}_F$) are used for training a classification model designed to distinguish between those operating conditions. The online FDD workflow follows the structure below:

- The fault detection is a binary classification problem where the classification model is trained to differentiate normal operating conditions from all the faulty conditions.
- The fault identification is a multi-class classification problem where the classification model assigns a fault class to each sample detected as faulty. In this case, the classification model is trained on $\tilde{\Omega}_F$. Fault detection and identification can also be done simultaneously with a single classification model if a class with normal operation is included with the fault classes in the multi-class classification problem.

In practice, data-driven online FDD can be a combination of unsupervised and supervised methods. The unsupervised scheme has been traditionally dominant for fault detection but fault identification using supervised methods is a powerful tool when historical data from both faults and normal operation are available. The conditions of application of supervised methods in the process industry are challenging since historical data from the faults are usually limited and not necessarily labelled. In addition, the varying operating conditions highlighted in Section 1.4 complicate the application of methods based on statistical learning.

Several research directions have been proposed to overcome those challenges for supervised learning in the process industry. Semi-supervised learning that incorporates both labelled and unlabelled samples for training of the model has been increasingly investigated [88, 89, 90, 91, 92]. Classification approaches able to cope with missing data [93] or uncertain labels [94] have been proposed.

Models in data-driven online FDD methods are usually trained offline using historical data, and then implemented online. Online adaptive methods that continually update the model as new data become available have been an active area of research [24, 95]. Updating the model continually is a way to deal with the evolving characteristics of the plant, e.g. due to the ageing of the assets or the introduction of new equipment.

Data sources for data-driven methods

The integration of data from sources other than the process measurements into online FDD is a recurrent research direction [95, 96], also identified by PRONTO (Section 1.2). The authors of [97] suggested the integration of electrical and low-pass filters measurements in the online FDD framework, for example for the detection of transient disturbances. Several works introduced methods to take into account plant connectivity and plant causality in online FDD. The authors of [98, 99] combined the causal map of the process with MSPM method for fault diagnosis, and [100] takes into account the causal dependencies for fault detection and fault isolation. The analysis of images or videos for online FDD is also a common practice [101]. The authors of [55] published a benchmark dataset from a multiphase flow facility with heterogeneous data types from various sources (process measurements, alarms and events, high-frequency measurements, videos) to promote and support the development of online FDD methods integrating various data sources.

The integration of alarms and alarm systems into online FDD has not been considered until recently. The authors of [29], also part of PRONTO, suggested a two-stage Bayesian

fusion framework to combine alarms from the alarm records and process measurements for fault detection and identification. Chapter 5 and Chapter 6 demonstrate the benefits of integrating statistical alarms, delay-timers and alarm thresholds into online FDD.

3.3.2 Preprocessing

In the context of online FDD, preprocessing refers to the preparation of historical process measurements for offline training of the model and to the preparation of the new samples of the process measurements for online fault detection, fault isolation, and fault identification. While Sub-section 2.4.2 listed three steps in the preprocessing stage (dealing with outliers, dealing with missing data, and normalizing the data), this sub-section discusses the state of the art in normalization, which is a central topic of Chapter 5. Two selected normalization approaches are presented, since they will serve as benchmark in Chapter 5.

Normalization for online fault detection and diagnosis

Process measurements associated with the variables X_m have various engineering units and scales and may need to be preprocessed so that they are dimensionless, to match the requirements of online FDD methods [86]. Normalization is generally based on the statistical properties of the process measurements used for training the method. The training set can contain process measurements from normal operation (i.e. $\tilde{\Omega}_{N0}$) or process measurements from faulty operation (i.e. $\tilde{\Omega}_F$). The process measurements in the test set Ω_{N_0} are normalized using the statistics of the training set $\tilde{\Omega}_{N0}$ and the process measurements in the test set Ω_F are normalized using the statistics of the training set $\tilde{\Omega}_F$.

Methods for fault detection and fault isolation are usually trained on $\tilde{\Omega}_{N0}$ in order to assess how abnormal a new measurement is. In this case, centering and scaling of the variance of the process measurements using the mean and standard deviation of $\tilde{\Omega}_{N0}$ is a common approach.

Methods for fault identification are usually trained on $\tilde{\Omega}_F$ to recognize the selected types of faults, but normalization practices vary. Mean-centering of the process measurements is a pre-requisite for many methods but scaling might not be necessary since not all methods require dimensionless measurements. In particular, fault identification methods using features learnt from the training data (such as methods based on deep learning [102, 103]) have been used without scaling the process measurements. When dimensionless measurements should be used, standardization can be based on the statistics of $\tilde{\Omega}_F$ but can also be based

on the statistics of $\check{\Omega}_{N0}$, in particular if the fault identification method is directly based on the output of fault detection method. In this case, the normalization statistics are the same as the ones used for fault detection, e.g. the mean and the standard deviation of $\check{\Omega}_{N0}$ [104, 105, 106].

Other approaches exist but remain rare. Reference [107] uses feature scaling where the extremum of the process measurements in $\check{\Omega}_{N0}$ are used to scale the process measurements between zero and one [107]. Reference [108] divides the mean-centered process measurement by the mean of the process measurements in $\check{\Omega}_{N0}$.

Selected normalization for fault detection and isolation

The most common normalization approach for fault detection and isolation is standardization using $\check{\Omega}_{N0}$ statistics from normal operation, which is referred to as N-standardization in the rest of the thesis. The N-standardized plant vector at time t_n is defined as $\tilde{\mathbf{x}}(t_n) = [\tilde{x}_1(t_n), \dots, \tilde{x}_M(t_n)]^T$ where:

$$\tilde{x}_m(t_n) = \frac{\dot{x}_m(t_n) - \mu_m(\check{\Omega}_{N0})}{\sigma_m(\check{\Omega}_{N0})} \quad (3.1)$$

where $\mu_m(\check{\Omega}_{N0})$ and $\sigma_m(\check{\Omega}_{N0})$ are respectively the mean and standard deviation of process variable X_m in $\check{\Omega}_{N0}$.

Selected normalization for fault identification

For fault identification, the most common normalization approach is F-standardization, which corresponds to a standardization using the standard deviation in faulty operation. The F-standardized plant vector at time t_n is defined as $\tilde{\mathbf{x}}(t_n) = [\tilde{x}_1(t_n), \dots, \tilde{x}_M(t_n)]^T$ where:

$$\tilde{x}_m(t_n) = \frac{\dot{x}_m(t_n) - \mu_m(\check{\Omega}_{N0})}{\sigma_m(\check{\Omega}_F)} \quad (3.2)$$

where $\mu_m(\check{\Omega}_{N0})$ is the mean of process variable X_m in $\check{\Omega}_{N0}$ and $\sigma_m(\check{\Omega}_F)$ is the standard deviation of process variable X_m in $\check{\Omega}_F$.

3.3.3 Variable selection

Online FDD in industrial processes typically involves the analysis of a large number of process variables [109]. Variable selection and feature extraction are the two main approaches for

reducing the number of variables, where a feature is defined as an individual measured property of the monitored system [110]. Variable selection chooses informative and discriminative variables for online FDD [111], while feature extraction applies a transformation to the original variables to highlight characteristics or reduce the dimension [111]. Historically, variable selection methods have not received the same attention as feature extraction methods in the process fault detection and diagnosis literature [109]. In particular, feature extraction has been preferred to variable selection for the task of fault detection since removing variables from the monitoring algorithm can increase chances of missed detection. However, [111] demonstrated that variable selection and feature extraction are complementary, and the recent review of [112] highlights a growing interest in variable selection. Variable selection is a central topic of Chapter 6.

A popular strategy for variable selection combines pre-selection of variables with an optimization problem, where the set of variables leading to the best performance of the online FDD method is retained. The work reported in [113] combined a multivariate extension of the mutual information criterion with discriminant analysis for fault identification. Other criteria aiming at highlighting variables with abnormal variations have been suggested: [114] identified the nonsteady faulty variables that are disturbed significantly using a stability factor, and [104] used an index describing the degree of abnormal variation for each variable. Genetic algorithms, which use bio-inspired operators (such as mutation or crossover) for the combinatorial search of the optimal subset of variables, have been used in [111]. Regularization techniques (such as least absolute shrinkage and selection operators or elastic net) for variable selection in regression analysis have also been used for discriminant analysis in the context of fault isolation [115]. Finally, the authors of [17] proposed to select the variables that correspond to the root causes of the faults to be identified.

3.3.4 Fault detection

Multivariate Statistical Process Monitoring (MSPM) methods have been popular for fault detection in the process industry because they offer a simple approach to deal with large numbers of highly correlated variables by building a model of lower dimension [87]. Principal Component Analysis (PCA) and Partial Least Squares (PLS) have been the first MSPM methods applied for fault detection [116, 117]. PCA and PLS both exploit the correlation structure of the process measurements in $\tilde{\Omega}_{N0}$. In the case of PCA, the correlation structure is used to define a set of linearly uncorrelated principal components. PLS is by default a

regression technique used to predict a variable Y (e.g. a quality variable for the product) from a set of process variables X_m . PLS finds a reduced dimension by maximizing the covariance between the predictors in the reduced dimension and the predicted Y . Variations in the subspace of the principal components (for PCA) or in the subspace of the predictors (for PLS) are used to defined statistical thresholds that will constitute the basis for detecting faults in the new samples.

Other popular MSPM methods are Independent Component Analysis (ICA) [118] and Canonical Variate Analysis (CVA) [119, 120]. ICA is an alternative to PCA that extracts components that are statistically independent using for example minimization of mutual information between the components. While basic PCA or ICA only considers the instantaneous correlation between variables, CVA exploits the serial correlation of the variables taking into account past samples. CVA finds linear combinations of the original variables (called canonical variables) that maximize the correlation between the set of past values and the set of future values of the process measurements in $\check{\Omega}_{N0}$. The canonical variables are used to derive a state-space model of the data in $\check{\Omega}_{N0}$, and variation in the subspace of the canonical variables are used to define statistical thresholds for the fault detection.

Extended versions of PCA, PLS, ICA, and CVA can be applied to nonlinear processes, non-Gaussian processes, dynamic processes, as well as time-varying and multimode processes [96]. Methods to deal with nonlinear processes include kernel transformations such as in kernel PCA [121] or transformations using neural networks [122]. Methods to deal with non-Gaussian processes include ICA or Gaussian mixture models such as in mixture PCA [123]. CVA or dynamic extensions of MSPM methods such as dynamic PCA [124] can be used for fault detection in dynamic processes. Recursive methods such as recursive PLS [125] have been introduced to deal with time-varying processes, i.e. processes that evolve due for example to degradation of the equipment. Multi-model methods, i.e. methods with one model for each operating mode, have been used for fault detection in multimode processes [126]. Another approach for fault detection in multimode processes is to use kernel transformations [127].

Supervised methods for fault detection are sometimes used as an alternative to MSPM methods. The fault detection can be done simultaneously with the fault identification, which means that there is a single classifier covering $N_C + 1$ classes, i.e. the normal class and the N_C faulty classes. This is the case in [128] that proposed the first application of artificial neural networks for fault detection and identification in the process literature. An alternative approach is to treat the fault detection as a separate binary classification problem with a

classifier trained to distinguish the normal class from the faulty class covering the N_C fault classes.

Chapter 5 demonstrates the impact of normalization on a traditional MSPM approach (with PCA) for fault detection on an oil and gas separation plant. Chapter 6 investigates the benefits of integrating statistical alarms in supervised online FDD where fault detection and identification are done simultaneously.

3.3.5 Fault isolation

Fault isolation looks for the process variables affected by the fault. The fault isolation stage is usually considered as an extension of the fault detection stage. It usually analyses the contribution of each variable in the fault detection indicators. Contribution plots summarize visually the contribution of each variable. For fault detection using traditional MSPM methods, determining the contribution of each variable to the detection indicators is equivalent to inverting the transformation from the space of the original variables to the modeling subspace. Inverting the transformation can be straightforward (for example with PCA or PLS [129]) or involve additional operations (for example with CVA [130]). In particular, deriving the contribution plots can be challenging when the MSPM method for fault detection includes nonlinear transformations that are hard to invert, such as in kernel PCA [131].

A recurrent problem of contribution plots is the smearing effect, meaning that for a fault affecting a single variable X_m , the contribution plots might show several contributing variables with no guarantee that X_m has the highest contribution. Reconstruction-based contributions have been proposed as an alternative to contribution plots to reduce the impact of the smearing effect [132]. The amount of reconstruction needed for each variable to minimize the fault detection indicator (i.e. the value that should be added or subtracted to each variable for the fault detection indicator to be minimal) is defined as the reconstruction-based contribution of that variable.

Chapter 5 performs fault isolation using reconstruction-based contribution analysis with PCA.

3.3.6 Fault identification

Fault identification determines the type of fault and is conducted using classification-based methods [104]. Classification-based methods assume that historical data from the faults are

available and new faults are identified in a supervised manner [104]. Many classifiers have been proposed to identify faults from the process measurements, including discriminant analysis [133, 134, 135], neural networks [102, 103, 128], support vector machines [133], dissimilarity-based classifiers [104], k -nearest neighbour classifiers [136], and random forest classifiers [137]. Those methods have been designed to identify faults based on continuous features, i.e. based on the values of the process measurements, but the variability in the process measurements from one fault occurrence to another is not addressed, especially in simulated case studies. Traditional fault identification methods also assume processes are stationary, but recent work addressed the issue of nonstationary fault characteristics through cointegration analysis [138].

Unlike the classifiers cited above, Bayesian classifiers can integrate heterogeneous data. Bayesian methods for fault identification have been introduced in the process literature through the work in [139] for control loop monitoring, which consists of detecting the underperforming control loops in the plant. The input data of the classifier were either discrete (e.g. values generated by control loop monitors) or continuous (e.g. through kernel density estimation of the probability density functions of the process measurements [140]). The work in [139] has been extended to fault identification problems in specific cases where Bayesian methods bring advantages over traditional methods, e.g. for fault identification in multimode processes [141] or for fault identification with heterogeneous data [29, 142]. The fault are identified using a Bayesian classifier combining asynchronous measurements in [142] or combining alarms from alarm records and process measurements in [29]. A Bayesian framework was used to combine the classification outcomes of various classifiers for fault identification in [143]. Classification approaches using graphical models based on Bayesian networks have also been proposed for fault identification [144, 145].

The advantages that a Bayesian classification framework offers for integrating heterogeneous data, and in particular discrete and continuous data, are exploited in Chapter 6 for combining process measurements and statistical alarms.

3.4 Root cause analysis in industrial processes

This section gives an overview of the state of the art in data-driven RCA methods and introduces twelve methods that will be compared in Chapter 7. Sub-section 3.4.6 also mentions comparative analyses of data-driven RCA methods published in the literature.

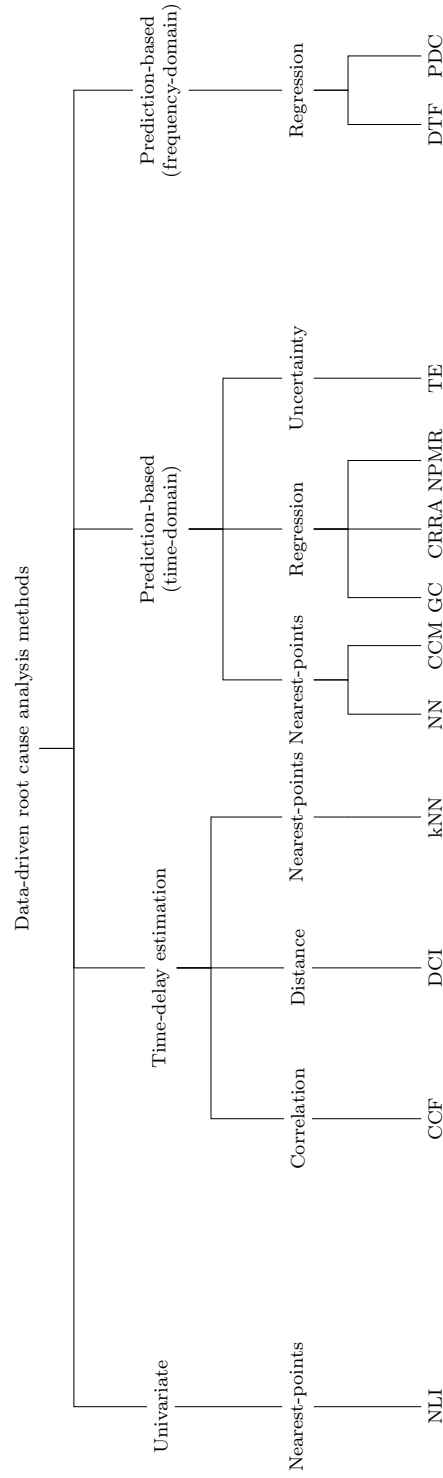


Figure 3.2: Family tree of methods for data-driven root cause analysis.

3.4.1 State of the art in root cause analysis

Data-driven RCA methods have been introduced in the literature as a fast and automatic means to identify the root cause or root causes of a process disturbance. The methods solely rely on process measurements, which makes them attractive for industrial practice. The simplicity of use of data-driven methods hides a certain complexity that can be an obstacle to the acceptance of the methods. Understanding the assumptions and the scope of applicability of each method is essential to make use of the results.

The field of data-driven RCA has been growing in the recent years with many publications proposing new methods or improvements to existing methods. Figure 3.2 shows the family tree of existing data-driven RCA methods, highlighting four branches of methods according to the RCA strategy adopted.

Univariate methods

Univariate methods rely on univariate properties of the process measurements. It is assumed that a given univariate property of the process measurements vary with the distance of the corresponding process variable from the root cause, and this property can be used to rank the process variables from the closest to the farthest from the root cause. The univariate property used by the nonlinearity index is the self-predictability of a measurement that is quantified based on the ability to predict ongoing segments of a measurement using similar past segments (nearest-points) of the measurement.

Time-delay estimation methods

Time-delay estimation methods are based on time-delay estimation between two process variables. Each pair of process variables is compared in the time dimension to determine a potential time-delay, and the process variable that precedes all the others is likely to be the root cause. The comparison of the process variables can be based on correlation analysis between the measurements, on the computation of a distance between the measurements, or on the ability to predict ongoing segments of one measurement using similar past segments (nearest-points) of the other measurement. An advantage of time-delay estimation methods is that they give the exact time-delay between process variables in addition to the root cause.

Prediction-based methods in the time-domain

Prediction-based methods in the time-domain assume that the predictability of an effect variable increases when a cause variable is included in the predictors. Therefore, if the predictability of Y is improved when taking X into account, X is said to cause Y . The variable that is not caused by any other variable is considered as the root cause. In this case, the prediction is based on methods in the time-domain. Most methods compare the univariate prediction of Y (i.e. based on past values of Y) to the bivariate prediction of Y based on both past values of X and Y . Some methods investigate the impact of removing a given variable X_m in the multivariate prediction of Y based on the M predictors $X_1, \dots, X_M$.

Prediction-based methods in the time-domain are divided according to the type of predictability criteria they use. Methods based on nearest-points use past segments of the measurement of X to predict future values of an ongoing segment of the measurement of Y , in particular the past segments that are the nearest to the ongoing segment according to a given distance metric. Methods based on regression fit a model to predict Y based on the predictors X_m and criteria describing the goodness of fit of the model with and without a certain predictor are used to determine whether this predictor causes Y . Methods based on uncertainty reduction rely on the comparison of the uncertainty about the values of Y with and without taking into account past values of X .

Prediction-based methods in the frequency-domain

Prediction-based methods in the frequency-domain also rely on the predictability improvement of Y when taking X into account, but the prediction uses methods in the frequency-domain to focus on specific frequencies. Prediction-based methods in the frequency-domain are based on regression and are multivariate.

Notations

In the following, X and Y are two process variables with measurements $x(t_n)$ and $y(t_n)$ at time instances $t_n, n = 1 \dots N$, where N is the number of samples. In the multivariate case, the definition is extended to M process variables X_m .

For each variable X , an embedded vector $\vec{x}(t_n)$ at time t_n is constructed as:

$$\vec{x}(t_n) = [x(t_n), x(t_{n-\tau}), \dots, x(t_{n-(L-1)\tau})] \quad (3.3)$$

where τ is the sampling constant, and L the embedding dimension. The embedded vector $\vec{y}(t_n)$ of Y at time t_n is constructed similarly.

3.4.2 Univariate methods

Univariate methods rely on univariate properties of the process measurements that are impacted by the propagation. The evolution of a given univariate property can serve as a tracer for the propagation of the disturbance and highlight possible root cause(s).

Nonlinearity index

Nonlinearity Index (NLI) assumes that systems in a process plant act as low-pass filters and that the nonlinearity of the time series of the process measurements can be used to retrace the root cause [146]. The more nonlinear the time series of X is, the closer to the root cause X should be, since low-pass filters cut the harmonics (and thus reduce the nonlinearity) of a time series. The method is based on univariate predictability using nearest neighbours information. The nearest neighbours of an embedded vector $\vec{x}(t_n)$ are the embedded vectors $\vec{x}(t_k)$ with the smallest distance $d(t_n, t_k) = \|\vec{x}(t_n) - \vec{x}(t_k)\|$, where the norm indicates the Euclidean norm. For each embedded vector $\vec{x}(t_n)$ at time t_n , K nearest neighbours are selected and identified through the indices $r(t_n, t_k)$, $k = 1 \dots K$.

An embedded vector $\vec{x}(t_n)$ can be used to predict a future value $x(t_{n+h})$, where h indicates the prediction horizon. The prediction value obtained from the nearest neighbours is named $x(t_{r(t_n, t_k)+h})$. The self-predictability of X can be measured as:

$$D_X(t_n) = \frac{1}{K} \sum_{k=1}^K |x(t_{n+h}) - x(t_{r(t_n, t_k)+h})| \quad (3.4)$$

A nonlinearity measure can be defined as:

$$\Gamma_X = \frac{1}{N_v} \sum_{n=1}^{N_v} \left(x(t_{n+h}) - \frac{1}{K} \sum_{k=1}^K x(t_{r(t_n, t_k)+h}) \right)^2 \quad (3.5)$$

where $N_v = N - (L - 1)\tau$ is the number of embedded vectors.

The first application of the nonlinearity index for root cause analysis of process disturbances was presented in [146] and [147] provided additional insights on how to tune the parameters. More recently, [148] pointed out a limitation of nonlinearity index: if a nonlinearity is introduced during the propagation of the disturbance (e.g. due to a saturation

effect in a level control with a maximum level), nonlinearity index is not uniformly decreasing with the propagation. In this case, root cause analysis with nonlinearity index can lead to spurious results.

3.4.3 Time-delay estimation methods

Time-delay estimation methods estimate the time-delay κ_{XY} between each pair of variables X and Y . The variable that precedes all the others is said to be the root cause.

Cross-correlation function

The Cross-Correlation Function (CCF) measures the linear correlation of two variables over time [149]. Cross-correlation function is estimated from two variables X and Y measured at discrete times $x(t_n)$ and $y(t_{n+\kappa})$, where κ indicates the lag:

$$\rho_{XY}(\kappa) = \frac{1}{N - \kappa} \sum_{n=1}^{N-\kappa} \frac{(x(t_n) - \mu_X)(y(t_{n+\kappa}) - \mu_Y)}{\sigma_X \sigma_Y} \quad (3.6)$$

The estimated time-delay $\hat{\kappa}_{XY}$ can be formulated as:

$$\hat{\kappa}_{XY} = \underset{\kappa_{\min} \leq \kappa \leq \kappa_{\max}}{\operatorname{argmax}} \quad \rho_{XY}(\kappa) \quad (3.7)$$

Cross-correlation function for root cause analysis of process disturbances was introduced in [149]. The authors of [150] used cross-correlation function for root cause analysis of oscillations in a paper-board machine and pointed out drawbacks of the method: it is difficult to determine which process variable is lagging behind due to the periodic nature of the oscillations and the bivariate aspect of cross-correlation function leads to redundant dependencies compared to the multivariate methods such as the ones based on regression.

Dynamic time warping causality index

Dynamic time warping Causality Index (DCI) uses a distance based on Dynamic Time Warping (DTW) algorithm [151, 152] to measure the similarity of two measurements over time [153]. DCI is estimated from two variables X and Y measured at discrete times $x(t_n)$ and $y(t_{n+\kappa})$, where κ indicates the lag:

$$DCI_{XY} = \frac{\min_{0 \leq \kappa \leq \kappa_{\max}} d_{XY}(\kappa)}{d_{XY}(0)} \quad (3.8)$$

where $d_{XY}(\kappa)$ indicates the distance between X and Y with lag κ computed using dynamic time warping, i.e.:

$$d_{XY}(\kappa) = d_{DTW}([x(t_1), \dots, x(t_{N-\kappa})], [y(t_{\kappa+1}), \dots, y(t_N)]) \quad (3.9)$$

The estimated time-delay $\hat{\kappa}_{XY}$ is the lag κ that minimizes $d_{XY}(\kappa)$.

Dynamic time warping causality index was introduced in [153] for root cause analysis of non-stationary disturbances. Dynamic time warping causality index assumes that the process measurements are similar in shape, so a clustering (also based on dynamic time warping) of the process measurements affected by the same disturbance is carried out beforehand. Within each cluster, the variable that precedes the others according to the time-delay estimation is considered as the root cause.

k nearest neighbours

k Nearest Neighbours (kNN) compares nearest neighbours of embedded vectors of two variables to measure the similarity of those two variables over time [154]. kNN is estimated from two variables X and Y measured at discrete times $x(t_n)$ and $y(t_{n+\kappa})$, where κ indicates the lag. As with nonlinearity index, the K nearest neighbours of the embedded vector $\vec{x}(t_n)$ are identified through their index $r(t_n, t_k)$ ($k = 1 \dots K$). $y(t_{n+\kappa})$ is estimated using the values $y(t_{r(t_n, t_k)+\kappa})$:

$$\hat{y}(t_{n+\kappa}) = \frac{1}{K} \sum_{k=1}^K y(t_{r(t_n, t_k)+\kappa}) \quad (3.10)$$

The mean sum of square deviations s_{XY} between estimated and real values of $y(t_{n+\kappa})$ can be used to estimate the time-delay between X and Y :

$$s_{XY}(\kappa) = \sum_{n=1}^N (\hat{y}(t_{n+\kappa}) - y(t_{n+\kappa}))^2 \quad (3.11)$$

A normalized quality criterion $\rho_{XY}(\kappa)$ is introduced for the time-delay estimation:

$$\rho_{XY}(\kappa) = \left(1 + \sqrt{\frac{s_{XY}(\kappa)}{N} \frac{K}{K+1}} \right)^{-1} \quad (3.12)$$

The expected value of the quality criterion is independent from K and N . The time-delay $\hat{\kappa}_{XY}$ can be estimated as:

$$\hat{\kappa}_{XY} = \underset{\kappa_{\min} \leq \kappa \leq \kappa_{\max}}{\operatorname{argmax}} \quad \rho_{XY}(\kappa) \quad (3.13)$$

kNN was introduced in [154] for root cause analysis of oscillations in a industrial hydro-cracker.

3.4.4 Prediction-based methods in the time-domain

Prediction-based methods rely on the principle of predictability improvement. If the predictability of Y is improved by including X in the predictors, X is said to cause Y . The predictability improvement is quantified by a causality measure $\eta_{X \rightarrow Y}$. X is said to cause Y if $\eta_{X \rightarrow Y} \geq \eta_\alpha$, where η_α indicates the significance threshold estimated using statistical methods discussed in Chapter 7. In this sub-section, the prediction relies on time-domain analysis, whereas the next sub-section focuses on prediction approaches in the frequency-domain.

Nearest neighbours

Nearest Neighbours (NN) uses the nearest neighbours of the embedded vectors of a variable X to assess if X can be used as a good predictor of Y [155]. The K nearest neighbours of $\vec{x}(t_n)$ are identified through their index $r(t_n, t_k)$ ($k = 1 \dots K$), and the K nearest neighbours of $\vec{y}(t_n)$ are identified through their index $s(t_n, t_k)$ ($k = 1 \dots K$). If the prediction value $y(t_{n+h})$ can be predicted by the values $y(t_{r(t_n, t_k)+h})$, X is considered a good predictor of Y . The predictability factor of Y given X is:

$$D_{Y|X}(t_n) = \frac{1}{K} \sum_{k=1}^K |y(t_{n+h}) - y(t_{r(t_n, t_k)+h})| \quad (3.14)$$

The dependency measure is calculated as:

$$\eta(Y|X) = \frac{1}{N_v \sum_{n=1}^{N_v} \frac{D_{Y|X}(t_n)}{D_Y(t_n)}} \quad (3.15)$$

where $D_Y(t_n)$ indicates the self-predictability of Y like in Eq. 3.4:

$$D_Y(t_n) = \frac{1}{K} \sum_{k=1}^K |y(t_{n+h}) - x(t_{s(t_n, t_k)+h})| \quad (3.16)$$

The causality measure is defined as:

$$\eta_{X \rightarrow Y} = \eta(X|Y) - \eta(Y|X) \quad (3.17)$$

The authors of [155] introduced nearest neighbours for root cause analysis of process disturbances comparing it to nonlinearity index since both methods rely on the concept of predictability using nearest neighbours information. Nearest neighbours was extended to multi-rate systems in [156] to enable integration of electro-low-pass filters measurements. The authors of [157] presented a procedure to integrate nearest neighbours with connectivity information and derived a multivariate version of nearest neighbours that can be used for direct causality analysis.

Convergent cross mapping

Convergent Cross Mapping (CCM) uses mutual predictability (or cross predictability) of the attractor manifolds of X and Y as a criterion to determine if X and Y are causally related [158]. In dynamical systems theory, two time series variables X and Y are considered causally related if they are generated from the same dynamic system, which means that they share a common attractor manifold M_s [158]. In this case, Takens embedding theorem states that the time-delayed embedding manifolds M_X of X (constructed using lagged values of X) and M_Y of Y (constructed using lagged values of Y) are diffeomorphic to the manifold of the system M_s [158]. This means that if one can use M_X to predict M_Y , it can be said that X cause Y .

The time-delayed embedding manifold (usually named shadow manifold) M_X of X is constructed such that its value $\mathbf{m}_X(t_n)$ at each time t_n ($n = (L - 1)\tau + 1, \dots, N$) is the embedded vector $\vec{\mathbf{x}}(t_n)$ of X at time t_n :

$$\mathbf{m}_X(t_n) = \vec{\mathbf{x}}(t_n) \quad (3.18)$$

The shadow manifold of M_Y of Y is constructed similarly, i.e. $\mathbf{m}_Y(t_n) = \vec{\mathbf{y}}(t_n)$. A one-to-one mapping can be found between M_X and M_Y , and convergent cross mapping predicts the value $y(t_{n_0})$ of Y at a new time t_{n_0} using the points in M_Y that are mapped to the nearest neighbours of $x(t_{n_0})$ in M_X . The $L + 1$ nearest neighbours of $\mathbf{m}_X(t_{n_0})$ in M_X are denoted as $\mathbf{m}_X(t_{n_i})$ for $i = 1 \dots L + 1$, with $\mathbf{m}_X(t_{n_1})$ being the nearest neighbour. $\mathbf{m}_Y(t_{n_0})$ is estimated using the mapping of the nearest neighbours $\mathbf{m}_X(t_{n_i})$ in M_Y , i.e. $\mathbf{m}_Y(t_{n_i})$:

$$\hat{\mathbf{m}}_Y(t_{n_0})|_{M_X} = \sum_{i=1}^{L+1} w_i \mathbf{m}_Y(t_{n_i}) \quad (3.19)$$

where:

$$w_i = \frac{\gamma_i}{\sum_{k=1}^{L+1} \gamma_k}, \text{ with } \gamma_k = \exp\left(-\frac{\mathbf{m}_X(t_{n_i}) - \mathbf{m}_X(t_{n_0})}{\mathbf{m}_X(t_{n_1}) - \mathbf{m}_X(t_{n_0})}\right) \quad (3.20)$$

If the number of samples N available for X and Y is large enough, the predicting ability of M_X for M_Y , expressed as the covariance between M_Y and $\hat{M}_Y|M_X$, converges to a constant $\rho_{X \rightarrow Y}$:

$$\lim_{N \rightarrow \infty} \text{cov}(M_Y, \hat{M}_Y|M_X) = \rho_{X \rightarrow Y} \quad (3.21)$$

The causality measure $\eta_{X \rightarrow Y}$ is defined as:

$$\eta_{X \rightarrow Y} = \rho_{X \rightarrow Y} - \rho_{Y \rightarrow X} \quad (3.22)$$

The authors of [159] proposed a direct application of convergent cross mapping to RCA of nonlinear oscillations in a refinery. Reference [160] refined the manifold construction and proposed an automatic method for parameter tuning. The authors of [161] proposed a method for choosing the embedding dimension.

Granger causality

Granger Causality (GC) relies on regression modeling [162]. Y can be modeled as an autoregressive model:

$$y(t_n) = \sum_{r=1}^R B_{Y,r} y(t_{n-r}) + \epsilon_Y(t_n) \quad (3.23)$$

where R is the model order, $B_{Y,r}$ are the autoregressive coefficients, and ϵ_Y is the prediction error. Y can also be modeled incorporating past values of both X and Y (autoregressive exogenous model):

$$y(t_n) = \sum_{r=1}^R [A_{YX,r} x(t_{n-r}) + A_{YY,r} y(t_{n-r})] + \epsilon_{Y|X}(t_n) \quad (3.24)$$

where $A_{YX,r}$ and $A_{YY,r}$ represent the regression coefficients. $A_{YX,r}$, $A_{YY,r}$ and $B_{Y,r}$ can be determined using the least squares approach. The model order R can be determined using the Akaike Information Criterion.

The variance of the prediction error ϵ_Y is used as a criterion to quantify the predictability improvement. When the variance of $\epsilon_{Y|X}(t_n)$ is smaller than the variance of $\epsilon_Y(t_n)$, the prediction of Y is improved by including past values of X . X is then said to Granger cause

Y . The Granger causality measure is computed as:

$$\eta_{X \rightarrow Y} = \ln\left(\frac{\text{var}(\epsilon_Y(t_n))}{\text{var}(\epsilon_{Y|X}(t_n))}\right) \quad (3.25)$$

The authors of [162] introduced Granger causality for root cause analysis of process disturbances. Granger causality is proposed as a natural step to identify the root cause(s) among the variables affected by the fault selected during the fault isolation stage [153, 163, 164]. The systematic procedure for fault diagnosis proposed in [153] suggests root cause analysis with Granger causality if the disturbance is stationary after a fault isolation with reconstruction-based contribution. In [163], the faulty variables are clustered using dynamic time warping and Granger causality is used to identify the root cause(s). Reference [164] focuses on incipient disturbances so the fault isolation is done with a moving average residual difference reconstruction contribution plot and Granger causality is applied on the first order difference of the process measurements to remove the non-stationarity. Other references propose to apply Granger causality after dimension reduction to overcome collinearity issues in the regression, for example after PLS [163] or after PCA [165]. The authors of [166] suggest a version of Granger causality based on Gaussian process regression for nonlinear or non-stationary cases.

Causal relationship based on residual analysis

Causal Relationship based on Residual Analysis (CRRA) is an alternative to Granger causality that uses a system identification approach for modeling Y instead of the traditional regression approach of Granger causality [167]. An autoregressive exogenous model $\mathcal{M}(X, Y, R)$ is fitted:

$$A(q)y(t_n) = B(q)x(t_{n-1}) + \epsilon(t_n) \quad (3.26)$$

$A(q) = 1 + a_1q^{-1} + \dots + a_Rq^{-R}$ and $B(q) = 1 + b_1q^{-1} + \dots + b_Rq^{-R}$ are polynomials with parameters to be estimated, q is the shift operator, $\theta = [a_1 \dots a_R \ b_1 \dots b_R]$ is the parameter vector and $\epsilon(t_n)$ denotes the residual terms. $\hat{y}(t_n|\hat{\theta})$ is the estimated output at instant t_n using the estimated parameter value $\hat{\theta}$. The residual are given by:

$$\epsilon(t_n) = y(t_n) - \hat{y}(t_n|\hat{\theta}) \quad (3.27)$$

Two tests are used to determine the model order: a whiteness test of the residual and a test of independence of the residual and the input. The independence test is also used to

estimate the causality. The outcome of the independence test $v_R(\mathcal{M})$ on the model $\mathcal{M}$ can be formulated as:

$$v_R(\mathcal{M}) = \begin{cases} 1, & \text{if } |\hat{\varphi}_{eu}(\kappa)| \in I_R, \text{ for all } \kappa < 0 \\ -1, & \text{otherwise} \end{cases} \quad (3.28)$$

where $\varphi_{eu}(\kappa) = \frac{1}{N\sigma_\epsilon\sigma_u} \sum_{n=1}^N \epsilon(t_n)u(t_{n-\kappa})$ is the sample cross-correlation. I_R corresponds to a predefined confidence interval depending on the order R of the model.

A second autoregressive exogenous model $\mathcal{M}(Y, X, R')$ is fitted where X is predicted from Y . The causality between X and Y is assessed based on the result independence tests of $\mathcal{M}(X, Y, R)$ and $\mathcal{M}(Y, X, R')$:

- If $v_R(\mathcal{M}(X, Y, R)) = 1$ and $v_{R'}(\mathcal{M}(Y, X, R')) = -1$, Y causes X .
- If $v_R(\mathcal{M}(X, Y, R)) = -1$ and $v_{R'}(\mathcal{M}(Y, X, R')) = 1$, X causes Y .
- Otherwise, there is no causality between X and Y .

Causal relationship based on residual analysis was introduced in [167] and tested for root cause analysis of oscillations on a thermoelectric power plant.

Nonparametric multivariate regression

Nonparametric Multivariate Regression (NPMR) is another regression-based method that uses a nonparametric approach [168, 169]. Consider a variable Y to be estimated from M predictors X_m . A global response surface of Y is built from its M predictors using a multiplicative kernel smoother. This is achieved by estimating each $y(t_n)$ from its local neighbourhood in the predictor space $X(t_n) = [x_1(t_n), x_2(t_n), \dots, x_M(t_n)]$. The influence of each predictor is defined by its corresponding tolerance of the kernel smoother $b_m (m = 1 \dots M)$ and the kernel tolerance vector is defined as $\mathbf{b} = [b_1, b_2, \dots, b_M]$. The local neighbourhood is defined as a weighted mean and the weights are estimated using a Gaussian kernel smoother:

$$w_m(t_{n'}) = \exp -\frac{1}{2}[(x_m(t_{n'}) - x_m(t_n))/b_m]^2 \quad (3.29)$$

The weights represent the distances of each predictor from the target point estimated with a Gaussian kernel weighting function with a tolerance b_m . The estimation of a target point

$y(t_n)$ is defined as :

$$\hat{y}(t_n) = \frac{\sum_{n'=1, n' \neq n}^N y(t_{n'}) (\prod_{m=1}^M w_m(t_{n'}))}{\sum_{n'=1, n' \neq n}^N (\prod_{m=1}^M w_m(t_{n'}))} \quad (3.30)$$

The casualty estimator from the nonparametric multivariate regression method can be applied to estimate Y using only its own embedded vector $\vec{\mathbf{y}}$ defined according to Eq. 3.3 or combining both embedded vectors $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$. Similarly to Granger causality, causal relationships can be established by computing the variance of the prediction errors. In case of a bivariate estimate, the causality measure is defined as:

$$\eta_{X_m \rightarrow Y} = \ln\left(\frac{\text{var}(\epsilon_{Y|Y})}{\text{var}(\epsilon_{Y|X_m, Y})}\right) \quad (3.31)$$

where X_m is the m th predictor, $\text{var}(\epsilon_{Y|Y})$ is the variance of the prediction error when only past values of Y are used as predictors, and $\text{var}(\epsilon_{Y|X_m, Y})$ is the variance of the prediction error when both past values of Y and X_m are used as predictors.

Nonparametric multivariate regression was introduced in [168, 169] for root cause analysis of oscillations on a paper-board machine.

Transfer entropy

Transfer Entropy (TE) measures the reduction in uncertainty in the values of Y when incorporating past values of X [84]. Consider two variables X and Y with embedded vectors $\vec{\mathbf{x}}(t_n)$ and $\vec{\mathbf{y}}(t_n)$ at time t_n . By contrast with other RCA methods presented in this section, the embedding dimensions of $\vec{\mathbf{x}}(t_n)$ and $\vec{\mathbf{y}}(t_n)$ for transfer entropy can be set to different values, namely L for $\vec{\mathbf{x}}(t_n)$ and L' for $\vec{\mathbf{y}}(t_n)$. The transfer entropy from X to Y can be calculated as:

$$T_{Y|X} = \sum_{n=(L-1)\tau+1}^{N-h} p(y(t_{n+h}), \vec{\mathbf{y}}(t_n), \vec{\mathbf{x}}(t_n)) \log \frac{p(y(t_{n+h})|\vec{\mathbf{y}}(t_n), \vec{\mathbf{x}}(t_n))}{p(y(t_{n+h})|\vec{\mathbf{y}}(t_n))} \quad (3.32)$$

$p(.,.)$ represents the joint probability distribution function and $p(.|.)$ represents the conditional probability density function. h is the prediction horizon. The transfer entropy causality measured is computed as:

$$\eta_{X \rightarrow Y} = T_{Y|X} - T_{X|Y} \quad (3.33)$$

The authors of [84] introduced transfer entropy for root cause analysis of disturbances in the process literature. Tuning of the parameters of transfer entropy, in particular of the prediction horizon h and the sampling constant of the embedded vectors τ , can be a challenge and has attracted attention from the academic community. h and τ depend on the time-delay between variables. The authors of [170] suggested the use of time-lagged mutual information to pre-select pairs of variables to be considered in the analysis and to determine the best time lag parameter for each pair of variable. Reference [171] proposed a modified version of transfer entropy with improved estimation of the time-delay. A systematic workflow to apply transfer entropy in the industry including rules of thumb for the tuning of the parameters is presented in [172].

Another issue of transfer entropy is the computing load of the estimation of the probability density functions. The transfer zero-entropy that does not require probability density functions estimation has been proposed to speed up the analysis [173]. The authors of [174] proposed a multivariate implementation of transfer entropy called normalized transfer entropy and used a symbolic dynamic-based implementation of normalized transfer entropy to speed up the computation and enable real-time applications. Taking into account plant connectivity is another solution to limit the number of pairwise analyses and reduce the computational load of the RCA [175, 176].

Another extension of transfer entropy, direct transfer entropy, has been proposed to identify direct causal links [177]. Finally, [178] proposed a method to remove transient changes from oscillatory disturbances before RCA with transfer entropy to prevent spurious results due to those transient phenomena.

3.4.5 Prediction-based methods in the frequency-domain

Prediction-based methods in the frequency-domain also rely on the principle of predictability improvement, but the prediction focuses on a specific frequency of the process measurements. Instead of looking at the causality measure $\eta_{X \rightarrow Y}$ of X on Y over the whole power spectrum, those methods investigate the causality measure $\eta_{X \rightarrow Y}(f)$ of X on Y at a given frequency f .

Directed transfer function and partial directed coherence

The Directed Transfer Function (DTF) and the Partial Directed Coherence (PDC) [179] are both extensions of Granger causality in the frequency-domain. The time series system can

be modeled according a M -dimensional autoregressive model of order R :

$$\begin{bmatrix} x_1(t_n) \\ \vdots \\ x_M(t_n) \end{bmatrix} = \sum_{r=1}^R \mathbf{A}_r \begin{bmatrix} x_1(t_{n-r}) \\ \vdots \\ x_M(t_{n-r}) \end{bmatrix} + \begin{bmatrix} \epsilon_1(t_n) \\ \vdots \\ \epsilon_M(t_n) \end{bmatrix} \quad (3.34)$$

where the $\mathbf{A}_r$ correspond to the $M \times M$ matrices of the model coefficients. The frequency response of the process can be obtained by application of the Z -transform and setting $z^{-1} = e^{-i2\pi f}$ (where i represents the imaginary unit):

$$\mathbf{x}(f) = \mathbf{H}(f)\mathbf{E}(f) \quad (3.35)$$

where $\mathbf{x}(f) = [x_1(f), \dots, x_M(f)]^T$, $\mathbf{E}(f) = [\epsilon_1(f), \dots, \epsilon_M(f)]^T$, and $\mathbf{H}(f)$ is the transfer function of the multivariate autoregressive model with the following relation to the model coefficients:

$$\mathbf{H}(f) = (\mathbf{I}_M - \mathbf{A}(f))^{-1} = \bar{\mathbf{A}}(f)^{-1} = (\mathbf{I}_M - \sum_{r=1}^R \mathbf{A}_r z^{-r})^{-1} \quad (3.36)$$

The directed transfer function causality measure between two variables X_{m_1} and X_{m_2} at a given frequency f is defined as:

$$\eta_{X_{m_1} \rightarrow X_{m_2}}(f) = \frac{H_{m_1, m_2}(f)}{\sqrt{\sum_{m=1}^M |H_{m_1, m}(f)|^2}} \quad (3.37)$$

The directed transfer function represents the normalized measure of the power contribution from X_{m_1} to X_{m_2} through both direct and indirect pathways.

The partial directed coherence measures only the direct causal contribution from X_{m_1} to X_{m_2} . The partial directed coherence causality measure is defined solely from $\bar{A}_{m_1 m_2}(f)$:

$$\eta_{X_{m_1} \rightarrow X_{m_2}}(f) = \frac{\bar{A}_{m_1 m_2}(f)}{\sqrt{\sum_{m=1}^M |\bar{A}_{m m_2}(f)|^2}} \quad (3.38)$$

The authors of [162] introduced spectral Granger causality (which is equivalent to partial directed coherence) at the same time as they introduced Granger causality in the time-domain for RCA of oscillations. Frequency-domain analysis with directed transfer function and partial directed coherence has been used to refine the results of the analysis with Granger

causality for RCA of oscillatory disturbances [150, 162, 179]. Granger causality identifies a large number of causal relations since it does not focus on a particular frequency, which can be useful to get an overview of the propagation path of the disturbance. The number of causal relations identified by directed transfer function is smaller than with Granger causality since directed transfer function focuses on causal relations (both direct and indirect) at a specific frequency, usually chosen as the frequency of the analysed oscillatory disturbance. For this reason, directed transfer function is a more precise method for isolating the root cause than Granger causality. Finally, partial directed coherence is used to highlight the direct dependencies at the specific frequency.

The authors of [180] used the fact that transfer entropy and Granger causality are equivalent for Gaussian variables to compute the information transfer in the frequency-domain using Granger causality.

3.4.6 Comparative analyses of data-driven root cause analysis methods

Some research works have proposed comparative analyses of selected data-driven RCA methods. This sub-section introduces the main studies from the literature but detailed discussion of the analyses is left for Chapter 7 that reviews and compares the characteristics of data-driven RCA methods.

Granger causality and transfer entropy, which are the most popular methods in the process industry, have been compared several times. The authors of [64] compared Granger causality and transfer entropy (as well as model-based methods such as adjacency matrices introduced in Sub-section 2.4.3) for RCA of oscillatory disturbances on an industrial case study. The authors of [181] provided a comparative analysis of the sensitivity of Granger causality and transfer entropy with regard to the characteristics of the input disturbance for two types of disturbances: oscillations and step disturbances. The authors varied the amplitude and the frequency of the input oscillation or the amplitude of the input step disturbance propagating through a simulated tank system and analysed the impact on the value of the causality measure between the input and the output of the system obtained with Granger causality and with transfer entropy. The effect of additive noise and the effect of a feedback loop (i.e. including a controller of the output variable) were also investigated. The same authors proposed a more systematic comparison of Granger causality and transfer entropy in [182] for RCA of oscillations to help the industrial practitioners in the choice of the

method. Five high-level criteria were suggested: the accuracy of the method, the automatability of the method, the interpretability of the method, the computational complexity of the method, and the applicability of the method for different process characteristics. Those criteria are discussed in Chapter 7.

The authors of [183] and [184] compared Granger causality, transfer entropy, and cross-correlation function on a set of simple simulated process systems (e.g. first-order or second-order systems, dead-time, feedback loop, square function) and showed that the performance of the methods vary depending on the type of system studied. The simulations highlighted that Granger causality, transfer entropy, and cross-correlation function can be complementary since they perform well on different types of systems. The authors suggested a causality measure combining the results of Granger causality, transfer entropy, and cross-correlation function, which would be reliable on a larger set of systems than the methods taken separately.

Reference [48] also used simple simulations to study the effect of dead-time, low-pass filtering, and additive noise during the propagation of the disturbance on the results of cross-correlation function, NN, and transfer entropy. Finally, [168] recorded and compared the results of Granger causality (combined with partial directed coherence and directed transfer function), transfer entropy, NN, and nonparametric multivariate regression that have been applied on the same industrial case of oscillations in a paper-board machine.

3.5 Summary of the chapter

A wide range of alarm data analysis methods have been proposed in the literature to mitigate the impact of alarm floods. Some of those methods aim to rationalize the alarm systems, other methods offer online support to the operator during the alarm floods, and some other methods investigate the root causes of the alarm floods. Section 3.2 suggests a distinction between methods applied to alarm sequences and methods applied to alarm series. Sequence mining methods have been used to compare alarm floods for online operator support or for alarm rationalization. By contrast, time series analysis methods have mostly been used to compare individual alarms for alarm rationalization or for root cause analysis. In particular, Section 3.2 highlights that only sequence mining methods have been applied to online applications such as online alarm flood classification, although some of those sequence mining methods were not initially designed as online methods.

Section 3.3 gave an overview of the data-driven methods for online FDD highlighting the learning strategies and the integration of new data sources. The integration of alarms and alarm systems into online FDD has not been exploited until now.

Section 3.4 highlighted the various strategies for data-driven RCA and presented twelve data-driven RCA methods covering those various strategies. A few comparative analyses have already been published and constitute a starting point for the comparative review of Chapter 7. Those existing comparative analyses have been focused on a limited set of methods (Granger causality and transfer entropy in particular), on a limited set of criteria (characteristics of the input disturbance or characteristics of the propagation of the disturbance), and on a limited set of disturbances (oscillatory disturbances in particular).

Chapter 4

A binary series approach to online alarm flood classification

4.1 Description of the chapter

Many data-driven methods exploiting historical alarm records have been proposed for mitigation of alarm floods. Section 3.2 distinguished sequence mining methods such as Modified Smith Waterman (MSW) [76] that consider alarm data as alarm sequences, and time series analysis methods that consider alarm data as alarm series, as defined in Sub-section 2.3.4. The review of Section 3.2 highlights that, to the best knowledge of the author, no one has used time series analysis methods to compare alarm floods to date. In particular, no time series analysis method has been proposed for online operator support during ongoing alarm floods. This chapter gives an example by means of a case study to show how a time series approach for online alarm flood classification can start to address this gap. The material presented in this chapter has been published in the following articles:

- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “Online alarm flood classification using alarm coactivations,” in *Proceedings of 10th ADCHEM Symposium, Shenyang, China, July 25-27, IFAC-PapersOnLine*, vol. 51, pp. 345–350, 2018
- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “Advances in alarm data analysis with a practical application to online alarm flood classification,” *Journal of Process Control*, vol. 79, pp. 56–71, 2019

Section 4.2 highlights why a time series approach for online alarm flood classification fills a gap for the industry. Section 4.3 introduces an example of binary series approach (where

binary series refers to a type of time series introduced in Sub-section 2.3.4) based on alarm coactivations, which is a metric to quantify alarms active at the same time. The time series approach is motivated through an industrial case study, and the performance of the time series method is compared with the performance of the modified Smith Waterman sequence mining method. Results are presented in Section 4.5.

4.2 Background

4.2.1 A gap in online alarm flood classification for the industry

All methods for alarm flood classification in the literature are sequence mining methods that emphasize the importance of the order of the alarms in the alarm sequences for the classification [60, 81]. Both studies [60] and [81] have been carried out on data generated from simulations. In practice, however, the order of the alarms can be subject to randomness as shown in Sub-section 4.2.3. The Smith Waterman algorithm was modified in [76] to attenuate the effect of the exact order of the alarms when the times of activation are close.

Like other sequence alignment methods, the Modified Smith Waterman (MSW) method was designed as a highly accurate method for similarity analysis of two alarm floods for offline applications such as alarm rationalization [76]. The main drawback of the MSW method is the computational complexity required to compare each pair of alarm floods. A fast computation time is critical for online applications. The authors of [80] suggested several modifications to the MSW method [76] to make it applicable online including the extraction of one alarm flood template per class (to minimize the number of comparisons), a sequence window filter, and an incremental dynamic programming strategy.

An example of a time series method is presented in Section 4.3. The method does not consider the order of the alarms in the alarm floods but considers the alarm coactivations, which is a metric to quantify alarms active at the same time, for the classification of ongoing alarm floods. Results from this chapter suggest that classification based on alarm coactivations performs well compared to the sequence mining methods when it comes to industrial data, while being easily implemented online. An industrial case study is used to motivate the time series approach.

4.2.2 A case study in alarm flood classification

The motivation for the binary series approach arises from the analysis of alarms from a typical industrial process automation system. The case study is the offshore oil and gas separation plant introduced in Sub-section 2.5.1, designed to separate crude oil, gas and condensates next to the well before export.

According to the IEC 62682 standard [21], the criterion to define an alarm flood episode corresponds to a minimum number of alarms over a time window of size T_w , defined as the alarm flood threshold ν_{α_1} . Following the standards [21], ν_{α_1} can be set for example to 10, and the size of the window T_w 10 minutes. Chattering alarms are removed using the guidelines in [59]. Consecutive alarm flood episodes are merged and a window of size T_w preceding the merged alarm flood episode is included in the alarm flood episode to cover the possible early alarms ahead of the main alarm flood. A window of size T_w following the merged alarm flood episode is also included, to capture possible late alarms. Therefore, the final alarm flood episode has a duration of $(n + 2)T_w$. In this way, the alarm flood can be represented as a time series.

The analysis focuses on a subset of alarms that are the most relevant for the analysis of the selected abnormal episodes, namely process alarms and trip events (which amount to 1473 alarm identifiers on the plant). Only process alarms with high alarm priority (i.e. greater than 750) and trip events are considered in the analysis, other alarms and other types of events that might appear on the HMI of the alarm system are ignored. For this reason, the alarm flood threshold ν_{α_1} is set to eight, slightly lower than the value of 10 recommended in the standards. The size of the window T_w is set to 10 minutes. 926 alarm floods are identified over the 382 days available for the analysis. Among the identified 926 alarm floods, 81 correspond to five classes of recurrent alarm floods which occurred more than five times during the period under analysis. These are described in Table 4.1.

Two classes are used to illustrate how time series analysis methods could be used for alarm flood classification. Class D and class E correspond to alarm floods in the produced water reinjection system, where used water is going through degasification before being reintroduced in the plant via two injection points. Both alarm floods in class D and class E manifest as pump trips in the system, leading to reduced flow of water to one of the injection points. The time-stamped alarm sequences of two alarm floods from class D are displayed in Table 4.2.

Table 4.1: Description of the selected types of alarm floods. The sequence length is expressed in number of alarms.

Class	Number of alarm floods	Average sequence length	Longest sequence length	Shortest sequence length
A	16	16.6	31	8
B	16	13.4	26	8
C	9	24.4	40	18
D	11	25.3	31	19
E	29	24.6	48	15

4.2.3 Alarm binary series

Two alarm flood episodes of class D (alarm flood 70 and alarm flood 78) represented as binary series are depicted in Figure 4.1. The corresponding time-stamped alarm sequences are available in Table 4.2. The visual inspection of those alarm floods leads to two observations.

Firstly, the visual inspection highlights the variability of alarms across two episodes of the same class. The four main informative components of the alarm sequences (the alarm identifiers contained in the sequence, the order of the alarms in the sequence, and the times of activation and deactivation of the alarms) are significantly modified from an episode to the other.

The other observation from Figure 4.1 is that the binary alarms are sparse. Figure 4.1 plots only the 71 binary alarm series that are active at some point in one of the alarm flood episodes of the same class (class D), but there are 1473 binary alarm series in total on the plant. In that regard, the input data is high dimensional and the information contained in each binary series is limited.

Therefore, the risk of overfitting when applying a classifier directly on the binary series is high. It is necessary to look for a transformation of the data that reduces the dimensionality and limits the impact of the randomness in the times and order of activation of the alarms.

4.3 Method development

This section introduces the suggested online alarm flood classification method, starting with the computation of the alarm coactivations.

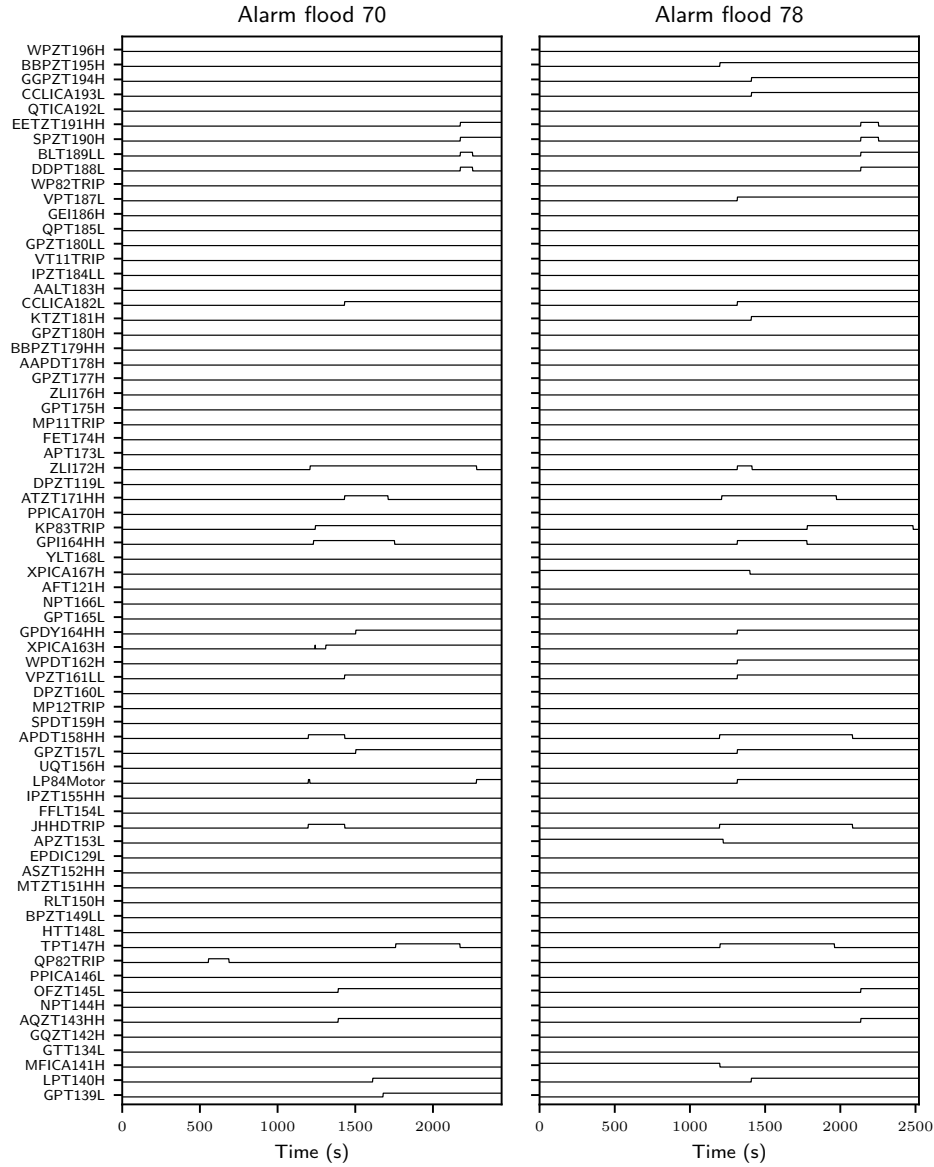


Figure 4.1: Binary alarm series for alarm flood 70 and alarm flood 78 belonging to class D.

Table 4.2: Time-stamped sequences of alarm floods 70 and 78 belonging to class D.

Alarm flood 70		Alarm flood 78	
Alarm tag	ACT time	Alarm tag	ACT time
HTT135L	06:14:30	GTT134L	22:45:40
AFICA130L	06:26:47	APT2051L	23:34:17
AFZAL115LL	06:26:47	AFZT127LL	23:34:30
AFICA116L	06:26:48	AFICA130L	23:34:30
AFZT115LL	06:26:48	AFZT114L	23:34:31
AFZT114L	06:26:48	AFZAL115LL	23:34:31
AFZT127LL	06:26:48	AFICA116L	23:34:31
APT121L	06:26:49	AFZT115LL	23:34:32
BP81TRIP	06:26:49	AFZT114LL	23:34:32
AP21TRIP	06:26:59	APZT2260LL	23:34:32
AP11TRIP	06:27:01	APT121L	23:34:32
CLICA128AH	06:27:14	BP81TRIP	23:34:33
CLT118H	06:27:23	AP21TRIP	23:34:43
EPDIC129H	06:27:31	AP11TRIP	23:34:44
DPZT119AL	06:28:58	APZT117L	23:34:46
APZT124H	06:33:42	CLICA128AH	23:34:57
APZT124HH	06:34:49	CLT118H	23:35:05
DPT131L	06:35:42	APZT2054L	23:36:35
DPT132L	06:36:12	DPZT119BL	23:37:02
DPT131LL	06:37:14	APZT124H	23:37:34
DPT132LL	06:38:48	APZT2054LL	23:37:47
AFICA116L	06:44:50	DPT131L	23:42:44
		DPT131LL	23:43:32
		DPT132L	23:44:26
		DPT132LL	23:47:16
		DPZT119AL	23:50:09
		DPZT119BL	23:50:09

4.3.1 Alarm coactivations

Alarm coactivations are quantified using the Jaccard similarity index originally defined in [185], expressed for binary data in [186], and also used for alarm similarity analysis in [58] and [65]. According to [58], the Jaccard similarity index between a pair of alarm binary series A_{m_1} and A_{m_2} (corresponding to the alarm identifiers e_{m_1} and e_{m_2} , $m_1, m_2 \in [1, M_{\mathcal{E}}]$) can be computed on a time interval $I_n = [t_{n_s}, t_{n_e}]$:

$$s_{A_{m_1}, A_{m_2}}(I_n) = \max_{l \in \mathcal{L}} \frac{\gamma_1(l)}{\gamma_1(l) + \gamma_2(l) + \gamma_3(l)} \quad (4.1)$$

where:

- $l \in \mathcal{L}$ is the lag to be considered between the two alarm series A_{m_1} and A_{m_2} . The set of lags $\mathcal{L}$ is chosen according to the dynamics of the process.
- $\gamma_1(l)$ is the number of observations where $a_{m_1}(t_n) = 1$ and $a_{m_2}(t_{n+l}) = 1$ where t_n ($t_{n_s} \leq t_n \leq t_{n_e-l}$) indicates the time-stamp,
- $\gamma_2(l)$ is the number of observations where $a_{m_1}(t_n) = 1$ and $a_{m_2}(t_{n+l}) = 0$ where t_n ($t_{n_s} \leq t_n \leq t_{n_e-l}$) indicates the time-stamp,
- $\gamma_3(l)$ is the number of observations where $a_{m_1}(t_n) = 0$ and $a_{m_2}(t_{n+l}) = 1$ where t_n ($t_{n_s} \leq t_n \leq t_{n_e-l}$) indicates the time-stamp.

An alarm coactivation matrix is defined for each alarm flood as the matrix $\mathbf{S}(I_n)$:

$$(S_{m_1, m_2}(I_n))_{m_1, m_2 \in [1, M_{\mathcal{E}}]} = \begin{cases} s_{A_{m_1}, A_{m_2}}(I_n) & \text{if } m_1 \neq m_2 \\ 1 & \text{if } m_1 = m_2 \text{ and } \|A_{m_1}\| > 0 \\ 0 & \text{if } m_1 = m_2 \text{ and } \|A_{m_1}\| = 0 \end{cases} \quad (4.2)$$

where A_{m_1} represents the alarm binary series associated with the alarm identifier $e_{m_1} \in \mathcal{E}$, with $\mathcal{E}$ the set of $M_{\mathcal{E}}$ (unique) alarm identifiers configured on the process plant. By convention, the similarity of an alarm series with itself $s_{A_{m_1}, A_{m_1}}(I_n)$ is set to one if A_{m_1} is activated at some point during the time interval I_n , to zero otherwise. Thus, the diagonal elements of the matrix represent the vector of alarms that appeared during the alarm flood.

The alarm coactivation matrix representations of the alarm floods of Table 4.2 are presented in Figure 4.2 and Figure 4.3. They represent the 71 alarms active during one of the alarm floods of class D. Figure 4.2 and Figure 4.3 highlight common patterns in the alarm coactivations of alarm floods 70 and 78. Since several alarms appear in alarm flood

70 but not alarm flood 78 (or vice versa), the training of the classifier must determine the statistically relevant columns when considering alarm coactivation matrices of alarm floods of the same class and alarm coactivation matrices of alarm floods of other classes.

4.3.2 Classification of the alarm floods

Classification involves calculation of an alarm coactivation matrix (Eq. 4.2) for each alarm flood interval I_n . The alarm coactivation matrices are computed on a time window covering the alarm flood episodes based on the binary alarm series and are then used for classification of ongoing alarm floods.

The alarm coactivation matrices are classified using a Support Vector Machine (SVM) classifier [187] appropriate for the classification of high-dimensional inputs. The alarm coactivation matrices being symmetric matrices, the classifier is applied to the lower triangular elements of the matrices $S_{m_1, m_2}(I_n)$ ($m_1 \in [1, M_\mathcal{E}], m_2 \in [1, m_1]$), which are stacked as an input vector $\mathbf{s}$:

$$\mathbf{s}(I_n) = [S_{1,1}(I_n)S_{2,1}(I_n) \dots S_{M_\mathcal{E},1}(I_n)S_{2,2}(I_n) \dots S_{M_\mathcal{E},2}(I_n) \dots S_{M_\mathcal{E},M_\mathcal{E}}(I_n)]' \quad (4.3)$$

As introduced in Sub-section 3.2.4, classification of alarm floods requires two stages: a detection stage and classification stage.

Classification stage

A SVM classifier is a linear discriminative classifier that uses a hyperplane to separate two classes of vectors. When the original vectors $\mathbf{s}$ are not linearly separable, the vectors $\mathbf{s}$ are mapped into a higher dimensional feature space using a mapping function ϕ and the linear classifier is constructed on the vectors $\phi(\mathbf{s})$. The mapping is implicitly given by the kernel function $\mathcal{K}(\mathbf{s}, \mathbf{s}')$ between two vectors $\mathbf{s}$ and $\mathbf{s}'$ defined as $\mathcal{K}(\mathbf{s}, \mathbf{s}') = \phi(\mathbf{s}) \cdot \phi(\mathbf{s}')$. Common kernel functions are:

- Linear kernel: $\mathcal{K}(\mathbf{s}, \mathbf{s}') = \mathbf{s}^T \mathbf{s}'$
- Polynomial kernel: $\mathcal{K}(\mathbf{s}, \mathbf{s}') = (\theta_0 \mathbf{s}^T \mathbf{s}' + \theta_1)^d$, where θ_0 and θ_1 are constants and d indicates the degree of the polynomial.
- Radial basis function kernel: $\mathcal{K}(\mathbf{s}, \mathbf{s}') = \exp\left(-\frac{\|\mathbf{s} - \mathbf{s}'\|^2}{2\beta^2}\right)$.

The estimated class $\hat{c}(I_n)$ of a vector $\mathbf{s}(I_n)$ to be classified is determined by the decision function $D(\mathbf{s})$ as:

$$\hat{c}(I_n) = \begin{cases} 1 & \text{if } D(\mathbf{s}(I_n)) > 0 \\ -1 & \text{otherwise} \end{cases} \quad (4.4)$$

The decision function $D(\mathbf{s})$ is defined as:

$$D(\mathbf{s}) = \sum_{\tilde{n}=1}^N c(I_{\tilde{n}}) b_{\tilde{n}} \mathcal{K}(\mathbf{s}(I_{\tilde{n}}), \mathbf{s}) + \rho \quad (4.5)$$

where the $\mathbf{s}(I_{\tilde{n}})$ ($\tilde{n} = 1, \dots, \tilde{N}$) are the $\tilde{N}$ training vectors and the $c(I_{\tilde{n}})$ their corresponding class labels. The $b_{\tilde{n}}$ and ρ are the parameters of the linear decision function. The parameters $b_{\tilde{n}}$ and the bias ρ are determined solving the optimization problem of Eq. 4.6 which is the dual problem of the optimization problem in Eq. 4.7 [187]:

$$\begin{aligned} \min_{\mathbf{b}} \quad & \frac{1}{2} \mathbf{b}^T \mathbf{Q} \mathbf{b} - \mathbf{1}_{\tilde{N}}^T \mathbf{b} \\ \text{subject to} \quad & \mathbf{c}^T \mathbf{b} = 0, \quad 0 \leq b_{\tilde{n}} \leq B, \quad \tilde{n} = 1, \dots, \tilde{N} \end{aligned} \quad (4.6)$$

where:

- $\mathbf{b}$ is the vector $\mathbf{b} = [b_1, b_{\tilde{N}}]$.
- $\mathbf{Q}$ is a $\tilde{N}$ by $\tilde{N}$ positive semi definite matrix, $Q_{\tilde{n}_1 \tilde{n}_2} \equiv c(I_{\tilde{n}_1}) c(I_{\tilde{n}_2}) \mathcal{K}(\mathbf{s}(I_{\tilde{n}_1}), \mathbf{s}(I_{\tilde{n}_2}))$.
- $\mathcal{K}(\mathbf{s}, \mathbf{s}') = \phi(\mathbf{s})^T \phi(\mathbf{s}')$ is the kernel function, and ϕ the mapping function.
- $\mathbf{1}_{\tilde{N}}$ is the full one vector of size $\tilde{N}$.
- $\mathbf{c}$ is the class vector $\mathbf{c} = [c(I_1), c(I_{\tilde{N}})]$.
- $B > 0$ is the upper bound.

The original optimization problem is:

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{b}, \xi} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + B \sum_{\tilde{n}=1}^{\tilde{N}} \xi_{\tilde{n}} \\ \text{subject to} \quad & c(I_{\tilde{n}}) (\mathbf{w}^T \phi(\mathbf{s}(I_{\tilde{n}})) + b) \geq 1 - \xi_{\tilde{n}}, \quad \xi_{\tilde{n}} \geq 0, \quad \tilde{n} = 1, \dots, \tilde{N} \end{aligned} \quad (4.7)$$

$\mathbf{w}$ is the normal vector to the separating hyperplane, $\mathbf{b}$ the offset of the hyperplane to the origin along the vector $\mathbf{w}$. A radial basis function kernel is used in the present work, i.e. $\mathcal{K}(\mathbf{s}, \mathbf{s}') = \exp(-\frac{\|\mathbf{s} - \mathbf{s}'\|^2}{2\beta^2})$.

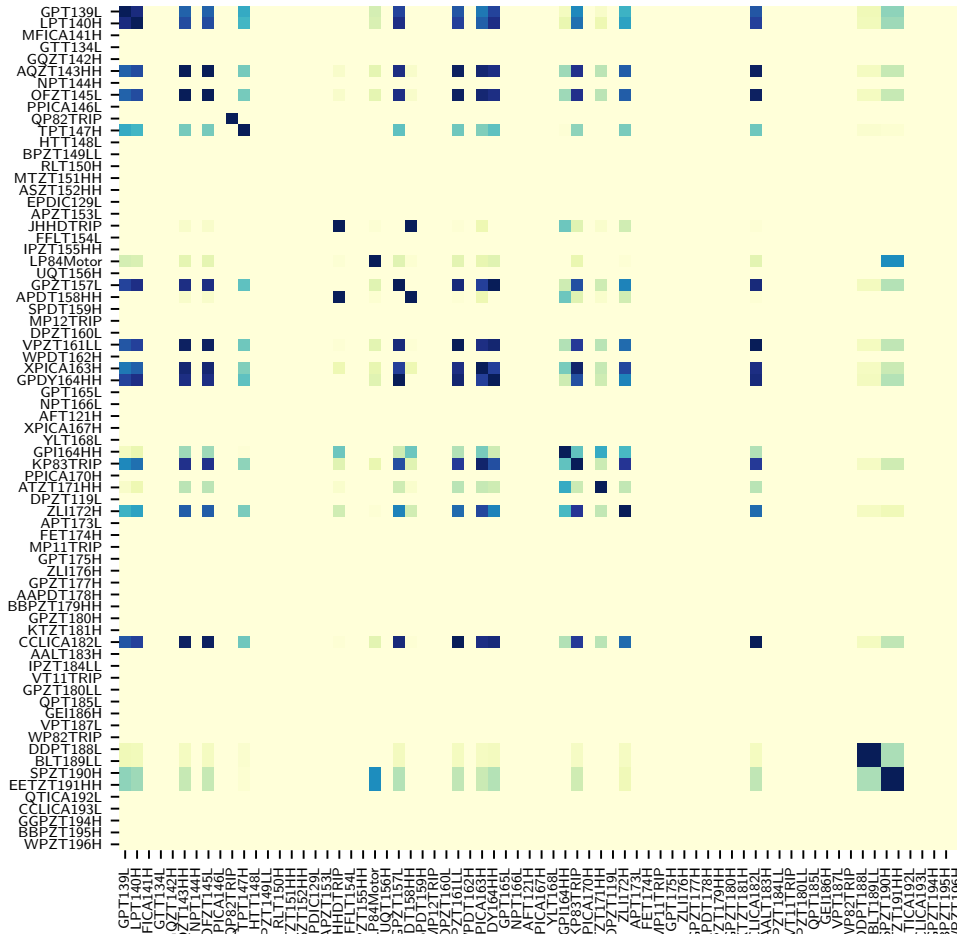


Figure 4.2: Alarm coactivation matrix for alarm flood 70 belonging to class D. Darker colours indicate higher coactivation values, with yellow corresponding to zero and black corresponding to one.

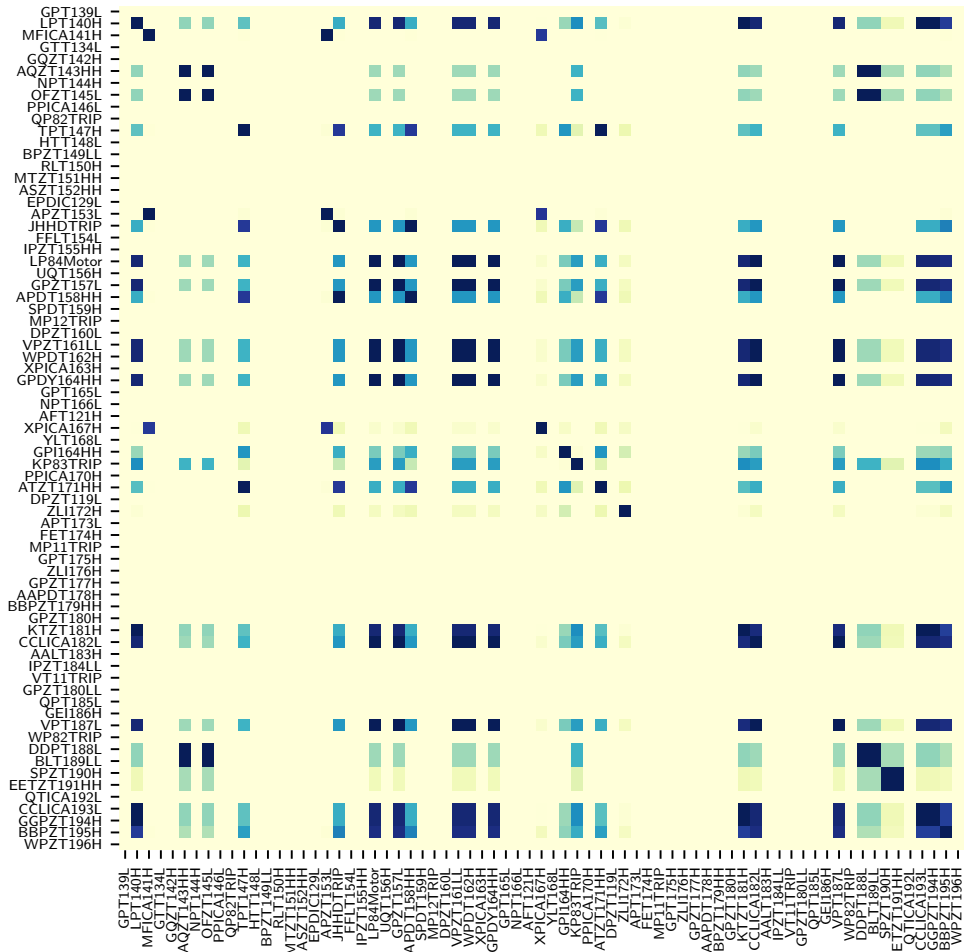


Figure 4.3: Alarm coactivation matrix for alarm flood 78 belonging to class D. Darker colours indicate higher coactivation values, with yellow corresponding to zero and black corresponding to one.

Since the classification of the alarm floods is a multi-class classification problem with N_C classes (N_C being the number of different types of alarm floods), $\frac{N_C(N_C-1)}{2}$ SVM classifiers are used in practice to separate each pair of classes. The final decision on the class c of the input vector $\mathbf{s}(I_n)$ is reached through a vote on the results of the $\frac{N_C(N_C-1)}{2}$ SVM classifiers.

Detection stage

The metric used for the detection is the probability of the classification result which is expressed as the posterior class probability defined by Platt's probabilistic output to the SVM as described in [188]. The posterior class probability $P(C = 1|\mathbf{s})$ is assumed to follow a sigmoid function of $\mathbf{s}$:

$$P(C = 1|\mathbf{s}) = \frac{1}{1 + \exp(\theta_1 D(\mathbf{s}) + \theta_0)} \quad (4.8)$$

where $D(\mathbf{s})$ the decision function of the SVM from Eq. 4.5 and θ_0 and θ_1 parameters calibrated on the training data as described in [188]. The multi-class equivalent $P(C = \hat{c}|\mathbf{s})$ of $P(C = 1|\mathbf{s})$ is computed using pairwise coupling as detailed in [189].

A probability threshold P_α is introduced as a cut-off parameter to determine if $P(C = \hat{c}(I_n)|\mathbf{s}(I_n))$ is high enough for $\mathbf{s}(I_n)$ to be considered as belonging to class $\hat{c}(I_n)$, or if it should be considered as part of a leftover class.

4.3.3 Online implementation

A flowchart for online implementation of the method for analysis of incoming alarm floods is presented in Figure 4.4. A window of size T_w (as introduced in Sub-section 4.2.2) is used to detect the periods when an alarm flood is developing. The size of the window T_w is tuned according to the dynamics of the alarm floods in the selected classes, as indicated in Sub-section 4.5.4.

Two thresholds are set: ν_{α_1} is the alarm flood threshold (as introduced in Sub-section 4.2.2), ν_{α_2} is an intermediate alarm threshold chosen such that $0 < \nu_{\alpha_2} < \nu_{\alpha_1}$. ν_{α_2} is introduced for the analysis to be triggered prior to the end of the alarm flood episode to provide operators with early diagnostics. Tuning of ν_{α_1} and ν_{α_2} is discussed in Sub-section 4.5.4.

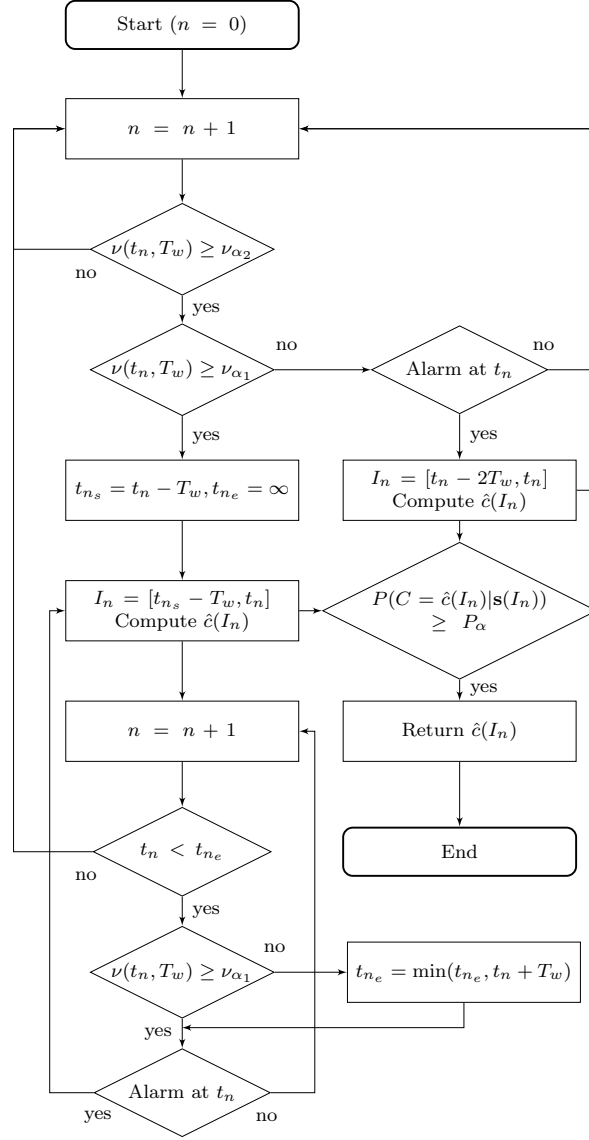


Figure 4.4: Flowchart of the proposed online implementation.

The alarm rate $\nu(t_n, T_w)$ is computed periodically at each time sample t_n on a window of size T_w , i.e. on $[t_n - T_w, t_n]$. When $\nu(t_n, T_w) \geq \nu_{\alpha_1}$ (first decision block in Figure 4.4), indicating that an alarm flood may be developing, $\mathbf{s}(I_n)$ is computed every time a new alarm occurs on the last interval of size $2T_w$, i.e. $I_n = [t_n - 2T_w, t_n]$. This corresponds to the window where $\nu(t_n, T_w) \geq \nu_{\alpha_1}$ and the window preceding it to include possible early alarms of the alarm flood.

When an alarm flood episode is detected (i.e. $\nu(t_n, T_w) \geq \nu_{\alpha_2}$, second decision block in Figure 4.4), the starting time t_{n_s} of the episode to consider is fixed to $t_n - T_w$. $\mathbf{s}(I_n)$ is computed every time a new alarm occurs ($I_n = [t_{n_s} - T_w, t_n]$ to include the possible early alarms) until T_w after the alarm rate drops below the alarm flood threshold (to include the possible late alarms), corresponding to the end time t_{n_e} .

Each $\mathbf{s}(I_n)$ is fed to the classifier. The classification output $\hat{c}(I_n)$ is presented to the operator if the probability $P(C = \hat{c}(I_n) | \mathbf{s}(I_n))$ returned by the SVM is greater than the detection threshold P_α .

4.4 Application to industrial case study

The performance of the alarm coactivation matrix classification is compared with the performance of a sequence alignment approach on the oil and gas separation plant. The sequence alignment approach is the MSW method [76] with a 1NN classifier [190], and the distance to the nearest neighbour is used as a metric for the detection stage for the MSW approach.

The comparison is done on the complete alarm flood episodes, i.e. offline analysis. The performance is assessed based on three criteria: the False Detection Rate (FDR), the Missed Detection Rate (MDR) and the accuracy of classification. FDR and MDR indicate the ability of the method to differentiate selected classes from leftover classes during the detection stage. The accuracy of classification indicates the ability to identify the correct class among the five selected in Table 4.1 during the classification stage. Sub-section 4.4.3 provides a criteria to assess the performance of the online implementation of the alarm coactivation matrix classification.

4.4.1 Assessment of performance of the detection stage

False detection rate. The FDR indicates the proportion of alarm floods from leftover classes that are identified as selected classes:

$$FDR = \frac{\delta_0}{\delta_0 + \delta_1} \quad (4.9)$$

where δ_0 is the number of alarm floods from leftover classes identified as selected classes, and δ_1 is the number of alarm floods from leftover classes identified as leftover classes.

For the FDR computation, the classifier is trained on the 81 selected alarm floods of Table 4.1 and tested on the 845 leftover alarm floods. The objective is to assess the ability of the method to deal with alarm flood classes that were not in the training set. Ideally, only alarm floods from selected classes should be detected, which means a low FDR.

Missed detection rate. The MDR indicates the proportion of alarm floods from selected classes that are not identified as such:

$$MDR = \frac{\delta'_0}{\delta'_0 + \delta'_1} \quad (4.10)$$

where δ'_0 is the number of alarm floods from selected classes identified as leftover classes, and δ'_1 is the number of alarm floods from selected classes identified as selected classes.

For the MDR computation, the classifier is trained on a set containing a random half of each of the five classes (in total 42 alarm floods), and tested on a set with the other halves (39 alarm floods). The objective is to assess the ability of the method to recognize alarm floods from classes that were in the training set. Ideally, all the alarm floods from the selected classes should pass the detection stage, which means a low MDR.

Section 4.5 analyse results for three configurations of the detection threshold that highlight the trade-off between FDR and MDR: one configuration corresponds to a priority given to a low MDR (called *low-MDR configuration*), one to a priority given to a low FDR (called *low-FDR configuration*), and one to a trade-off between MDR and FDR (called *trade-off configuration*).

4.4.2 Assessment of performance of the classification stage

The accuracy of classification is computed as the average accuracy of classification on 50 independent experiments using random splits between training and test data (using half for

training and half for testing) on the 81 alarm coactivation matrices of the selected classes. Using an average accuracy on 50 random splits mitigates the variability in the accuracy value due to the small number of observations.

4.4.3 Assessment of performance of the online implementation

The ability of the proposed method to classify ongoing alarm floods as one of the selected classes is evaluated through the probability $P(C = \hat{c}(I_n) | \mathbf{s}(I_n))$ of the classification outcome $\hat{c}(I_n)$ that can be compared with the detection threshold P_α . ν_{α_1} is maintained to 8 as indicated in Section 4.2.2 and ν_{α_2} is set to 4, i.e. half of the alarm flood threshold ν_{α_1} . The classifier is trained with the same training set as the one used for the MDR test, and tested on the same test set as the one used for the MDR test containing 39 alarm floods from the selected classes. In addition, the classifier is tested on 39 random alarm floods from leftover classes.

4.5 Results

4.5.1 Results of the detection stage

Table 4.3 presents the FDR and MDR values for three configurations introduced in Subsection 4.4.1. The performance of the alarm coactivation matrix method with SVM classification rivals the one of the sequence alignment method both in term of accuracy of classification and detection rates for the low-MDR and trade-off configurations, which correspond to usual industrial applications. The high value of the MDR for alarm coactivation matrix in the low-FDR configuration shows the limit of the detection threshold based on the posterior probability of the classification outcome. This configuration corresponds to a very high probability threshold P_α (here 0.8), which means that only the alarm floods with a very high classification probability outcome are classified as one of the selected classes. The low-FDR configuration is not adapted to industrial systems where variability in the alarm identifiers contained in the alarm floods and in the dynamics of activation and deactivation of the alarms can be observed.

Table 4.3: Results of FDR and MDR tests.

	Modified Smith Waterman		Alarm coactivation matrix	
	FDR (%)	MDR (%)	FDR (%)	MDR (%)
Low-MDR	22.1	0.0	17.2	0.0
Trade-off	7.5	7.7	8.8	7.7
Low-FDR	0.5	46.1	0.9	82.0

Table 4.4: Average accuracy of classification (offline analysis).

Method	Accuracy (%)
Modified Smith Waterman	91.2
Alarm coactivation matrix	92.2

4.5.2 Results of the classification stage

Table 4.4 summarizes the results of the accuracy tests. As for the detection, the alarm coactivation matrix method reaches a performance comparable to the MSW method in terms of accuracy of the classification. Table 4.4 shows that the accuracy of classification of both methods does not reach 100% since alarm floods from different classes have a number of alarms in common. For instance, one cannot expect the alarm floods of class D and class E both dealing with abnormal situations in the produced water reinjection system to be always distinguished because they involve the same alarms (e.g. low outlet flow of pump P21 AFICA116L) and the trips of the same assets (e.g. pump P21).

4.5.3 Results of the online implementation

This sub-section considers the validity of the online implementation of the alarm coactivation matrix method. Two criteria with regard to the online implementation of the method are studied: the ability to classify ongoing alarm floods, and the computation time of the method.

Classification of ongoing alarm floods

Figure 4.5, Figure 4.6 and Figure 4.7 illustrate the performance of the method for classification of ongoing alarm floods. Figure 4.5 corresponds to the low-MDR configuration (i.e.

Table 4.5: Online results for selected classes.

	Low-MDR	Trade-off	Low-FDR
Number of DIC	40	25	0
Number of alarm floods with DIC	5	4	0
Number of alarm floods with DIC at t_{n_e}	2	2	0
Number of DCC	270	204	29
Number of alarm floods with DCC	37	35	9
Number of alarm floods with DCC at t_{n_e}	37	34	8
Average progress of the sequence before DCC (in %)	67.3	77.1	93.6

$P_\alpha = 0.5$), Figure 4.6 corresponds to the trade-off configuration (i.e. $P_\alpha = 0.6$) and Figure 4.7 corresponds to the low-FDR configuration (i.e. $P_\alpha = 0.8$). In each figure, the top plot represents the classification of the 39 alarm floods from the selected classes and the bottom plot represents the classification of the 39 alarm floods from the leftover classes. Each point corresponds to a classification, triggered every time a new alarm (or simultaneous alarms) occurs. Black points indicate that the detection has not occurred (i.e. that the probability $P(C = \hat{c}(I_n)|s(I_n))$ of the classification output $\hat{c}(I_n)$ has not crossed the detection threshold P_α), red points indicate that the detection has occurred but the classification is incorrect (designated as scenario DIC for Detected but Incorrectly Classified), green points indicate that the detection has occurred and the classification is correct (designated as scenario DCC for Detected and Correctly Classified). The evolution of the ongoing alarm flood is expressed in term of ratio of alarms triggered to the total number of alarms in an alarm flood. The final classification is given at the end of the alarm flood episode interval t_{n_e} , shortly after all the alarms of the alarm flood appeared. The classification results at t_{n_e} correspond to the offline results (Table 4.3).

Tables 4.5 and 4.6 summarize the online classification results for the selected classes and for the leftover classes. In Table 4.5, the number of DIC scenarios describes the risk of receiving a diagnosis that does not correspond to the actual ongoing situation. This affects five alarm floods in the low-MDR configuration and four alarm floods in the trade-off configuration, but only two alarm floods in both configurations when all the alarms of the sequences have triggered. The low-FDR configuration reduces the number of DIC scenarios. In Table 4.6, the number of DIC scenarios indicates the number of leftover classes falsely detected as selected classes.

Table 4.6: Online results for leftover classes.

	Low-MDR	Trade-off	Low-FDR
Number of DIC	63	11	0
Number of alarm floods with DIC	4	3	0
Number of alarm floods with DIC at t_{n_e}	4	1	0

The low-FDR configuration ensures that no alarm flood is falsely detected in this subset, while the trade-off configuration offers a compromise with only one alarm flood eventually falsely detected. The number of DCC scenarios in Table 4.5 indicates the ability of the method to recognize and correctly classify the ongoing alarm flood. The low-MDR and trade-off configurations recognize and correctly classify respectively 37 and 34 alarm floods, and the correct classification occurs in average at 67.3% of the sequence in the low-MDR configuration and at 77.1% of the sequence in the trade-off configuration. The performance of the low-FDR configuration is much lower than the other two configurations since it only detects and correctly classifies 8 alarm floods, after 93.6% of the sequence has appeared.

Computation time

The *response time* is the time between an alarm occurrence and the time when the method returns a classification. The average response time of the alarm coactivation matrix method with a C# implementation on a 64 bit Windows PC with Intel(R) Core(TM) i7-6820 2.70 GHz CPU and 16.0 GB memory is 1.5 ms, which is satisfactory compared to the one second resolution of the alarms on the plant. The response time analysis of the MSW method in [80] shows that an optimized version of the method for online analysis would also have a response time below one second, and thus would also be applicable in this case. However, while the MSW method requires the modifications presented in [79] (for the template extraction) and in [80] (for the sequence window filter and an incremental dynamic programming strategy) to be implemented online, the method based on alarm coactivations has been designed directly for online applications.

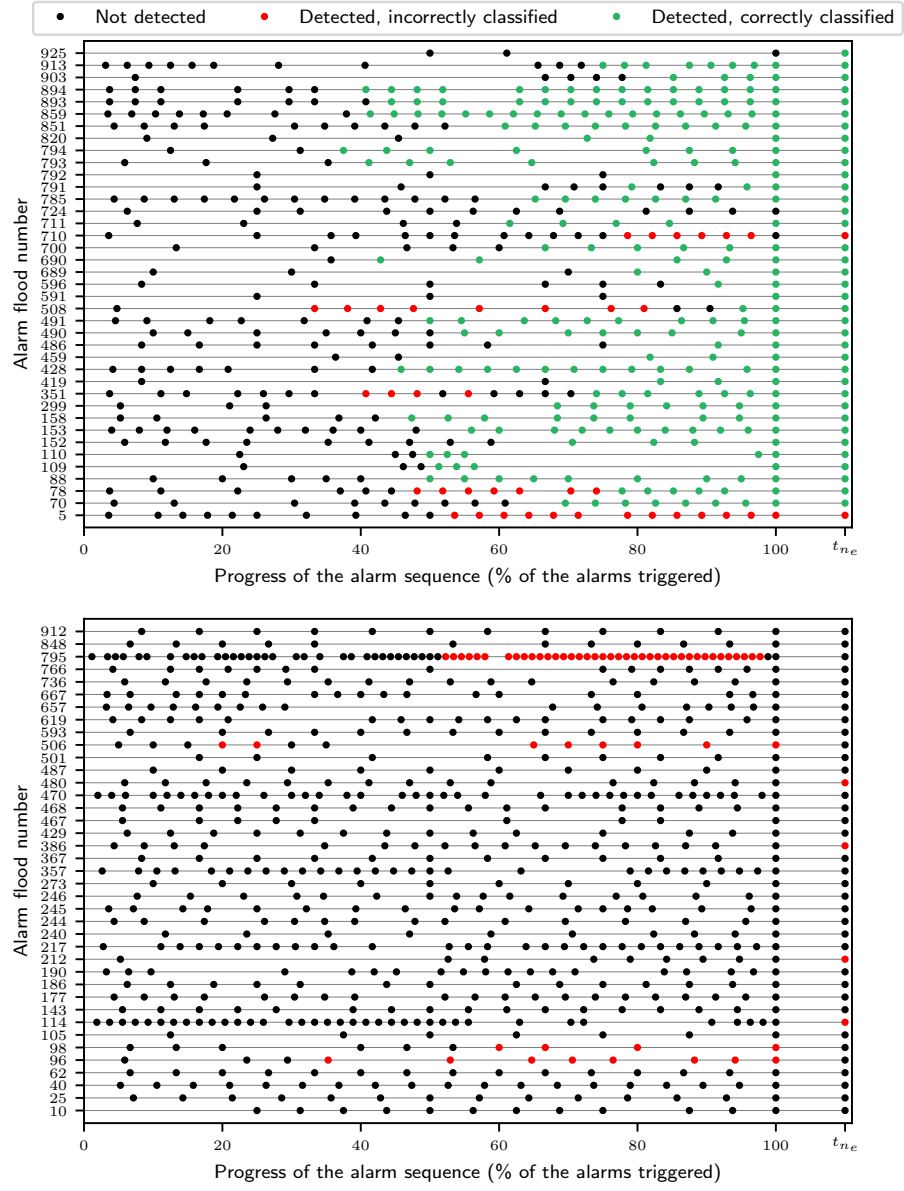


Figure 4.5: Classification outcomes for ongoing alarm floods from selected classes (top) and leftover classes (bottom) in the low-MDR configuration ($P_\alpha = 0.5$).

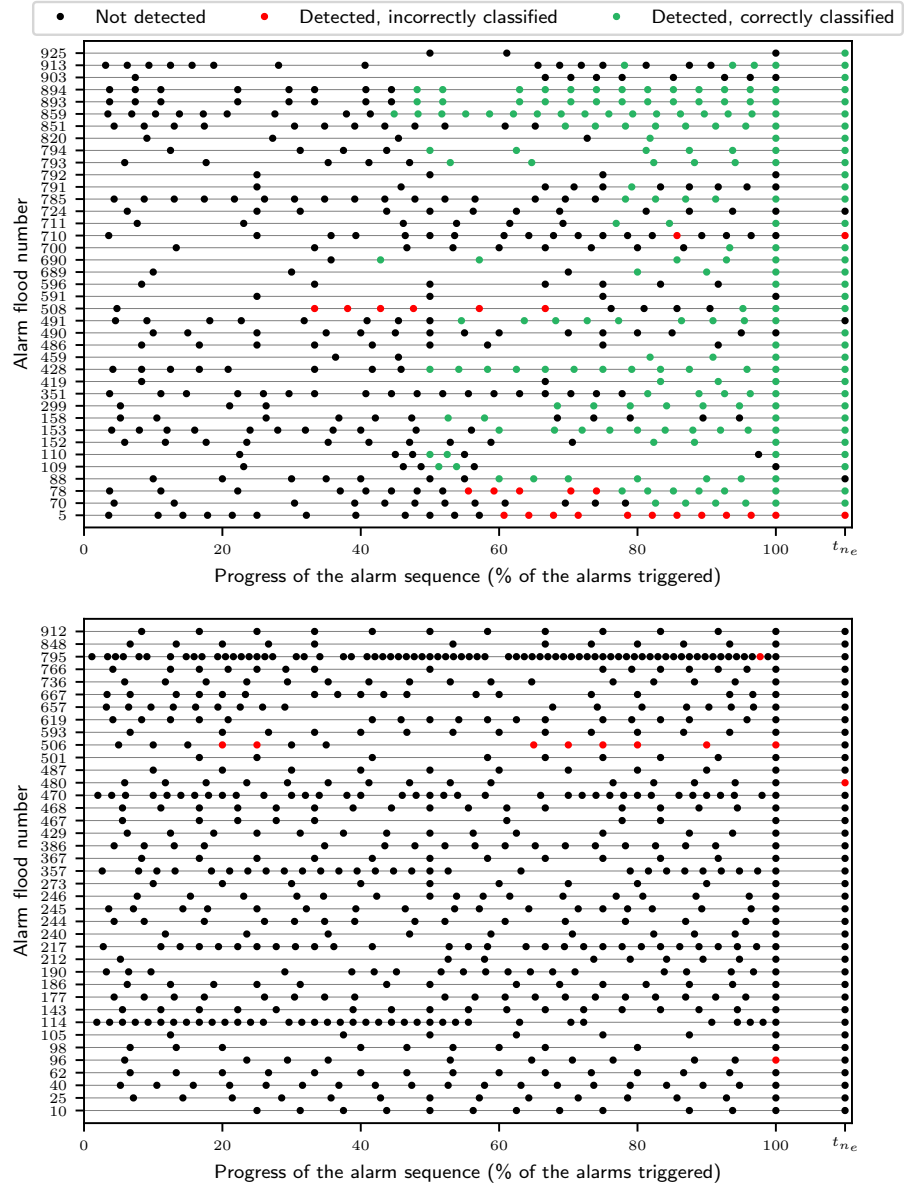


Figure 4.6: Classification outcomes for ongoing alarm floods from selected classes (top) and leftover classes (bottom) in the trade-off configuration ($P_\alpha = 0.6$).

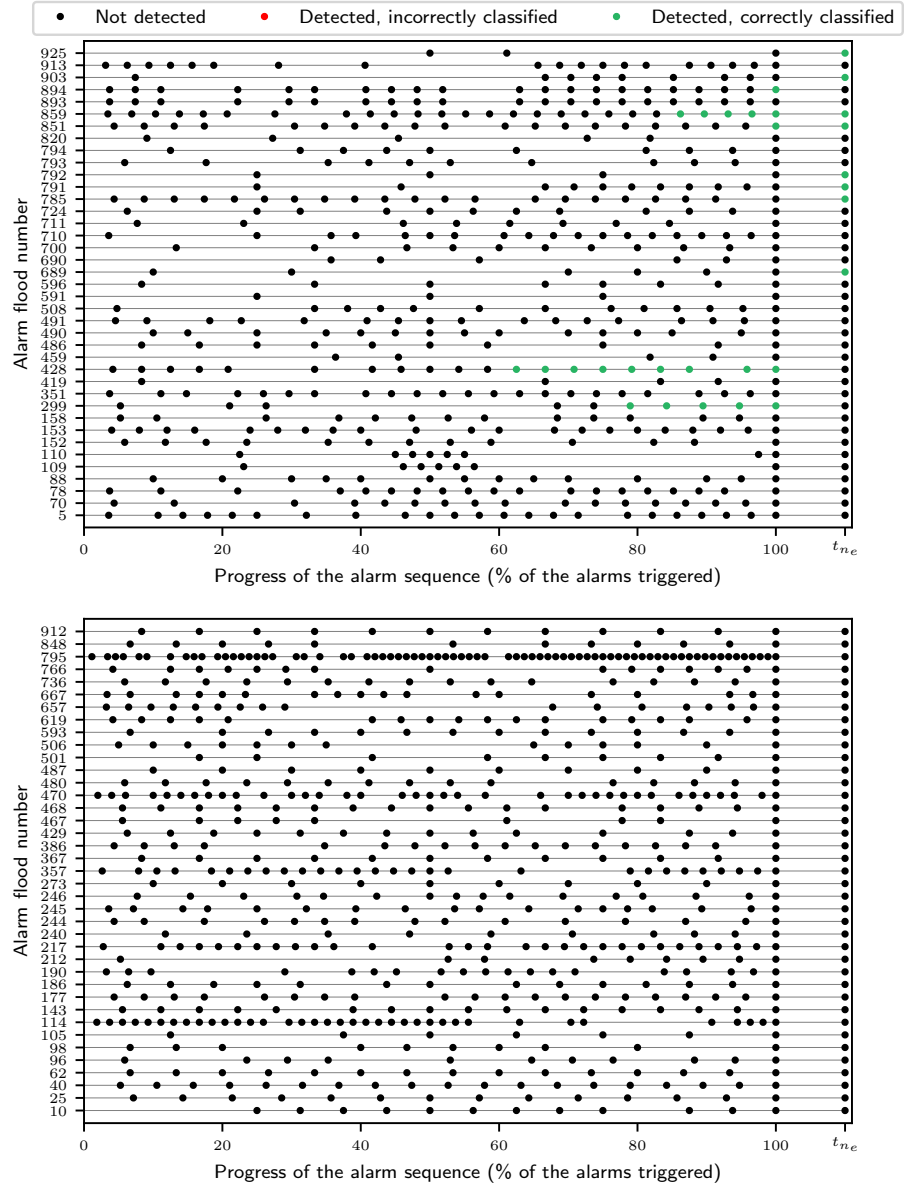


Figure 4.7: Classification outcomes for ongoing alarm floods from selected classes (top) and leftover classes (bottom) in the low-FDR configuration ($P_\alpha = 0.8$).

4.5.4 Analysis of the parameters

A practical implementation requires default values of parameters such as the detection threshold, the size of the window, and the alarm thresholds. These depend on the application. Recommendations are given below.

The detection threshold corresponds to a probability threshold P_α for the SVM classification and determines the confidence required on the output of classification before it is presented to the operator. Three configurations are proposed in this chapter: a low-MDR configuration ($P_\alpha = 0.5$), a low-FDR configuration ($P_\alpha = 0.8$), and a trade-off configuration ($P_\alpha = 0.6$). The trade-off configuration should be used, or alternatively the low-MDR configuration if an early diagnosis is preferred and false detections can be handled. The low-FDR configuration is not recommended since it provides a late diagnosis and misses most of the known situations.

The size of the window T_w corresponds to the reference time interval based on which the alarm flood episodes are determined. T_w is tuned according to the dynamics of activation of the alarms in the alarm floods of the selected classes. T_w should match the average duration of the alarm floods in the selected classes, which is around ten minutes for the oil and gas separation plant case study.

The alarm threshold ν_{α_1} determines how alarm floods are defined and has to be chosen in relation to the size of the window T_w . As for T_w , the default value for ν_{α_1} should be chosen according to the alarm floods in the selected classes used for training of the classifier. While T_w should match the average duration of the alarm floods in the selected classes, ν_{α_1} should match the minimum number of alarms over T_w in the alarm floods of the selected classes, for new episodes of the same classes to be considered in the online implementation. In the case of the oil and gas separation plant, the minimum number of alarms per ten minutes in the alarm floods of the selected classes is set to eight, as explained in Sub-section 4.2.2.

The alarm threshold ν_{α_2} determines how early the online classification begins as the first alarms of the alarm floods trigger. ν_{α_2} should be tuned based on the value of ν_{α_1} . A value of ν_{α_2} around $\nu_{\alpha_1}/2$ is recommended to avoid the online classification being triggered too early and too frequently. In the case of the oil and gas separation plant, ν_{α_1} is set to eight (corresponding to eight alarms per ten minutes for defining the alarm floods), so ν_{α_2} is set to four.

4.6 Summary of the chapter

This chapter shows by means of a case study how a binary series approach can address a gap in online alarm flood classification in the industry. While not considering the order of the alarms in the alarm floods like previous sequence mining methods, the proposed method based on alarm coactivations performs well compared to an established similarity-based classification method with MSW sequence alignment. The MSW method requires additional steps to be implemented online (alarm flood template extraction, sequence window filter, incremental dynamic programming strategy), whereas the method based on alarm coactivations has been designed directly for online applications following the example of online FDD methods. Traditional online FDD methods detect and classify faults based on multiple continuous time series of the process measurements; the method based on alarm coactivations suggest a classification of alarm floods based on multiple alarm binary series.

The ability of the alarm coactivation matrix method to detect selected classes, the accuracy of classification and the ability to classify ongoing alarm floods are demonstrated on an offshore oil and gas separation plant.

Chapter 5

Integrating alarm design in online fault detection and diagnosis through alarm-range normalization

5.1 Description of the chapter

Preprocessing is a prerequisite for applying online FDD methods on industrial process measurements, and normalization is a key element of preprocessing [86]. Normalization is usually based on statistical indicators inferred from the process measurements used for training the online FDD methods, but limited availability of historical data for training also impacts the normalization.

This chapter investigates an alternative normalization that uses the alarm settings defined during the alarm design stage. The proposed Alarm-Range (AR) normalization uses the difference between the low and the high alarm thresholds for each process variable to situate the variations of the related process measurement in the context of the safe operational range. The impact of significant variations (with regard to the alarm thresholds) on the online FDD is accentuated compared to the impact of noisy variations due to sensor noise or process noise. This chapter provides a study of the effect of AR-normalization on fault detection, fault isolation, and fault identification on reference methods in the industry, namely Principal Component Analysis (PCA) for fault detection and fault isolation, and nearest neighbour classification for fault identification. The analysis is carried on the industrial case study of an offshore oil and gas separation plant where process measurements

and alarm thresholds could be extracted. The material presented in this chapter has been published in the following articles:

- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “On improving fault detection and diagnosis using alarm-range normalisation,” in *Proceedings of 10th SAFEPROCESS Symposium, Warsaw, Poland, August 29-31, IFAC-PapersOnLine*, vol. 51, pp. 1227–1232, 2018
- M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, “Integration of alarm design in fault detection and diagnosis through alarm-range normalization,” *Submitted to Control Engineering Practice*

Section 5.2 gives additional information on the integration of alarm systems in online FDD. Section 5.3 details the normalization, the fault detection, isolation, and identification methods used for the analysis. Section 5.4 demonstrates the benefits of AR-normalization in the case study from the oil and gas separation plant. The results of the study are presented and discussed in Section 5.5.

5.2 Background

5.2.1 Integrating alarm systems in online fault detection and diagnosis

Alarms are the foremost indicators for operators when it comes to detecting and identifying ongoing abnormal situations on the plant. Improving alarm management based on analysis of the process measurements is a common research direction [26]. Several methods have been proposed to design the best alarm thresholds based on process measurements using a univariate approach [191] or a multivariate approach [192]. Alarm configuration and alarms triggered by abnormality propagation are other problems for which solutions based on the analysis of process measurements have been proposed [26].

Methods for analysis of alarm records can help investigating abnormal situations [33] such as alarm floods, conditions during which the alarm rate is greater than the operator can effectively manage [21]. Online alarm flood classification methods [30, 60, 61, 80, 81] rely on approaches that are close to traditional fault detection and identification approaches [33], but only consider alarm records.

Preliminary analysis of alarm records can be applied to speed up root cause analysis based on process measurements. The authors of [27] clustered similar abnormal situations based on their alarms before proceeding to root cause analysis on process measurements. The authors of [28] conducted a preliminary root cause analysis on alarm records to identify process variables that could be root cause candidates and reduce the scope of root cause analysis on process measurements.

Statistical alarms are a particular case since they are generated using statistical indicators of the process measurements such as the standard deviation in normal operation. Statistical alarms can be substituted for process measurements to reduce the computational load, in particular for root cause analysis with transfer entropy [57, 67, 70]. Statistical alarms can also be substituted for process measurements when doing variable selection for fault detection and identification to limit the impact of noisy variations of the process measurements [32].

Direct integration of process measurements and alarm records for fault detection and identification has been addressed in [29] where a two-stage Bayesian classifier is used. The authors of [55] have released a dataset from a multiphase flow rig to underpin future research into the integration of heterogeneous data for fault detection and identification, in particular alarm records and process measurements.

5.2.2 Motivation for the method

Normalization can have a high impact on the performance of FDD methods. Existing normalization approaches such as standardization are dependent on the availability of historical process measurements to estimate appropriate statistical indicators (e.g. standard deviation) for each process measurement. In practice, finding normal operation data $\tilde{\Omega}_{N0}$ or faulty data $\tilde{\Omega}_F$ that can be used for training the FDD model is a difficult and time-consuming task.

For $\tilde{\Omega}_{N0}$, it is challenging to find a time interval where all process measurements are fault-free and where all process measurements present representative variations. Variations during the fault-free interval do not necessarily give a good indication of the acceptable range of variations of the process measurements. For example, variations of process measurements under feedback control are usually small and may not provide robust statistical indicators for the normalization (e.g. for N-standardization introduced in Sub-section 3.3.2). Section 5.4 gives an example of industrial case study where the process measurements present a low variability during normal operation. The case study considers the water reinjection section of the offshore oil and gas separation plant introduced in Sub-section 2.5.1, where

the water flows are controlled to ensure the safe operation of the plant. Sub-sections 5.5.2 and 5.5.3 show how traditional N-standardization combined with the low variability of the measurements in $\tilde{\Omega}_{N0}$ can lead to spurious results in fault detection and fault isolation.

F-standardization (introduced in Sub-section 3.3.2) also depends on the faulty process measurements available for training in $\tilde{\Omega}_F$. The patterns of the process measurements during fault occurrences in the training set might not be exactly the same as the patterns of the process measurements in new occurrences, and a change in amplitude of variation of a measurement from the training set to the test set can be amplified by the standardization. Sub-section 5.5.4 illustrates how F-standardization combined with changing fault characteristics can lead to spurious fault identification.

The proposed AR-normalization is intended as a normalization approach that should facilitate and speed up the work of the engineers in charge of implementing FDD methods in industrial environments. AR-normalization scales process measurements using the alarm range (the difference between the upper and lower alarm thresholds). Such scaling is independent of the statistics of the training data and can be used both for fault detection and isolation and for fault identification. Therefore, the engineers do not need to consider the variability of the measurements in the choice of the training data and can focus on isolating the intervals of normal operation and the fault intervals for $\tilde{\Omega}_{N0}$ and $\tilde{\Omega}_F$.

5.3 Method development

5.3.1 Alarm-range normalization

AR-normalization uses absolute alarm thresholds of a process variable as a reference for the normalization of the corresponding process measurement. The difference between the high absolute alarm threshold h_m and the low absolute alarm threshold l_m of a process variable X_m can be used to give an indication of the acceptable amplitude of variation of the corresponding process measurement. The AR-normalized plant vector at time t_n is defined as $\tilde{\mathbf{x}}(t_n) = [\tilde{x}_1(t_n), \dots, \tilde{x}_M(t_n)]^T$ where:

$$\tilde{x}_m(t_n) = \frac{\dot{x}_m(t_n) - \mu_m(\tilde{\Omega}_{N0})}{h_m - l_m} \quad (5.1)$$

5.3.2 Fault detection and isolation

PCA modeling

The fault detection and isolation in this chapter is based on Principal Component Analysis (PCA) [117]. PCA is a linear dimensionality reduction technique that determines the set of orthogonal vectors (called loading vectors) ordered by the amount of variance explained in the loading vectors direction. The PCA model is trained with the training set $\tilde{\Omega}_{N0}$ of process measurements from normal operation. The $\tilde{N}$ normalized plant vectors $\tilde{\mathbf{x}}(t_{\tilde{n}})$ in $\tilde{\Omega}_{N0}$ are stacked into a matrix $\tilde{\mathbf{X}}_{\tilde{N}} \in \mathbb{R}^{\tilde{N} \times M}$, and the loading vectors are computed via the singular value decomposition [129]:

$$\tilde{\mathbf{X}}_{\tilde{N}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (5.2)$$

where $\mathbf{U} \in \mathbb{R}^{\tilde{N} \times \tilde{N}}$ and $\mathbf{V} \in \mathbb{R}^{M \times M}$ are orthogonal matrices, and the matrix $\mathbf{\Sigma} \in \mathbb{R}^{\tilde{N} \times M}$ contains the non-negative real singular values λ_m of decreasing magnitude on the diagonal and zero off-diagonal elements.

Fault detection

Each new normalized plant vector $\tilde{\mathbf{x}}(t_n)$ at time t_n is tested with the PCA model to determine if it should be considered as normal or faulty. Hotelling's T^2 statistic and the Q statistic can be used to detect faulty plant vectors [129]. The value of T^2 at time t_n can be computed as:

$$T^2(t_n) = \tilde{\mathbf{x}}(t_n)^T \mathbf{P} \mathbf{\Sigma}_{\mathbf{M}_a}^{-2} \mathbf{P}^T \tilde{\mathbf{x}}(t_n) = \tilde{\mathbf{x}}(t_n)^T \mathbf{D} \tilde{\mathbf{x}}(t_n) \quad (5.3)$$

where $\mathbf{D} = \mathbf{P} \mathbf{\Sigma}_{\mathbf{M}_a}^{-2} \mathbf{P}^T$, $\mathbf{\Sigma}_{\mathbf{M}_a}$ contains the first M_a rows and columns of $\mathbf{\Sigma}$ and $\mathbf{P} \in \mathbb{R}^{M \times M_a}$ is the matrix of the loading vectors associated with the M_a largest singular values.

The portion of the observation space corresponding to the $M - M_a$ smallest singular values can be monitored by the Q statistic:

$$Q(t_n) = \mathbf{r}(t_n)^T \mathbf{r}(t_n) = \tilde{\mathbf{x}}(t_n)^T \mathbf{C} \tilde{\mathbf{x}}(t_n) \quad (5.4)$$

where $\mathbf{r}(t_n) = (\mathbf{I}_M - \mathbf{P} \mathbf{P}^T) \tilde{\mathbf{x}}(t_n)$ is the residual vector at time t_n , $\mathbf{I}_M$ is the identity matrix of size M , and $\mathbf{C} = (\mathbf{I}_M - \mathbf{P} \mathbf{P}^T)$.

$\tilde{\mathbf{x}}(t_n)$ is considered as faulty if $T^2(t_n) \geq T_\alpha^2$ or if $Q(t_n) \geq Q_\alpha$, where T_α^2 is the T^2 statistic threshold and Q_α the Q statistic threshold, and α indicates the significance level of the thresholds. T_α^2 and Q_α are computed based on percentiles of the values of T^2 and Q for the process measurements in $\tilde{\Omega}_{N0}$. In the following, $T_{1\%}^2$ and $Q_{1\%}$ correspond to the thresholds for $\alpha = 1\%$, and $T_{0.1\%}^2$ and $Q_{0.1\%}$ correspond to the thresholds for $\alpha = 0.1\%$.

Fault isolation

If $\tilde{\mathbf{x}}(t_n)$ is detected as faulty, the contribution of each process variable X_m to $T^2(t_n)$ and $Q(t_n)$ is investigated through reconstruction-based contribution analysis [132]. The reconstruction-based contribution $f_{T^2,m}(t_n)$ of X_m to $T^2(t_n)$ is:

$$f_{T^2,m}(t_n) = \frac{(\xi_m^T \mathbf{D} \tilde{\mathbf{x}}(t_n))^2}{d_{mm}} \quad (5.5)$$

where ξ_m is the m th column of the m by m identity matrix and d_{mm} is the m th diagonal element of the matrix $\mathbf{D}$ defined in sub-section 5.3.2.

The reconstruction-based contribution $f_{Q,m}(t_n)$ of X_m to $Q(t_n)$ is:

$$f_{Q,m}(t_n) = \frac{r_m^2}{c_{mm}} \quad (5.6)$$

where r_m is the m th row of the residual vector $\mathbf{r}(t_n)$ and c_{mm} is the m th diagonal element of the matrix $\mathbf{C}$ defined in sub-section 5.3.2.

5.3.3 Fault identification

If $\tilde{\mathbf{x}}(t_n)$ is detected as faulty, $\tilde{\mathbf{x}}(t_n)$ is compared to normalized plant vectors $\tilde{\mathbf{x}}(t_{\tilde{n}})$ of the training set of process measurements from faulty operation $\tilde{\Omega}_F$ to determine if $\tilde{\mathbf{x}}(t_n)$ belongs to one of the fault classes c covered by the classifier. The estimated fault class $\hat{c}(t_n)$ corresponding to $\tilde{\mathbf{x}}(t_n)$ is identified using a popular classifier for fault identification in the industry [108], the 1-Nearest Neighbour (1NN) classification algorithm. 1NN associates $\tilde{\mathbf{x}}(t_n)$ with the class $c(t_{\tilde{n}_1})$ of the nearest neighbour $\tilde{\mathbf{x}}(t_{\tilde{n}_1})$ of $\tilde{\mathbf{x}}(t_n)$ in $\tilde{\Omega}_F$:

$$\tilde{\mathbf{x}}(t_{\tilde{n}_1}) = \arg \min_{\tilde{\mathbf{x}}(t_{\tilde{n}}) \in \tilde{\Omega}_F} \sqrt{\sum_{m=1}^M (\tilde{x}_m(t_n) - \tilde{x}_m(t_{\tilde{n}}))^2} \quad (5.7)$$

Table 5.1: Selected faults in the PWRI section.

Fault	Description	Number of occurrences
1	Low suction pressure without pump trip	2
2	Low suction pressure with P21 trip	3
3	Fuel change	8
4	P21 trip due to vibrations	2

5.4 Application to industrial case study

5.4.1 Description of the industrial case study

The industrial case study is the offshore oil and gas separation plant described in Sub-section 2.5.1. In this chapter, the analysis focuses on the Produced Water Reinjection (PWRI) section described in Figure 2.6. Recurrent faults have been observed by experts in the PWRI section over a period of four months. Four different types of faults have been labelled and are listed in Table 5.1:

- Fault 1: the fault starts with a pressure drop in the water coming into the PWRI section, observable through the drop in the suction pressures of P11 and P21. The situation goes back to normal after a few samples.
- Fault 2: the fault starts with a pressure drop in the water coming into the PWRI section, observable through the drop in the suction pressures of P11 and P21. The low suction pressure triggers a trip of P21.
- Fault 3: the fault starts with a change in the type of fuel used for the pumps of the PWRI section (from diesel to gas, or vice versa). The fuel change triggers a drop of pressure and flow in the whole water system that can lead to a trip of P11 or P21 (or both).
- Fault 4: the fault starts with a trip of P21 due to high vibrations in the motor of the pump.

Table 5.2 lists the 16 process variables in the PWRI section. For each process variable, the table indicates the high and low alarm thresholds defined in the alarm system, as well as the alarm range. The table also indicates the standard deviation of each process variable in normal operation and during the selected faults.

Table 5.2: Process variables, alarm thresholds, and standard deviations in the PWRI section.

Original tag	Description	Unit	l_m	h_m	$h_m - l_m$	$\sigma_m(\dot{\Omega}_{N0})$	$\sigma_m(\dot{\Omega}_F)$
CLY128A	Water level degassing drum	%	0.00	90.00	90.00	1.04	11.34
CPZT151	Suction pressure P11	barg	0.40	7.00	6.60	0.02	0.26
CPZT112	Suction pressure P21	barg	0.60	7.00	6.40	0.01	0.31
AFT121	Water flow to water injection tank 2	$\text{m}^3 \text{h}^{-1}$	250.00	900.00	650.00	1.75	344.85
AFZT114	Inlet flow P22	$\text{Sm}^3 \text{h}^{-1}$	470.00	900.00	430.00	2.07	278.77
AFZT139	Water flow to water injection tank 2	$\text{m}^3 \text{h}^{-1}$	200.00	900.00	700.00	1.96	290.08
AFZT115	Discharge flow P21	$\text{m}^3 \text{h}^{-1}$	340.00	750.00	410.00	3.46	285.05
ALT148	Water level of the injection rundown tank	%	10.00	95.00	85.00	0.10	0.08
APT123	Suction pressure P12 and P22	barg	7.10	40.00	32.90	0.01	3.73
APT146	Discharge pressure P22	barg	90.00	193.00	103.00	0.10	57.36
APZT141	Discharge pressure P11	barg	10.00	23.00	13.00	0.00	7.57
APZT154	Suction pressure P12	barg	6.50	23.00	16.50	0.01	1.24
APZT143	Discharge pressure P12	barg	8.00	170.00	162.00	0.09	14.13
APZT117	Discharge pressure P21	barg	10.00	23.00	13.00	0.02	12.37
APZT140	Suction pressure P22	barg	6.50	23.00	16.50	0.02	3.67
APZT142	Discharge pressure P22	barg	85.00	215.00	130.00	0.05	59.64

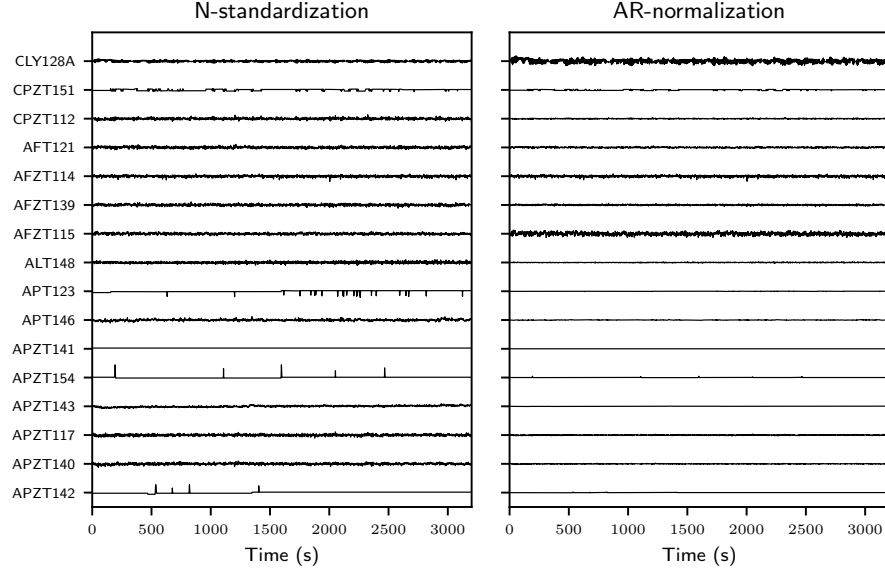


Figure 5.1: Process measurements in the training set $\check{\Omega}_{N0}$ normalized with N-standardization and AR-normalization. The scale is $[-20, 20]$ for each N-standardized measurement and $[-0.1, 0.1]$ for each AR-normalized measurement.

5.4.2 Design of experiment

Common normalization practices in online FDD such as N-standardization and F-standardization were introduced in Sub-section 3.3.2. AR-normalization is compared to N-standardization for fault detection and isolation, and AR-normalization is compared to F-standardization for fault identification. N-standardization uses the standard deviation $\sigma_m(\check{\Omega}_{N0})$ computed on the training set of process measurements from normal operation $\check{\Omega}_{N0}$, while F-standardization uses the standard deviation $\sigma_m(\check{\Omega}_F)$ computed on the training set of process measurements from faulty operation $\check{\Omega}_F$. AR-normalization uses the difference between the high alarm threshold h_m and the low alarm threshold l_m . All three normalizations use the mean $\mu_m(\check{\Omega}_{N0})$ computed during normal operation.

Fault detection and isolation

Figure 5.1 represents the N-standardized process measurements and the AR-normalized process measurements during normal operation in the training set $\check{\Omega}_{N0}$. The values of $\sigma_m(\check{\Omega}_{N0})$ for each process variable are indicated in Table 5.2.

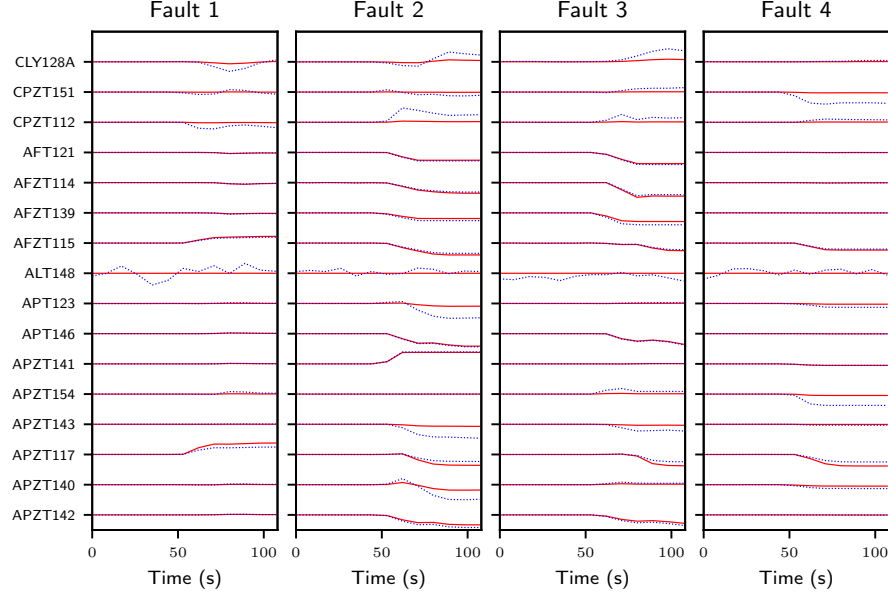


Figure 5.2: Process measurements in the training set $\tilde{\Omega}_F$ normalized with F-standardization (blue) and AR-normalization (red). The scale is $[-4, 4]$ for each F-standardized measurement and $[-2, 2]$ for each AR-normalized measurement.

A PCA model is built on the N-standardized measurements and another PCA model is built for the AR-normalized measurements. APZT141 is not included in the PCA model since it does not move during normal operation. The number of loading vectors is set to 10 (i.e. $M_a = 10$).

The fault detection is done on a test set of measurements from normal operation Ω_{N0} . The training set $\tilde{\Omega}_F$ used for training the fault identification model is also used for testing the detection and isolation on one occurrence of each type of fault. The reconstruction-based contribution plots for T^2 and Q are analysed for each of those faults.

Fault identification

The training set $\tilde{\Omega}_F$ containing one occurrence of each type of fault listed in Table 5.1 is used for training the classifier for the fault identification. Figure 5.2 represents the normalized process measurements during each type of fault in $\tilde{\Omega}_F$. The blue dotted lines correspond to the F-standardized measurements, and the red lines correspond to the AR-normalized measurements. The values of $\sigma_m(\tilde{\Omega}_F)$ for each process variable are indicated in Table 5.2.

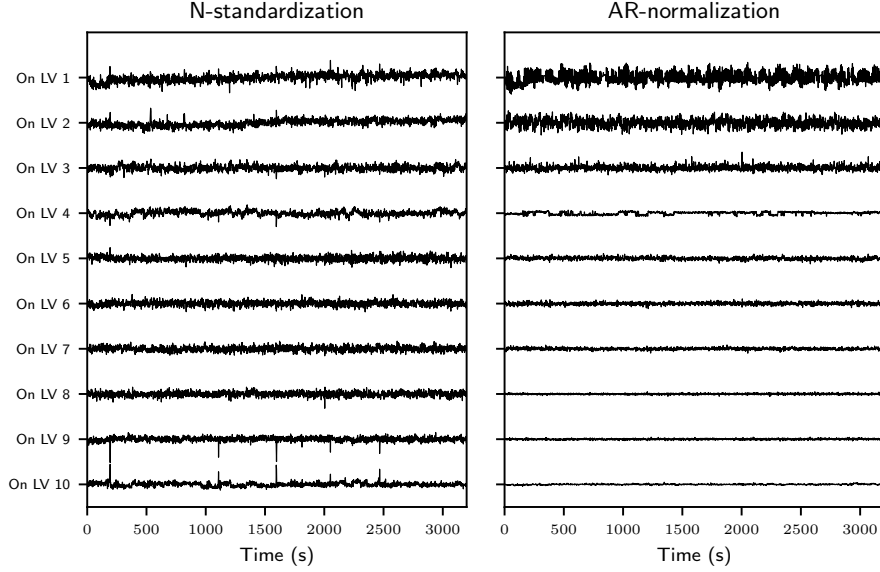


Figure 5.3: Projection of the process measurements of $\check{\Omega}_{N0}$ on the first ten Loading Vectors (LV) with N-standardization and AR-normalization. The scale is $[-10, 10]$ for each projection on the loading vectors with N-standardization and $[-0.05, 0.05]$ for each projection on the loading vectors with AR-normalization.

The classifier is tested on a test set of data from the remaining fault occurrences, Ω_F . The new fault occurrences are detected using the PCA model with AR-normalization, and a classification is proposed for each sample as soon as the fault is detected. The fault identification focuses on the four types of faults of Table 5.1 and does not consider unknown types of faults.

5.5 Results

5.5.1 PCA modeling

Figure 5.3 shows the projection of the process measurements of the training set $\check{\Omega}_{N0}$ on the first ten loading vectors of the PCA with N-standardization and with AR-normalization. Figure 5.4 indicates the cumulative variance explained by the loading vectors of the PCA with N-standardization (in blue) and with AR-normalization (in red) on the training set $\check{\Omega}_{N0}$. The slope is steeper for AR-normalization than for N-standardization in Figure 5.4. This shows that the variance explained by the loading vectors associated with the largest

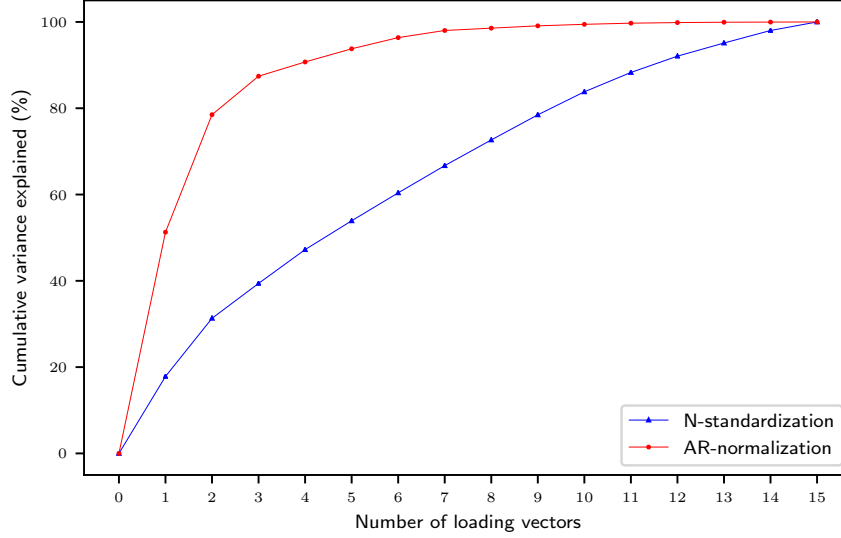


Figure 5.4: Cumulative variance explained by the loading vectors on $\tilde{\Omega}_{N0}$ with N-standardization and AR-normalization.

singular values is larger for AR-normalization than for N-standardization, and that the variance explained by the loading vectors associated with the smallest singular values is negligible for AR-normalization.

The PCA model with AR-normalization emphasizes the variations that are the most representative compared to the safe operation range defined by the alarm thresholds. This has two benefits. The first benefit is to build a model capturing the representative variations and limiting the impact of small variations (compared to the alarm thresholds) that may not give a good indication of the variability of the measurements in the normal operation data available for training. The second benefit is to give more weight to process measurements with larger variations compared to their alarm threshold, since those process measurements are the most likely to reach their alarm thresholds. Therefore, each of those variations should have a high impact on the detection statistic T^2 .

Capturing small variations in the model can have a negative impact. N-standardization gives the same weight to all process measurements in $\tilde{\Omega}_{N0}$ since they are all scaled to unit-variance. Process measurements with small variations in $\tilde{\Omega}_{N0}$ have the same weight as the other ones, even though the variations may only represent minor fluctuations. Therefore, any move of one of those process measurements has a major impact on the detection statistic

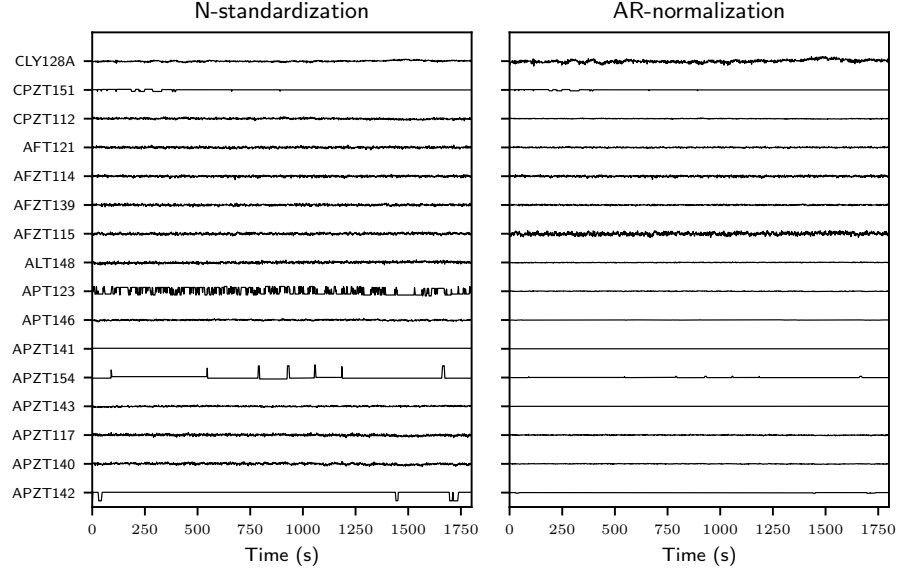


Figure 5.5: Process measurements in the test set Ω_{N0} normalized with N-standardization and AR-normalization. The scale is $[-20, 20]$ for each N-standardized measurement and $[-0.1, 0.1]$ for each AR-normalized measurement.

T^2 . This can trigger false detections and can introduce a bias in the contribution plots when it comes to fault isolation. Not monitoring the process measurements with small variations can lead to missed detections, but those process measurements do not need to be included in the PCA model and can be monitored by the Q statistic.

5.5.2 Fault detection

The fault detection is tested on a test set of data from normal process operation Ω_{N0} , plotted in Figure 5.5. The values of the T^2 statistic are plotted in Figure 5.6 for N-standardization and AR-normalization, and the values of the Q statistic are plotted in Figure 5.7 for N-standardization and AR-normalization. For each plot, two detection thresholds are represented, for $\alpha = 1\%$ and for $\alpha = 0.1\%$. Table 5.3 shows the false detection rates for each normalization approach with T^2 , with Q , and with T^2 OR Q where a plant vector is detected as faulty if it is faulty with T^2 or with Q . The false detection rate is expressed as the ratio of plant vectors in the test set Ω_{N0} detected as faulty compared to the total number of plant vectors in Ω_{N0} . The false detection rates are given for the detection thresholds with $\alpha = 1\%$ and for the detection thresholds with $\alpha = 0.1\%$. $T^2_{0.1\%}$ and $Q_{0.1\%}$ are retained for the rest

Table 5.3: False detection rates on the test set Ω_{N0} with N-standardization and AR-normalization.

α	T^2		Q		T^2 OR Q	
	1%	0.1%	1%	0.1%	1%	0.1%
N-standardization	0.058	0.007	0.111	0.012	0.118	0.019
AR-normalization	0.003	0.000	0.067	0.012	0.070	0.012

of the chapter as they provide a better robustness with regard to fault detection for both N-standardization and AR-normalization.

The false detection rates for T^2 with AR-normalization are much lower than the false detection rates for T^2 with N-standardization. For example, the T^2 statistic with N-standardization in Figure 5.6 presents a series of spikes that lead to false detections. Those spikes corresponds to small variations in APZT154 and APZT142 that had a small variability in the training set $\tilde{\Omega}_{N0}$. The T^2 statistic with AR-normalization is more robust as it does not depend as much on those process measurements.

Some of the variations in APZT154 and APZT142 are captured by the Q statistic with AR-normalization, but the impact is not as high as with N-standardization since the false detection rates for Q are still smaller with AR-normalization than with N-standardization (Table 5.3). Generally speaking, the false detection rates with T^2 OR Q indicate that the fault detection is more robust with AR-normalization than with N-standardization.

Figure 5.8 shows the T^2 and Q statistics for one occurrence of each fault with the corresponding detection thresholds $T^2_{0.1\%}$ and $Q_{0.1\%}$. The blue lines correspond to the scores with N-standardization, the red lines correspond to the scores with AR-normalization. All four faults are well detected with T^2 and Q for both N-standardization and AR-normalization, although AR-normalization offers a better detection performance than N-standardization, in particular for Fault 1. The deviation in the discharge pressure of P21 (APZT117) that occurs in Fault 1 is considered to be more abnormal compared to the alarm range than compared to the standard deviation in the training set, and T^2 with AR-normalization detects the fault before T^2 with N-standardization.

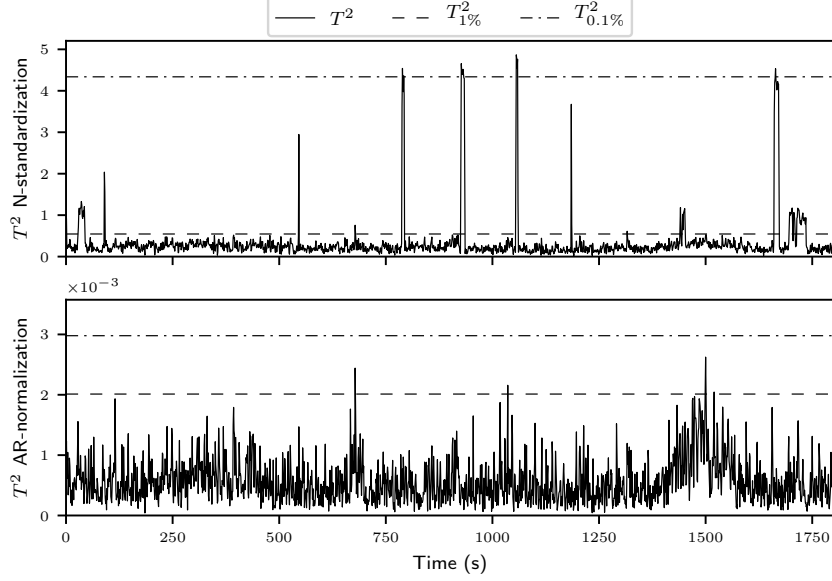


Figure 5.6: T^2 statistic for the test set Ω_{N0} with N-standardization (top) and AR-normalization (bottom).

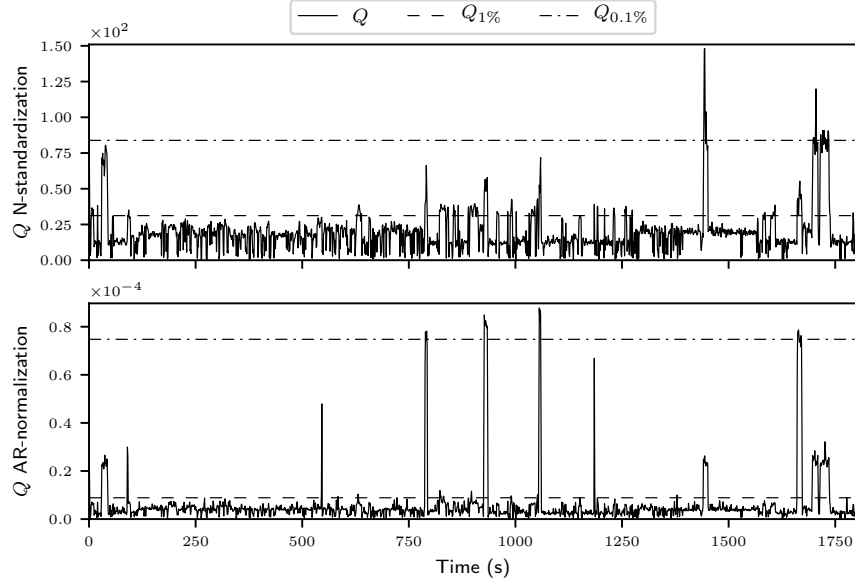


Figure 5.7: Q statistic for the test set Ω_{N0} with N-standardization (top) and AR-normalization (bottom).

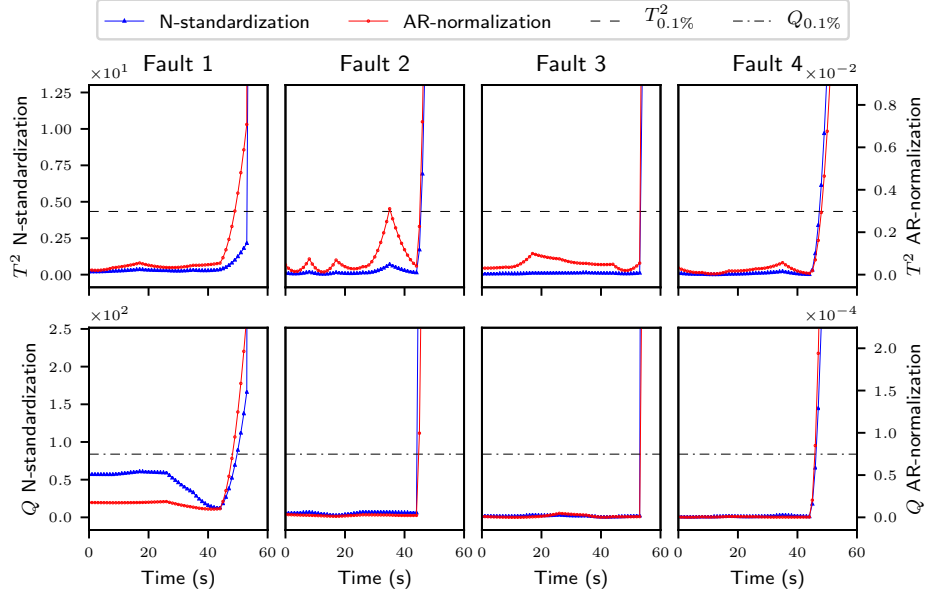


Figure 5.8: T^2 and Q statistics with N-standardization (blue) and with AR-normalization (red) for one occurrence of each type of fault in $\tilde{\Omega}_F$.

5.5.3 Fault isolation

Figure 5.9 shows the T^2 reconstruction-based contributions with N-standardization (in blue) and with AR-normalization (in red) for one occurrence of each type of fault, and Figure 5.10 shows the Q reconstruction-based contributions with N-standardization (in blue) and with AR-normalization (in red). The reconstruction-based contributions are plotted in the early stage of the fault, as soon as it is detected, to identify the first variables affected.

For the first three faults, the T^2 reconstruction-based contribution plots with N-standardization and with AR-normalization are comparable. APZT117 (discharge pressure of P21) and APZT140 (suction pressure of P22) are identified as the most contributing variables in Fault 1 since the pressure drop in the incoming water has more effect on P21 than on P11. The pressure drop in the incoming water also affects APZT117 and APZT140 in Fault 2 but the trip of P21 has an immediate effect on the discharge pressures of P22 (APZT142) and P12 (APZT143), whose T^2 contributions exceed the ones of APZT117 and APZT140. For Fault 3, N-standardization and AR-normalization both highlight the T^2 contribution of AFZT139, the water flow to the water injection tank 2. While the whole water system is affected by the drop in pressure and flow caused by the fuel change, the effect on AFZT139 has more

importance because the water flow to water injection tank 2 is the key controlled variable of the system. Any variation in this flow is considered as highly abnormal, both compared to statistical variations in the training set $\tilde{\Omega}_{N0}$ and compared to the alarm thresholds.

However, the T^2 contributions with N-standardization and with AR-normalization differ for Fault 4. APZT154 (suction pressure of P12) is identified as the most contributing variable to Fault 4 with N-standardization but has a minor T^2 contribution with AR-normalization. This is because APZT154 presents small variations in the training set $\tilde{\Omega}_{N0}$. The high contribution of this variable hides the real root cause of the fault. Fault 4 starts with a trip of P21 (due to high vibrations in the motor of the pump), so the process variables affiliated to this pump should be highlighted as root causes. The highest contribution to the T^2 statistic with AR-normalization for this fault is APZT117 which is the discharge pressure of P21. Therefore, Fault 4 is an example of how AR-normalization improves the fault isolation by highlighting in priority the process measurements with large variations compared to their alarm thresholds.

The Q contributions of Figure 5.10 confirm some of the variables isolated using the T^2 contributions (e.g. APZT117 and APZT140 for Fault 1 and Fault 4 and APZT142 for Fault 2) and highlight some variables that are not captured in T^2 such as the discharge flow of P21 (AFZT115) in Fault 1.

5.5.4 Fault identification

The fault identification is tested on a test set Ω_F of data containing new occurrences of faults 1 to 4. The fault occurrences of Ω_F are plotted in Figure 5.11. Each column corresponds to a fault occurrence in the test set, e.g. Fault 2-1 and Fault 2-2 correspond respectively to the first and second occurrences of Fault 2 in Ω_F . As for the plot of the training set $\tilde{\Omega}_F$, red lines indicate the AR-normalized process measurements and blue dotted lines indicate the F-standardized process measurements.

Figure 5.12 and Figure 5.13 show the classification results respectively with F-standardization and AR-normalization. For each figure, the detection is done using the T^2 statistic of the PCA with AR-normalization. The classification outcome is displayed as soon as the fault is detected. Each point indicates a new sample appearing every second. Black points indicate samples that are not detected as faulty, and the other colours indicate the classification outcome.

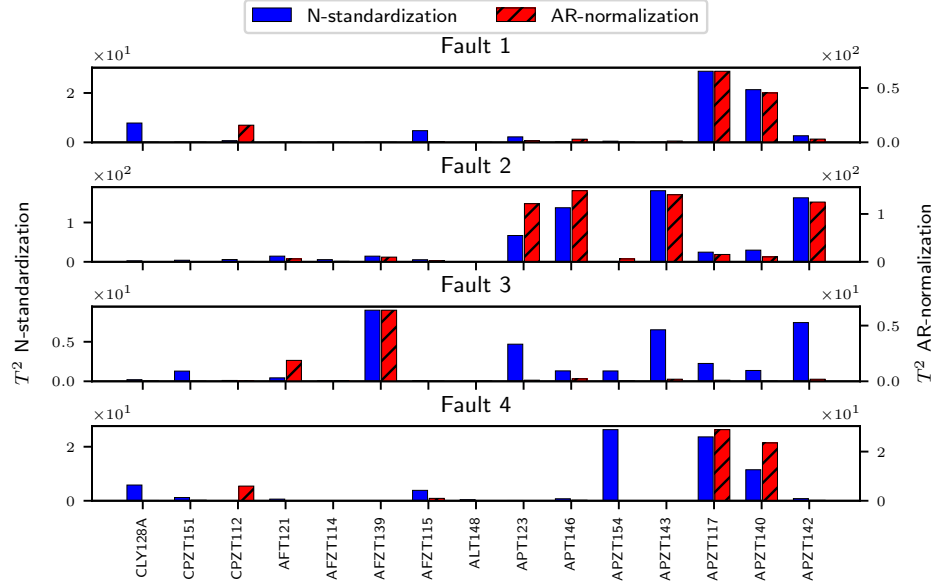


Figure 5.9: T^2 reconstruction-based contribution plots with N-standardization (blue) and with AR-normalization (red) for one occurrence of each type of fault in $\tilde{\Omega}_F$.

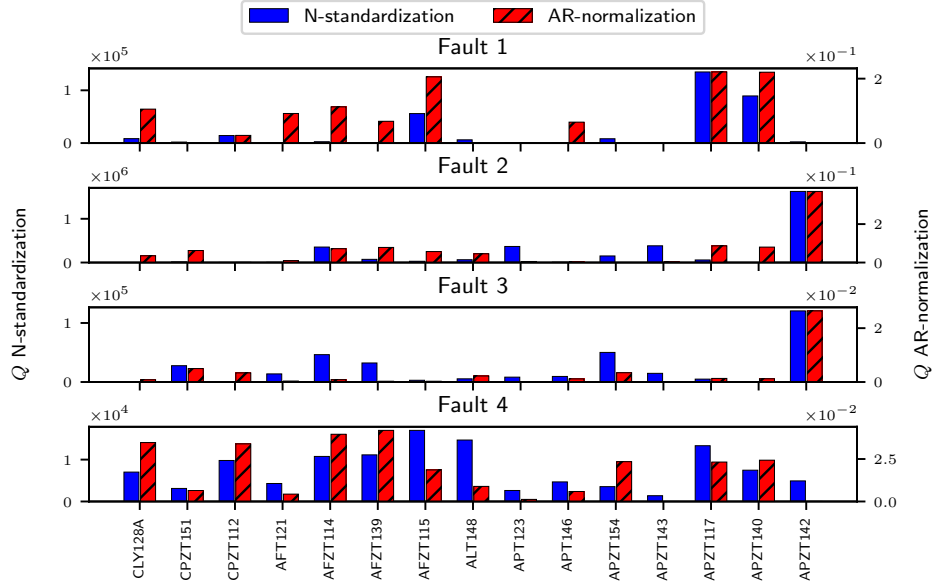


Figure 5.10: Q reconstruction-based contribution plots with N-standardization (blue) and with AR-normalization (red) for one occurrence of each type of fault in $\tilde{\Omega}_F$.

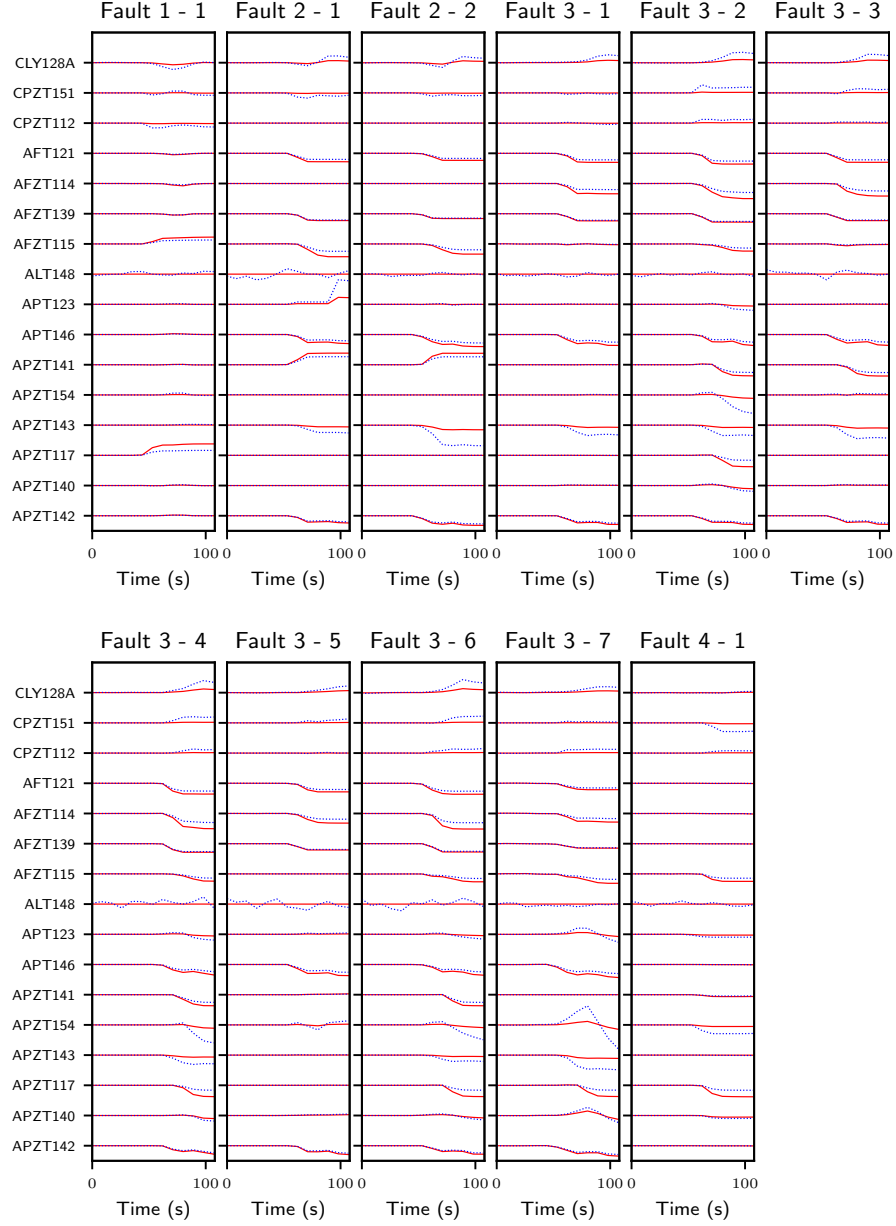


Figure 5.11: Process measurements in the test set Ω_F normalized with F-standardization (blue) and AR-normalization (red). The scale is $[-6, 6]$ for each F-standardized measurement and $[-2, 2]$ for each AR-normalized measurement.

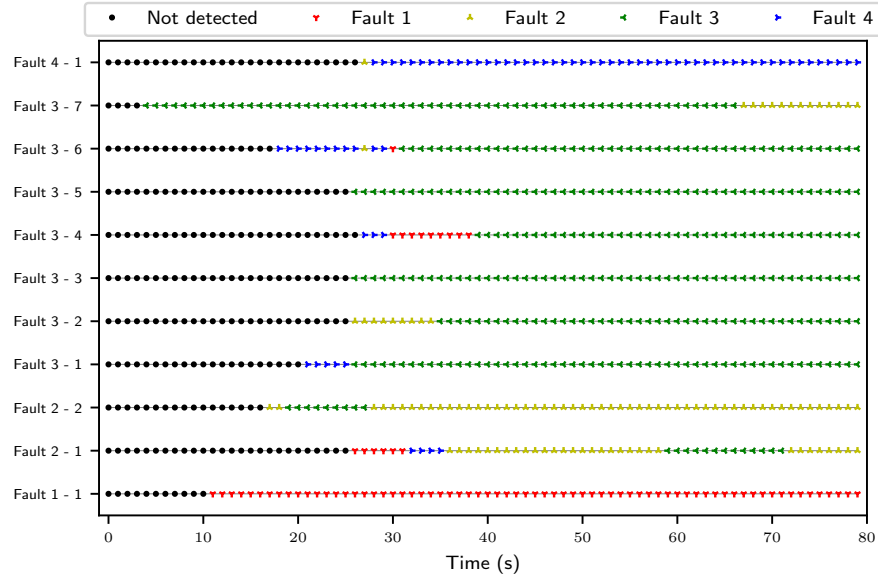


Figure 5.12: Classification outcomes with 1NN and F-standardization.

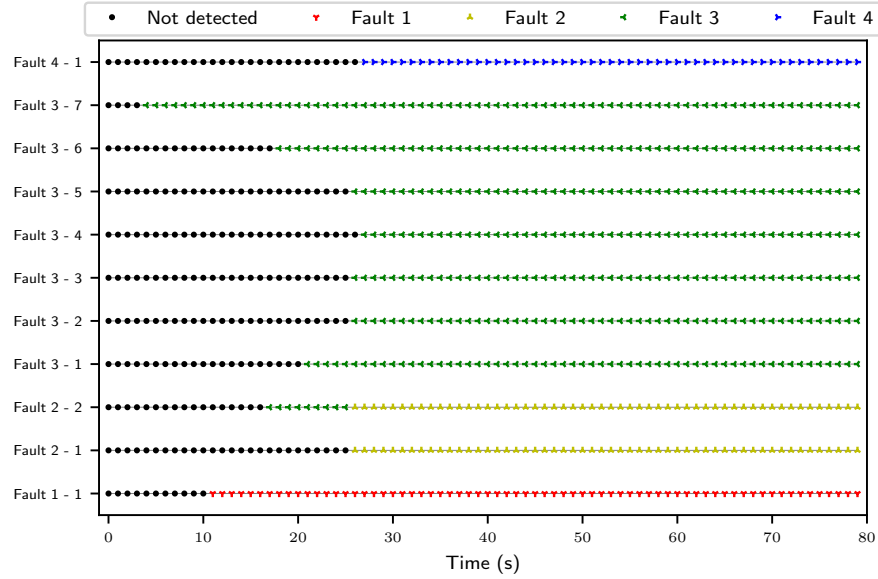


Figure 5.13: Classification outcomes with 1NN and AR-normalization.

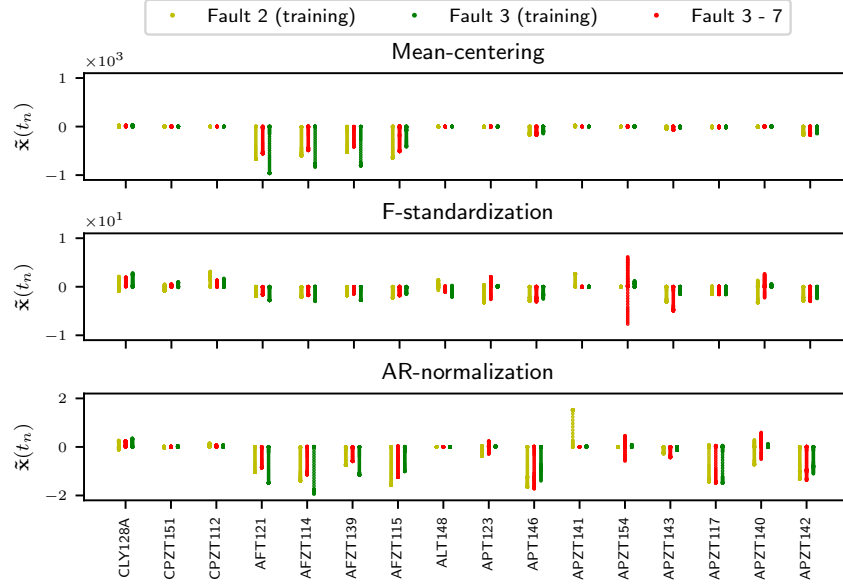


Figure 5.14: Normalized plant vectors $\tilde{\mathbf{x}}(t_n)$ for the training occurrence of Fault 2 (yellow) and Fault 3 (green), and for Fault 3-7 (red). Each point corresponds to the value of each process variable in $\tilde{\mathbf{x}}(t_n)$ at a certain time t_n during the development of the fault.

Classification with AR-normalization outperforms classification with F-standardization in both early stage classification and steady-state classification. For early stage classification, the classifier with AR-normalization correctly identifies faults 2-1, 2-2, 3-1, 3-2, 3-4, 3-6 and 4-1 before the classifier with F-standardization. For steady-state classification, the classifier with AR-normalization identifies all faults correctly in their steady-state while the classifier with F-standardization classifies Fault 2-1 as Fault 3 and Fault 3-7 as Fault 2.

Fault 3-7 illustrates the advantages of AR-normalizations over F-standardization. This occurrence of Fault 3 is classified as Fault 2 in steady-state with F-standardization, but is correctly classified with AR-normalization. Figure 5.14 gives more details about the classification for this fault. The figure shows the values taken by each process variable during the training occurrence of Fault 2 (in yellow) and during the training occurrence of Fault 3 (in green), as well as the values taken during Fault 3-7 (in red). The values are displayed for F-standardization and AR-normalization. The mean-centered values are also displayed for comparison.

The main discriminating process variable between Fault 2 and Fault 3 is APZT141, which is the discharge pressure of P11. In the case of Fault 2, APZT141 increases significantly

since the trip of P21 causes P11 to pump more water. In Fault 3, all the pumps of the PWRI section are affected by the fuel change and the flow and pressure in the water system go down. In this case, the discharge pressure of P11 stays constant or decreases.

With mean-centering or F-standardization, the variations of APZT141 are small compared to the variations of other process variables. AR-normalization shows that the variations of APZT141 are actually significant (in Fault 2) compared to its alarm thresholds, and the weight of the variations of other process variables is reduced. As a result, the behaviour of APZT141 has a major impact on the classification outcome with AR-normalization, and Fault 3-7 is correctly classified. With the other normalizations, Fault 3-7 is misclassified as Fault 2 since the amplitudes of the drop in the flows (AFT121, AFZT114, AFZT139, and AFZT115) are closer to the amplitude of the drops in Fault 2 than in Fault 3.

5.6 Summary of the chapter

Recent work has demonstrated the benefits of integrating alarm records in fault detection and diagnosis. This chapter affirms that alarm settings defined during the alarm design stage based on engineering and safety considerations can also be exploited to improve fault detection, fault isolation and fault identification based on process data. A new normalization of the process measurements based on their alarm thresholds is proposed. Online fault detection and diagnosis methods are dependent on the availability of process measurements, and the current normalization methods such as standardization are also impacted by the choice of the training set. AR-normalization is independent of the statistics of the training data and takes into account the acceptable range of variation of each process measurement. The oil and gas separation plant case study showed how traditional standardization approaches could lead to spurious results, and how AR-normalization improved the robustness of popular methods to noise in the measurements for both for fault detection and isolation and for fault identification.

Chapter 6

Fault detection and identification combining process measurements and statistical alarms

6.1 Description of the chapter

Fault identification assumes that the types of faults, or classes of faults, are well-defined and have been encountered in the past so that historical fault occurrences can be used as a reference. Classification-based methods are trained to recognize patterns in the process measurements based on historical fault occurrences of the same class, where *patterns* refer to regularities in the data or in the characteristics of the data. In industrial systems, patterns in the process measurements for a given class of fault are subject to variability from one fault occurrence to another [31]. This variability can be due to factors independent from the fault such as noise and external disturbances in the process measurements that are not related to the ongoing fault. The variability can also be due to the fault characteristics, for example with various fault amplitudes.

This chapter presents a fault detection and identification method designed to deal with noise and with variability in the patterns of the process measurements. The method uses a combination of process measurements and statistical alarms generated from statistical processing of the process measurements. Two levels of alarms are defined according to the position of the alarm threshold as defined in Sub-section 2.3.2: level-1 statistical alarms (low severity threshold) and level-2 statistical alarms (high severity threshold). The most infor-

mative variables with respect to the classification problem are selected using the minimal-Redundancy-Maximal-Relevance (mRMR) criterion [193]. The mRMR criterion uses level-2 alarms instead of the continuous measurements in order to reduce the impact of small variations in the process measurements on the variable selection and to limit the effect of noise and external disturbances in the classification. The classification method for fault detection and identification also relies on statistical alarms, offering a Bayesian fusion of the outcome of a classifier based on level-1 alarms and the outcome of a classifier based on process measurements. The Bayesian classifier on level-1 alarms offers a good robustness to noise during the development of the fault. The Bayesian classifier on process measurements is used to identify the fault during later stages. The material presented in this chapter has been published in the following articles:

- M. Lucke, X. Mei, A. Stief, M. Chioua, and N. F. Thornhill, “Variable selection for fault detection and identification based on mutual information of alarm series,” in *Proceedings of 12th DYCOPS Symposium, Florianópolis, Brazil, April 23 - 26, IFAC-PapersOnLine*, vol. 52, pp. 673–678, 2019
- M. Lucke, A. Stief, M. Chioua, J. R. Ottewill, and N. F. Thornhill, “Fault detection and identification combining process measurements and statistical alarms,” *Control Engineering Practice*, vol. 94, p. 104195, 2020

Section 6.2 provides preliminary information about variable selection and fault detection and identification in the process literature. Sub-section 6.3.1 presents the variable selection procedure and Sub-section 6.3.2 presents the fault detection and identification method with Bayesian fusion of process measurements and level-1 alarms. The benefits of integrating statistical alarms in fault detection and identification are illustrated on a simulation example in Section 6.4 and demonstrated on data from a multiphase flow facility in Sections 6.5 and 6.6.

6.2 Background

6.2.1 Variable selection with mRMR criterion for process measurements

Variable selection aims at selecting the X_m that should be integrated in the fault detection and diagnosis. Variable selection based on mutual information of the process measurements

proposed by [113] and discussed in Sub-section 3.3.3 has become a benchmark in the literature. However, the analytical formulations of the univariate and multivariate mutual information of [113] assume that the faults are stationary and that the process variables follow a Gaussian distribution during each fault. Faults are generally non-stationary in industrial systems and process variables evolve during the faults so that the Gaussian assumption does not apply. For this reason, a data-driven rather than analytical estimation of mutual information is preferred.

Multivariate data-driven estimation of mutual information is challenging because it assumes the estimation of joint probabilities that are heavy to compute. An alternative is to use pairwise comparisons taking into account a relevance criterion between each variable and the class, which assesses how the variables are related to the various classes of faults, and a redundancy criterion between each pair of variables that assesses how both variables are related. The authors of [194] benchmarked multiple combinations of relevance criterion and redundancy criterion with different classifiers. The combination of a redundancy criterion and relevance criterion based on mutual information was introduced in [193] as the minimal-Redundancy-Maximal-Relevance (mRMR) criterion. This sub-section presents the direct application of the mRMR criterion to process measurements.

The mutual information between two random variables quantifies the mutual dependence of the two variables [195]. In the context of variable selection, the simplest criterion is the univariate mutual information $I(X_m; C)$ between a continuous process variable X_m and the discrete class variable C . The higher the value of $I(X_m; C)$, the more relevant X_m is considered for the classification. It can be expressed as:

$$I(X_m; C) = \sum_{c \in C} \int_{X_m} P(x_m, c) \log \frac{P(x_m, c)}{P(x_m)P(c)} dx_m \quad (6.1)$$

where x_m and c represent the values that X_m and C can take. The probability distributions are computed using the extension of the nearest neighbours estimator between a continuous and a discrete variable [196].

The purpose of variable selection for classification is to find a set $\mathcal{X}_M = \{X_1, \dots, X_M\}$ of M variables X_m that have the largest dependency on the class C . In this chapter, the values c of C corresponds to the fault classes, including normal operation. The max-dependency criterion is defined as:

$$\max_{\mathcal{X}_M} d(\mathcal{X}_M, C), \quad d = I(\mathcal{X}_M; C) \quad (6.2)$$

where:

$$I(\mathcal{X}_M; C) = \sum_{c \in C} \int_{X_1} \dots \int_{X_M} P(x_1, \dots, x_M, c) \log \frac{P(x_1, \dots, x_M, c)}{P(x_1, \dots, x_M)P(c)} dx_1 \dots dx_M \quad (6.3)$$

Since the joint probability distributions $P(x_1, \dots, x_M, c)$ are difficult to estimate in practice, the max-dependency criterion is approximated using simplified criteria such as the mRMR criterion [193]. The max-relevance criterion [193] is an approximation of the dependency criterion in Eq. 6.2 considering the mutual information between each variable X_m and the class C :

$$\max_{\mathcal{X}_M} D(\mathcal{X}_M, C), D(\mathcal{X}_M, C) = \frac{1}{|\mathcal{X}_M|} \sum_{X_m \in \mathcal{X}_M} I(X_m; C) \quad (6.4)$$

The min-redundancy criterion [193] considers the redundancy in the information of selected variables in a pairwise manner:

$$\min_{\mathcal{X}_M} R(\mathcal{X}_M), R(\mathcal{X}_M) = \frac{1}{|\mathcal{X}_M|^2} \sum_{X_m \in \mathcal{X}_M} \sum_{X_{m'} \in \mathcal{X}_M} I(X_m; X_{m'}) \quad (6.5)$$

While extending the redundancy analysis to more variables using multivariate mutual information is possible, it requires more data for the estimation and increases the computational load. Since the number of fault occurrences available for training are limited, the pairwise approximation of [193] is preferred. The max-relevance criterion and the min-redundancy criterion are combined as the mRMR criterion:

$$\max_{\mathcal{X}_M} \Phi(D(\mathcal{X}_M, C), R(\mathcal{X}_M)), \Phi(D(\mathcal{X}_M, C), R(\mathcal{X}_M)) = D(\mathcal{X}_M, C) - R(\mathcal{X}_M) \quad (6.6)$$

$I(X_m; X_{m'})$ is computed using the nearest neighbours estimator explained in [197]. $I(X_m; C)$ is computed using the extension of the nearest neighbours estimator between a continuous and a discrete variable described in [196].

In practice, an incremental search is performed to find the near-optimal set of variables defined by $\Phi(D(\mathcal{X}_M, C), R(\mathcal{X}_M))$. Assuming we know the set of $M-1$ variables $\mathcal{X}_{M-1}$, the M th variable is selected from the set of remaining variables $\mathcal{X} - \mathcal{X}_{M-1}$ (where $\mathcal{X}$ indicates the set of all the variables X_m) as:

$$\max_{X_{m'} \in \mathcal{X} - \mathcal{X}_{M-1}} \left[I(X_{m'}; C) - \frac{1}{M-1} \sum_{X_m \in \mathcal{X}_{M-1}} I(X_{m'}; X_m) \right] \quad (6.7)$$

A *variable ranking* ordering the best variables for the classification can be derived from the sets $\mathcal{X}_m$. The variable in $\mathcal{X}_1$ is ranked as the best variable for the classification, the variable in $\mathcal{X}_2 - \mathcal{X}_1$ is ranked as the second best, and the variable in $\mathcal{X} - \mathcal{X}_{M_{\mathcal{X}}-1}$ is ranked last (where $M_{\mathcal{X}}$ indicates the total number of variables).

6.2.2 Motivation for the method

In process plant, the number of fault occurrences available for training a classifier is limited and process measurements can have a high variability from one fault occurrence to another. Therefore, classifiers in FDD methods should be robust, i.e. able to classify accurately new fault occurrences in spite of noise. This chapter suggests the use of statistical alarms as defined in Sub-section 2.3.2 to improve the robustness of the fault detection and identification to noise and variability in the process measurements.

Recent works highlighted the benefits of integrating alarms in fault detection and identification to improve the accuracy of the methods as well as their robustness to variability in the process measurements. The authors of [29] combined alarms from alarm records with process measurements using a two-stage Bayesian classifier to improve the accuracy of the classification by integrating a new source of information. The work in [31] (presented in Chapter 5) proposed a normalization of process measurements based on their alarm thresholds extracted from the alarm system to situate the variations of the measurements in the context of their safe operating range.

Statistical alarms are generated based on statistical processing of process measurements (Sub-section 2.3.2). Statistical alarms have been used as a substitute for process measurements to reduce the computational load of methods based on probability density estimation, in particular for root cause analysis with transfer entropy [57, 67, 70]. In the current work, statistical alarms are used as a substitute for process measurements in variable selection to limit the impact of the small variations in the estimation of the mutual information. Statistical alarms are also used in combination with process measurements for fault detection and identification using a two-stage Bayesian classification approach similar to the one in [29]. However, while [29] extracts only discrete features from the process measurements, the present chapter suggests a combination of discrete and continuous features. Statistical alarms are used as features for the detection and identification during the early stage of the fault since they are designed to indicate deviations from the normal operation while limiting

the number of transitions due to chattering. The values of the process measurements are used to provide a more accurate fault identification during the late stage of the fault.

As highlighted in Sub-section 3.3.6, the Bayesian framework offers a simple approach to integrate continuous and discrete features in the classification. A discrete Bayesian classifier gives probabilities of the classification outcome based on the values of the alarms after estimation of the discrete distributions on the training set. A continuous Bayesian classifier gives probabilities of the classification outcome based on the values of the process measurements after estimation of the probability density functions using kernel density estimation on the training set. The final decision is taken based on a fusion of the outcomes of both classifiers.

6.3 Method development

6.3.1 Variable selection with mRMR on level-2 alarms

In order to reduce the impact of small variations on the variable selection, and to limit the effects of noise and external disturbances on the classification, the mRMR criterion [193] is applied to level-2 alarms $\bar{\bar{A}}_m$ (i.e. the statistical alarms with high severity threshold $\bar{\bar{k}}\sigma_m$ defined in Sub-section 2.3.2) rather than to the original process measurements. The accuracy of the classifier is defined as the proportion of observations $\mathbf{x}(t_n)$ of the plant vector $\mathbf{X}$ that are correctly classified. The mutual information $I(\bar{\bar{A}}_m; C)$ based on $\bar{\bar{A}}_m$ can be computed as a discrete sum:

$$I(\bar{\bar{A}}_m; C) = \sum_{c \in C} \sum_{\bar{\bar{a}}_m \in \bar{\bar{A}}_m} P(\bar{\bar{a}}_m, c) \log \frac{P(\bar{\bar{a}}_m, c)}{P(\bar{\bar{a}}_m)P(c)} \quad (6.8)$$

The incremental search of Sub-section 6.2.1 for the mRMR criterion based on the level-2 alarms $\bar{\bar{A}}_m$ becomes:

$$\max_{\bar{\bar{A}}_{m'} \in \bar{\bar{A}} - \bar{\bar{A}}_{M-1}} \left[I(\bar{\bar{A}}_{m'}; C) - \frac{1}{M-1} \sum_{\bar{\bar{A}}_m \in \bar{\bar{A}}_{M-1}} I(\bar{\bar{A}}_{m'}; \bar{\bar{A}}_m) \right] \quad (6.9)$$

where $\bar{\bar{A}}_{M-1}$ indicates the set of the best $M-1$ variables $\bar{\bar{A}}_m$ and $\bar{\bar{A}}$ the set of all variables $\bar{\bar{A}}_m$. Both $I(\bar{\bar{A}}_{m'}; C)$ and $I(\bar{\bar{A}}_{m'}; \bar{\bar{A}}_m)$ are computed as discrete sums. As for variable selection with mRMR on process measurements, a variable ranking is derived from the sets $\bar{\bar{A}}_m$.

6.3.2 Fault detection and identification with Bayesian fusion of process measurements and level-1 alarms

Bayesian classification with level-1 alarms

The plant vector of level-1 statistical alarms $\bar{\mathbf{A}}$ is defined as the vector of the level-1 alarms $\bar{A}_m$ (i.e. the statistical alarms with low severity threshold $\bar{k}\sigma_m$ defined in Subsection 2.3.2) associated with the M selected variables $X_m \in \mathcal{X}_M$, taking value $\bar{\mathbf{a}}(t_n) = [\bar{a}_1(t_n), \bar{a}_2(t_n), \dots, \bar{a}_M(t_n)]$ at time t_n . The probability that an observation $\bar{\mathbf{a}}(t_n)$ is associated with class c is computed as:

$$P(C = c | \bar{\mathbf{A}} = \bar{\mathbf{a}}(t_n)) = \frac{P(\bar{\mathbf{a}}(t_n) | c)}{P(\bar{\mathbf{a}}(t_n))} P(c) = \frac{P(\bar{\mathbf{a}}(t_n) | c)}{\sum_{c' \in \mathcal{C}} P(\bar{\mathbf{a}}(t_n) | c')} P(c) \quad (6.10)$$

where $P(c)$ is the prior probability of class c , $P(c | \bar{\mathbf{a}}(t_n))$ is the posterior probability of class c , $P(\bar{\mathbf{a}}(t_n) | c)$ is the likelihood, and $P(\bar{\mathbf{a}}(t_n))$ is the probability of observing the observation $\bar{\mathbf{a}}(t_n)$. Since all of the values are discrete, the probabilities are computed as discrete sums on the training data.

Bayesian classification with process measurements

The plant vector $\mathbf{X}$ is defined as the vector of the M selected variables $X_m \in \mathcal{X}_M$, taking value $\mathbf{x}(t_n) = [x_1(t_n), x_2(t_n), \dots, x_M(t_n)]$ at time t_n . The probability that a given observation $\mathbf{x}(t_n)$ is associated with class c can be computed using Bayes' rule as:

$$P(C = c | \mathbf{X} = \mathbf{x}(t_n)) = \frac{P(\mathbf{x}(t_n) | c)}{P(\mathbf{x}(t_n))} P(c) \quad (6.11)$$

where $P(c)$ is the prior probability of class c , $P(c | \mathbf{x}(t_n))$ is the posterior probability of class c (i.e. once the value $\mathbf{x}(t_n)$ at time t_n has been observed), $P(\mathbf{x}(t_n) | c)$ is the likelihood (i.e. the probability of the value $\mathbf{x}(t_n)$ assuming class c), and $P(\mathbf{x}(t_n))$ is the probability of observing the value $\mathbf{x}(t_n)$.

The probability density function $f(\mathbf{x})$ of $\mathbf{X}$ is estimated using kernel density estimation [198]. The estimated probability density function $\hat{f}(\mathbf{x}, \mathbf{H})$ is a summation of kernel functions centered around the values $\mathbf{x}(t_{\tilde{n}})$ of the plant vectors used for training ($\tilde{n} = 1 \dots \check{N}$, $\check{N}$ being

the number of values used for training):

$$\hat{f}(\mathbf{x}, \mathbf{H}) = \frac{1}{\tilde{N}} \sum_{\tilde{n}=1}^{\tilde{N}} |\mathbf{H}|^{-\frac{1}{2}} \mathcal{K}' \left(\mathbf{H}^{-\frac{1}{2}} (\mathbf{x} - \mathbf{x}(t_{\tilde{n}})) \right) \quad (6.12)$$

$\mathbf{H}$ is a symmetric positive definite matrix called the bandwidth matrix, $\mathcal{K}'$ is a M -variate kernel function such that $\int \mathcal{K}'(\mathbf{x}) d\mathbf{x} = 1$. A popular choice for the kernel function is the M -variate normal density function:

$$\mathcal{K}'(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{M}{2}}} \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{x}\right) \quad (6.13)$$

which gives the kernel density estimate of $f(\mathbf{x})$:

$$\hat{f}(\mathbf{x}, \mathbf{H}) = \frac{1}{\tilde{N} \sqrt{(2\pi)^M |\mathbf{H}|}} \sum_{\tilde{n}=1}^{\tilde{N}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{x}(t_{\tilde{n}}))^T \mathbf{H}^{-1} (\mathbf{x} - \mathbf{x}(t_{\tilde{n}}))\right) \quad (6.14)$$

The ratio $\frac{P(\mathbf{x}(t_n)|c)}{P(\mathbf{x}(t_n))}$ of Eq. 6.11 can thus be expressed with the kernel estimates of the probability densities $f(\mathbf{x}|c)$ and $f(\mathbf{x})$ ¹:

$$P(C = c | \mathbf{X} = \mathbf{x}(t_n)) = \frac{\hat{f}(\mathbf{x}(t_n)|c)}{\hat{f}(\mathbf{x}(t_n))} P(c) = \frac{\hat{f}(\mathbf{x}(t_n)|c)}{\sum_{c' \in C} \hat{f}(\mathbf{x}(t_n)|c')} P(c) \quad (6.15)$$

Each probability density function $f(\mathbf{x}|c)$ corresponding to a class c is estimated using the bandwidth matrix $\mathbf{H}(c)$ defined as the diagonal matrix of $h(c)\sigma_m$:

$$\mathbf{H}(c) = h(c) \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_M \end{bmatrix} \quad (6.16)$$

where $h(c)$ is a positive constant tuned for each class c and σ_m is the standard deviation during normal operation for the variable X_m . The choice of $h(c)$ in the bandwidth matrix is discussed in Sub-section 6.6.3.

¹As explained in [199], Eq. 6.15 is not strictly correct. The probability that a continuous-valued random variable exactly equals any value is zero. Instead we speak about the variable lying within an interval of size Δ , and for a very small constant Δ , $P(\mathbf{x}(t_n)) \approx f(\mathbf{x}(t_n))\Delta$. The Δ factors cancel out in the numerator and denominator of Eq. 6.11.

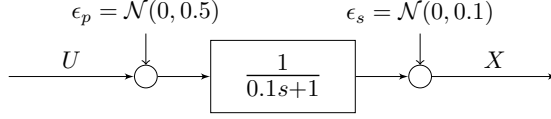


Figure 6.1: System of the simulation example.

Bayesian fusion

The final classification is taken based on the decision level fusion of the discrete and the continuous classifiers. The estimated class $\hat{c}_f(t_n)$ for a given observation $\mathbf{x}(t_n)$ of the plant vector $\mathbf{X}$ and a given observation $\bar{\mathbf{a}}(t_n)$ of the corresponding plant vector of level-1 statistical alarms $\bar{\mathbf{A}}$ is computed merging the posterior probabilities $P(C = c | \bar{\mathbf{A}} = \bar{\mathbf{a}}(t_n))$ of the discrete classifier and the posterior probabilities $P(C = c | \mathbf{X} = \mathbf{x}(t_n))$ of the continuous classifier:

$$\begin{aligned} \hat{c}_f(t_n) &= \operatorname{argmax}_{c \in C} \left(P(c | \mathbf{x}(t_n)) P(c | \bar{\mathbf{a}}(t_n)) \right) \\ &= \operatorname{argmax}_{c \in C} \left(\frac{\hat{f}(\mathbf{x}(t_n) | c)}{\sum_{c' \in C} \hat{f}(\mathbf{x}(t_n) | c')} P(c) \frac{P(\bar{\mathbf{a}}(t_n) | c)}{\sum_{c' \in C} P(\bar{\mathbf{a}}(t_n) | c')} P(c) \right) \end{aligned} \quad (6.17)$$

In this case, the prior probability $P(c)$ is chosen as uniform (i.e. all fault classes c have the same prior probabilities $P(c)$) so the expression can be simplified:

$$\hat{c}_f(t_n) = \operatorname{argmax}_{c \in C} \left(\frac{\hat{f}(\mathbf{x}(t_n) | c)}{\sum_{c' \in C} \hat{f}(\mathbf{x}(t_n) | c')} \frac{P(\bar{\mathbf{a}}(t_n) | c)}{\sum_{c' \in C} P(\bar{\mathbf{a}}(t_n) | c')} \right) \quad (6.18)$$

6.4 Tests of the method in simulation

6.4.1 Simulation example

The benefits of integrating level-1 alarms in the classification are demonstrated in a simple case study. The system is a first order systems with input U and output X , described in Figure 6.1. Process noise ϵ_p and sensor noise ϵ_s are added according to Figure 6.1. Level-1 alarms $\bar{A}$ are generated from X according to Eq. 2.1 with $\bar{k} = 3$. T_{on} is set to 5 for the on-delay timer and T_{off} is set to 10 for the off-delay timer.

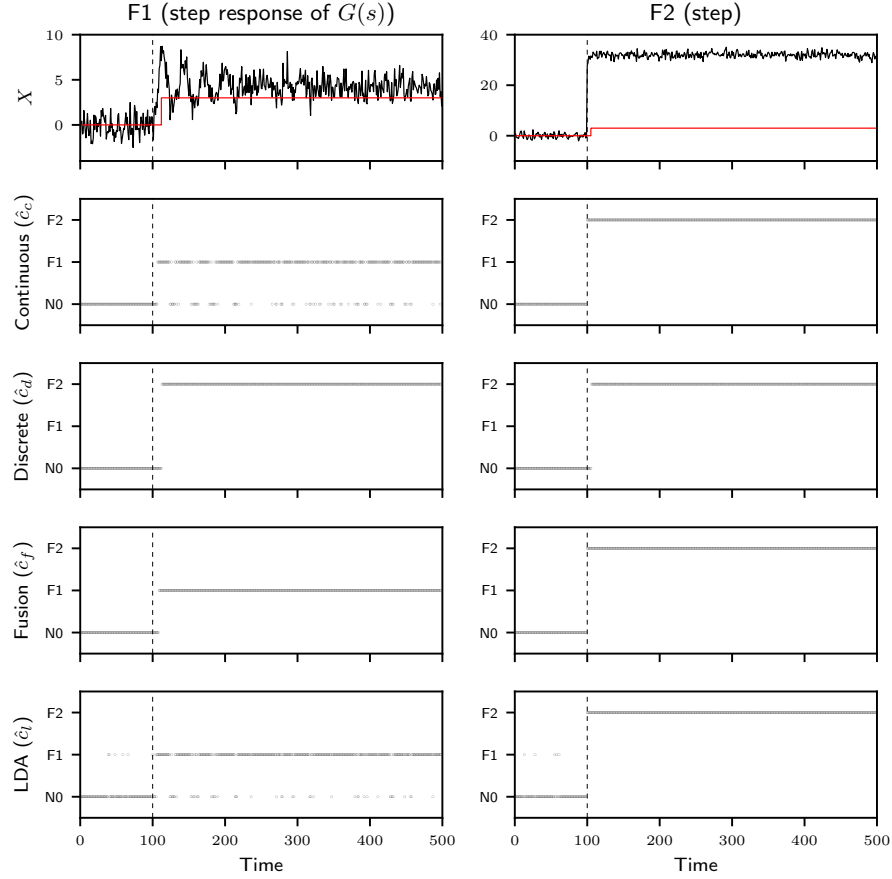


Figure 6.2: Standardized measurements, level-1 alarms (in red), and classification outcomes with continuous Bayesian classifier, discrete Bayesian classifier, fusion of both Bayesian classifiers, and LDA on the simulation example. The measurement X is standardized based on the standard deviation during normal operation, and the amplitude of the level-1 alarm indicates the level-1 alarm threshold ($\bar{k} = 3$). The vertical dotted lines indicate the beginning of the faults.

Two types of faults are introduced as input U of the system after 100 samples of normal operation and the fault detection and identification is done based on the output X of the system. Two occurrences of each type of fault with various amplitudes are generated (one for training the classifier and one for testing) and inserted as input U of the system. The two faults are:

- Fault 1 (F1): U is the unit step response of a second order system $G(s) = \frac{1}{20s^2 + 0.5s + 1}$ followed by a static gain. The training occurrence is generated with a static gain of 3, and the test occurrence is generated with a static gain of 2.
- Fault 2 (F2): U is a step of amplitude 20 for the training occurrence and amplitude 15 for the test occurrence.
- Normal operation (N0): all samples before the introduction of the faults constitute the data from normal operation.

6.4.2 Results

Figure 6.2 summarizes the results. The plots in the first row shows the outputs X during the test occurrence of each fault and the corresponding alarms $\bar{A}$ in red. The vertical line indicates the beginning of the fault after 100 samples. The following plots show respectively the classification outcome $\hat{c}_c$ with the continuous Bayesian classifier, the classification outcome $\hat{c}_d$ with the discrete Bayesian classifier, and the classification outcome $\hat{c}_f$ with the fusion of both classifiers. The plots in the last row show the classification outcome $\hat{c}_l$ of a Linear Discriminant Analysis (LDA) classifier on the process measurements for comparison. LDA is a traditional method for fault detection and identification on process measurements [133] but is not adapted to deal with discrete features, like most fault detection and identification methods in the literature. The benefits of integrating statistical alarms in fault detection and identification are demonstrated through a comparison with a traditional fault detection and identification method on process measurements only. The accuracies of the classifiers are listed in Table 6.1.

6.4.3 Comparison of the methods

The classification outcome of the continuous Bayesian classifier in Figure 6.2 shows that the continuous classifier cannot distinguish clearly between normal operation N0 and F1 due to the noisy behaviour of X during fault F1. When it comes to fault F1, the discrete classifier

Table 6.1: Classification accuracies on the simulation example.

	Accuracy
Continuous	0.934
Discrete	0.594
Fusion	0.991
LDA	0.920

based on alarm $\bar{A}$ (in red) offers a better separation between normal operation and the fault than the continuous classifier based on X . The on-delay and off-delay timers limit the number of transitions in the state of the alarm and $\bar{A}$ shows a clear transition between N0 and F1 once the threshold has been crossed for T_{on} consecutive samples. Thus, the discrete classifier offers a better distinction between normal operation and the faults, although it is not able to distinguish F1 from F2 since it classifies both faults as F2.

The fusion of both classifiers offers both a clear distinction between normal operation and the faults, and between F1 and F2. The delay in the alarm activation introduced by the on-delay timer has a limited impact on the detection of F1 and F2 with the fusion of both classifiers: the jump in value of X during F1 and F2 gives a high fault probability in the continuous classifier, and the outcome of the fusion of both classifiers indicates a fault even before the alarm activates.

As a comparison, the outcome of the LDA classifier in Figure 6.2 presents the same characteristics as the outcome of the continuous Bayesian classifier. The LDA classifier is unable to distinguish clearly F1 from normal operation.

6.5 Application to industrial case study

6.5.1 The multiphase flow facility

The case study is the multiphase flow facility described in Sub-section 2.5.2. The system is operated at two different points and data from normal operation are gathered for each operating point. Operating point A corresponds to an air flow rate of $120 \text{ Sm}^3 \text{ h}^{-1}$ and a water flow rate of 0.1 kg s^{-1} . Operating point B corresponds to an air flow rate of $150 \text{ Sm}^3 \text{ h}^{-1}$ and a water flow rate of 0.5 kg s^{-1} . Three types of fault were induced successively

at each operating point. The data are separated into four classes, the data from the three faults and the data from normal operation:

- Air blockage (F1): valve V11 was gradually closed to simulate a developing blockage in the input airline.
- Air leakage (F2): valve V10 was gradually opened so that the air is partially leaked out to the atmosphere.
- Diverted flow (F3): bypass valve U39 was gradually opened so that the mixed flow is partially led straight to the riser and partially led into the horizontal pipeline before joining the riser. Fault F3 is an incipient fault, and thus more difficult to detect than faults F1 and F2.
- Normal operation (N0): all samples before the introduction of the faults constitute the data from normal operation.

The analysis focuses on the 17 process variables listed in Table 2.4. The variables are mean-centered at each operating point. The mean-centering of X_m is done offline using the mean in normal operation for each operating point. Fault identification for faults developing during transitions between operating points is considered as a separate research question that requires a classifier able to deal with highly non-stationary dynamics. Figure 6.3 shows the standardized measurements in operating point A, and Figure 6.4 shows the standardized measurements in operating point B.

The level-1 statistical alarms $\bar{A}_m$ and the level-2 statistical alarms $\bar{\bar{A}}_m$ for operating point A are generated using σ_m in operating point A, and $\bar{A}_m$ and $\bar{\bar{A}}_m$ for operating point B are generated using σ_m in operating point B. According to Sub-section 2.3.2, $\bar{k}$ is set to 3 for the generation of level-1 alarms and $\bar{\bar{k}}$ is set to 5 for the generation of level-2 alarms. T_{on} is set to 30 for the on-delay timer and T_{off} is set to 60 for the off-delay timer.

6.5.2 Case study

Variable selection and training of the classifier are performed on the data from one operating point and tested on the data from the other operating point. Two scenarios are presented:

- Scenario AB: variable selection and training on data from operating point A and testing on data from operating point B. The training set in scenario AB includes normal operation data from operating point A and one occurrence of each fault in operating point A, pictured in Figure 6.3. The test set in scenario AB includes normal

operation data from operating point B and one occurrence of each fault in operating point B, pictured in Figure 6.4.

- Scenario BA: variable selection and training on operating point B and testing on operating point A. The training set in scenario BA includes normal operation data from operating point B and one occurrence of each fault in operating point B, pictured in Figure 6.4. The test set in scenario BA includes normal operation data from operating point A and one occurrence of each fault in operating point A, pictured in Figure 6.3.

A variable ranking is computed using the mRMR criterion on the level-2 alarms corresponding to the process measurements of the training set in each scenario. As a comparison, another variable ranking is computed using the mRMR criterion on process measurements (Sub-section 6.2.1).

All the sets $\mathcal{X}_M$ corresponding to the M best ranked variables ($M = 1 \dots M_{\mathcal{X}}$ with $M_{\mathcal{X}}$ the total number of variables) are assessed for each variable ranking. The classification accuracy with the Bayesian classifier of Sub-section 6.3.2 applied to the variables in $\mathcal{X}_M$ is computed for each $\mathcal{X}_M$. The classification accuracy, which indicates the proportion of plant vectors that are correctly classified, is computed on the training dataset and on the test dataset:

- *Classification accuracy on the training dataset.* The accuracy of the classifier is computed on the training dataset using a three-fold cross-validation strategy. The training dataset is divided in three parts that are used successively as test set while the two remaining parts are used for training, and the final accuracy represents the average accuracy of the three tests.
- *Classification accuracy on the test dataset.* The classification accuracy is computed on the whole test set. The classification accuracy on the test dataset shows the robustness of each variable ranking (i.e. the relevance of the variable ranking for classification of new fault occurrences in spite of noise) since it is evaluated on another occurrence of each fault and new data from normal operation.

As for the simulation example, the Bayesian classifier is compared to a traditional discriminant analysis classifier for fault detection and identification, in this case a Quadratic Discriminant Analysis (QDA) classifier. QDA is the quadratic extension of linear discriminant analysis used in Section 6.4, which could not be used on the multiphase flow facility because the faults are not linearly separable in this case. As in Section 6.4, QDA is applied on the process measurements only, since it is not adapted to discrete features.

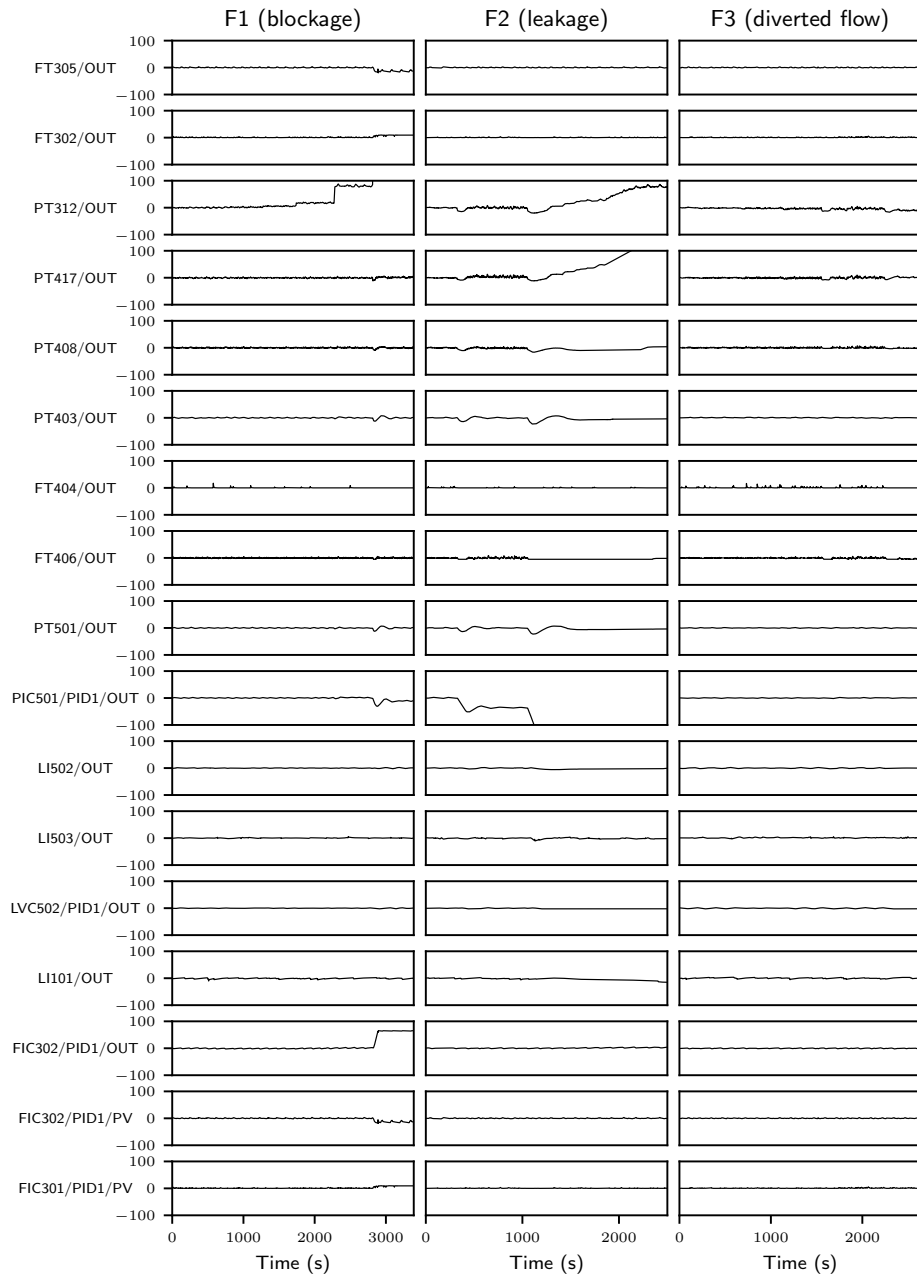


Figure 6.3: Standardized measurements captured during each fault introduced in operating point A.

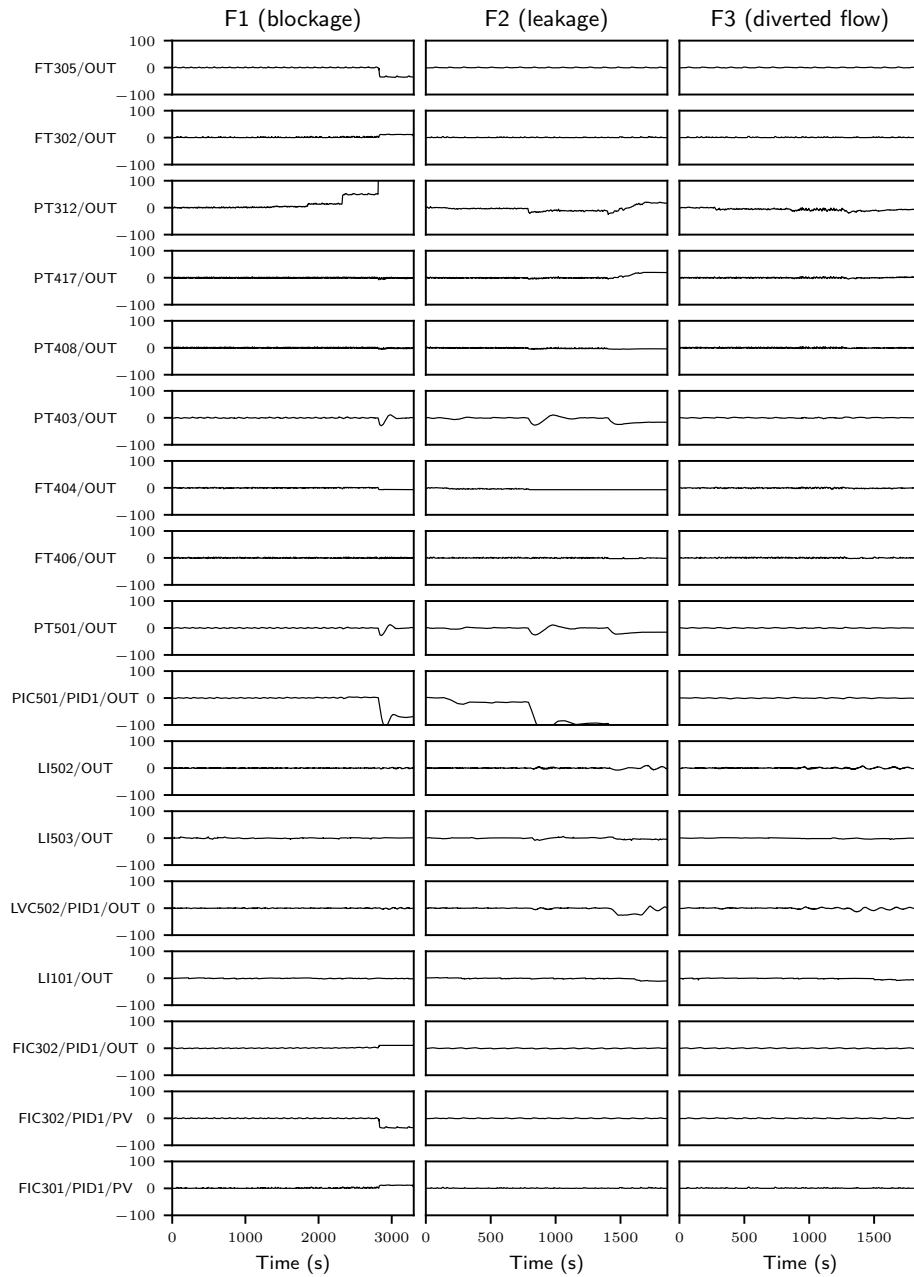


Figure 6.4: Standardized measurements captured during each fault introduced in operating point B.

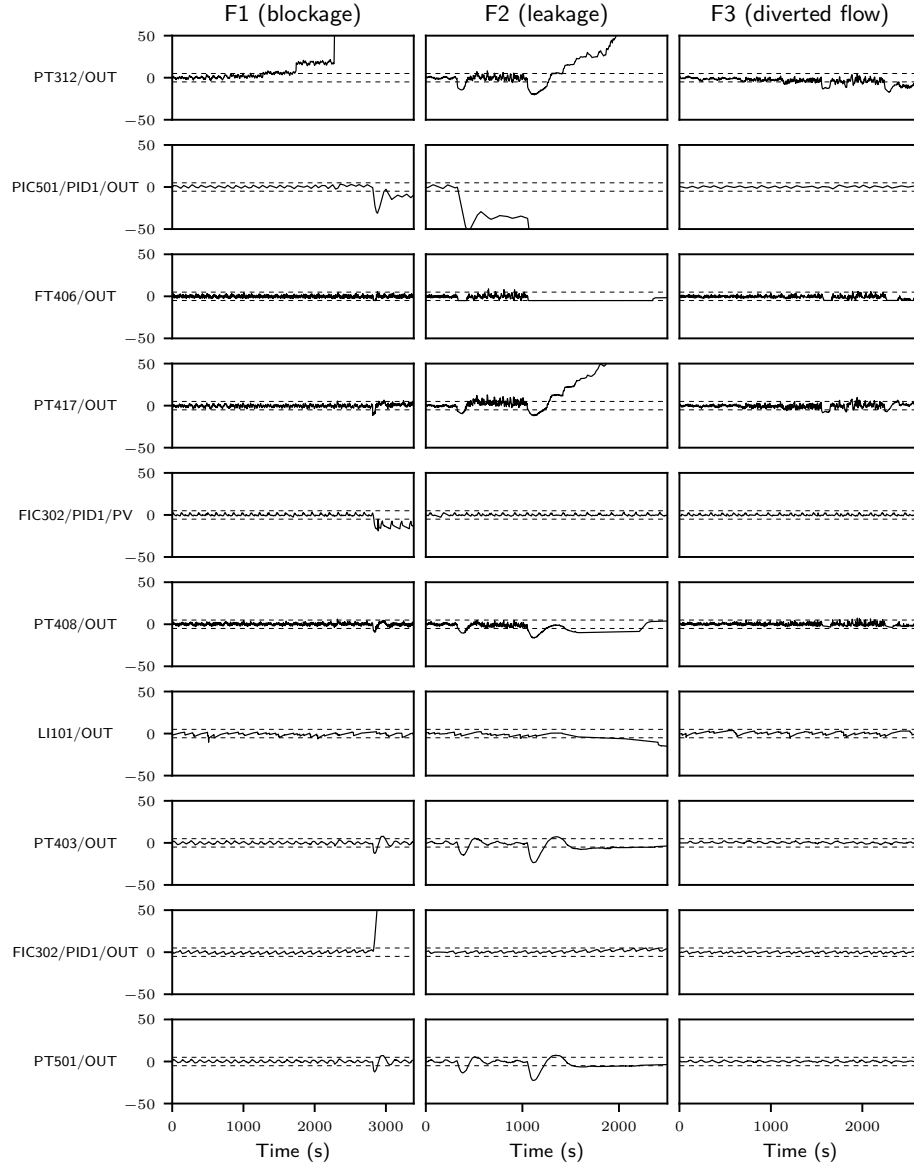


Figure 6.5: Standardized process measurements of the first ten ranked variables with mRMR on level-2 alarms in the training set for scenario AB. The variables are ranked from top to bottom, with PT312/OUT being ranked as the most relevant variable. Each column corresponds to a fault occurrence in the training set. The dotted lines indicate the level-2 alarm thresholds ($\bar{k} = 5$).

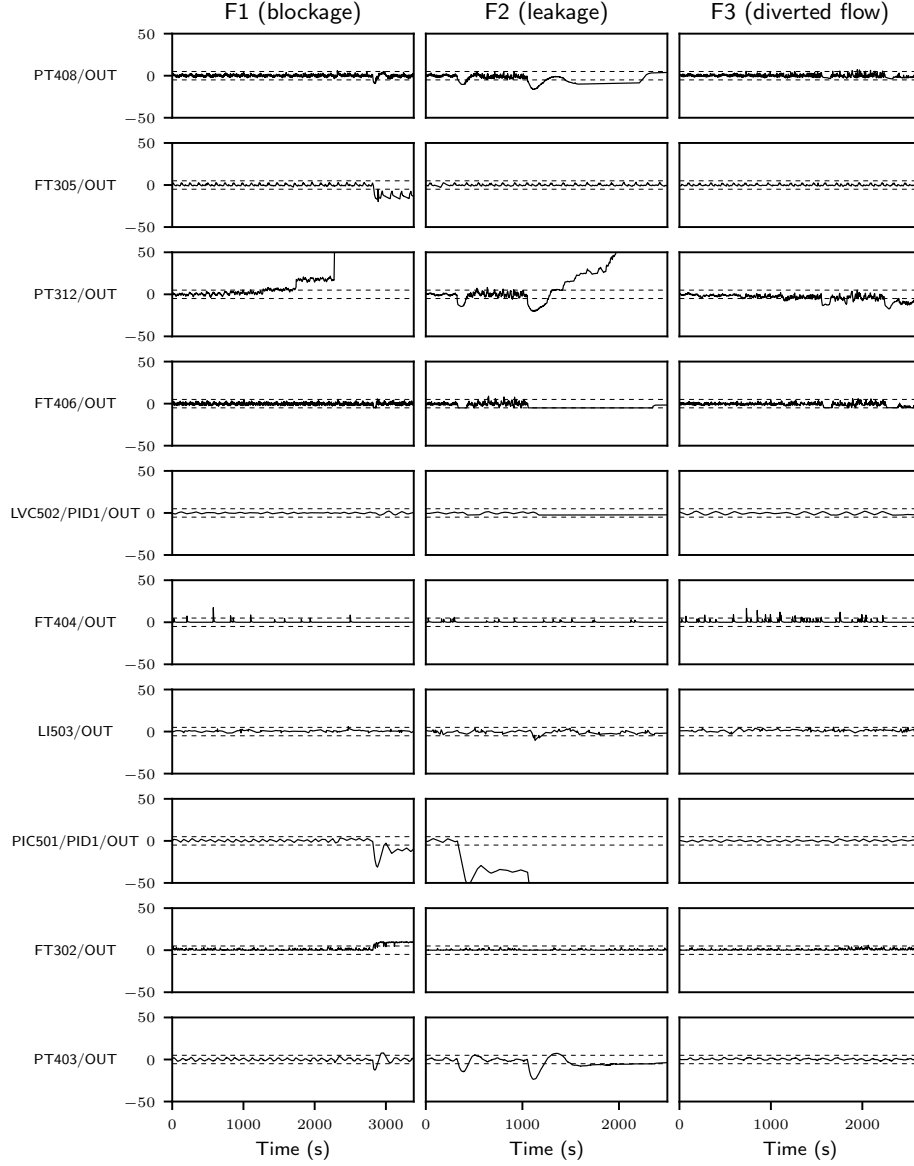


Figure 6.6: Standardized process measurements of the first ten ranked variables with mRMR on process measurements in the training set for scenario AB. The variables are ranked from top to bottom, with PT408/OUT being ranked as the most relevant variable. Each column corresponds to a fault occurrence in the training set. The dotted lines indicate the level-2 alarm thresholds ($\bar{k} = 5$).

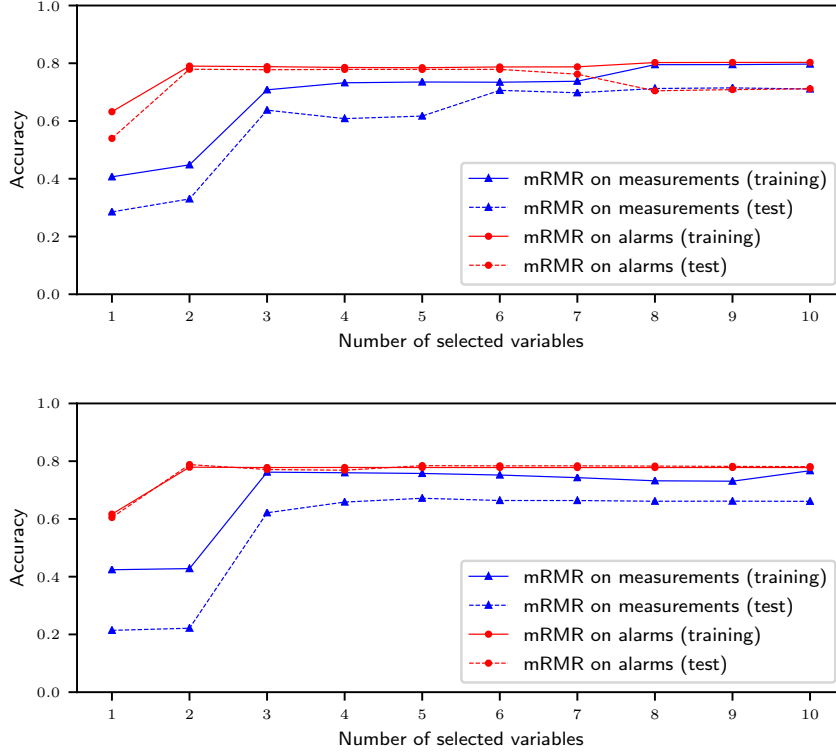


Figure 6.7: Classification accuracy (as defined in Sub-section 6.5.2) on training and test data for scenario AB (top) and scenario BA (bottom) for the first 10 variables with mRMR applied to process measurements (blue) and level-2 alarms (red).

6.6 Results of the case study

6.6.1 Variable selection

This sub-section illustrates how variable selection on level-2 alarms $\bar{A}_m$ prioritizes variables with large variations compared to variable selection on the process measurements X_m , and how it directly impacts the robustness of the classification results with regard to variability in the operating conditions.

Variable ranking

Figure 6.5 shows the standardized process measurements of the first ten ranked variables with mRMR on level-2 alarms in the training set for scenario AB. For comparison, Figure 6.6

shows the standardized process measurements of the first ten ranked variables with mRMR on process measurements in the training set for the same scenario. The dotted lines indicate the level-2 alarm thresholds with $\bar{k} = 5$, and each column depicts one of the faults. The ranked variables with mRMR on level-2 alarms and with mRMR on process measurements for scenario BA are available respectively in Figure A.1 and Figure A.2 in Appendix A.

The first two ranked variables in Figure 6.5 show that mRMR on the level-2 alarms prioritizes variables with large variations beyond the level-2 alarm thresholds. By contrast, mRMR on the continuous measurements selects variables independently from the alarm thresholds, and the top two ranked variables in Figure 6.5 present variations that are mostly within the alarm thresholds, i.e. variations that stay close to the normal range. Therefore, noise and external disturbances have a stronger effect on the patterns of the measurements of the Figure 6.6 than on the patterns of the measurements of the Figure 6.5.

Classification accuracies of the sets of variables

Figure 6.7 shows the classification accuracies as defined in Sub-section 6.5.2 for scenario AB (top plot) and scenario BA (bottom plot). The solid lines indicate the classification accuracy on the training set, and the dotted lines indicate the classification accuracy on the test set. The classification accuracies are computed for the various sets $\mathcal{X}_M$ of the M best ranked variables (with M from 1 to 10) that are summarized in Table A.1 for scenario AB and Table A.2 for scenario BA, both available in Appendix A. The red lines use the variable ranking with mRMR on level-2 alarms corresponding to the ranking pictured in Figure 6.5 for scenario AB and Figure A.1 for scenario BA. The blue lines use the variable ranking with mRMR on process measurements corresponding to the ranking pictured in Figure 6.6 for scenario AB and Figure A.2 for scenario BA.

Figure 6.7 highlights the contrast between classification accuracy with mRMR on level-2 alarms (red lines) and classification accuracy with mRMR on process measurements (blue lines). The classification accuracy when using level-2 alarms increases quickly for the first ranked variables and becomes flat or starts decreasing after a certain number of variables (in this case, two variables) since additional variables bring noise that impact the classification. By contrast, the classification accuracy when using process measurements increases slowly for the first variables, which means that the first ranked variables with mRMR on process measurements are not able to discriminate between the different classes as the variables selected with mRMR on level-2 alarms. In addition, the difference between the accuracy on the

training data and the accuracy on the test data is larger when using mRMR on the process measurements than when using mRMR on the alarms. This shows that there is overfitting in the classification with variables selected with mRMR on the process measurements, and that the corresponding classifier is not robust.

Selected set of variables

The set $\mathcal{X}_2$ of the two highest-ranked variables with mRMR on the level-2 alarms is selected as best set of variables based on the classification accuracy on the training and test data in scenarios AB and BA. $\mathcal{X}_2$ is used for fault detection and identification in the next sections.

The two variables in $\mathcal{X}_2$ are relevant with regard to the considered faults. The highest-ranked variable, PT312, corresponds to the air delivery pressure, measured just after the valves V10 and V11 that are used to induce the blockage and leakage faults, and just before the bypass valve U39 used to induce the diverted flow fault. The second variable, PIC501/PID1, refers to the air outlet valve of the three-phase separator. As the controller output of the valve opening, PIC501/PID1 gives an indication of the value of the outlet air flow, which is not directly measured in the facility. The blockage and the leakage fault lead to a reduction in the quantity of air introduced in the mixed flow, which triggers a closure of the air outlet valve of the three-phase separator to maintain the air pressure of the three-phase separator at its setpoint.

6.6.2 Fault detection and identification

This sub-section investigates Bayesian classification of the selected variables, based on fusion of a continuous classifier on process measurements X_m and a discrete classifier on the corresponding level-1 alarms $\bar{A}_m$.

Table 6.2: Classification accuracies on the test dataset (as defined in Sub-section 6.5.2) for scenarios AB and BA with the selected set of variables $\mathcal{X}_2$.

	Scenario AB	Scenario BA
Continuous	0.678	0.72
Discrete	0.719	0.676
Fusion	0.779	0.788
QDA	0.702	0.740

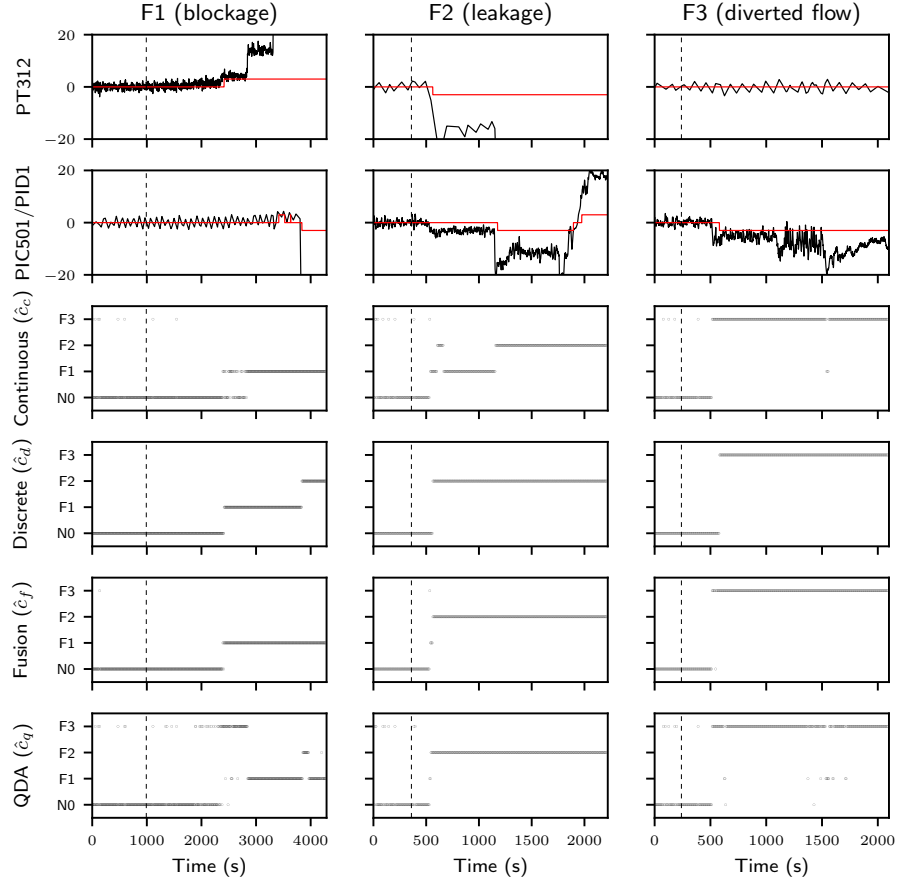


Figure 6.8: Standardized measurements, level-1 alarms (in red), and classification outcomes with continuous Bayesian classifier, discrete Bayesian classifier, fusion of both Bayesian classifiers, and QDA (scenario AB). The measurements are standardized based on the standard deviations during normal operation, and the amplitude of the level-1 alarms indicates the level-1 alarm thresholds ($\bar{k} = 3$). The vertical dotted lines indicate the beginning of the faults.

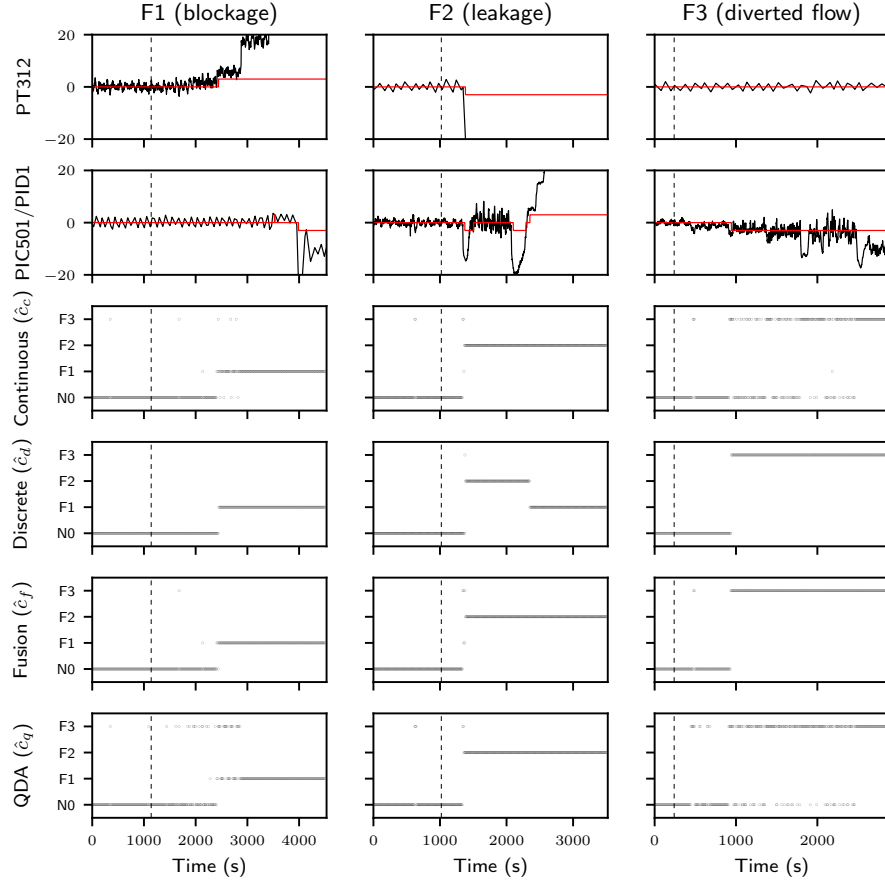


Figure 6.9: Standardized measurements, level-1 alarms (in red), and classification outcomes with continuous Bayesian classifier, discrete Bayesian classifier, fusion of both Bayesian classifiers, and QDA (scenario BA). The measurements are standardized based on the standard deviations during normal operation, and the amplitude of the level-1 alarms indicates the level-1 alarm thresholds ($\bar{k} = 3$). The vertical dotted lines indicate the beginning of the faults.

Classification results

Figure 6.8 summarizes the classification outcome for scenario AB, i.e. the classification outcome for the faults in operating point B when process measurements in operating point A are used for training the classifier. The plots of the first two rows show the standardized selected variables (in black) and the corresponding level-1 alarms (in red) for faults F1, F2, and F3. The vertical dotted lines indicate the beginning of the faults: the samples before the vertical line correspond to normal operation (N0) in operating point B. The bottom plots show respectively the classification outcome $\hat{c}_c$ with the continuous Bayesian classifier, the classification outcome $\hat{c}_d$ with the discrete Bayesian classifier, and the classification outcome $\hat{c}_f$ with the fusion of both classifiers. The plots in the last row show the classification outcome $\hat{c}_q$ of the quadratic discriminant analysis classifier on the process measurements for comparison. Figure 6.9 summarizes the classification outcome for scenario BA.

The classification accuracies corresponding to Figure 6.8 and 6.9 are indicated in Table 6.2. The accuracy indicates the proportion of observations of the plant vector correctly classified. The confusion matrices are presented in Table 6.3. The confusion matrices show the proportion of observations of the plant vector classified as each class.

Comparison of the methods

This sub-section discusses and compares the classification results of the considered methods:

- The continuous Bayesian classifier on the process measurements.
- The discrete Bayesian classifier on level-1 alarms.
- The Bayesian fusion of the continuous and the discrete classifiers.
- A traditional discriminant analysis classifier, in this case QDA.

The third rows of Figure 6.8 and Figure 6.9 show that the continuous Bayesian classifier is able to classify all faults correctly after a certain time. However, the detection occurs late and the classifier has troubles clearly distinguishing the faults from normal operation during their development, in particular for F3 as shown in the third row of Figure 6.9.

On the other hand, the discrete Bayesian classifier shows a clear and early detection of the faults due to the activation of the alarms, but may not be able to distinguish between faults during the late stage of the fault when all the alarms have reached the same values. For example, the discrete classifier classifies F1 as F2 in the fourth row of Figure 6.8 or F2

Table 6.3: Confusion matrices on the test dataset for scenarios AB and BA with the selected set of variables $\mathcal{X}_2$. The confusion matrices show the proportion of observations of the plant vector classified as each class. The columns correspond to the predicted classes and the rows correspond to the actual classes.

		Scenario AB				Scenario BA			
		N0	F1	F2	F3	N0	F1	F2	F3
Continuous	N0	0.94	0.00	0.00	0.06	0.99	0.00	0.00	0.01
	F1	0.50	0.50	0.00	0.01	0.38	0.60	0.00	0.02
	F2	0.09	0.30	0.61	0.01	0.13	0.00	0.86	0.01
	F3	0.15	0.01	0.00	0.84	0.51	0.00	0.00	0.49
Discrete	N0	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	F1	0.43	0.43	0.14	0.00	0.38	0.62	0.00	0.00
	F2	0.11	0.00	0.89	0.00	0.14	0.47	0.39	0.00
	F3	0.18	0.00	0.00	0.82	0.27	0.00	0.00	0.73
Fusion	N0	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	F1	0.43	0.57	0.00	0.00	0.38	0.62	0.00	0.00
	F2	0.09	0.01	0.89	0.00	0.13	0.01	0.85	0.01
	F3	0.15	0.00	0.00	0.85	0.26	0.00	0.00	0.74
QDA	N0	0.95	0.00	0.00	0.05	0.97	0.00	0.00	0.03
	F1	0.37	0.41	0.04	0.18	0.33	0.56	0.00	0.11
	F2	0.09	0.01	0.90	0.00	0.13	0.00	0.87	0.00
	F3	0.15	0.04	0.00	0.80	0.36	0.00	0.00	0.64

as F1 in the fourth row of Figure 6.9 at the end of the faults because the the alarms have reached the same values in F1 and in F2.

The fifth rows of Figure 6.8 and Figure 6.9 demonstrate that the fusion of both classifiers brings together the advantages of both the continuous and the discrete classifiers. There is a clear detection and identification of the fault as soon an alarm gets activated (for example with F2 in Figure 6.8), as for the discrete classifier. The faults are still identified correctly when the values of the alarms lead to indetermination (for example with F1 in Figure 6.8), due to the continuous classifier. This has an impact on the classification accuracy, and Table 6.2 illustrates the gain in classification accuracy with the Bayesian fusion compared to both the continuous and discrete classifier, in scenario AB and scenario BA. The confusion matrices of Table 6.3 demonstrate that the Bayesian fusion improves the classification for all

the faults compared to the continuous and discrete classifier, and that misclassification with the Bayesian fusion only corresponds to late detection of the faults, i.e. to the faults being classified as normal in the beginning. All three faults have been introduced gradually in the system and their development is very slow in the beginning (in particular for F1), which explains why the fault detection occurs late. Specific methods for incipient fault detection should be applied if earlier detection is required.

In comparison, the QDA classifier on the process measurements presents the same characteristics as the continuous Bayesian classifier. As for the continuous Bayesian classifier, the QDA classifier classifies all faults correctly after a certain time in the sixth rows of Figure 6.8 and Figure 6.9. However, QDA cannot clearly separates fault classes when the level of noise is high compared to the pattern of the fault. This is the case for example with F3 which is regularly classified as normal in the sixth row of Figure 6.8, or with F1 which is regularly classified as F3 in the sixth rows of Figure 6.8 and Figure 6.9. In addition, the QDA classifier classifies several samples from normal operation as F3 in Figure 6.8 and Figure 6.9, which leads to false detections. Limiting the number of false detections is another advantage of the discrete Bayesian classifier, which explains why the fusion of the discrete and continuous classifiers presents very few false detections in scenario AB and BA, as shown in Figure 6.8 and Figure 6.9.

6.6.3 Analysis of the parameters

Alarm thresholds

The role of level-1 alarms is to indicate any deviation from the normal operating range. For this reason, the objectives for designing level-1 alarms and for designing a fault detection system are similar, and the level-1 alarms can directly be used for fault detection. It is common practice to set the alarm threshold to $3\sigma_m$ (σ_m indicates the standard deviation in normal operation) following basic statistical process control rules for a Gaussian variable. In this chapter, the variables are assumed to be Gaussian during normal operation and the $3\sigma_m$ threshold is retained.

The role of level-2 alarms is to indicate a severe deviation from the normal operating range. In this chapter, level-2 alarms are used to select only the variable with severe deviations from the normal operating range. The variables whose level-2 alarms can best discriminate fault classes are good candidates for a robust classification, since the effect of noise and external disturbances compared to the fault pattern is limited on those variables.

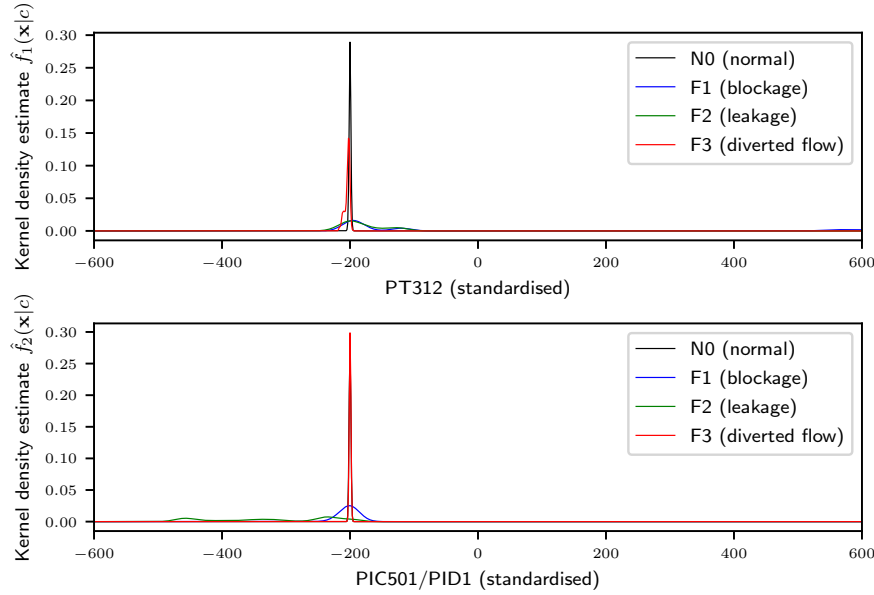


Figure 6.10: Kernel density estimates of the probability density functions of the standardised variables for normal operation N0 and faults F1, F2, and F3 in operating point A.

This variable selection approach explains why the level-1 alarms are also helpful, to some extent, in discriminating fault classes for the fault identification. Similarly to the level-1 alarm threshold, the level-2 alarm threshold can be set as a multiple of σ_m ($5\sigma_m$ is common practice).

Delay timers

Delay timers are set to reduced the number of alarm transitions not to overload the operators. T_{on} and T_{off} for the on-delay and off-delay timers are set during the alarm design stage along with the alarm thresholds. From a safety perspective, the on-delay timer has a higher impact than the off-delay timer since it delays the activation of the alarm [44], which can be an issue in case of real fault. For this reasons, T_{on} is usually taken smaller than T_{off} . In this chapter, $T_{off} = 2T_{on}$ for both the simulation and the industrial case study. T_{on} and T_{off} also depends on the dynamics of the process variable. For the industrial case study, an on-delay timer of 30 seconds and an off-delay timer of 60 seconds are used.

Kernel bandwidth for the continuous Bayesian classifier

The continuous Bayesian classifier assumes the estimation of the probability density functions of the selected variables (Eq. 6.15). The estimated probability density functions are used to assign the probability of each fault class for each new plant vector. In this chapter, the constant $h(c)$ in the bandwidth matrix $\mathbf{H}(c)$ (Eq. 6.16) is used to adjust the smoothness of the estimated probability density functions.

A common method for choosing a good value of the kernel bandwidth in kernel density estimation is cross-validation [200]. In this chapter, a value of $h(c)$ is determined for each fault class c using three-fold cross-validation on the training dataset corresponding to fault c . The training dataset corresponding to fault c is divided in three parts that are used successively to test the fit of the kernel density estimate while the two remaining parts are used for computing the kernel density estimate. The fit of the kernel density estimate is computed using the log likelihood of the data that was not used for computing the kernel density estimate, and the value of $h(c)$ maximizing the log likelihood is chosen for the estimation of $\hat{f}(\mathbf{x}|c)$ in the continuous classifier.

The kernel density estimates $\hat{f}_1(\mathbf{x}|c)$ and $\hat{f}_2(\mathbf{x}|c)$ of the probability density functions of the two selected standardised variables are shown in Figure 6.10 for each fault occurrence. In this case, h is set to 1 for N0 and F3, and to 15 for F2 and F3.

6.7 Summary of the chapter

Classification-based methods are popular for fault detection and identification but their implementation in the industry is limited by noise and variability of process measurements from one fault occurrence to another. This chapter demonstrated how statistical alarms could be used to improve the robustness of the fault detection and identification to the variability in the process measurements. A variable selection stage identifies the most relevant variables for the classification based on the corresponding statistical alarms with high severity threshold, focusing on variables with large variations where the impact of noise and external disturbances is limited. The classification stage uses a Bayesian framework combining the outcome of a discrete classifier based on statistical alarms with low severity threshold and the outcome of a continuous classifier based on process measurements. The discrete classifier limits the effect of noise for fault detection and identification during the early stage of the fault since the number of alarm transitions are limited by the delay timers. The continuous

classifier ensures that the fault is identified correctly during later stages of the fault when alarm values only are not sufficient to distinguish faults, and the final classification outcome is based on the fusion of the discrete and continuous classifiers.

Chapter 7

Review of existing methods for root cause analysis of oscillatory and non-oscillatory disturbances

7.1 Description of the chapter

A *disturbance* refers to a deviation of the variables of the system from their desired values or trends [46]. A disturbance might propagate throughout a plant because of interconnections between assets and processes. Offline diagnosis of a disturbance covers Root Cause Analysis (RCA), where the root cause or root causes of the disturbance are investigated, and propagation path analysis, where the propagation of the disturbance is retraced. The hypothesis of a RCA method is that a disturbance has a root cause that can be discovered. This chapter examines data-driven methods for RCA of disturbances.

Data-driven RCA of disturbances in industrial processes has been a popular research topic in the last twenty years, and many methods have been proposed. One aim of this chapter is to provide a comparative review of the existing RCA analysis methods and clarify their applicability with regard to:

- The type of the disturbance studied. In particular, a distinction is made in the work presented in this chapter between oscillatory disturbances and non-oscillatory disturbances..

- The properties of the system through which the disturbance propagates. The propagation of the disturbance has an impact on the results of data-driven RCA methods since the deviation of the variables from their desired trends that defines the disturbance evolves as the disturbance propagates through the plant. The evolution of the deviation of the variables during the propagation of the disturbance depends on the properties of the systems through which the disturbance propagates.

Most methods have been designed and tested to deal with oscillatory disturbances. While oscillatory disturbances are common, non-oscillatory disturbances also exist. However, they have received less attention in the literature. A second aim of this chapter is to discuss the applicability of existing RCA methods to non-oscillatory disturbances and to give suggestions on how to proceed in this case, based on the outcome of the comparative review.

Section 7.2 defines the criteria that are important for the choice of the RCA method and that will be used as a basis for the comparative review of the methods. The criteria cover the properties of the method, the impact of the systems through which the disturbance propagates on the results of the method, and the types of disturbances covered by the method. Section 7.3 contains the comparative review of data-driven RCA methods.

Building on the outcome of the comparative review, Section 7.4 provides guidelines for RCA of persistent disturbances. Sub-section 7.4.1 focuses on oscillatory disturbances, already well-studied in the literature, and Sub-section 7.4.2 extends the guidelines to RCA of non-oscillatory persistent disturbances. The topic of RCA of transient disturbances is tackled in Section 7.5. RCA of transient disturbances is an unsolved problem in the process industry, and Section 7.5 uses the comparative review of RCA methods to suggest an approach for RCA of transient disturbances.

7.2 Criteria for the comparative review

7.2.1 Properties of the methods

Section 3.4 introduced twelve data-driven RCA methods from the literature that will be studied in the comparative review. Those methods belong to four branches depending on the RCA strategy adopted:

- *Univariate methods* rely on univariate properties of the process measurements, and the root cause is determined based on the values of the univariate properties.

- *Time-delay estimation methods* assess the delay between X and Y , and the variable that precedes all the others is considered as the root cause.
- *Prediction-based methods* assess the effect of X in the prediction of Y to determine if X causes Y , and the variable that is not caused by any other variable is considered as the root cause. Prediction-based methods are divided into two branches according to the prediction approach:
 - Prediction-based methods in the *time-domain*.
 - Prediction-based methods in the *frequency-domain*.

This sub-section presents the properties of the methods covered in the comparative review. The number of variables included in the scope of the RCA is not considered in the comparative review. The selected methods investigate the causal relation between each pair of variables (with the exception of univariate methods), so increasing the number of variables will impact the computational load of each method by increasing exponentially the number of pairs of variables.

Computational load

Root cause analysis can easily become computationally heavy, and the computational load of the method is an important factor to consider. The computational load of the method covers both the computation of the causality measures and the significance test.

Parameters

The automatability of RCA methods is an important factor for their use in industry [182], and parameter tuning is a critical step for automating RCA. The number of parameters and the complexity of their tuning are considered.

Significance test

Interpretability of RCA methods can be limited and the reliability of the results is an important factor towards the acceptance of the methods. Significance tests assess whether given values of the causal dependencies between process measurements indicate causality, for example by means of significance thresholds. The nature of the significance tests and their reliability are considered.

Multivariate and direct causality

Sub-section 2.4.3 distinguished direct causality that implies a direct causal relation between two variables and indirect causality that implies a causal relation between two variables through one or several intermediate variables. Indirect causality is pervasive in highly interconnected systems. Not differentiating between direct and indirect causality can lead to dense causality graphs where important information is lost. A common way to highlight direct causality between two variables X and Y is to look at the influence of X on Y conditioning on a third variable Z that could be the common cause of both X and Y . This requires a multivariate method, i.e. a method able to take into account the effect of several variables X_m on a variable Y such as methods based on multivariate prediction. Therefore, the multivariate property of a method is related to its ability to study direct causality, and both the ability of the methods to deal with multivariate cases and to distinguish direct causality from indirect causality are considered.

Data requirements

The data requirements for the methods refer to two aspects in particular: the number of samples and the excitation of the measurements. The number of samples N available for the RCA has an influence on the choice of the method since the minimum value for N assumed by the different methods vary. The excitation of the measurements refers to the amount of variability present in the measurements available for the RCA. In addition to the number of samples, specific types of methods have requirements in terms of excitation of the data available.

7.2.2 Impact of the systems' properties on the results of the methods

The impact of the systems met during the disturbance propagation on the results of RCA methods have not been studied in a systematic manner. Nevertheless, several comparative analyses studied specific system properties as mentioned in Sub-section 3.4.6. Reference [48] investigated the effect of dead-time, low-pass filtering, and additive noise on the results of cross-correlation function, nearest neighbours, and transfer entropy. Reference [183] looked at the effect of simple simulated blocks (such as first and second order systems, feedback loops, dead-time, and squaring function) on the results of cross-correlation function, Granger

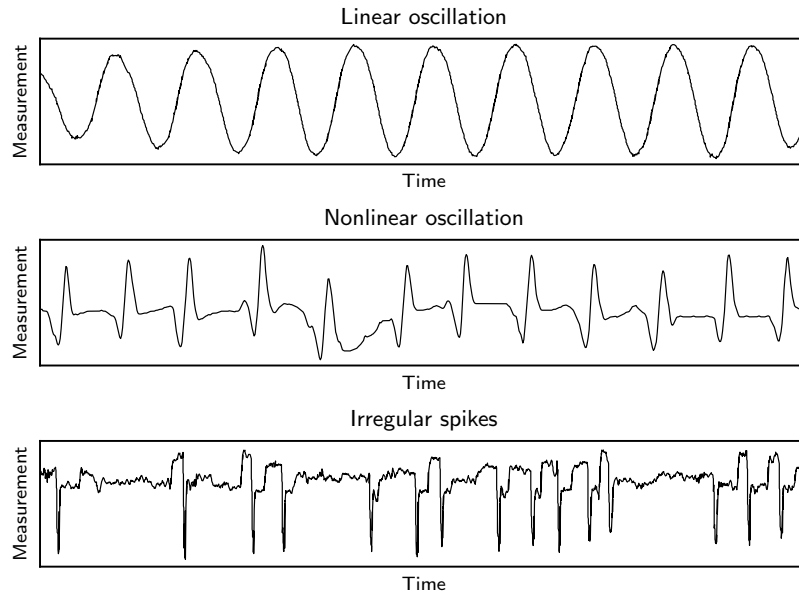


Figure 7.1: Examples of persistent disturbances obtained from [48] using the software developed by [201].

causality, and transfer entropy. Reference [181] studied the impact of noise and feedback loop on the results Granger causality and transfer entropy. This chapter extends the analysis started in [48], [183], and [181].

This sub-section selects and defines common properties that describe the systems through which the disturbance propagates in order to analyse the impact of those properties on the results of the methods.

Noise

Noise can refer to process noise and sensor noise [181]. Process noise is the noise that enters the process. Sensor noise is the additive noise in output of the process induced by the sensor when measuring the variable.

Dead-time

The dead-time of a system introduces a time-delay between the input and the output of the system. A distinction between short and long dead-times is made [183].

Low-pass filter

Many components in the process industry act as low-pass filters [8]. Therefore, the effect of low-pass filtering on the results of RCA methods is considered.

Nonlinearity

The two main sources of nonlinearity during the propagation of a disturbance on a process plant are [202]: a nonlinearity in the nature of the process, or a nonlinearity in the actuator (e.g. control valve) including nonlinear characteristics and nonlinear faults in control valves (e.g. stiction, saturation, deadband and hysteresis). Nonlinearities due to additional disturbances entering the loop [202] are not considered. Process nonlinearities and nonlinear valve characteristics are called *soft nonlinearities* because of their nature of being locally linear in the presence of small excitations. Nonlinear faults in control valves, such as stiction, saturation, deadband, backlash and hysteresis, are called *hard nonlinearities* because of their dominant nonlinear characteristics. The effect of soft nonlinearities and hard nonlinearities in the propagation of the disturbance on the results of RCA methods are considered separately.

7.2.3 Types of disturbances

Sub-section 2.4.1 introduced the distinction between incipient, persistent, and transient disturbances. This chapter focuses on persistent disturbances and transient disturbances. A distinction is made between oscillatory persistent disturbances and non-oscillatory persistent disturbances. The types of disturbances retained for the review are listed below.

Oscillatory disturbances

Two common oscillatory disturbances are considered: linear oscillations and nonlinear oscillations, both defined in Sub-section 2.4.1. As introduced in Sub-section 2.4.1, linear oscillations can result from controller tuning, poor process and control system design, or linear external disturbances. Nonlinear oscillations can result from valve nonlinearity, poor process and control system design, or nonlinear external disturbances. Figure 7.1 shows examples of oscillations obtained from [48] using the software developed by [201]. The linear oscillation of Figure 7.1 is generated by interactions between control loops due to poor con-

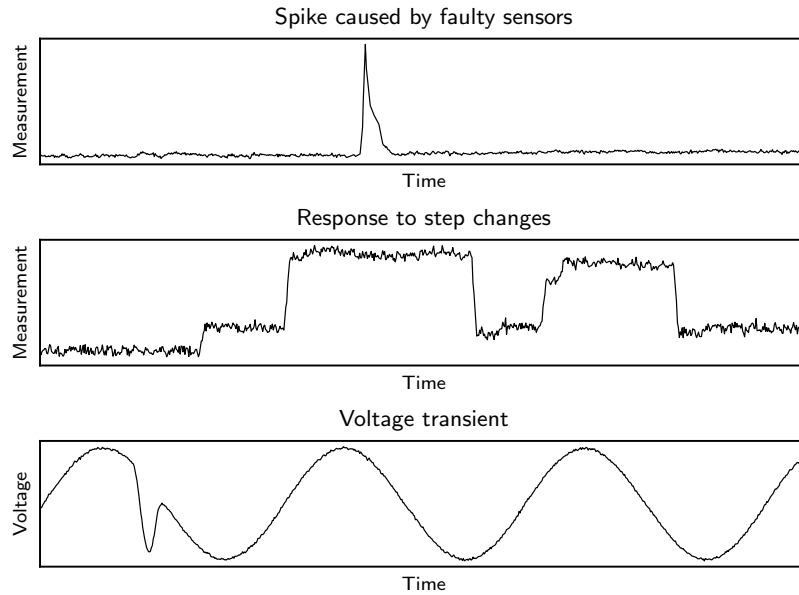


Figure 7.2: Examples of transient disturbances obtained from [97] using the software developed by [201].

troller tuning. The nonlinear oscillation of Figure 7.1 occurs in a distillation column when flow surges appear in an upstream reactor due to a pressure setpoint set too high.

Non-oscillatory persistent disturbances

Non-oscillatory persistent disturbances are also common in the process industry. Irregular persistent spikes are a typical example that has been considered for RCA in process literature. Figure 7.1 shows an example of irregular persistent spikes obtained from [48] using the software developed by [201]. The irregular persistent spikes of Figure 7.1 are due to inert gas inflow in a reactive distillation column.

Transient disturbances

Many types of transient disturbances exist on process plants. Step changes are the most common example, and RCA of step changes has been treated in a few publications [153, 181]. Figure 7.2 shows examples of transient disturbances obtained from [97] using the software developed by [201].

Table 7.1: Review of the properties of the methods.

	Univariate	Time-delay estimation				Prediction-based (time-domain)						Prediction-based (frequency)	
	Nearest-points	Correlation	Distance	Nearest-points	Nearest-points		Regression			Uncertainty	Regression		
Properties of the method:	NLI	CCF	DCI	kNN	NN	CCM	GC	CRRR	NPMR	TE	DTF	PDC	
Computational load	Medium	Low	Medium	Medium	Medium	Medium	Low	Low	Medium	High	Medium	Medium	
Main parameters	h, K, L	None	None	K	h, K, L, τ	L, τ	R	R	$L, \tau, \mathbf{b}$	h, L, L', τ	R	R	
Complexity of parameters tuning	Medium	None	None	Low	Medium	Medium	Low	Low	High	High	Low	Low	
Significance test	Surrogates	Heuristic	Heuristic	Statistical	Surrogates	Statistical	Statistical	Statistical and heuristic	Surrogates	Surrogates	Surrogates	Surrogates	
Significance test reliability	Low	Low ^a	Low	Medium	Low	Medium	Medium	Medium	Low	Low	Low	Low	
Multivariate	No	No	No	No	No ^b	No	Yes	Yes	Yes	No ^b	Yes	Yes	
Direct causality	No	No	No	No	No ^b	No	Yes	Yes	Yes	No ^b	No	Yes	
Number of samples	Medium	Low	Low	Low	Medium	Low	High	High	High	High	High	High	
Data excitation	Medium	Low	Low	Low	Medium	Medium	High	High	High	Medium	High	High	

^a For oscillations in particular

^b Extension exists

7.3 Comparative review of existing data-driven root cause analysis methods

This section presents an overview of the comparative review of data-driven RCA methods. The details of the review are available in Appendix B.

7.3.1 Review of the properties of the methods

Table 7.1 summarizes the comparative review of the properties of the methods defined in Sub-section 7.2.1. The details of the comparative review are presented in Section B.1 of Appendix B.

Computational load

The computational load of the method covers both the computation of the causality measure and the significance test. The computation load for calculating the causality measure depends on the complexity of the method. Cross-correlation function, which computes the linear correlation between the measurements to estimate the time-delay, has a low computational load. By contrast, transfer entropy that relies on multidimensional probability density function estimation has a high computational load. For regression-based methods, the computational load depends on the complexity of the model. Granger causality, causal relationship based on residual analysis, directed transfer function, and partial directed coherence present the advantage of relying on simple autoregressive models. Increasing the complexity of the model to deal for example with nonlinear processes (Sub-section 7.3.2) increases the computational load of the method.

The significance test can have an impact on the computational load of the method, especially if the significance test is based on surrogates. Surrogate testing requires the computation of the causality measure on a set of new measurements (the surrogates) generated from the original measurements. The significance threshold is estimated based on the distribution of the values of the causality measures computed on the surrogates. For this reason, surrogate testing multiplies the computational load of the method. By contrast, statistical significance tests based on known distributions or heuristic significance tests have a minor impact on the computational load.

Parameters

Time-delay estimation methods usually have few parameters, which makes them easy to implement. Methods based on nearest-points require the number of nearest neighbours K to be set, and the literature has proposed guidance on how to choose the value of K (e.g. equal to the number of cycles in the analysed oscillatory disturbance for nearest-neighbours [155]). Methods based on nearest-points compare embedding vectors, so parameters of the embedded vectors such as the embedding dimension L and the sampling constant τ should also be set. Nonlinearity index and nearest neighbours that investigate the predictability of the measurements also require a prediction horizon h to be set, although in practice it is recommended to set $h = \tau$ [48]. L and τ have an impact on the results of the method and can be difficult to tune in practice. Therefore, the complexity of parameter tuning for methods based on nearest-points is described as moderate.

Transfer entropy also relies on embedded vectors, so L and τ should also be set ($h = \tau$ is recommended [48]). In addition, the embedding dimension L' of Y should be tuned separately. The tuning of transfer entropy is often described as complex since the effect of the parameters on the probability density function estimation (and thus on the causality measure) is high, in particular the effect of τ .

The complexity of parameter tuning for regression-methods depends on the complexity of the model. Granger causality, causal relationship based on residual analysis, directed transfer function, and partial directed coherence rely on simple autoregressive models where the main parameter is the model order R . R can easily be determined using Akaike information criterion or Bayesian information criterion. Increasing the complexity of the model to deal for example with nonlinear processes increases the complexity of the parameter tuning.

Significance test

Significance tests can rely on significance thresholds determined based on distributions generated from surrogates or based on known statistical distributions (e.g. F -distribution for Granger causality). The value of the causality measure is considered significant if it is above the significance threshold. Heuristic significance tests have also been used, relying on rules of thumb derived from experiments or from the literature (e.g. tests derived from system identification theory for causal relationship based on residual analysis [167]).

The reliability of the significance test is hard to estimate in practice, but publications have pointed out the drawbacks of some tests. Reference [168] highlighted the limited

reliability of significance tests based on surrogates. Reference [150] pointed out that the directionality test of cross-correlation function failed in case of oscillations since the minimum and the maximum of the cross-correlation function were of similar magnitude.

For regression-based methods, the fit of the model must be checked in addition to the significance of the causality measure. The fit of the autoregressive model in Granger causality (and partial directed coherence and directed transfer function) can be assessed using the consistency criterion [167], and the results of the causality analysis are considered reliable only if the quality of the model is good enough (for example if the consistency of the model is greater than 80% according to [167]). For nonparametric multivariate regression, the fit of the model is assessed using the residual sum of squares and the total sum of squares. For causal relationship based on residual analysis, two tests derived from system identification theory, the whiteness test of the residuals and the independence test of the residual and the input, serve both as significance test for the causality measure and as test for the fit of the model.

Multivariate and direct causality

Nonlinearity index is a univariate method, and the values of the nonlinearity index are not sufficient to conclude about direct causality. If two variables X_2 and X_3 are not directly causally related but have a common cause X_1 (scenario of Figure 2.4), the values of the nonlinearity index for X_2 and X_3 do not show that X_2 and X_3 are not directly causally related. Similarly, time-delay estimation methods cannot highlight direct causality, since the time-delay between X_2 and X_3 does not show that X_2 and X_3 are not directly causally related.

Prediction-based methods are bivariate by default, except regression-based methods that can be multivariate if more than one predictor is included in the model. Therefore, regression-based methods can easily be used to investigate direct causality by including intermediate variables in the predictors, without increasing the complexity of the method. Two popular prediction-based methods, nearest neighbours and transfer entropy, have been extended for multivariate analysis to study direct causality but this has a large impact on the computational load, in particular for transfer entropy.

Data requirements

Methods based on regression require a large number of samples for the modeling to be reliable, e.g. no fewer than 2000 samples for Granger causality according to [64]. In addition, the measurements should be sufficiently excited for the modeling [64]. Transfer entropy relies on the estimation of probability density functions and thus also requires a large number of samples as well, i.e. 2000 samples [84]. The quality of the probability density function estimation also depends on the excitation of the data, a high excitation being preferred. Prediction-based methods based on nearest-points require less samples than regression-based methods or transfer entropy, with a minimum number of samples estimated to 400 for nearest neighbours [155] and 100 for convergent cross mapping [159]. Nonlinearity index that also relies on the concept of predictability requires a minimum number of 300 samples [147]. By contrast, the number of samples required by time-delay estimation methods is low, e.g. $N = 50$ or four times the time-delay for cross-correlation function [149].

7.3.2 Review of impact of the systems' properties on the results of the methods

Table 7.2 summarizes the impact of the systems through which the disturbance propagates on the results of the methods based on the criteria defined in Sub-section 7.2.2. The details of the review are presented in Section B.2 of Appendix B.

Noise

Regression-based methods present a good ability to deal with noise compared with other methods such as methods based on nearest-points. The effect of noise on the results of regression-based methods can be limited by reducing the overfitting of the model. By contrast, methods based on nearest-points rely on the shape of the embedded vectors as defined in Eq. 3.3, and the effect of noise on those shapes is high. Time-delay estimation methods are also affected by noise. The authors of [48] highlighted the high effect of noise on the results of cross-correlation function compared with nearest neighbours or transfer entropy, and dynamic time warping causality index requires de-noising beforehand for the computation of the dynamic time warping distance to be effective [153].

The estimation of the probability density functions for transfer entropy is also affected by noise, but not as much as nearest-points methods since the method to estimate the

probability density functions (e.g. histograms or kernel density estimation) can be tuned to adapt to the level of noise.

Dead-time

Time-delay estimation methods assume that the process measurements are separated by a time-delay and thus that the propagation includes a certain dead-time. Therefore, those methods are not affected by the presence of a dead-time. Methods that need to take into account the time-delay in the parameters τ and h are affected by the dead-time, since the impact of those parameters on the results is high and the perfect tuning can be difficult to find. For short dead-times, methods based on autoregressive modeling do not need to specify a time-delay since the time-delay can be included in the model order. If the dead-time increases, a time-delay parameter between each variable needs to be introduced for the autoregressive modeling.

Low-pass filter

Nonlinearity index relies on the principle that process equipment acts as low-pass filters and decrease the nonlinearity of the process measurements during the propagation. Therefore, the effect of low-pass filtering is not only beneficial but required for the method to be effective. However, low-pass filtering can have a negative effect on the results of other types of methods.

Low-pass filtering smooths the patterns of the process measurements. This can make the exact time-delay estimation more difficult to estimate as demonstrated with cross-correlation function in [48]: the low-pass filters decrease the sharpness of the peaks and the time-delay estimation is less precise. This also has an impact on results of methods based on nearest-points since those methods are based on the shape of the embedded vectors that are affected by the low-pass filtering.

The authors of [48] demonstrated that transfer entropy was better at dealing with low-pass filtering than cross-correlation function and nearest neighbours because the probability density functions are less affected by the low-pass filtering than the shape of the embedded vectors. Indeed, the dimensions of the embedded vectors in transfer entropy are small in practice (i.e. $L, L' \leq 2$) compared to dimensions of the embedded vectors in nearest neighbours, so the shape of the embedding vectors and the phase-coupling are more important in nearest neighbours than in transfer entropy.

Table 7.2: Review of impact of the systems through which the disturbance propagates on the results of the methods.

	Univariate	Time-delay estimation				Prediction-based (time-domain)						Prediction-based (frequency)	
	Nearest-points	Correlation	Distance	Nearest-points	Nearest-points		Regression			Uncertainty	Regression		
Efficiency of the method with:	NLI	CCF	DCI	kNN	NN	CCM	GC	CRRA	NPMR	TE	DTF	PDC	
Process noise	Low	Low	Low	Low	Low	Medium	High	High	High	Medium	High	High	
Sensor noise	Low	Low	Low	Low	Low	Medium	High	High	High	Medium	High	High	
Short dead-time	High	High	High	High	Medium ^a	Medium ^b	High	High	Medium ^b	Medium ^a	High	High	
Large dead-time	High	High	High	High	Medium ^a	Medium ^b	Medium ^c	Medium ^c	Medium ^b	Medium ^a	Medium ^c	Medium ^c	
Low-pass filter	High	Low	Medium	Medium	Medium	Medium	Medium	Medium	Medium	High	Medium ^d	Medium ^d	
Soft nonlinearity	Low	Low	Medium	Medium	High	High	Low ^e	Low	High	High	Low ^e	Low ^e	
Hard nonlinearity	Not applicable	Not applicable	Low	Low	Medium	Medium	Not applicable ^e	Not applicable	Medium	Medium	Not applicable ^e	Not applicable ^e	

^a h and τ must be set correctly
^b τ must be set correctly
^c A time-delay parameter must be introduced
^d Depends on cut-off frequency
^e Extension exists

Depending on the cut-off frequency, the low-pass filters can also have an impact on prediction-based methods in the frequency-domain. The causal dependencies at frequencies above the cut-off frequency should be very limited since the low-pass filter attenuates those frequencies.

Nonlinearity

Nonlinearities in the propagation of the disturbance have an impact on the results of most RCA methods, and whether the nonlinearity is soft or hard influences the amplitude of this impact. Soft nonlinearities are locally linear in the presence of small excitations, so all RCA methods could potentially give accurate results in case of small disturbances. Hard nonlinearities in the propagation of the disturbance require that the RCA method can deal with nonlinear relations between the measurements.

The methods that are the most affected by a nonlinearity in the disturbance propagation are nonlinearity index, cross-correlation function, and regression-based methods using linear modeling approaches. The authors of [148] demonstrated that nonlinearities in the propagation path of the disturbance introduce nonlinearity other than that of the disturbance, and could lead to spurious results. Cross-correlation function analyses the linear correlation between two measurements to estimate the time-delay, which is effective only if the nonlinearity can be linearly approximated. Regression-based methods using linear autoregressive models (Granger causality, causal relationship based on residual analysis, directed transfer function, partial directed coherence) are also affected when there are nonlinearities in the propagation. Extensions of Granger causality, directed transfer function, and partial directed coherence using nonlinear modeling approaches (such as nonlinear autoregressive models) have been proposed, especially for cases with hard nonlinearities. Dynamic time warping causality index is also affected by nonlinearities since the dynamic time warping distance between the lagged measurements is not reliable if the lagged measurements have a nonlinear relation. Reference [153] underlines that dynamic time warping causality index should not be applied in case of hard nonlinearities.

Methods that perform well in presence of nonlinearities include nonparametric multivariate regression, transfer entropy, and prediction-based methods using nearest-points such as nearest neighbours and convergent cross mapping. The modelling approach of nonparametric multivariate regression is compatible with hard nonlinearities. Transfer entropy relies on probability density functions estimated from embedded vectors that do not assume a linear

relation between the measurements [84]. Nearest neighbours and convergent cross mapping use a nonlinear prediction approach that is also compatible with hard nonlinearities [157, 160]. kNN time-delay estimation relies on the same nonlinear prediction approach as nearest neighbours, but a nonlinear relation between the two measurements can make the exact time-delay estimation more difficult.

7.3.3 Review of the types of disturbances covered by the methods

Table 7.3 summarizes the types of disturbances covered by the methods following the distinctions introduced in Sub-section 7.2.2. The details of the types of disturbances covered by the methods are presented in Section B.3 of Appendix B.

Oscillatory disturbances

Table 7.3 makes a distinction between linear oscillations and nonlinear oscillations. Nonlinearity index is applicable on nonlinear oscillations, although [148] pointed out the limited reliability of nonlinearity index due to possible external sources of nonlinearity. However, nonlinearity index is not applicable on linear oscillations because linear oscillations do not contain harmonics that can be used for the nonlinearity analysis. Prediction-based methods using nearest-points such as nearest neighbours and convergent cross mapping are more efficient for RCA of nonlinear oscillations than for RCA of linear oscillations. Indeed, [155] highlighted that the phase-coupling which occurs in measurements with nonlinear oscillations is best utilized with nearest neighbours than with methods such as transfer entropy. The dimensions of the embedded vectors in transfer entropy are small for the multivariate probability density function estimation to be tractable, whereas nearest neighbours uses larger values of L that allow the embedded vectors to give a better representation of the phase-coupling.

Time-delay estimation methods are not adapted for RCA of oscillatory disturbances (both linear and nonlinear) because the periodicity of the measurements makes the estimation of the exact time-delay difficult.

RCA of oscillatory disturbance with Granger causality alone leads to a high number of causal dependencies, which is useful for propagation path analysis but not for root cause analysis [168]. However, the results of Granger causality can be refined with an analysis in the frequency-domain at the frequency of the oscillation with partial directed coherence and directed transfer function. Reference [168] pointed out that the accuracy of the results

of Granger causality, partial directed coherence and directed transfer function combined for RCA of oscillatory disturbances was high, comparable to the accuracy of the results of nearest neighbours, and higher than the accuracy of the results of transfer entropy. Reference [168] also highlighted that nonparametric multivariate regression provided highly accurate results for RCA of oscillatory disturbances.

Non-oscillatory persistent disturbances

Irregular persistent spikes are a typical example of a non-oscillatory persistent disturbance, which has been used as a case study for RCA in the literature [48]. In general, methods applicable on nonlinear oscillations are applicable on irregular persistent spikes, but exceptions exist.

Prediction-based methods in the frequency-domain are not appropriate for irregular persistent spikes since the spectrum of measurements with irregular persistent spikes is not localized, and there is no advantage in analysing the causal dependencies at each frequency separately. On the contrary, time-delay estimation methods are more appropriate for RCA of irregular persistent spikes than for RCA of oscillations since the measurements are not periodic, which makes the estimation of the time-delay easier.

Reference [48] highlighted that transfer entropy performed well for RCA of irregular persistent spikes, better than nearest neighbours or cross-correlation function. nearest neighbours did not detect causal dependencies far from the root cause, which shows that the distortion of the shape of the measurements during the propagation of the disturbance has an impact on methods based on nearest-points. The distortion of the shape is more visible on irregular persistent spikes than on oscillations, which makes methods such as nearest neighbours or convergent cross mapping less efficient for RCA of irregular persistent spikes than for RCA of nonlinear oscillations.

Transient disturbances

A transient disturbance is a short-lasting deviation from desired operation that seldom repeats itself within the time horizon of analysis, as introduced in Sub-section 2.4.1. Step changes are an example of transient disturbance, which has been used as a case study for RCA in the literature [153, 181].

Prediction-based methods in the frequency-domain are not appropriate for transient disturbances since the spectrum of measurements with transient disturbances is not localized.

Methods based on nearest-points can also be difficult to apply since transient disturbances seldom repeat themselves, which means that the number of nearest neighbours available is limited. For prediction-based methods using nearest-points such as nearest neighbours or convergent cross mapping, this means that the number of nearest neighbours K that can be used for studying the predictability of Y based on embedded vectors of X is low. Therefore, the reliability of nearest neighbours and convergent cross mapping is impacted. Both nearest neighbours and nonlinearity index use the self-predictability of X based on past embedded vectors of X , so the disturbance must repeat itself for the method to be applicable. For this reason, nonlinearity index, nearest neighbours and convergent cross mapping are not recommended for RCA of transient disturbances.

kNN also relies on nearest-points, but the comparison of X and Y is used to estimate the time-delay between the corresponding measurements. kNN can be applied with $K = 1$ if the disturbance does not repeat itself. Generally speaking, time-delay estimation methods can be used for RCA of transient disturbances, and the time-delay estimation works well even if the disturbance does not repeat itself. Reference [183] pointed out that cross-correlation function was more reliable than transfer entropy or Granger causality for RCA of a transient weld disturbance in a glass forming process, and the authors of [153] introduced dynamic time warping causality index for RCA of transient disturbances with a case study on a step change. In addition, the number of samples required by time-delay estimation methods is low so the short-lasting aspect of transient disturbances is not an issue.

Prediction-based methods with regression can also be applied for RCA of transient disturbances. Reference [183] underlines the good performance of Granger causality for RCA of a transient weld disturbance in an industrial glass forming process. In this case, Granger causality is performing better than transfer entropy, which finds causal dependencies between all variables affected by the disturbance and is thus unable to determine the root cause. However, Granger causality and regression-based methods in general require a minimum number of samples N for the modeling, which can be an obstacle since transient disturbances are short-lasting. In addition, the measurements should be sufficiently excited for the modeling to be reliable. This is not the case if the transient disturbance corresponds to a small deviation from desired operation.

Table 7.3: Review of types of disturbances covered by the methods.

	Univariate	Time-delay estimation			Prediction-based (time-domain)					Prediction-based (frequency)		
		Nearest-points	Distance	Nearest-points	Nearest-points		Regression			Uncertainty	Regression	
Efficiency of the method with:	Nearest-points	DCI	CCF	kNN	NN	CCM	GC	CRR	NPMR	TE	DTF	PDC
Linear oscillation	Not applicable	Low	Low	Low	Medium	Medium	Medium (or high ^a)	Medium	Medium	Medium	High ^a	High ^a
Nonlinear oscillation	Low	Low	Low	Low	High	High	Medium (or high ^a)	Medium	High	Medium	High ^a	High ^a
Non-oscillatory persistent disturbance (e.g. irregular spikes)	Low	Medium	Medium	Medium	Medium	Medium	Medium	Medium	High	High	Low	Low
Transient disturbance (e.g. step change)	Low	Medium	Medium	Medium	Low	Low	Medium ^b	Medium ^b	Medium ^b	Low	Very low	Very low

^a When using Granger causality, directed transfer function, and partial directed coherence combined

^b If the number of samples available is large enough and the measurements are sufficiently excited

7.4 Root cause analysis of persistent disturbances

This section discusses guidelines for RCA of persistent disturbances based on the results of the comparative review. Sub-section 7.4.1 focuses on RCA of oscillatory disturbances. RCA of oscillatory disturbances has been extensively discussed in process literature and Sub-section 7.4.1 uses recommendations from the literature updated in the light of the comparative review of Section 7.3. Sub-section 7.4.2 extends the analysis to RCA of non-oscillatory persistent disturbances.

7.4.1 Root cause analysis of oscillatory disturbances

Existing data-driven methods have been designed for RCA of oscillatory disturbances. RCA of oscillatory disturbances is complex since the starting time of the oscillations can be hard to locate in the measurements, and the periodic nature of the disturbance makes it difficult to identify the measurement that precedes the others. For those reasons, data-driven methods have been designed (or imported from other fields) to automatically capture the causal dependencies in the measurements that could not be spotted by humans.

References [64], [182] and [168] proposed guidelines for RCA of oscillatory disturbances. RCA of oscillatory disturbances usually starts with grouping the variables that are oscillating at a common frequency [64]. Both [182] and [168] recommend starting with a simple linear method to get a first overview of causal dependencies. Granger causality can be used if the number of samples available N is large enough. The results of Granger causality can be refined with an analysis in the frequency-domain at the frequency of the oscillation with directed transfer function and partial directed coherence, where directed transfer function is used to find the root cause based on causal dependencies at the given frequency and partial directed coherence is used to highlight direct causal dependencies. If N is small, cross-correlation function can give a first indication of the possible root cause, although it might not be efficient on oscillatory disturbances.

If the system contains nonlinearities or if the linear method provides unsatisfactory results, methods able to deal with nonlinearities should be applied [168]. The choice of the method depends of the number of samples available. If N is large, [182] recommends transfer entropy, although [168] recommends to avoid transfer entropy especially when the system is large since transfer entropy usually finds many causal dependencies and direct transfer entropy has a very high computational load. Nonparametric multivariate regression

is recommended by [168] instead, if N is large. If N is below 2000 samples (but above 400 samples), nearest neighbours should be preferred.

In any case, if N is too small and the above mentioned methods do not provided satisfactory results, methods based on adjacency matrices can be an alternative to data-driven RCA methods [64]. Methods based on adjacency matrices assume information about process connectivity is available and not too complex. Regardless, it is advised to verify the final RCA results using process knowledge or site personnel if available [168].

7.4.2 Root cause analysis of non-oscillatory persistent disturbances

Root cause analysis of non-oscillatory persistent disturbances has not been widely treated in the literature. The irregular persistent spikes studied in [48] are a notable exception. Sub-section 7.3.3 pointed out that most methods applicable on oscillatory disturbances were applicable on non-oscillatory persistent disturbances, with some exceptions. Consequently, guidelines for RCA of non-oscillatory persistent disturbances can be adapted from the guidelines for RCA of oscillatory disturbances from the previous sub-section.

The root cause analysis should start with a clustering of the measurements to make sure the selected variables are the variables affected by the disturbance studied. As in sub-section 7.4.1, cross-correlation function or Granger causality should be used as a first linear analysis. Cross-correlation function should be tested in priority since it performs well on non-oscillatory disturbances and is very fast to implement. If N is large, Granger causality can also be used, although in the case of non-oscillatory disturbances Granger causality cannot be complemented by an analysis in the frequency-domain.

If the system contains nonlinearities or if the linear method provides unsatisfactory results, methods able to deal with nonlinearities should be applied. Nonlinear modeling approaches such as nonparametric multivariate regression offer a good performance if N is large but tuning the model is complex. Prediction-based methods using nearest-points (such as nearest neighbours) or transfer entropy provide alternative nonlinear prediction approaches to regression-based methods using complex modeling approaches. Nearest neighbours that rely on the shape of the measurements can present a limited performance if the shape of the measurements is distorted during the disturbance propagation, as it is the case with low-pass filtering of irregular persistent spikes [48]. In comparison with nearest neighbours, transfer entropy is not as affected by the distortion of the measurements. Tuning of the parameters of transfer entropy requires an accurate estimation of the time-delays, which is easier on

non-oscillatory disturbances than on oscillatory disturbances. Still, transfer entropy can find many causal dependencies, which makes it difficult to determine the root cause if the system is large.

7.5 Root cause analysis of transient disturbances

The problem of root cause analysis of transient disturbances has not been addressed in a systematic manner in process literature. This section uses the analysis of the comparative review to propose an approach to RCA of transient disturbances.

7.5.1 Root cause analysis of transient disturbances with existing methods

Transient disturbances are short-lasting and seldom repeat themselves. Consequently, locating the starting time of a transient disturbance is easier than locating the starting time of an oscillatory disturbance, which is an important factor when it comes to RCA of transient disturbances. If the starting time of the disturbance can be located, the measurement or the measurements where the disturbance appears first are likely to be the root cause(s).

Sub-section 7.3.3 pointed out two types of existing data-driven RCA methods that could be applicable for RCA of transient disturbances: prediction-based methods using regression (in the time-domain), and time-delay estimation methods. Methods based on regression correspond to the category of methods that look for causal dependencies in the measurements, while time-delay estimation methods try to highlight the measurement affected by the disturbance that precedes the others.

Methods based on regression require a model to be fitted, which can be a complex task. Model fitting assumes a large number of samples available and a sufficient excitation of the measurements, which might not be the case for transient disturbances. In addition, the distortion of the shape of the measurements as the disturbance propagates can be hard to capture for linear models, and advanced modeling techniques may be needed with complex parameter tuning and high computational load.

By contrast, time-delay estimation methods rely on the assumption that the root cause is the variable on which the disturbance is measured first. Time-delay estimation methods are easier to apply on transient disturbances than on oscillating disturbances. Time-delay

estimation methods are usually fast to implement, easy to interpret for industrial practitioners, and do not have the constraints of modeling approaches with regards to number of samples or excitation of the measurements. Nevertheless, existing time-delay estimation methods are affected by the effects of the disturbance propagation such as nonlinearities, low-pass filtering, and noise. The next sub-sections illustrate the limitations of the time-delay estimation approach for RCA of transient disturbances and suggest a new approach using methods for detection of transient disturbances.

7.5.2 Limitations of the time-delay estimation approach

Among the existing data-driven RCA methods reviewed in this chapter, time-delay estimation methods emerged as the most appropriate for RCA of transient disturbances. Reference [183] compared cross-correlation function, Granger causality, and transfer entropy for RCA of transient weld disturbances in a glass forming process and pointed out that cross-correlation function was the most reliable method in that case. Nevertheless, existing time-delay estimation methods are dependent on the shape of the measurements. This sub-section uses the weld disturbance of [183] to illustrate the limits of the time-delay estimation approach.

Figure 7.3 shows a transient weld disturbance obtained from [183] using the software developed by [201]. A glass cylinder is fed with constant speed into an electrical arc furnace and is pulled out as a beam with a controllable speed (pulling speed) to maintain the diameter of the beam (upper diameter) and other properties of the beam at their setpoints. The control structure is complex and detailed in [203]. Figure 7.3 focuses on four variables: the temperature of the cylinder, the upper diameter, the pulling speed, and the final diameter of the solidified glass (lower diameter).

A transient disturbance in the cylinder temperature enters the process and propagates to the upper diameter and to the pulling speed which is controlling the upper diameter. For this reason, the upper diameter and the pulling speed are affected simultaneously and the causal influence is bidirectional. The disturbance eventually propagates to the lower diameter at the end of the process.

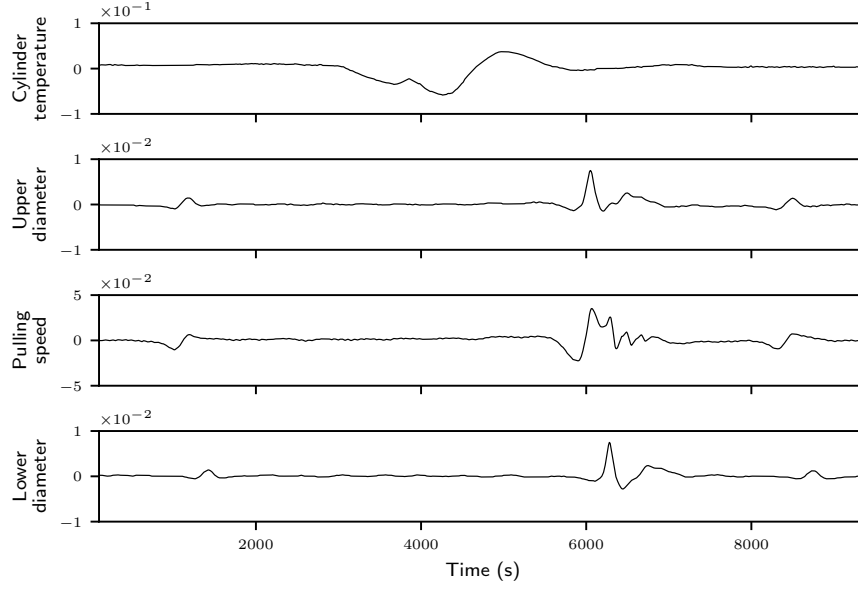


Figure 7.3: Transient weld disturbance in a glass forming process from [183]. The sampling period is 10 seconds.

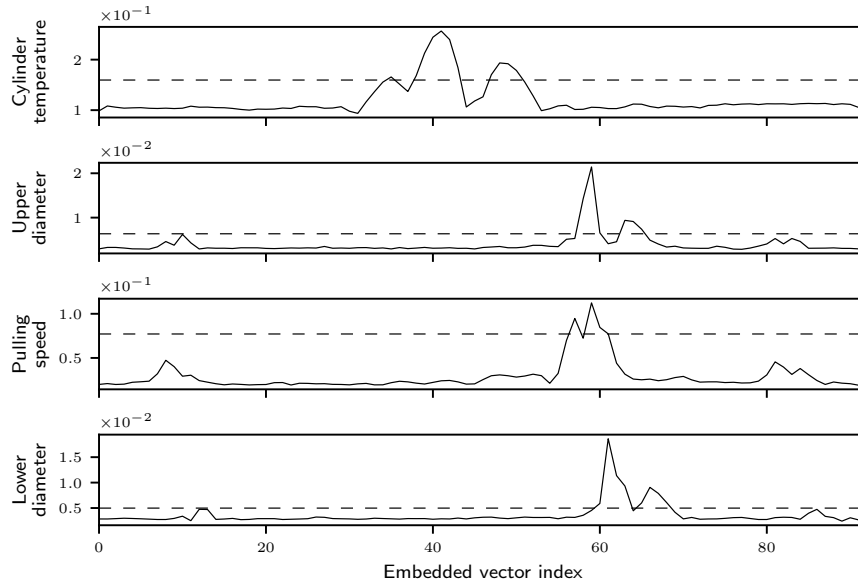


Figure 7.4: Anomaly indices for the transient weld disturbance of Figure 7.3. The embedding granularity is set to 1, the embedding dimension is set to 15, the embedding step is set to 10, and the number of nearest neighbours is set to 75. The dotted lines indicate the detection threshold.

Table 7.4: Estimated time-delays with cross-correlation function on the transient weld disturbance of Figure 7.3. The values are indicated in number of samples. The sampling period is 10 seconds.

	Cylinder temperature	Upper diameter	Pulling speed
Upper diameter	134		
Pulling speed	142	3	
Lower diameter	164	23	67

Reference [183] pointed out that cross-correlation function successfully identified the cylinder temperature as the root cause. Table 7.4 indicates the time-delays estimated with cross-correlation function on the disturbance of Figure 7.3, and Table 7.5 shows the corresponding maximum correlation values ¹. The estimated time-delays demonstrate that the cylinder temperature is first affected by the disturbance, followed by the upper diameter and the pulling speed (separated only by three samples), and finally by the pulling speed. However, the maximum correlation values listed in Table 7.5 show that cross-correlation function is affected by the distortion of the measurements during the propagation of the disturbance. In particular, the measurement of the cylinder temperature and the measurement of the upper diameter present different shape. The maximum time-delayed cross-correlation between those two measurements is 0.178, which is barely above the significance threshold for the correlation index computed according to [149] (0.175 in this case).

7.5.3 Root cause analysis using methods for detection of transient disturbances

Sub-section 7.5.1 highlighted that if the starting time of a transient disturbance can be located, the measurement or measurements where the disturbance appears first are likely to be the root cause(s). Therefore, methods for detection of transient disturbances could be used to identify the measurements where the disturbance appears first. Detection of transient disturbances has been well-studied in the literature, and methods for detection of transient disturbances do not rely on the comparison of the shape of the measurements like time-delay estimation methods. This sub-section demonstrates the concept of RCA based

¹The time-delays and the maximum correlation values have been computed by the author of this thesis using the data extracted with the software developed by [201] based on the original plots from [183]. The maximum correlation values and the corresponding significance threshold are computed using the method introduced in [149].

Table 7.5: Maximum time-delayed cross-correlation values with cross-correlation function on the transient weld disturbance of Figure 7.3. The significance threshold for the correlation index is 0.175.

	Cylinder temperature	Upper diameter	Pulling speed
Upper diameter	0.178		
Pulling speed	0.294	0.673	
Lower diameter	0.211	0.931	0.473

on disturbance detection using the method for detection of transient disturbances developed in [204].

The method of [204] compares embedded vectors of a measurement with each other to identify anomalous segments in the measurement. An anomaly index is defined for each embedded vector as the Euclidean distance to the k -th nearest neighbour of this embedded vector, where the nearest neighbours correspond to other non-overlapping embedded vectors of the measurement. A detection threshold is defined based on the distribution of the values of the anomaly indices for all the embedded vectors to determine if a given embedded vector is abnormal. Abnormal embedded vectors are considered to be affected by transient disturbances.

Figure 7.4 shows the values of the anomaly indices for the embedded vectors computed on the measurements of the transient weld disturbance of Figure 7.3. The dotted lines indicate the detection threshold computed according to [204]. Table 7.6 indicates the index of the first abnormal embedded vector for each variable affected by the transient weld disturbance, which indicates when the disturbance is detected in each variable. Abnormal embedded vectors are first detected in the measurement of the cylinder temperature, then in the measurement of the pulling speed and in the measurement of the upper diameter, and eventually in the measurement of the lower diameter. This corresponds to the propagation

Table 7.6: Index of the first abnormal embedded vector for each variable affected by the transient weld disturbance of Figure 7.3.

	Cylinder temperature	Upper diameter	Pulling speed	Lower diameter
Embedded vector index	35	58	57	60

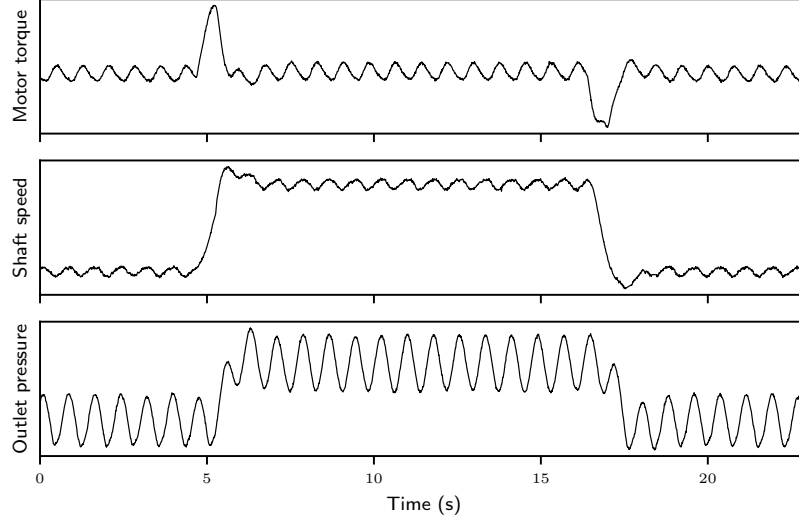


Figure 7.5: Response to step changes in the shaft speed setpoint with overlapping oscillatory disturbance on a compressor rig from [97].

of the disturbance, and the chronology of the disturbance detection can be used to determine the root cause.

The method of [204] can also be applied for transient disturbance detection in more complex cases. The method was originally applied for detecting step changes in a compressor rig in [97]. In this case, the measurements are oscillating in normal operation and the anomaly index is used to detect abnormal segments from the normal oscillating segments. Figure 7.5 shows a selection of the measurements analysed in [97], extracted using the software developed by [201]. In this example as well, the anomaly index can be used to locate the starting time of the disturbance and determine the root cause. The step changes in the shaft speed setpoint have an impact on the motor torque first, then on the shaft speed, and eventually on the outlet pressure. Values of the anomaly indices can be found in [97].

7.6 Summary of the chapter

Multiple data-driven root cause analysis methods have been proposed, corresponding to various root cause analysis strategies. Section 3.4 suggested a classification of those methods

based on the strategy adopted, and this chapter provides a comparative review of the selected methods and strategies. Criteria for assessing RCA methods are defined in Section 7.2 with a focus on two aspects: the impact of the systems met during the disturbance propagation on the results of the method, and the type of disturbance on which the method can be used.

Since most methods have been developed for oscillatory disturbances, this chapter investigates two types of disturbances in particular: non-oscillatory persistent disturbances, and transient disturbances. Non-oscillatory persistent disturbances share characteristics with oscillatory disturbances and Section 7.4 provides recommendations for RCA of non-oscillatory persistent disturbances based on the outcome of the comparative review and on existing guidelines for RCA of oscillatory disturbances. By contrast with oscillatory disturbances, transient disturbances are short-lasting and seldom repeat themselves. This has a large impact on the applicability of most of the existing RCA methods.

Section 7.5 highlights that, of the existing methods, time-delay estimation methods are the most appropriate for RCA of transient disturbances. However, existing time-delay estimation methods require the measurements to present similar shapes, which might not be the case due to the propagation effects discussed in the comparative review. Section 7.5 suggests an alternative approach based on disturbance detection that does not involve comparing the shapes of the measurements. The suggested approach uses methods for detection of transient disturbance from the literature to detect in which measurements the disturbance appears first. The approach is demonstrated on a transient weld disturbance from the literature.

Chapter 8

Critical evaluation and suggestions for future work

8.1 Description of the chapter

This chapter discusses the contribution of the thesis based on the three research objectives introduced in Section 1.5 and suggests research directions for extending the contribution. Section 8.2 lists the merits and limitations of the approaches proposed to answer each of the research objectives, and assesses the contribution of the thesis to the main motivation, which is increasing the applicability of data-driven Fault Detection and Diagnosis (FDD) methods in industrial sites. Section 8.3 identifies future work for each of the research objectives and discusses future research directions in the field of data-driven FDD.

8.2 Evaluation of the achievements of the objectives

8.2.1 Objective 1: online FDD approach to alarm data analysis

Methods for analysis of alarm data have been initially designed for offline applications such as alarm rationalization (i.e. removing redundant alarms) to match the use cases of alarm management. Rationalization of the alarm system is done periodically on the plant, for example every few months, based on statistical analysis of the alarm records generated during the previous months. Methods for offline analysis of alarm data have been recently extended as online methods to provide real-time support to the operators, for example during

alarm floods where the rate of alarms is greater than the operators can manage. Drawing a parallel between fault identification and online alarm flood classification, the first objective of this thesis was to evaluate an online FDD approach to the problem of online alarm flood classification.

Description of the approach

The state of the art in alarm data analysis in Chapter 3 distinguished sequence mining methods that apply to alarm sequences from time series analysis methods that apply to alarm series. The state of the art highlighted that alarm data analysis methods for online decision-support were all sequence mining methods, initially designed as offline methods, which can make their online implementation in the plant challenging. Chapter 4 suggested a time series analysis approach to online alarm flood classification inspired from the traditional online FDD approach that could be implemented online directly. The time series analysis approach of Chapter 4 relies on the online classification of multiple alarm binary series, when traditional fault identification relies on the online classification of continuous process measurements. Since the alarm binary series are sparse and the exact dates of activations of the alarms are subject to randomness, Chapter 4 suggested to classify the alarm floods based on matrices quantifying the alarm coactivations (i.e. the number of samples where two alarms are active at the same time) on a given time interval defined by a sliding window in the online implementation.

Achievements and merits of the approach

Chapter 3 proposed a new structure of the field of alarm data analysis and highlighted the bias towards sequence mining methods for online operator support. The taxonomy with regard to the type of application and with regard to the alarm data format offers a new perspective on alarm data analysis methods. The taxonomy has been published as part of a journal paper [33] and is attracting attention. It was published in July 2019, and by October had become the fifth most downloaded paper of the previous 90 days in the Journal of Process Control.

Chapter 4 gives an example of time series analysis method for online alarm flood classification that can be applied online directly. By contrast with sequence mining methods for alarm flood classification, the method of Chapter 4 does not require additional computations to be implemented online. This computational simplicity facilitates the industrial applica-

tion of the method. The method has been tested on industrial data from several customers of ABB, including the oil and gas separation plant of Chapter 4. An offline implementation of the method to detect and classify alarm flood episodes from historical alarm records was integrated in ABB's Analytics platform, as described in an article published in ABB Review [205]. The integration of the online implementation of the method in ABB's platform faces challenges due to software architecture and data access that could not be treated within the time-frame of this project.

The method received attention from the research community with a keynote talk at ADCHEM 2018 [30] and an invited journal paper [33], and from the industry with several software demonstrations and the article in ABB Review [205].

Limitations of the approach

Some technical limitations are inherent to the method of Chapter 4, for example the number of historical alarm flood episodes required for training the classifier. The method of Chapter 4 belongs to the category of alarm flood classification methods that rely on statistical classification of numerical feature vectors, as presented in Sub-section 3.2.4. Those methods use statistical classifiers that learn from the feature vectors and that require a certain number of feature vectors per class for training. This can be a limiting factor for industrial applications if the number of alarm flood occurrences available for training is low. The industrial case study of Chapter 4 used a minimum of five alarm floods occurrences per class for training. A solution to overcome this issue when the number of alarm floods per class for training is low (e.g. one alarm flood per class) would be to substitute the support vector machine classifier of Chapter 4 with a simpler classifier that requires few training episodes such as a nearest neighbour classifier, even though it may decrease the accuracy of the classification.

More generally, online alarm data analysis methods designed for online operator support face challenges due to their practical implementation on site. Real-time alarm data analysis is a recent field of research and industrial alarm systems need to be prepared for it. Software architecture and data access have been mentioned as issues that prevented the implementation of the method on site. For this reason, tests have been done on online scenarios created from the alarm records of the industrial case studies, but the method could not be implemented on site in real-time data.

8.2.2 Objective 2: integration of alarm systems into online FDD

Objective 2 was to bring alarm systems and alarm management into online FDD to improve the robustness of data-driven online FDD methods with regard to the variability in the process measurements and in the operating conditions. This means integrating data (such as alarm records), contextual information (such as alarm settings), and practices (such as alarm processing techniques) into online FDD.

Description of the approach

This thesis demonstrated by means of two examples of how alarm systems and alarm management could be integrated into online FDD.

Chapter 5 gave an example of how contextual information available in the alarm master database could be used to improve the preprocessing of the process measurements. Chapter 5 suggests a normalization of the process measurements based on the corresponding alarm thresholds extracted from the alarm settings, the alarm-range normalization, instead of traditional standardization approaches.

Chapter 6 gave an example of how one type of alarm data, statistical alarms, and one type of alarm management practice, delay-timers, could be used to improve the robustness of the variable selection, fault detection and fault identification with regard to the variability in the process measurements. The statistical alarms help selecting the most relevant variables for online FDD during the modeling stage, and the statistical alarms are used in combination with the process measurements for fault detection and identification.

Achievements and merits of the approach

By contrast with existing normalization approaches, alarm-range normalization does not depend on historical data available. Finding a time interval where all process measurements present representative variations for the normalization can be difficult and time-consuming. Conversely, choosing an time interval where process measurements do not present the same level of variability can lead to spurious results in the fault detection and diagnosis as highlighted in Chapter 5. Alarm-range normalization can be used as a reference normalization regardless of the application (i.e. fault detection, fault isolation, or fault identification), since alarm-range normalization is independent from the choice of the training set. In addition, alarm-range normalization places variations of the process measurements in the context of

the safe operation of the plant. It brings contextual information into the data-driven methods and increases the reliability of FDD methods by reducing the impact of noisy variations.

Alarm-range normalization was tested on industrial data from the oil and gas separation plant, where alarm thresholds could be extracted from the alarm master database. The method was presented at SAFEPROCESS 2018 [31] and a journal paper was submitted [35].

Both the variable selection and the two-stage Bayesian fusion improve the robustness to noise of the FDD. The variable selection prioritizes process measurements with large variations compared to the level of noise to limit the impact of noise in the classification. The discrete Bayesian classifiers based on statistical alarms benefits from the effect of the delay-timers and limits the number of false detections and missed detections due to noise in the measurements. The variable selection method of Chapter 6 is simple and computationally efficient since the probability density functions can be computed as discrete sums. Therefore, the method can be applied to industrial systems with a large number of variables.

The variable selection approach of Chapter 6 was presented at DYCOPS 2019 [32] and the general framework for integrating statistical alarms in fault detection and identification, including the variable selection and the Bayesian fusion, was published in a journal article [34].

Both the method of Chapter 5 and the method of Chapter 6 require low effort for industrial implementation. Alarm thresholds are directly available in the master alarm database, and statistical alarms and delay-timers have been designed beforehand on the plant.

Limitations of the approach

Both the method of Chapter 5 and the method of Chapter 6 depend on the availability of data or information from the alarm systems. Alarm-range normalization is limited to variables for which alarm thresholds have been defined. This is not the case for variables without alarms, or for variables with rate-of-change alarms for example. If such variables should be integrated into the FDD, statistical alarm thresholds would need to be computed. The variable selection and the Bayesian classification of Chapter 6 depend on the design of the statistical alarms and on the tuning of the delay-timers. If such information is not readily available in the alarm system, it would have to be determined based on historical process measurements.

The problem of varying operating conditions and multimode operation has not been addressed in Chapter 5 and Chapter 6. In its current form, alarm-range normalization does not exploit information related to adaptive alarms where the threshold is modified to adapt to the process, including multimode alarming where various thresholds are defined to match the various operating regimes. An improved version of alarm-range normalization would need to take into account the modification of the alarm thresholds to adapt to the changing operating conditions. The Bayesian classification method is not able to distinguish faults and transitions between operating regimes. This is a research problem for FDD methods based on process measurements alone, but in this case the design of the statistical alarms may also change according to the operating regime. The Bayesian classification method should be able to recognize (or compensate for) transitions between operating regimes and identify the fault classes in the new operating regime in spite of the possibly modified patterns in the process measurements and in the statistical alarms.

8.2.3 Objective 3: RCA, from oscillatory to non-oscillatory disturbances

Objective 3 was to determine the field of applicability of existing data-driven root cause analysis methods to guide industrial practitioners and facilitate the application of those methods. Since most methods have been designed for oscillatory disturbances, this includes clarifying the procedure for RCA of oscillatory disturbances and proposing a procedure for RCA of non-oscillatory disturbances.

Description of the approach

Existing data-driven methods are reviewed in Chapter 7 according to three types of criteria: the properties of the method, the impact of the systems through which the disturbance propagates on the results of the method, and the types of disturbances covered by the method. The comparative review is used to derive guidelines for RCA of oscillatory and non-oscillatory disturbances in the industry. In particular, Chapter 7 suggests a new approach for RCA of transient disturbances based on methods for detection of transient disturbances.

Achievements and merits of the approach

Chapter 3 suggested a new structure of the field of data-driven RCA identifying main strategies and branches of methods. The structure is used as a basis for the comparative review, highlighting similarities among methods of the same branch. Chapter 7 is the first transversal comparative review of data-driven RCA methods in process literature. Several works proposed comparative analyses of a limited set of methods (usually two or three), or studied the properties or propagation effects on a limited set of methods. Chapter 7 used the results of those analyses and generalized the outcome. Chapter 7 presented a structured approach to the problem of RCA of non-oscillatory disturbances that has not been addressed in the literature. Chapter 7 highlighted the limitations of existing data-driven RCA methods for transient disturbances and extended the state of the art by suggesting an alternative to time-delay estimation methods.

The comparative review of Chapter 7 is targeting primarily engineers working on FDD in industry, for example in companies such as ABB. It is intended as a tool to increase the awareness and the understanding of data-driven RCA methods in the industry to promote their use. The focus on transient disturbances answers a need for methods to automate or facilitate RCA of transient disturbances in industrial processes or industrial machinery that was observed by the author during discussions with service engineers at ABB. The material of Chapter 7 will be made available as self-standing material in a journal article or technical report.

By contrast with existing methods such as time-delay estimation methods, the transient disturbance detection approach does not rely on the comparison of the shapes of the measurements. Therefore, the proposed approach is not affected by the distortion of the shape during the propagation, which was the main limitation of existing methods. The transient disturbance detection approach uses established methods from the field of transient disturbance detection. Chapter 7 demonstrated how a method for transient disturbance detection from the literature could directly be applied for RCA on an industrial glass forming process.

The transient disturbance detection approach of Chapter 7 is a practical suggestion to provide FDD engineers in industry with a solution for RCA of transient disturbances. The method also works on transient disturbances in oscillating measurements, which is important for applications such as RCA in electrolow-pass filters machinery.

Limitations of the approach

The comparative review of Chapter 7 only gives high-level indications on RCA methods and strategies but does not give a detailed procedure. It does not answer categorically the critical question for the industrial practitioner, i.e. whether the results of a given RCA method can be trusted on a given industrial case. Joint qualitative and quantitative analysis would be required to provide a clear answer to this question, which is a direction for future work discussed in Sub-section 8.3.3. In addition, transient disturbances can take a wide variety of shapes. The comparative review relies on generic properties of transient disturbances and does not discuss specific shapes. It could be interesting to constitute a taxonomy of the types of transient disturbances and discuss root cause analysis case by case.

The transient disturbance detection approach assumes that the variables studied are all affected by the same disturbance. A preliminary disturbance clustering stage is a traditional way to make sure that all variables are affected by the same disturbance. However, clustering of transient disturbances can be challenging due to the distortion of the shapes of the measurements during the propagation. Clustering of transient disturbances is discussed as future work in Sub-section 8.3.3.

8.2.4 Contribution to the main motivation of the thesis

Increasing the reliability of data-driven FDD methods to improve the applicability and the acceptance of FDD methods in the industry was the main motivation of this project. The first two objectives investigated the integration of alarm management and online FDD to increase the applicability of FDD methods. The contribution to Objective 1 demonstrated that the new field of online alarm data analysis for operator support could be treated as an online FDD problem. The proposed time series approach to online alarm flood classification could be implemented online while state of the art sequence mining methods rely on complex extensions of offline methods. The contribution to Objective 2 gave specific examples of the integration alarm data, alarm settings, and alarm practices into online FDD. The use of alarm thresholds reduced the chances of spurious FDD results due to statistical normalization and facilitated the work of FDD engineers with regard to the choice of the training data. The use of statistical alarms and delay-timers improved the robustness of variable selection, fault detection, and fault identification to noise and variability in the process measurements across fault occurrences. Both contributions to Objective 1 and Objective 2 represent practical methods that require limited implementation effort. The method of Chapter 4 has

been integrated into ABB’s Analytics platform, and the methods of Chapter 5 and Chapter 6 rely on the integration of data and information readily available in the alarm systems into widely-used FDD methods. All methods have been validated on industrial datasets.

Objective 3 aimed at increasing the reliability of data-driven RCA methods by clarifying the scope of application of the various methods. The state of the art of Chapter 3 and the comparative review of Chapter 7 constitute good material for introducing data-driven RCA methods to engineers in industry and for popularizing those methods. The disturbance detection approach to RCA of transient disturbances is a practical answer to a gap observed in industry, and can be applied for RCA in industrial processes or in electrolow-pass filters machinery.

8.3 Future work and future research directions

8.3.1 Future work for Objective 1

Recent research work on online alarm data analysis methods paves the way for more real-time support to operators based on alarms and events. However, selecting alarms and events to be included in the analysis is still an open problem, and deriving a method for alarms and events selection would be an important next step. Online alarm flood classification methods can take into account alarms, but also on other types of events. The number of alarms and events available on an process plant is very high, and the complexity of the classification must be limited by the constraints of the online implementation. The industrial case study in Chapter 4 only considers alarms with a high priority and trip events that indicate the trip of a piece of equipment. Low-priority alarms and other types of events have been discarded after discussion with the engineers. The selection of the alarms and events could be automated using data-driven methods (such as variable selection methods in FDD) and process knowledge (such as alarm priorities, event type or plant connectivity).

8.3.2 Future work for Objective 2

This thesis highlighted the benefits of integrating alarm systems and alarm management into FDD. Chapter 5 demonstrated the use of one type of alarm settings (alarm thresholds) for normalization, and Chapter 6 demonstrated the use of one type of alarm (statistical

alarms) and one type of practice (delay-timers) for variable selection and fault detection and identification. A natural way to extend the work would be:

- To integrate other types of alarm settings in FDD methods. For example, alarm priorities from the master alarm database could be used to weight the importance of each process measurement in the fault detection and diagnosis.
- To integrate other types of alarms in FDD methods. For example, FDD methods could use the conditions defined for adaptive alarms to adapt to the operating conditions of the plant. Generation mechanism of rate-of-change alarms could also be used to bring a temporal dimension in the FDD method without implementing dynamic methods that can be heavy and lack robustness on industrial systems.
- To integrate other types of practices in FDD methods. For example, alarm deadbands defined in the alarm system could be used to reduce the number of false detections.

8.3.3 Future work for Objective 3

The comparative review of data-driven RCA methods gave a qualitative outlook of the applicability of the various methods. The approach could be extended to provide a comprehensive decision-support tool for engineers in the industry. A first step to provide a comprehensive decision-support tool would be to combine the qualitative analysis started in the comparative review with quantitative analysis on the process measurements available to decide if the results of a given RCA method are reliable in a given industrial case. The qualitative analysis should investigate the system through which the disturbance propagates and the type of disturbance, as done in the comparative review. The quantitative analysis should consider the process measurements available. Significance testing is a first step towards the assessment of the reliability of the results, but several works have pointed out the limited reliability of existing significance tests.

Sub-section 8.2.3 underlined that the transient disturbance detection approach of Chapter 7 requires the variables studied to be all affected by the same disturbance. Developing a method for clustering of transient disturbances able to deal with the distortion of the shape of the measurements during the disturbance propagation would be a possible solution. Another solution would be to integrate the results of the disturbance detection approach with information regarding the plant connectivity. If the disturbance clustering is too challenging, information about plant connectivity can be used to determine if transient disturbances detected in several measurements correspond to the same transient disturbance propagating

in the process. By assessing the physical distance between the measurements affected by the disturbances and taking into account the time constants of the process, one can estimate if the disturbances are related or not. In the same perspective, the disturbance detection approach can be used to highlight the propagation path of a transient disturbance.

8.3.4 Future research directions in data-driven fault detection and diagnosis

This thesis advocated the integration of alarm systems and alarm management into FDD and this section identified ways to bring this integration further, by considering other types of events into online alarm data analysis methods or by integrating more data and contextual information from the alarm systems into FDD methods. More generally, the integration of data from multiple sources is a fruitful research direction for data-driven FDD that needs to be pursued, in particular when considering the role of FDD in dealing with unexpected events.

Chapter 2 mentioned process measurements, alarms and events, and plant connectivity as the key data sources for this project. The thesis gave evidence of the benefits of integrating process measurements and alarms and events in online FDD, and identified the integration of process measurements and plant connectivity as a next step for RCA of transient disturbances. The integration of plant connectivity with process measurements for FDD and the integration of plant connectivity with alarms and events for alarm management have already been investigated. All three sources should now be considered together to provide a more accurate picture of the status of the plant and increase the reliability of the fault diagnosis. Other sources of information such as the production planning containing the planned changes in operation regime should also be considered to anticipate transitions between regimes and adapt the FDD methods accordingly.

Autonomous process plants have been described as the next technological step in the digitalization of the process industry [206]. One major challenge to the fully autonomous plant is the risk of unexpected events for which the algorithms are not prepared, and which might have hazardous consequences. The role of the operators is critical in dealing with unexpected events, and FDD methods must become key actors during those events, as a decision-support tool for the operators or as decision support-tool for the algorithms of the autonomous plant. For this reason, FDD research should focus more on dealing with unseen events that cannot be learnt from historical data. In this case, other sources of contextual

information become critical and the consideration of all possible data sources in the plant is necessary. Traditional data sources such as process measurements and alarms and events should be combined with process knowledge and risk studies (e.g. hazard and operability study) to understand the possible consequences of a given plant status.

8.4 Summary of the chapter

This chapter summarized the contribution of the thesis with regard to the initial research objectives, highlighting the merits and limitations of the chosen approaches. Chapter 4 presented a binary series method for online classification of alarm floods that is adapted for online industrial implementations but that requires several occurrences of each class of alarm flood for training. Chapter 5 showed that a normalization of process measurements based on alarm thresholds improved the reliability of the FDD, although the normalization approach did not exploit fully the information available in the alarm settings. Chapter 6 proposed a variable selection method and a Bayesian classification method considering statistical alarms that increase the robustness of FDD to variability and noise in the process measurements. The comparative review of Chapter 7 provided a high-level qualitative decision-support tool for engineers dealing with data-driven RCA, but the decision-support tool should be complemented with quantitative analysis. Chapter 7 also introduced a new disturbance detection approach to RCA of transient disturbances that does not rely on the shape of process measurements, but the approach assumes that the process measurements are all affected by the same disturbance.

Future work has been identified based on the limitations of the proposed approaches. Online classification of alarm floods could benefit from more efficient classification methods that would focus on the relevant alarms and events and require few alarm floods for training. The integration of alarm systems into FDD could be taken further by looking at the advantages of considering other types of alarms, other types of alarm settings, or other types of alarm management practices. The choice of a data-driven method for RCA of a given process disturbance can be guided with qualitative recommendations, but more work needs to be done on quantifying the reliability of the results on a given dataset. RCA of transient disturbances should be investigated in more details, and the proposed disturbance detection approach to RCA of transient disturbances is only applicable if the detected disturbances are related with each other. Methods based on clustering of transient disturbances or methods based on process knowledge could be investigated.

Chapter 9

Conclusion

Fault Detection and Diagnosis (FDD) in industrial processes emerged from the process control community as a decision-support tool for operators during abnormal situations. Data-driven FDD methods, traditionally based on process measurements, are the fastest to set up but their implementation on site can be difficult due to availability and variability of the process measurements. The key motivation of this project was to increase the reliability of data-driven FDD methods to extend their applicability in the process industry.

Alarms and alarm systems, which belong to the process safety community, follow objectives similar to the ones of FDD since they indicate abnormal situations to the operators to trigger a timely response. Still, FDD and alarm systems have been considered separately until now. The first objective of this thesis was to evaluate an online FDD approach for the analysis of alarms in real-time to provide support to the operators during abnormal situations. The second objective of the thesis was to integrate data and contextual information from the alarm systems into traditional online FDD methods to increase the reliability of the methods.

Root Cause Analysis (RCA) is a key step of the offline fault diagnosis stage. Data-driven methods that automate the RCA based on process measurements exist, but those methods are not widely used in industry due to a lack of clarity regarding their conditions of application. The third objective of the thesis was to provide a decision-support tool for engineers dealing with RCA of process disturbances to promote the use of data-driven RCA methods in the industry, in particular for transient disturbances that are short and seldom repeat themselves.

9.1 Conclusion on Objective 1

This thesis suggested a new view on the field of alarm data analysis, distinguishing offline applications for traditional alarm management and recent online applications for operator support. The thesis also differentiated sequence mining methods that apply to alarm sequences and time series analysis methods that apply to alarm series. While sequence mining methods inherited from alarm management may be difficult to implement online, the time series approach to online alarm flood classification suggested in this thesis is inspired from online FDD approaches. The suggested approach was applied successfully on a real industrial example and was implemented in ABB's Analytics platform. Both the review of the field of alarm data analysis and the suggested alarm flood classification approach were received positively by the research community with a keynote talk and a widely-read journal article.

9.2 Conclusion on Objective 2

This thesis gave two practical examples where the integration of data and contextual information from the alarm systems improved the reliability of online FDD methods. Alarm thresholds extracted from the alarm settings were used as a reference for normalizing the process measurements, which reduced the chances of spurious FDD results due to miscalibrated statistical normalization. Statistical alarms with the corresponding delay-timers (that delay the activation of the alarms until the threshold has been crossed for a number of consecutive samples) were combined with process measurements to improve the robustness to noise of FDD methods. Both examples require low implementation effort since the information about alarm thresholds, statistical alarms, and delay-timers is readily available in the alarm systems, and the examples have been validated on two industrial cases. Practical suggestions for further integration with other types of alarms, other types of alarm settings, and other types of alarm management practices have also been mentioned.

9.3 Conclusion on Objective 3

The decision-support tool for engineers dealing with RCA of process disturbances takes the form of a qualitative comparative review of data-driven methods, clarifying the applicability of each method according to the properties of the system through which the disturbance

propagates and according to the properties of the disturbance. The thesis proposed guidelines for RCA of oscillatory and non-oscillatory disturbances, and suggested a new practical method for RCA of transient disturbances to answer the need of engineers in industry. The suggested method uses an existing transient disturbance detection approach and does not rely on the comparison of the shape of the measurements, which increases the robustness of the method to the distortion of the shape of the measurements as the disturbance propagates. The suggested method was validated on an example of industrial transient disturbance from the literature.

9.4 Final comments

The research reported in this thesis was conducted both in a university and in an industrial research centre, and bridging the gap between the state of the art in data-driven FDD literature and industrial implementation has been the first goal. All research objectives of the thesis aimed at increasing the applicability of data-driven FDD methods on site by improving their reliability while keeping the implementation effort low. It implied developing methods appropriate for online implementation, improving the robustness of existing methods to variability in the measurements, clarifying their field of application, and suggesting new practical methods for unsolved problems in industry. The writing of the thesis was done such that it brings useful insights to FDD engineers working in industry, with new perspectives on the state of the art in the literature and practical applications on real industrial cases.

The thesis identified the integration of alarm management and FDD as a promising research direction to improve reliability of data-driven methods and gave practical examples to support this point. This research direction has been received positively by both the alarm management and the FDD research communities, and has attracted interest from the industry. In that sense, the thesis not only contributed to bridging the gap between FDD in literature and FDD in industry, but also contributed to extending the state of the art in both academic and industrial FDD.

Appendix A

Variable selection in the multiphase flow facility

This appendix provides complementary material regarding variable selection in the industrial case study of Chapter 6.

Figure A.1 shows the standardized process measurements of the first ten ranked variables with mRMR on level-2 alarms in the training set for scenario BA. Figure A.2 shows the standardized process measurements of the first ten ranked variables with mRMR on process measurements in the training set for the same scenario.

Table A.1 and Table A.2 describe the process variables included in the sets of variables $\mathcal{X}_M$ (with M from 1 to 10) respectively in scenario AB and scenario BA.

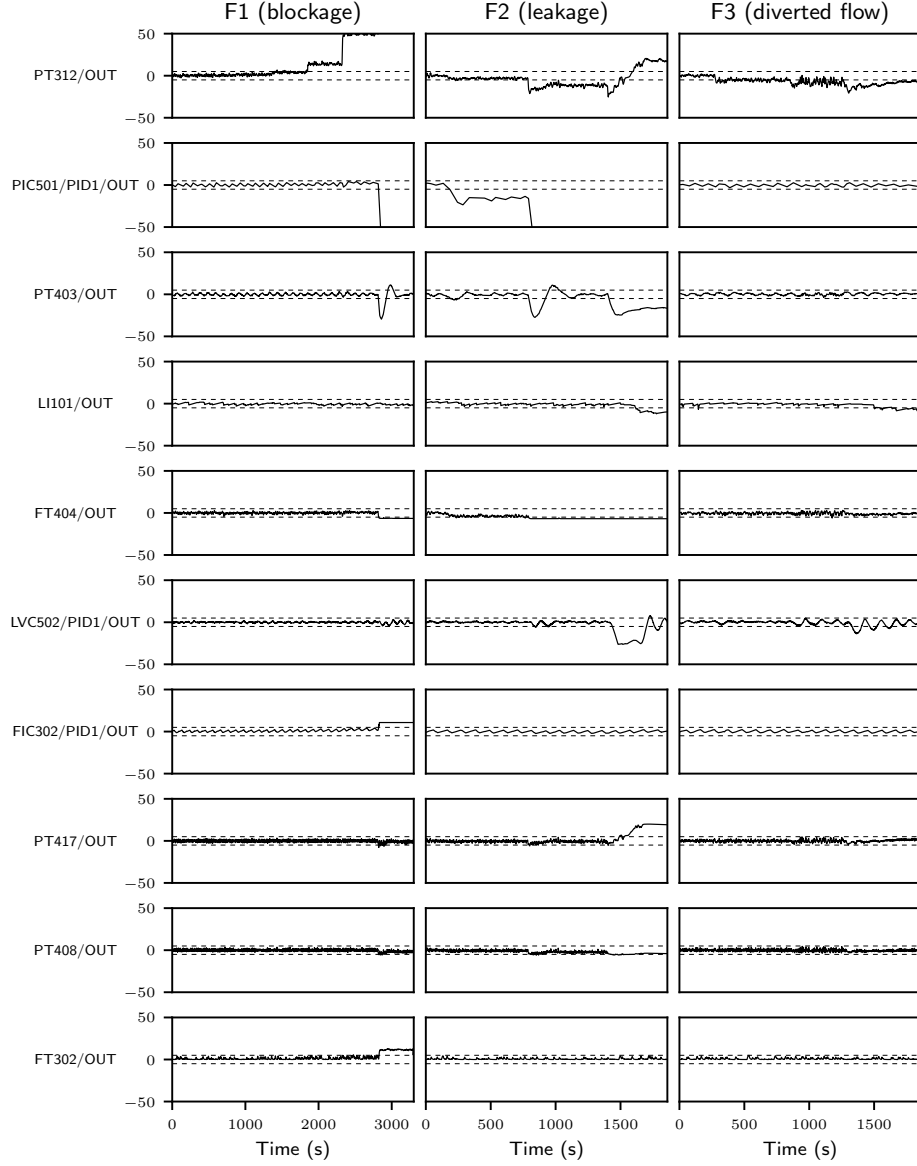


Figure A.1: Standardized process measurements of the first ten ranked variables with mRMR on level-2 alarms in the training set for scenario BA. The variables are ranked from top to bottom, with PT312/OUT being ranked as the most relevant variable. Each column corresponds to a fault occurrence in the training set. The dotted lines indicate the level-2 alarm thresholds ($\bar{k} = 5$).

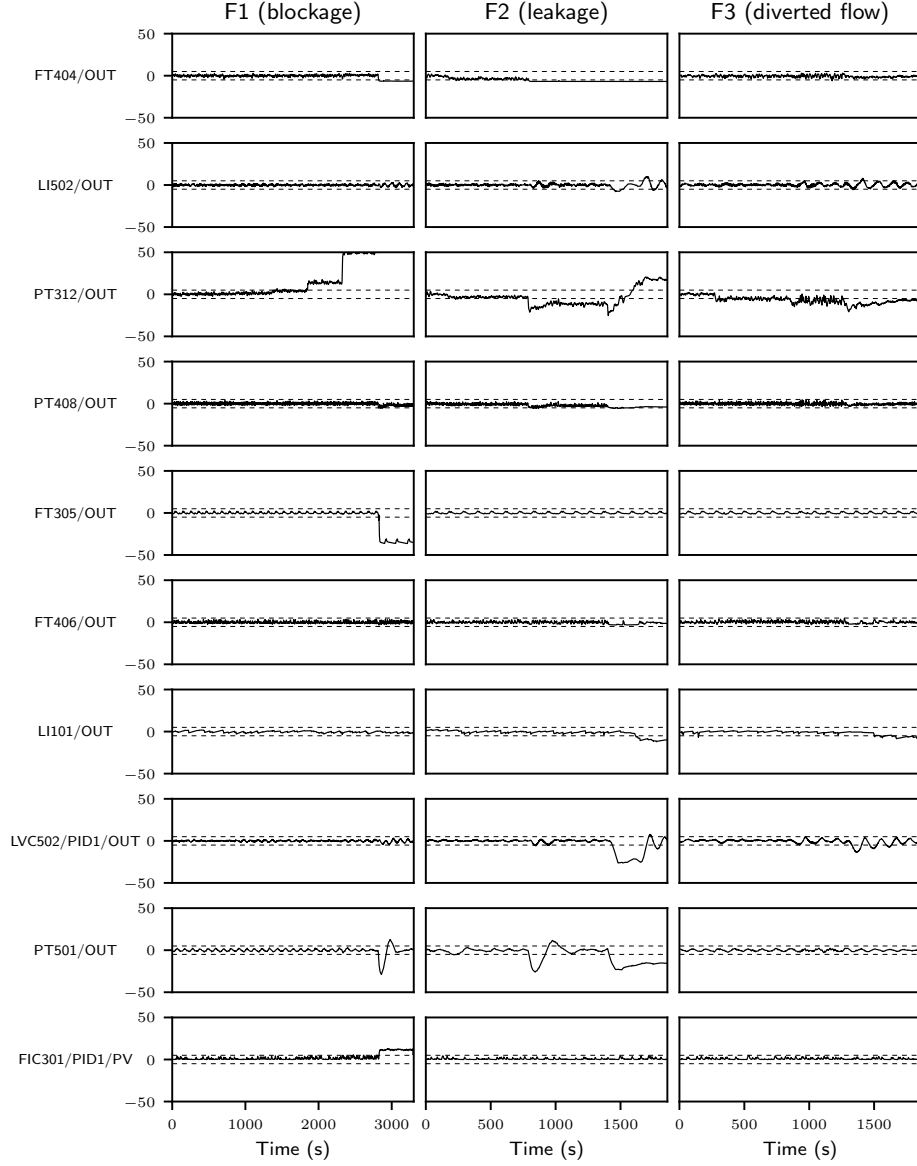


Figure A.2: Standardized process measurements of the first ten ranked variables with mRMR on process measurements in the training set for scenario BA. The variables are ranked from top to bottom, with FT404/OUT being ranked as the most relevant variable. Each column corresponds to a fault occurrence in the training set. The dotted lines indicate the level-2 alarm thresholds ($\bar{k} = 5$).

Table A.1: Sets of selected variables $\mathcal{X}_M$ with mRMR on level-2 alarms and with mRMR on process measurements in scenario AB, with M from 1 to 10. The bullet points indicate the variables that are included in each set $\mathcal{X}_M$.

	mRMR on level-2 alarms										mRMR on process measurements									
	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$	$\mathcal{X}_5$	$\mathcal{X}_6$	$\mathcal{X}_7$	$\mathcal{X}_8$	$\mathcal{X}_9$	$\mathcal{X}_{10}$	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$	$\mathcal{X}_5$	$\mathcal{X}_6$	$\mathcal{X}_7$	$\mathcal{X}_8$	$\mathcal{X}_9$	$\mathcal{X}_{10}$
FT305/OUT												•	•	•	•	•	•	•	•	•
FT302/OUT																			•	•
PT312/OUT	•	•	•	•	•	•	•	•	•	•			•	•	•	•	•	•	•	•
PT417/OUT			•		•	•	•	•	•	•										
PT408/OUT					•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PT403/OUT								•	•	•										•
FT404/OUT																•	•	•	•	•
FT406/OUT			•	•	•	•	•	•	•	•				•	•	•	•	•	•	•
PT501/OUT										•										
PIC501/PID1/OUT		•	•	•	•	•	•	•	•	•								•	•	•
LI502/OUT																				
LI503/OUT																		•	•	•
LVC502/PID1/OUT															•	•	•	•	•	•
LI101/OUT							•	•	•	•										
FIC302/PID1/OUT									•	•										
FIC302/PID1/PV					•	•	•	•	•	•										
FIC301/PID1/PV																				

Table A.2: Sets of selected variables $\mathcal{K}_M$ with mRMR on level-2 alarms and with mRMR on process measurements in scenario BA, with M from 1 to 10. The bullet points indicate the variables that are included in each set $\mathcal{K}_M$.

	mRMR on level-2 alarms										mRMR on process measurements									
	$\mathcal{K}_1$	$\mathcal{K}_2$	$\mathcal{K}_3$	$\mathcal{K}_4$	$\mathcal{K}_5$	$\mathcal{K}_6$	$\mathcal{K}_7$	$\mathcal{K}_8$	$\mathcal{K}_9$	$\mathcal{K}_{10}$	$\mathcal{K}_1$	$\mathcal{K}_2$	$\mathcal{K}_3$	$\mathcal{K}_4$	$\mathcal{K}_5$	$\mathcal{K}_6$	$\mathcal{K}_7$	$\mathcal{K}_8$	$\mathcal{K}_9$	$\mathcal{K}_{10}$
FT305/OUT																				
FT302/OUT										•										
PT312/OUT	•	•	•	•	•	•	•	•	•	•			•	•	•	•	•	•	•	•
PT417/OUT								•	•	•										
PT408/OUT									•	•			•	•	•	•	•	•	•	•
PT403/OUT			•	•	•	•	•	•	•	•										
FT404/OUT					•	•	•	•	•	•	•									
FT406/OUT																•	•	•	•	•
PT501/OUT																			•	•
PIC501/PID1/OUT		•	•	•	•	•	•	•	•	•										
LI502/OUT											•	•	•	•	•	•	•	•	•	•
LI503/OUT																				
LVC502/PID1/OUT						•	•	•	•	•								•	•	•
LH101/OUT				•	•	•	•	•	•	•							•	•	•	•
FIC302/PID1/OUT							•	•	•	•										
FIC302/PID1/PV																				
FIC301/PID1/PV																				•

Appendix B

Review of existing root cause analysis methods

B.1 Review of the properties of the methods

B.1.1 Univariate methods

Nonlinearity index

- *Computational load*: The computational load for calculating nonlinearity index is low but the load is medium if surrogate testing is included.
- *Parameters*: The parameters of nonlinearity index are the prediction horizon h , the number of nearest neighbours K , and the embedding dimension L . Methods to tune the parameters are provided in [147], so complexity of tuning is moderate.
- *Significance test*: The significance threshold is a three-sigma threshold determined based on nonlinearity index values on surrogates generated using the Adjusted Amplitude Fourier Transform (AAFT) [147, 155]. This is the same method as the one used for the significance test of nearest neighbours, which has a low reliability according to [168].
- *Multivariate and direct causality*: Nonlinearity index is a univariate method only, direct causality cannot be assessed.
- *Data requirements*: Reference [147] recommends a minimum of 12 cycles with 25 samples per cycle. This corresponds to a moderate number of samples (comparable to the number of samples required by nearest neighbours) and to a moderate excitation of the measurements (with regard to the number of times the disturbance should repeat itself).

B.1.2 Time-delay estimation methods

Cross-correlation function

- *Computational load*: The computational load for calculating cross-correlation function is low since it relies on the linear correlation between two measurements.
- *Parameters*: Cross-correlation function does not have any parameter.
- *Significance test*: Significance thresholds for the correlation index and the directionality index are expressed as functions of N (the number of samples) based on heuristic tests on random sequences with uniform and normal distributions [149]. In case of oscillations, the directionality test failed (since the minimum and the maximum of the cross-correlation function were of similar magnitude) and only the significance threshold for the correlation index was used [150].
- *Multivariate and direct causality*: Cross-correlation function is a bivariate method only, direct causality cannot be assessed.
- *Data requirements*: The authors of [149] recommend a minimum value of N of 50 samples or four times the time-delay, so the number of samples required to apply cross-correlation function is low. There is no constraint regarding the excitation of the measurements.

Dynamic time warping causality index

- *Computational load*: The computational load of dynamic time warping causality index is medium since it requires the computation of the dynamic time warping distance between each pair of variable for multiple time lags.
- *Parameters*: Dynamic time warping causality index does not have any parameter.
- *Significance test*: Dynamic time warping causality index is a normalized index and a significance threshold is proposed arbitrarily as 0.99 [153], so the significance test has a limited reliability.
- *Multivariate and direct causality*: Dynamic time warping causality index is a bivariate method only, direct causality cannot be assessed.
- *Data requirements*: As other time-delay estimation methods, the number of samples required by dynamic time warping causality index is low, and there is no constraint regarding the excitation of the measurements.

k nearest neighbours

- *Computational load*: The computational load of kNN is moderate [154].
- *Parameters*: kNN has one parameter, the number of nearest neighbours K , and [154] provides a method for tuning of K .
- *Significance test*: The significance level is determined based on the distribution (assumed Gaussian) of a measure defined as the quality of the results [154].

- *Multivariate and direct causality:* kNN is a bivariate method only, direct causality cannot be assessed.
- *Data requirements:* As other time-delay estimation methods, the number of samples required by kNN is low. If the disturbance does not repeat itself, kNN can be used with $K = 1$, so the constraints on the repeatability of the disturbance in the measurements are not as high as for nearest neighbours or nonlinearity index.

B.1.3 Prediction-based methods in the time-domain

Nearest neighbours

- *Computational load:* Nearest neighbours has a moderate computational load according to [168].
- *Parameters:* The parameters of nearest neighbours are the prediction horizon h , the number of nearest neighbours K , and the embedding dimension L . Complexity of tuning is moderate.
- *Significance test:* The significance threshold is a three-sigma threshold determined based on values of the nearest neighbours causality measure on surrogates generated using AAFT [147, 155]. The authors of [168] pointed out the low reliability of this significance test.
- *Multivariate and direct causality:* The initial version of nearest neighbours is not multivariate but [157] extends nearest neighbours to a multivariate directionality measure taking into account the effect of X on Y via m intermediate variables. Reference [157] uses the multivariate extension of nearest neighbours to exclude indirect dependencies.
- *Data requirements:* The authors of [155] recommend a minimum of 400 samples, so the minimum number of samples is moderate. As for nonlinearity index, nearest neighbours requires a minimum number of repetitions of the disturbance for the self-predictability to be reliable, so the minimum excitation of the measurements is set to moderate.

Convergent cross mapping

- *Computational load:* The computational load of convergent cross mapping is moderate [160, 207].
- *Parameters:* The parameters of convergent cross mapping are the embedding dimension L and the time-lags τ [159, 160]. Methods have been proposed for tuning such as the automatic determination of L [159, 160] and the autocorrelation function for τ [160].
- *Significance test:* In [159], the significance test is a Z -test with Fisher's z -transformation and a confidence level of 0.01. Reference [161] uses Monte Carlo simulations to determine a significance threshold.

- *Multivariate and direct causality:* Convergent cross mapping is a bivariate method only, direct causality cannot be assessed.
- *Data requirements:* Convergent cross mapping can deal with short time series according to [158], and [159] recommends a minimum of 100 samples. Convergent cross mapping is a prediction-based methods relying on nearest-points as nearest neighbours, so the minimum excitation of the measurements is moderate.

Granger causality

- *Computational load:* The computational load of the autoregressive modeling is low according to [168].
- *Parameters:* Granger causality has one parameter, the model order R , which can easily be tuned using Akaike information criterion or Bayesian information criterion [207].
- *Significance test:* The significance test is a F -statistical test since the Granger causality measure is F -distributed [162, 207]. The fit of the model also needs to be considered (e.g. using the consistency criterion [167]) since Granger causality is only reliable if the model quality is good enough (i.e. consistency greater than 80% [167]).
- *Multivariate and direct causality:* Multivariate regression can be used and direct causality can be assessed using the multivariate formulation [167].
- *Data requirements:* A large number of samples is required for the modeling according to [64], mentioning a minimum of 2000 samples. The modeling also requires a high excitation of the measurements [64].

Causal relationship based on residual analysis

- *Computational load:* As for Granger causality, the computational load of the autoregressive modeling is low.
- *Parameters:* Causal relationship based on residual analysis has one parameter, the model order R , which is increased iteratively in the method proposed by [167]. Therefore, parameter tuning is easy.
- *Significance test:* The significance test is based on two thresholds from system identification theory [167]: the confidence interval for the whiteness test of the residuals (Sub-section 3.4.5) is based on the chi-squared distribution, and the confidence interval for the independence test of the residual and the input (Sub-section 3.4.5) is based on rules of thumb from system identification literature. The whiteness test of the residuals and the independence test of the residual and the input also serve as indicators for the fit of the model [167].
- *Multivariate and direct causality:* System identification methods are multivariate. Causal relationship based on residual analysis is used in the bivariate case, except to investigate direct causality [167].
- *Data requirements:* As for Granger causality, a large number of samples and a high excitation of the measurements are required.

Nonparametric multivariate regression

- *Computational load*: The computational load of nonparametric multivariate regression is moderate according to [168].
- *Parameters*: The parameters of nonparametric multivariate regression are the kernel tolerance vector $\mathbf{b}$, the time-delay vector τ , and the embedding dimension L [168]. Reference [168] proposed methods to tune the parameters but tuning is complex, in particular for $\mathbf{b}$. τ can be estimated using the autocorrelation function, L can be estimated using the method proposed by [208], and $\mathbf{b}$ is selected for each pair of variables based on the model fit.
- *Significance test*: The significance threshold is a statistical threshold determined based on values of the nonparametric multivariate regression causality measure on surrogates generated using AAFT [168]. This is the same method as the one used for the significance test of nearest neighbours, which has a low reliability according to [168]. The fit of the model is assessed using the residual sum of squares and the total sum of squares.
- *Multivariate and direct causality*: Nonparametric multivariate regression is multivariate by default, so direct causality can be assessed [168].
- *Data requirements*: As for other regression-based methods, a large number of samples and a high excitation of the measurements are required.

Transfer entropy

- *Computational load*: The computational load of transfer entropy is high according to [168].
- *Parameters*: The parameters of transfer entropy are the prediction horizon h , the sampling constant of the embedded vectors τ , and the embedding dimensions L and L' . Tuning of the parameters is complex [172], and several methods have been proposed: [84] proposed default values for h , τ , L and L' , and [172] completed the methodology with heuristic rules to tune those parameters. The authors of [170] used lagged mutual information to determine h and τ .
- *Significance test*: The significance threshold is a six-sigma threshold determined based on values of the transfer entropy causality measure on surrogates generated using iterative AAFT [84]. The authors of [168] pointed out the low reliability of this significance test.
- *Multivariate and direct causality*: Transfer entropy is bivariate but direct transfer entropy has been proposed to take into account intermediate variables and assess direct causality [177].
- *Data requirements*: Transfer entropy requires a high number of samples, and reference [84] recommends a minimum of 2000 samples. The measurements should be sufficiently excited for the estimation of the probability density functions to be reliable.

B.1.4 Prediction-based methods in the frequency-domain

Directed transfer function

- *Computational load*: The computational load of the autoregressive modeling is low, but the computational load of directed transfer function is high if surrogate testing is included.
- *Parameters*: Directed transfer function has one parameter, the model order R , which can easily be tuned using Akaike information criterion or Bayesian information criterion [179, 207].
- *Significance test*: The significance threshold is a statistical threshold determined based on values of the directed transfer function causality measure on surrogates. Reference [179] uses a 95th-percentile threshold based on values of the directed transfer function causality measure on surrogates generated with the full Causal Fourier Transform (CFT). Reference [209] uses a three-sigma threshold and generates surrogates using iterative AAFT. As for other significance tests based on surrogates, the reliability is low [168].
- *Multivariate and direct causality*: The autoregressive modeling is multivariate by default [209] but directed transfer function does not distinguish direct and indirect causality.
- *Data requirements*: As for Granger causality, a large number of samples and a high excitation of the measurements are required.

Partial directed coherence

- *Computational load*: The computational load of the autoregressive modeling is low, but the computational load of partial directed coherence is high if surrogate testing is included.
- *Parameters*: Partial directed coherence has one parameter, the model order R , which can easily be tuned using Akaike information criterion or Bayesian information criterion [162, 179, 207].
- *Significance test*: The significance threshold is a statistical threshold determined based on values of the partial directed coherence causality measure on surrogates. Reference [179] uses a 95th-percentile threshold based on values of the partial directed coherence causality measure on surrogates generated with full CFT. As for other significance tests based on surrogates, the reliability is low [168].
- *Multivariate and direct causality*: The autoregressive modeling is multivariate by default and partial directed coherence is able to detect direct causality [179].
- *Data requirements*: As for Granger causality, a large number of samples and a high excitation of the measurements are required.

B.2 Review of the impact of the systems' properties on the results of the methods

B.2.1 Univariate methods

Nonlinearity index

- *Noise*: As for nearest neighbours, the effect of sensor noise and process noise on the results of nonlinearity index is high because it has a direct impact on the shape of the embedded vectors and thus on the estimation of the nearest-neighbours.
- *Dead-time*: Dead-time does not have any effect on the results of the method since nonlinearity index is univariate.
- *Low-pass filter*: Nonlinearity index assumes that the disturbance propagates through low-pass filters that attenuate the harmonics of the measurements. Therefore, the effect of low-pass filters is not only acceptable but required for nonlinearity index to be applicable.
- *Nonlinearity*: Nonlinearities induced by the systems during the propagation of the disturbance have an impact on the values of nonlinearity index and can lead to spurious results for RCA [148]. Therefore, nonlinearity index is not recommended when soft or hard nonlinearities exist during the propagation.

B.2.2 Time-delay estimation methods

Cross-correlation function

- *Noise*: Reference [48] highlighted the high effect of noise on the results of cross-correlation function, compared to nearest neighbours or transfer entropy.
- *Dead-time*: Dead-time is not only acceptable but required for time-delay estimation methods to be effective, including cross-correlation function.
- *Low-pass filter*: Reference [48] highlighted the high effect of low-pass filters on the results of cross-correlation function, compared to nearest neighbours or transfer entropy.
- *Nonlinearity*: Cross-correlation function computes the linear correlation, so the method is not applicable when nonlinearities appear in the propagation. The authors of [183] demonstrated that cross-correlation function is not able to detect causality in presence of a squaring function, which is a soft nonlinearity. Therefore, cross-correlation function is not adapted in case of soft or hard nonlinearities.

Dynamic time warping causality index

- *Noise*: Reference [153] pointed out the high effect of noise on the results of dynamic time warping causality index, since de-noising is recommended beforehand.

- *Dead-time*: Dead-time is not only acceptable but required for time-delay estimation methods to be effective, including dynamic time warping causality index.
- *Low-pass filter*: Dynamic time warping causality index can deal with the effect of low-pass filters during the propagation. Nevertheless, dynamic time warping causality index relies on dynamic time warping alignment of the measurements which is based on their shapes, and the shapes are smoothed by the low-pass filtering. The efficiency of dynamic time warping causality index is impacted.
- *Nonlinearity*: According to [153], dynamic time warping causality index is not applicable in case of hard nonlinearity since the shape of the disturbance is not conserved during the propagation. Dynamic time warping is not able to align measurements with different shapes.

k nearest neighbours

- *Noise*: As for nearest neighbours, process noise and sensor noise have a high effect on the results of kNN.
- *Dead-time*: Dead-time is not only acceptable but required for time-delay estimation methods to be effective, including kNN.
- *Low-pass filter*: kNN can deal with the effect of low-pass filters during the propagation. Nevertheless, kNN relies on computation of nearest-neighbours based on the shapes of the measurements, and the shapes are smoothed by the low-pass filtering. The efficiency of kNN is impacted.
- *Nonlinearity*: kNN relies on the same nonlinear prediction approach as nearest neighbours, but a nonlinear relation between the two measurements can make the exact time-delay estimation more difficult.

B.2.3 Prediction-based methods in the time-domain

Nearest neighbours

- *Noise*: Reference [48] highlighted the high effect of process noise and sensor noise on the results of nearest neighbours. The noise has a direct impact on the shape of the embedded vectors, and thus on the estimation of the nearest-neighbours.
- *Dead-time*: The dead-time has no effect as long as τ and h are tuned correctly [48], but tuning can be difficult.
- *Low-pass filter*: According to [48], nearest neighbours is better at dealing with low-pass filters than cross-correlation function but not as good as transfer entropy, because the shapes of the embedded vectors are smoothed by the low-pass filtering.
- *Nonlinearity*: According to [157, 168], nearest neighbours can be applied in case of both soft and hard nonlinearities.

Convergent cross mapping

- *Noise*: As for other nearest-points methods, results of convergent cross mapping are affected by process and sensor noise, and [160] mentions failed causality detection because of the noise level in the measurements. Still, convergent cross mapping has been applied on noisy time series in the non-process literature [210].
- *Dead-time*: The dead-time has no effect as long as the time-lag parameter τ is set correctly [160].
- *Low-pass filter*: Convergent cross mapping is a method based on nearest-points, as nearest neighbours, so it is compatible with low-pass filters.
- *Nonlinearity*: According to [160], convergent cross mapping can deal with soft and hard nonlinearities.

Granger causality

- *Noise*: Reference [183] highlighted the low effect of process noise and sensor noise on the results of Granger causality, since overfitting is limited in the autoregressive model.
- *Dead-time*: Short dead-times can be automatically handled by choosing a relevant model order R [183]. For long dead-times, a time-delay parameter should be introduced.
- *Low-pass filter*: Reference [183] highlighted the ability of Granger causality to deal with low-pass filters.
- *Nonlinearity*: The basic version of Granger causality relies on a linear autoregressive model, which is not adapted in presence of nonlinearities. The authors of [183] highlighted that Granger causality was unable to detect causality in presence of a squaring function, so Granger causality is expected to have a low efficiency in presence of soft nonlinearities, and not be applicable in presence of hard nonlinearities. In both cases, nonlinear extensions of Granger causality based for example on nonlinear autoregressive models proposed in the literature and should be used.

Causal relationship based on residual analysis

- *Noise*: Causal relationship based on residual analysis relies on the same type of model as Granger causality, so the effect of process noise and sensor noise on the results of the method are low, as for Granger causality.
- *Dead-time*: As for Granger causality, short dead-times can be dealt with by choosing an appropriate model order, but a time-delay parameter should be introduced for long dead-times.
- *Low-pass filter*: As for Granger causality, causal relationship based on residual analysis is able to deal with low-pass filters.

- *Nonlinearity*: As for Granger causality, causal relationship based on residual analysis relies on a linear autoregressive model, so causal relationship based on residual analysis is not adapted in presence nonlinearities [167].

Nonparametric multivariate regression

- *Noise*: The effect of noise is low since overfitting in the model is addressed using leave-one-out cross-validation [168].
- *Dead-time*: Dead-times do not affect the results of nonparametric multivariate regression as long as the time-lag parameter τ is set correctly [168].
- *Low-pass filter*: As other regression-based approaches, nonparametric multivariate regression can deal with low-pass filtering.
- *Nonlinearity*: According to [168], nonparametric multivariate regression can be applied to both linear and nonlinear systems without restriction on the type of nonlinearity.

Transfer entropy

- *Noise*: Reference [183] highlighted the moderate effect of process noise and sensor noise on the results of transfer entropy. Noise has a direct impact on the estimation of probability density function, although the sensibility of the estimation method (e.g. histogram or kernel density estimation) can be tuned to deal with noise.
- *Dead-time*: Dead-times do not affect the results of transfer entropy as long as h and τ are set correctly [48, 183]. However, h and τ have a strong impact on the probability density estimations, so tuning can be difficult.
- *Low-pass filter*: Low-pass filters have a small impact on the results of transfer entropy according to [48].
- *Nonlinearity*: Transfer entropy can be applied in nonlinear cases [84].

B.2.4 Prediction-based methods in the frequency-domain

Directed transfer function

- *Noise*: As for Granger causality, the effect of sensor and process noise on the autoregressive modeling is low.
- *Dead-time*: As for Granger causality, short dead-times can be dealt with by choosing an appropriate model order, but a time-delay parameter should be introduced for long dead-times.
- *Low-pass filter*: As for Granger causality, the autoregressive model can deal with low-pass filters. However, the cut-off frequency of the filter has an impact on the causality analysis in the frequency-domain since the filter cuts all frequencies above the cut-off frequency. No causal dependency should appear above the cut-off frequency.

- *Nonlinearity*: As for Granger causality, the linear autoregressive modeling approach is not applicable in presence of nonlinearities and nonlinear extensions of Granger causality in the frequency-domain should be used.

Partial directed coherence

- *Noise*: As for Granger causality, the effect of sensor and process noise on the autoregressive modeling is low.
- *Dead-time*: As for Granger causality, short dead-times can be dealt with by choosing an appropriate model order, but a time-delay parameter should be introduced for long dead-times.
- *Low-pass filter*: As for Granger causality, the autoregressive model can deal with low-pass filters. However, the cut-off frequency of the filter has an impact on the causality analysis in the frequency-domain since the filter cuts all frequencies above the cut-off frequency. No causal dependency should appear above the cut-off frequency.
- *Nonlinearity*: As for Granger causality, the linear autoregressive modeling approach is not applicable in presence of nonlinearities and nonlinear extensions of Granger causality in the frequency-domain should be used.

B.3 Review of the types of disturbances covered by the methods

B.3.1 Univariate methods

Nonlinearity index

- *Oscillating disturbance*: Nonlinearity index is not applicable for linear oscillations since the method is not applicable on linear time series. Nonlinearity index is used for RCA of nonlinear oscillations in [146, 147, 148], although [148] pointed out the limited reliability of nonlinearity index due to possible external sources of nonlinearity.
- *Non-oscillatory persistent disturbance*: The principle of self-predictability is used in [155] for RCA of irregular persistent spikes with nearest neighbours, so nonlinearity index is also applicable for RCA of irregular persistent spikes.
- *Transient disturbance*: Nonlinearity index is not applicable on transient disturbances since transient disturbances do not repeat themselves, so nonlinearity index cannot use past embedded vectors of X to predict future values of X (principle of self-predictability).

B.3.2 Time-delay estimation methods

Cross-correlation function

- *Oscillating disturbance*: According to [149] that tests cross-correlation function on both linear and nonlinear oscillations, cross-correlation function has difficulties to estimate the time-delay on periodic disturbances, in particular for linear oscillations.
- *Non-oscillatory persistent disturbance*: Cross-correlation function has been applied for RCA of irregular persistent spikes in [48] with better performance than on oscillations, although cross-correlation function misses certain causal dependencies due to the distortion of the shape of the measurements during the disturbance propagation.
- *Transient disturbance*: Cross-correlation function is applicable on transient disturbances as other time-delay estimation methods. Reference [183] point out that cross-correlation function is more reliable than Granger causality and transfer entropy for RCA of a transient weld disturbance in a glass forming process.

Dynamic time warping causality index

- *Oscillating disturbance*: As other time-delay estimation methods, dynamic time warping causality index can be applied for RCA of oscillating disturbances but the time-delay estimation is hard due to the periodic nature of the disturbance.
- *Non-oscillatory persistent disturbance*: As for cross-correlation function, dynamic time warping causality index is applicable on irregular persistent spikes with better performance than on oscillations.
- *Transient disturbance*: As other time-delay estimation methods, dynamic time warping causality index is applicable on transient disturbances. In particular, dynamic time warping causality index was introduced in [153] for RCA of a step disturbance.

k nearest neighbours

- *Oscillating disturbance*: As other time-delay estimation methods, kNN can be applied for RCA of oscillating disturbances but the time-delay estimation is hard due to the periodic nature of the disturbance. kNN is applied on oscillations in [154].
- *Non-oscillatory persistent disturbance*: As cross-correlation function, kNN is applicable on irregular persistent spikes with better performance than on oscillations.
- *Transient disturbance*: As other time-delay estimation methods, kNN is applicable on transient disturbances. However, in this case, the number of nearest-neighbours K should be set to one since the disturbance does not repeat itself, which means that only one embedded vector of X can be used to predict future values of Y .

B.3.3 Prediction-based methods in the time-domain

Nearest neighbours

- *Oscillating disturbance*: Reference [48] highlighted that nearest neighbours was applicable on linear oscillations, with medium efficiency. Reference [155] demonstrated that nearest neighbours was performing better than transfer entropy and cross-correlation function for RCA of nonlinear oscillations because the phase-coupling which occurs in nonlinear measurements is best utilized with nearest neighbours.
- *Non-oscillatory persistent disturbance*: Reference [155] used nearest neighbours for RCA of irregular persistent spikes, with moderate efficiency (compared to TE) since nearest neighbours did not detect the causal dependencies far from the root cause. This is due to the distortion of the shape of the measurements during the disturbance propagation. The authors of [156] also applied nearest neighbours on persistent disturbances in electrolow-pass filters systems.
- *Transient disturbance*: As for nonlinearity index, nearest neighbours relies on the principle of self-predictability that uses past embedded vectors of X to predict future values X . This approach requires the disturbance to repeat itself, and transient disturbances seldom repeat themselves.

Convergent cross mapping

- *Oscillating disturbance*: Convergent cross mapping is applicable for both linear and nonlinear oscillations [159]. Convergent cross mapping can be compared to nearest neighbours in terms of efficiency since both methods rely on nearest-points, which means that convergent cross mapping is performing well for RCA of nonlinear oscillations in particular, where phase-coupling is important.
- *Non-oscillatory persistent disturbance*: As for nearest neighbours, convergent cross mapping is applicable on irregular persistent spikes but is affected by the distortion of the shape of the measurements during the propagation of the disturbance.
- *Transient disturbance*: As for nearest neighbours, convergent cross mapping can also be applied for RCA of transient disturbances but the performance of the method is limited by the fact that transient disturbances seldom repeat themselves.

Granger causality

- *Oscillating disturbance*: Granger causality is used for RCA of linear oscillations in [181] and for RCA of nonlinear oscillations in [179]. Reference [168] highlighted that the accuracy of the results of the Granger causality combined with directed transfer function and partial directed coherence was high RCA of oscillations.
- *Non-oscillatory persistent disturbance*: Regression-based methods in the time-domain such as Granger causality can be applied for RCA of irregular persistent spikes as they can be applied for RCA of oscillating disturbances. The performance of the method is

similar on irregular persistent spikes and on oscillations since the modeling is facilitated by the repeatability of the disturbance but the model does not exploit the periodicity of the measurements. However, in the case of irregular persistent spikes, the results of Granger causality cannot be refined with an analysis in the frequency-domain since the spectrum of the measurements with irregular persistent spikes is not localized.

- *Transient disturbance*: Reference [181] demonstrated that Granger causality was reliable for RCA of step disturbances of all amplitudes, performing better than transfer entropy. Reference [183] highlights the good performance of Granger causality for RCA of a transient weld disturbance in a glass forming process.

Causal relationship based on residual analysis

- *Oscillating disturbance*: Causal relationship based on residual analysis is applicable for linear oscillations and nonlinear oscillations [167].
- *Non-oscillatory persistent disturbance*: As for Granger causality, causal relationship based on residual analysis can be applied with moderate efficiency for RCA of irregular persistent spikes.
- *Transient disturbance*: As for Granger causality, causal relationship based on residual analysis can be applied with moderate efficiency for RCA of transient disturbances.

Nonparametric multivariate regression

- *Oscillating disturbance*: Nonparametric multivariate regression is applicable for linear oscillations and nonlinear oscillations, performing well, in particular for nonlinear oscillations [168].
- *Non-oscillatory persistent disturbance*: Regression-based methods in the time-domain such as nonparametric multivariate regression can be applied for RCA of irregular persistent spikes as they can be applied for RCA of oscillating disturbances. The performance of the method is similar on irregular persistent spikes and on oscillations since the modeling is facilitated by the repeatability of the disturbance but the model does not exploit the periodicity of the measurements.
- *Transient disturbance*: Regression-based methods in the time-domain such as nonparametric multivariate regression can be applied for RCA of irregular persistent spikes as they can be applied for RCA of oscillating disturbances. Granger causality can be applied with moderate efficiency for RCA of irregular persistent spikes. The repeating nature of the irregular persistent spikes is helpful for the modeling, but the model does not rely on the periodicity of the measurements. As for oscillations, the performance of Granger causality is moderate but the results of Granger causality cannot be refined with an analysis with in the frequency-domain since the spectrum of the disturbance is not localized.

Transfer entropy

- *Oscillating disturbance*: Transfer entropy is applied for RCA of linear oscillations with good performance in [48, 181], although [181] highlighted that the results of Granger causality were more reliable. Transfer entropy is applied for RCA of nonlinear oscillations with good performance in [84, 148], although [84] highlighted that nearest neighbours was performing better since it exploited phase-coupling better.
- *Non-oscillatory persistent disturbance*: Reference [84] used transfer entropy for RCA of irregular persistent spikes, where transfer entropy performed better than nearest neighbours or cross-correlation function.
- *Transient disturbance*: Reference [181] demonstrated that transfer entropy was applicable for RCA of step disturbances but that transfer entropy was not able to detect causality for steps of small amplitude, by contrast with Granger causality. Reference [183] applied transfer entropy for RCA of another type of transient disturbance with limited efficiency since transfer entropy finds causal dependencies between all variables affected by the disturbance and is unable to deduce the root cause.

B.3.4 Prediction-based methods in the frequency-domain

Directed transfer function

- *Oscillating disturbance*: Directed transfer function, as other methods in the frequency-domain, is particularly adapted for RCA of oscillations, both linear and nonlinear. Directed transfer function is used for RCA of nonlinear oscillations in [179, 209], and [168] highlighted that directed transfer function combined with Granger causality and partial directed coherence provided highly accurate results for RCA of oscillations.
- *Non-oscillatory persistent disturbance*: Directed transfer function, as other methods in the frequency-domain, is not adapted for RCA of irregular persistent spikes since the spectrum of measurements with such disturbances is not localized.
- *Transient disturbance*: Directed transfer function, as other methods in the frequency-domain, is not adapted for RCA of transient disturbances since the spectrum of measurements with transient disturbances is not localized.

Partial directed coherence

- *Oscillating disturbance*: Partial directed coherence, as other methods in the frequency-domain, is particularly adapted for RCA of both linear and nonlinear oscillations. Directed transfer function is used for RCA of linear oscillations in [162] and for RCA of nonlinear oscillations in [162, 179, 180]. Reference [168] highlighted that partial directed coherence combined with Granger causality and directed transfer function provided highly accurate results for RCA of oscillations.

- *Non-oscillatory persistent disturbance*: Partial directed coherence, as other methods in the frequency-domain, is not adapted for RCA of irregular persistent spikes since the spectrum of measurements with such disturbances is not localized.
- *Transient disturbance*: Partial directed coherence, as other methods in the frequency-domain, is not adapted for RCA of transient disturbances since the spectrum of measurements with transient disturbances is not localized.

Bibliography

- [1] PRONTO Innovative Training Network, “PRONTO website: www.h2020pronto.eu.”
- [2] PRONTO Innovative Training Network, “PRONTO grant agreement,” tech. rep., 2015.
- [3] E. L. Russell, L. H. Chiang, and R. D. Braatz, *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes*. Advances in Industrial Control, London: Springer, 2000.
- [4] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. CRC Press, 1998.
- [5] R. Isermann, “Model-based fault-detection and diagnosis - Status and applications,” *Annual Reviews in Control*, vol. 29, no. 1, pp. 71–85, 2005.
- [6] R. Isermann and P. Ballé, “Trends in the application of model-based fault detection and diagnosis of technical processes,” *Control Engineering Practice*, vol. 5, no. 5, pp. 709–719, 1997.
- [7] R. J. Patton, “Fault-tolerant control systems: The 1997 situation,” in *IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes*, pp. 1033–1054, 1997.
- [8] N. F. Thornhill and A. Horch, “Advances and new directions in plant-wide disturbance detection and diagnosis,” *Control Engineering Practice*, vol. 15, no. 10, pp. 1196–1206, 2007.
- [9] F. Yang and D. Xiao, “Progress in root cause and fault propagation analysis of large-scale industrial processes,” *Journal of Control Science and Engineering*, vol. 2012, 2012.
- [10] V. Venkatasubramanian, “Abnormal events management in complex process plants : Challenges and opportunities in intelligent supervisory control,” in *Proceedings Foundations of Computer-Aided Process Operations (FOCAPO2003)*, pp. 117–132, 2003.
- [11] T. Stauffer and P. Clarke, “Using alarms as a layer of protection,” *Process Safety Progress*, vol. 35, no. 1, pp. 76–83, 2016.
- [12] C. Aldrich and L. Auret, *Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods*. London: Springer London, 2013.
- [13] E. C. Silva, “Why are major accidents still occurring?,” *Process Safety Progress*, vol. 35, no. 3, pp. 253–257, 2016.

- [14] Y. Dai, H. Wang, F. Khan, and J. Zhao, "Abnormal situation management for smart chemical process operation," *Current Opinion in Chemical Engineering*, vol. 14, pp. 49–55, 2016.
- [15] F. Eljack and M. K. Kazi, "Process safety and abnormal situation management," *Current Opinion in Chemical Engineering*, vol. 14, pp. 35–41, 2016.
- [16] D. Macdonald, *Practical hazops, trips and alarms*. Newnes, 2004.
- [17] Y. Shu, L. Ming, F. Cheng, Z. Zhang, and J. Zhao, "Abnormal situation management: Challenges and opportunities in the big data era," *Computers and Chemical Engineering*, vol. 91, pp. 104–113, 2016.
- [18] S. J. Qin, "Process data analytics in the era of big data," *AIChE Journal*, vol. 60, no. 9, pp. 3092–3100, 2014.
- [19] International Society of Automation, *ANSI/ISA-18.2-2016 Management of Alarm Systems for the Process Industries*. 2016.
- [20] OPC Foundation, *Alarms and Events Custom Interface Standard, Version 1.10*. 2002.
- [21] IEC (International Electrotechnical Commission), *Management of Alarm Systems for the Process Industries. IEC 62682*. IEC, 2014.
- [22] N. Md Nor, C. R. Che Hassan, and M. A. Hussain, "A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems," *Reviews in Chemical Engineering*, vol. AOP, 2019.
- [23] M. Alauddin, F. Khan, S. Imtiaz, and S. Ahmed, "A bibliometric review and analysis of data-driven fault detection and diagnosis methods for process systems," *Industrial and Engineering Chemistry Research*, vol. 57, no. 32, pp. 10719–10735, 2018.
- [24] M. S. Reis, R. D. Braatz, and L. H. Chiang, "Big data: Challenges and future research directions," *Chemical Engineering Progress*, vol. 112, no. 3, pp. 46–50, 2016.
- [25] B. R. Hollifield and E. Habibi, *Alarm Management: A Comprehensive Guide: Practical and Proven Methods to Optimize the Performance of Alarm Management Systems*. ISA, 2011.
- [26] J. Wang, F. Yang, T. Chen, and S. L. Shah, "An overview of industrial alarm systems: Main causes for alarm overloading, research status, and open problems," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 1045–1061, 2016.
- [27] V. Rodrigo, M. Chioua, T. Hagglund, and M. Hollender, "Causal analysis for alarm flood reduction," *Proceedings of 11th DYCOPS Symposium, Trondheim, Norway, June 6-8, 2016, IFAC-PapersOnLine*, vol. 49, no. 7, pp. 723–728, 2016.
- [28] W. Hu, T. Chen, S. L. Shah, and M. Hollender, "Cause and effect analysis for decision support in alarm floods," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 13940–13945, 2017.
- [29] A. Stief, J. R. Ottewill, R. Tan, and Y. Cao, "Process and alarm data integration under a two-stage Bayesian framework for fault diagnostics," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 1220–1226, 2018.
- [30] M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, "Online alarm flood classification using alarm coactivations," in *Proceedings of 10th ADCHEM Symposium, Shenyang, China, July 25-27, IFAC-PapersOnLine*, vol. 51, pp. 345–350, 2018.

- [31] M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, "On improving fault detection and diagnosis using alarm-range normalisation," in *Proceedings of 10th SAFEPROCESS Symposium, Warsaw, Poland, August 29-31, IFAC-PapersOnLine*, vol. 51, pp. 1227–1232, 2018.
- [32] M. Lucke, X. Mei, A. Stief, M. Chioua, and N. F. Thornhill, "Variable selection for fault detection and identification based on mutual information of alarm series," in *Proceedings of 12th DYCOPS Symposium, Florianópolis, Brazil, April 23 - 26, IFAC-PapersOnLine*, vol. 52, pp. 673–678, 2019.
- [33] M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, "Advances in alarm data analysis with a practical application to online alarm flood classification," *Journal of Process Control*, vol. 79, pp. 56–71, 2019.
- [34] M. Lucke, A. Stief, M. Chioua, J. R. Ottewill, and N. F. Thornhill, "Fault detection and identification combining process measurements and statistical alarms," *Control Engineering Practice*, vol. 94, p. 104195, 2020.
- [35] M. Lucke, M. Chioua, C. Grimholt, M. Hollender, and N. F. Thornhill, "Integration of alarm design in fault detection and diagnosis through alarm-range normalization," *Submitted to Control Engineering Practice*.
- [36] B. G. Lipták, *Instruments Engineers' Handbook*. CRC Press, 2003.
- [37] F. Yang, P. Duan, S. L. Shah, and T. Chen, *Capturing Connectivity and Causality in Complex Industrial Processes*. Springer, 2014.
- [38] R. S. Mah, *Chemical Process Structure and Information Flows*. Butterworths, 1990.
- [39] M. R. Maurya, R. Rengaswamy, and V. Venkatasubramanian, "Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets," *Engineering Applications of Artificial Intelligence*, vol. 17, pp. 501–518, aug 2004.
- [40] M. Fedai and R. Drath, "CAEX - A neutral data exchange format for engineering data," *ATP International Automation Technology*, vol. 1, no. 3, pp. 43–51, 2005.
- [41] J. Thambirajah, L. Benabbas, M. Bauer, and N. F. Thornhill, "Cause-and-effect analysis in chemical processes utilizing XML, plant connectivity and quantitative process history," *Computers and Chemical Engineering*, vol. 33, no. 2, pp. 503–512, 2009.
- [42] D. H. Rothenberg, *Alarm Management for Process Control: A Best-practice Guide for Design*. Momentum Press, 2009.
- [43] E. Naghoosi, I. Izadi, and T. Chen, "Estimation of alarm chattering," *Journal of Process Control*, vol. 21, no. 9, pp. 1243–1249, 2011.
- [44] N. A. Adnan, I. Izadi, and T. Chen, "On expected detection delays for alarm systems with deadbands and delay-timers," *Journal of Process Control*, vol. 21, no. 9, pp. 1318–1331, 2011.
- [45] F. Yang, S. L. Shah, D. Xiao, and T. Chen, "Improved correlation analysis and visualization of industrial alarm data," *ISA Transactions*, vol. 51, pp. 499–506, 2012.
- [46] J. R. Couper, *Chemical Process Equipment : Selection and Design*. Elsevier, 2005.
- [47] M. Jelali and B. Huang, *Detection and Diagnosis of Stiction in Control Loops*. Springer Science & Business Media, 2010.

- [48] M. Bauer, *Data-driven Methods for Process Analysis*. Diss. University of London, 2005.
- [49] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, “A review of process fault detection and diagnosis. Part II: Qualitative models and search strategies,” *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 313–326, 2003.
- [50] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, “A review of process fault detection and diagnosis. Part III: Process history based methods,” *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 327–346, 2003.
- [51] Z. Ge, “Review on data-driven modeling and monitoring for plant-wide industrial processes,” *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 16–25, 2017.
- [52] J. Pearl, *Causality*. Cambridge University Press, 2009.
- [53] N. Deo, *Graph Theory with Applications to Engineering and Computer Science*. Prentice-Hall, 1974.
- [54] A. Stief, R. Tan, Y. Cao, and J. R. Ottewill, “Analytics of heterogeneous process data: Multiphase flow facility case study,” *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 363–368, 2018.
- [55] A. Stief, R. Tan, Y. Cao, J. R. Ottewill, N. F. Thornhill, and J. Baranowski, “A heterogeneous benchmark dataset for data analytics: Multiphase flow facility case study,” *Journal of Process Control*, vol. 79, pp. 41–55, 2019.
- [56] P. Goel, A. Datta, and M. S. Mannan, “Industrial alarm systems: Challenges and opportunities,” *Journal of Loss Prevention in the Process Industries*, vol. 50, pp. 23–36, 2017.
- [57] W. Yu and F. Yang, “Detection of causality between process variables based on industrial alarm data using transfer entropy,” *Entropy*, vol. 17, no. 8, pp. 5868–5887, 2015.
- [58] S. R. Kondaveeti, I. Izadi, S. L. Shah, T. Black, and T. Chen, “Graphical tools for routine assessment of industrial alarm systems,” *Computers and Chemical Engineering*, vol. 46, pp. 39–47, 2012.
- [59] K. Ahmed, I. Izadi, T. Chen, D. Joe, and T. Burton, “Similarity analysis of industrial alarm flood data,” *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 2, pp. 452–457, 2013.
- [60] S. Charbonnier, N. Bouchair, and P. Gayet, “Fault template extraction to assist operators during industrial alarm floods,” *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 32–44, 2016.
- [61] S. Charbonnier, N. Bouchair, and P. Gayet, “A weighted dissimilarity index to isolate faults during alarm floods,” *Control Engineering Practice*, vol. 45, pp. 110–122, 2015.
- [62] J. Zhu, C. Wang, C. Li, X. Gao, and J. Zhao, “Dynamic alarm prediction for critical alarms using a probabilistic model,” *Chinese Journal of Chemical Engineering*, vol. 24, no. 7, pp. 881–885, 2016.
- [63] G. Dorgo and J. Abonyi, “Sequence mining based alarm suppression,” *IEEE Access*, vol. 6, pp. 15365–15379, 2018.

- [64] P. Duan, T. Chen, S. L. Shah, and F. Yang, "Methods for root cause diagnosis of plant-wide oscillations," *AIChE Journal*, vol. 60, no. 6, pp. 2019–2034, 2014.
- [65] Z. Yang, J. Wang, and T. Chen, "Detection of correlated alarms based on similarity coefficients of binary data," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 4, pp. 1014–1025, 2013.
- [66] W. Hu, J. Wang, and T. Chen, "A new method to detect and quantify correlated alarms with occurrence delays," *Computers and Chemical Engineering*, vol. 80, pp. 189–198, 2015.
- [67] W. Hu, J. Wang, T. Chen, and S. L. Shah, "Cause-effect analysis of industrial alarm variables using transfer entropies," *Control Engineering Practice*, vol. 64, pp. 205–214, 2017.
- [68] J. Wang, H. Li, J. Huang, and C. Su, "A data similarity based analysis to consequential alarms of industrial processes," *Journal of Loss Prevention in the Process Industries*, vol. 35, pp. 29–34, 2015.
- [69] J. Wang, H. Li, J. Huang, and C. Su, "Association rules mining based analysis of consequential alarm sequences in chemical processes," *Journal of Loss Prevention in the Process Industries*, vol. 41, pp. 178–185, 2016.
- [70] J. Su, D. Wang, Y. Zhang, F. Yang, Y. Zhao, and X. Pang, "Capturing causality for fault diagnosis based on multi-valued alarm series using transfer entropy," *Entropy*, vol. 19, no. 12, p. 663, 2017.
- [71] M. Fullen and O. Niggemann, "Validation of similarity measures for industrial alarm flood analysis," in *IMPROVE - Innovative Modelling Approaches for Production Systems to Raise Validatable Efficiency*, no. August, Springer Vieweg, 2018.
- [72] W. Hu, T. Chen, and S. L. Shah, "Detection of frequent alarm patterns in industrial alarm floods using itemset mining methods," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7290–7300, 2018.
- [73] B. Vogel-Heuser, D. Schütz, and J. Folmer, "Criteria-based alarm flood pattern recognition using historical data from automated production systems (aPS)," *Mechatronics*, vol. 31, pp. 89–100, 2015.
- [74] G. Dorgo, P. Pigler, M. Haragovics, and J. Abonyi, "Learning operation strategies from alarm management systems by temporal pattern mining and deep learning," *Computer Aided Chemical Engineering*, vol. 43, no. June, pp. 1003–1008, 2018.
- [75] J. Folmer, F. Schuricht, and B. Vogel-Heuser, "Detection of temporal dependencies in alarm time series of industrial plants," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 19, pp. 1802–1807, 2014.
- [76] Y. Cheng, I. Izadi, and T. Chen, "Pattern matching of alarm flood sequences by a modified Smith-Waterman algorithm," *Chemical Engineering Research and Design*, vol. 91, no. 6, pp. 1085–1094, 2013.
- [77] W. Hu, J. Wang, and T. Chen, "A local alignment approach to similarity analysis of industrial alarm flood sequences," *Control Engineering Practice*, vol. 55, pp. 13–25, 2016.

- [78] C. Guo, W. Hu, S. Lai, F. Yang, and T. Chen, "An accelerated alignment method for analyzing time sequences of industrial alarm floods," *Journal of Process Control*, vol. 57, pp. 102–115, 2017.
- [79] S. Lai and T. Chen, "A method for pattern mining in multiple alarm flood sequences," *Chemical Engineering Research and Design*, vol. 117, pp. 831–839, 2017.
- [80] S. Lai, F. Yang, and T. Chen, "Online pattern matching and prediction of incoming alarm floods," *Journal of Process Control*, vol. 56, pp. 69–78, 2017.
- [81] G. Dorgo, P. Pigler, and J. Abonyi, "Understanding the importance of process alarms based on the analysis of deep recurrent neural networks trained for fault isolation," *Journal of Chemometrics*, no. December 2017, pp. 1–18, 2018.
- [82] M. Schlegelburg, L. Christiansen, N. F. Thornhill, and A. Fay, "A combined analysis of plant connectivity and alarm logs to reduce the number of alerts in an automation system," *Journal of Process Control*, vol. 23, no. 6, pp. 839–851, 2013.
- [83] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, pp. 461–464, 2000.
- [84] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Finding the direction of disturbance propagation in a chemical process using transfer entropy," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 1, pp. 12–21, 2007.
- [85] Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi, and L. Cazzanti, "Similarity-based classification: Concepts and algorithms," *Journal of Machine Learning Research*, vol. 10, pp. 747–776, 2009.
- [86] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: The role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.
- [87] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6414–6428, 2014.
- [88] I. Monroy, R. Benitez, G. Escudero, and M. Graells, "A semi-supervised approach to fault diagnosis for chemical processes," *Computers and Chemical Engineering*, vol. 34, no. 5, pp. 631–642, 2010.
- [89] S. Zhong, Q. Wen, and Z. Ge, "Semi-supervised Fisher discriminant analysis model for fault classification in industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 138, pp. 203–211, 2014.
- [90] Z. Ge, S. Zhong, and Y. Zhang, "Semisupervised kernel learning for FDA model and its application for fault classification in industrial processes," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1403–1411, 2016.
- [91] L. Jiang, Z. Ge, and Z. Song, "Semi-supervised fault classification based on dynamic sparse stacked auto-encoders model," *Chemometrics and Intelligent Laboratory Systems*, vol. 168, pp. 72–83, 2017.
- [92] J. Zheng, H. Wang, Z. Song, and Z. Ge, "Ensemble semi-supervised Fisher discriminant analysis model for fault classification in industrial processes," *ISA Transactions*, 2019.

- [93] M. Askarian, G. Escudero, M. Graells, R. Zarghami, F. Jalali-Farahani, and N. Mostoufi, "Fault diagnosis of chemical processes with incomplete observations: A comparative study," *Computers and Chemical Engineering*, vol. 84, pp. 104–116, 2016.
- [94] M. Askarian, R. Benítez, M. Graells, and R. Zarghami, "Data-based fault detection in chemical processes: Managing records with operator intervention and uncertain labels," *Expert Systems with Applications*, vol. 63, pp. 35–48, 2016.
- [95] K. Severson, P. Chaiwatanodom, and R. D. Braatz, "Perspectives on process monitoring of industrial systems," *Annual Reviews in Control*, vol. 42, pp. 190–200, 2016.
- [96] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Industrial and Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543–3562, 2013.
- [97] I. M. Cecílio, *On Extending Process Monitoring and Diagnosis to the Electrical and Mechanical Utilities: An Advanced Signal Analysis Approach*. Diss. Imperial College London, 2014.
- [98] L. H. Chiang, B. Jiang, X. Zhu, D. Huang, and R. D. Braatz, "Diagnosis of multiple and unknown faults using the causal map and multivariate statistics," *Journal of Process Control*, vol. 28, pp. 27–39, 2015.
- [99] L. H. Chiang and R. D. Braatz, "Process monitoring using causal map and multivariate statistics: Fault detection and identification," *Chemometrics and Intelligent Laboratory Systems*, vol. 65, no. 2, pp. 159–178, 2003.
- [100] B. Jiang, X. Zhu, D. Huang, and R. D. Braatz, "Canonical variate analysis - based monitoring of process correlation structure using causal feature representation," *Journal of Process Control*, vol. 32, pp. 109–116, 2015.
- [101] C. Duchesne, J. J. Liu, and J. F. MacGregor, "Multivariate image analysis in the process industries: A review," *Chemometrics and Intelligent Laboratory Systems*, vol. 117, pp. 116–128, 2012.
- [102] Z. Zhang and J. Zhao, "A deep belief network based fault diagnosis model for complex chemical processes," *Computers and Chemical Engineering*, vol. 107, pp. 395–407, 2017.
- [103] H. Wu and J. Zhao, "Deep convolutional neural network model based chemical process fault diagnosis," *Computers and Chemical Engineering*, vol. 115, pp. 185–197, 2018.
- [104] C. Tong and A. Palazoglu, "Dissimilarity-based fault diagnosis through ensemble filtering of informative variables," *Industrial and Engineering Chemistry Research*, vol. 55, no. 32, pp. 8774–8783, 2016.
- [105] Z. B. Zhu and Z. H. Song, "Fault diagnosis based on imbalance modified kernel Fisher discriminant analysis," *Chemical Engineering Research and Design*, vol. 88, no. 8, pp. 936–951, 2010.
- [106] S. Gajjar and A. Palazoglu, "A data-driven multidimensional visualization technique for process fault detection and diagnosis," *Chemometrics and Intelligent Laboratory Systems*, vol. 154, pp. 122–136, 2016.
- [107] Y. Shu and J. Zhao, "Fault diagnosis of chemical processes using artificial immune system with vaccine transplant," *Industrial and Engineering Chemistry Research*, vol. 55, no. 12, pp. 3360–3371, 2016.

- [108] R. E. Vargas, C. J. Munaro, P. M. Ciarelli, and J. C. De Araujo, "Proposal for two classifiers of offshore naturally flowing wells events using k-nearest neighbors, sliding windows and time multiscale," in *2017 6th International Symposium on Advanced Control of Industrial Processes, AdCONIP 2017*, pp. 209–214, IEEE, 2017.
- [109] L. Ming and J. Zhao, "Review on chemical process fault detection and diagnosis," in *2017 6th International Symposium on Advanced Control of Industrial Processes, AdCONIP 2017*, pp. 457–462, IEEE, 2017.
- [110] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [111] K. Ghosh, M. Ramteke, and R. Srinivasan, "Optimal variable selection for effective statistical process monitoring," *Computers and Chemical Engineering*, vol. 60, pp. 260–276, 2014.
- [112] F. A. P. Peres and F. S. Fogliatto, "Variable selection methods in multivariate statistical process control: A systematic literature review," *Computers and Industrial Engineering*, vol. 115, pp. 603–619, 2018.
- [113] S. Verron, T. Tiplica, and A. Kobi, "Fault detection and identification with a new feature selection based on mutual information," *Journal of Process Control*, vol. 18, no. 5, pp. 479–490, 2008.
- [114] C. Zhao and F. Gao, "Critical-to-fault-degradation variable analysis and direction extraction for online fault prognostic," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 842–854, 2017.
- [115] T. H. Kuang, Z. Yan, and Y. Yao, "Multivariate fault isolation via variable selection in discriminant analysis," *Journal of Process Control*, vol. 35, pp. 30–40, 2015.
- [116] J. V. Kresta, J. F. Macgregor, and T. E. Marlin, "Multivariate statistical monitoring of process operating performance," *The Canadian Journal of Chemical Engineering*, vol. 69, no. 1, pp. 35–47, 1991.
- [117] J. F. MacGregor and T. Kourti, "Statistical process control of multivariate processes," *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, 1995.
- [118] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno, "Monitoring independent components for fault detection," *AIChE Journal*, vol. 49, no. 4, pp. 969–976, 2003.
- [119] A. Negiz and A. Çilinar, "Statistical monitoring of multivariable dynamic processes with state-space models," *AIChE Journal*, vol. 43, no. 8, pp. 2002–2020, 1997.
- [120] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 81–93, 2000.
- [121] J.-M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chemical Engineering Science*, vol. 59, pp. 223–234, 2004.
- [122] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE Journal*, vol. 37, no. 2, pp. 233–243, 1991.

- [123] J. Chen and J. Liu, "Mixture principal component analysis models for process monitoring," *Industrial and Engineering Chemistry Research*, vol. 38, no. 4, pp. 1478–1488, 1999.
- [124] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 179–196, 1995.
- [125] B. S. Dayal and J. F. Macgregor, "Recursive exponentially weighted PLS and its applications," *Science*, vol. 7, no. 3, pp. 169–179, 1997.
- [126] S. J. Zhao, J. Zhang, and Y. M. Xu, "Monitoring of processes with multiple operating modes through multiple principle component analysis models," *Industrial and Engineering Chemistry Research*, vol. 43, no. 22, pp. 7025–7035, 2004.
- [127] R. Tan, T. Cong, N. F. Thornhill, J. R. Ottewill, and J. Baranowski, "Statistical monitoring of processes with multiple operating modes," *IFAC-PapersOnLine*, vol. 52, no. 1, pp. 635–642, 2019.
- [128] J. C. Hoskins, K. M. Kaliyur, and D. M. Himmelblau, "Fault diagnosis in complex chemical plants using artificial neural networks," *AIChE Journal*, vol. 37, no. 1, pp. 137–141, 1991.
- [129] E. L. R. Braatz, L. H. Chiang and R. D., *Fault Detection and Diagnosis in Industrial Systems*. Advanced Textbooks in Control and Signal Processing, London: Springer London, 2001.
- [130] B. Jiang, D. Huang, X. Zhu, F. Yang, and R. D. Braatz, "Canonical variate analysis-based contributions for fault identification," *Journal of Process Control*, vol. 26, pp. 17–25, 2015.
- [131] R. Tan and Y. Cao, "Deviation contribution plots of multivariate statistics," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 833–841, 2019.
- [132] C. F. Alcalá and S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol. 45, no. 7, pp. 1593–1600, 2009.
- [133] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines," *Computers and Chemical Engineering*, vol. 28, no. 8, pp. 1389–1401, 2004.
- [134] X. B. He, W. Wang, Y. P. Yang, and Y. H. Yang, "Variable-weighted Fisher discriminant analysis for process fault diagnosis," *Journal of Process Control*, vol. 19, no. 6, pp. 923–931, 2009.
- [135] Y. Zeng, Z. Jia, W. Liang, and S. Gu, "Fault diagnosis based on variable-weighted separability-oriented subclass discriminant analysis," *Computers and Chemical Engineering Chemical Engineering*, vol. 129, p. 106514, 2019.
- [136] W. Zhu, W. Sun, and J. Romagnoli, "Adaptive k-nearest-neighbor method for process monitoring," *Industrial and Engineering Chemistry Research*, vol. 57, no. 7, pp. 2574–2586, 2018.
- [137] Z. Chai and C. Zhao, "Enhanced random forest with concurrent analysis of static and dynamic nodes for industrial fault classification," *IEEE Transactions on Industrial Informatics*, vol. PP, 2019.

- [138] Y. Hu and C. Zhao, "Fault diagnosis with dual cointegration analysis of common and specific nonstationary fault variations," *IEEE Transactions on Automation Science and Engineering*, vol. PP, pp. 1–11, 2019.
- [139] B. Huang, "Bayesian methods for control loop monitoring and diagnosis," *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 40, no. 5, pp. 29–38, 2007.
- [140] R. Gonzalez and B. Huang, "Control-loop diagnosis using continuous evidence through kernel density estimation," *Journal of Process Control*, vol. 24, no. 5, pp. 640–651, 2014.
- [141] Q. Jiang, B. Huang, and X. Yan, "GMM and optimal principal components-based Bayesian method for multimode fault diagnosis," *Computers and Chemical Engineering*, vol. 84, pp. 338–349, 2016.
- [142] Q. Jiang, B. Huang, S. X. Ding, and X. Yan, "Bayesian fault diagnosis with asynchronous measurements and its application in networked distributed monitoring," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 10, pp. 6316–6324, 2016.
- [143] K. Tidriri, T. Tiplica, N. Chatti, and S. Verron, "A generic framework for decision fusion in fault detection and diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 71, pp. 73–86, 2018.
- [144] S. Verron, T. Tiplica, and A. Kobi, "Fault diagnosis with Bayesian networks: Application to the Tennessee Eastman process," in *Proceedings of the IEEE International Conference on Industrial Technology*, pp. 98–103, IEEE, 2006.
- [145] W. Yu and C. Zhao, "Online fault diagnosis for industrial processes with Bayesian network-based probabilistic ensemble learning strategy," *IEEE Transactions on Automation Science and Engineering*, vol. PP, pp. 1–11, 2019.
- [146] N. F. Thornhill, J. W. Cox, and M. A. Paulonis, "Diagnosis of plant-wide oscillation through data-driven analysis and process understanding," *Control Engineering Practice*, vol. 11, no. 12, pp. 1481–1490, 2003.
- [147] N. F. Thornhill, "Finding the source of nonlinearity in a process with plant-wide oscillation," *IEEE Transactions on Control Systems Technology*, vol. 13, no. 3, pp. 434–443, 2005.
- [148] B. Lindner, M. Chioua, J. W. Groenewald, L. Auret, and M. Bauer, "Diagnosis of oscillations in an industrial mineral process using transfer entropy and nonlinearity index," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 1409–1416, 2018.
- [149] M. Bauer and N. F. Thornhill, "A practical method for identifying the propagation path of plant-wide disturbances," *Journal of Process Control*, vol. 18, no. 7-8, pp. 707–719, 2008.
- [150] R. Landman and S.-L. Jämsä-Jounela, "Data-driven causal analysis and its application on a large-scale board machine," in *Nordic Process Control Workshop 2013*, 2013.
- [151] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.
- [152] D. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," *KDD-94 Workshop on Knowledge Discovery in Databases*, vol. 398, no. 16, pp. 359–370, 1994.

- [153] G. Li, S. J. Qin, and T. Yuan, "Data-driven root cause diagnosis of faults in process industries," *Chemometrics and Intelligent Laboratory Systems*, vol. 159, pp. 1–11, 2016.
- [154] M. Stockmann, R. Haber, and U. Schmitz, "Source identification of plant-wide faults based on k nearest neighbor time delay estimation," *Journal of Process Control*, vol. 22, no. 3, pp. 583–598, 2012.
- [155] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Nearest neighbors methods for root cause analysis of plantwide disturbances," *Industrial and Engineering Chemistry Research*, vol. 46, no. 18, pp. 5977–5984, 2007.
- [156] I. M. Cecilio, J. R. Ottewill, and N. F. Thornhill, "Determining the propagation path of a disturbance in multi-rate process and electromechanical systems," *Control Engineering Practice*, vol. 49, pp. 187–193, 2016.
- [157] R. Landman and S. L. Jämsä-Jounela, "Fault propagation analysis by implementing nearest neighbors method using process connectivity," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 5, pp. 2058 – 2067, 2019.
- [158] G. Sugihara, R. May, H. Ye, C. H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *Science*, vol. 338, no. 6106, pp. 496–500, 2012.
- [159] M. F. Aftab, M. Hovd, and S. Sivalingam, "Convergent cross mapping (CCM) based approach for isolating the source of plant-wide disturbances," in *Proceedings of 1st Annual IEEE Conference on Control Technology and Applications, CCTA 2017*, pp. 1492–1498, 2017.
- [160] L. Luo, F. Cheng, T. Qiu, and J. Zhao, "Refined convergent cross-mapping for disturbance propagation analysis of chemical processes," *Computers and Chemical Engineering*, vol. 106, pp. 1–16, 2017.
- [161] Y. Wang, F. Hu, Y. Cao, X. Yuan, and C. Yang, "Improved CCM for variable causality detection in complex systems," *Control Engineering Practice*, vol. 83, pp. 67–82, 2019.
- [162] T. Yuan and S. J. Qin, "Root cause diagnosis of plant-wide oscillations using Granger causality," *Journal of Process Control*, vol. 24, no. 2, pp. 450–459, 2014.
- [163] T. Yuan, G. Li, Z. Zhang, and S. J. Qin, "Deep causal mining for plant-wide oscillations with multilevel Granger causality analysis," in *Proceedings of the American Control Conference*, pp. 5056–5061, 2016.
- [164] H. Cheng, J. Wu, Y. Liu, and D. Huang, "A novel fault identification and root-causality analysis of incipient faults with applications to wastewater treatment processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 188, pp. 24–36, 2019.
- [165] U. Ahmed, D. Ha, S. Shin, N. Shaukat, U. Zahid, and C. Han, "Estimation of disturbance propagation path using principal component analysis (PCA) and multivariate Granger causality (MVGCC) techniques," *Industrial and Engineering Chemistry Research*, vol. 56, no. 25, pp. 7260–7272, 2017.
- [166] H. S. Chen, Z. Yan, Y. Yao, T. B. Huang, and Y. S. Wong, "Systematic procedure for Granger-causality-based root cause diagnosis of chemical process faults," *Industrial and Engineering Chemistry Research*, vol. 57, no. 29, pp. 9500–9512, 2018.

- [167] V. M. Marques, C. J. Munaro, and S. L. Shah, "Detection of causal relationships based on residual analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 4, pp. 1525–1534, 2015.
- [168] R. Landman, *Data-based Causality Analysis by Exploiting Process Connectivity Information*. Aalto University publication series, Doctoral Dissertations, 61/2019, 2019.
- [169] R. Landman and S.-L. Jämsä-Jounela, "Hybrid causal analysis combining a nonparametric multiplicative regression causality estimator with process connectivity information," *Control Engineering Practice*, vol. 93, pp. 1–11, 2019.
- [170] E. Naghoosi, B. Huang, E. Domlan, and R. Kadali, "Information transfer methods in causality analysis of process variables with an industrial application," *Journal of Process Control*, vol. 23, no. 9, pp. 1296–1305, 2013.
- [171] Y. Shu and J. Zhao, "Data-driven causal inference based on a modified transfer entropy," *Computers and Chemical Engineering*, vol. 57, pp. 173–180, 2013.
- [172] B. Lindner, L. Auret, and M. Bauer, "A systematic workflow for oscillation diagnosis using transfer entropy," *IEEE Transactions on Control Systems Technology*, vol. PP, pp. 1–12, 2019.
- [173] P. Duan, F. Yang, S. L. Shah, and T. Chen, "Transfer zero-entropy and its application for capturing cause and effect relationship between variables," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 855–867, 2015.
- [174] B. Rashidi, D. S. Singh, and Q. Zhao, "Data-driven root-cause fault diagnosis for multivariate non-linear processes," *Control Engineering Practice*, vol. 70, pp. 134–147, 2018.
- [175] R. Landman and S. L. Jämsä-Jounela, "Hybrid approach to causal analysis on a complex industrial system based on transfer entropy in conjunction with process connectivity information," *Control Engineering Practice*, vol. 53, pp. 14–23, 2016.
- [176] Q. X. Zhu, Q. Q. Meng, P. J. Wang, and Y. L. He, "Novel causal network modeling method integrating process knowledge with modified transfer entropy: A case study of complex chemical processes," *Industrial and Engineering Chemistry Research*, vol. 56, no. 48, pp. 14282–14289, 2017.
- [177] P. Duan, F. Yang, T. Chen, and S. L. Shah, "Direct causality detection via the transfer entropy approach," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 6, pp. 2052–2066, 2013.
- [178] B. Zhou, M. Chioua, M. Bauer, J. C. Schlake, and N. F. Thornhill, "Improving root cause analysis by detecting and removing transient changes in oscillatory time series with application to a 1,3-Butadiene process," *Industrial and Engineering Chemistry Research*, vol. 58, no. 26, pp. 11234–11250, 2019.
- [179] R. Landman, J. Kortela, Q. Sun, and S. L. Jämsä-Jounela, "Fault propagation analysis of oscillations in control loops using data-driven causality and plant connectivity," *Computers and Chemical Engineering*, vol. 71, pp. 446–456, 2014.
- [180] S. Xu, M. Baldea, T. F. Edgar, W. Wojsznis, T. Blevins, and M. Nixon, "Root cause diagnosis of plant-wide oscillations based on information transfer in the frequency domain," *Industrial and Engineering Chemistry Research*, vol. 55, no. 6, pp. 1623–1629, 2016.

- [181] B. Lindner, L. Auret, and M. Bauer, "Investigating the impact of perturbations in chemical processes on data-based causality analysis. Part 2: Testing Granger causality and transfer entropy," in *IFAC-PapersOnLine*, vol. 50, pp. 3275–3280, 2017.
- [182] B. Lindner, L. Auret, M. Bauer, and J. W. Groenewald, "Comparative analysis of Granger causality and transfer entropy to present a decision flow for the application of oscillation diagnosis," *Journal of Process Control*, vol. 79, pp. 72–84, 2019.
- [183] C. Kühnert, *Data-driven Methods for Fault Localization in Process Technology*. KIT Scientific Publishing, 2013.
- [184] C. Kühnert and J. Beyerer, "Data-driven methods for the detection of causal structures in process technology," *Machines*, vol. 2, no. 4, pp. 255–274, 2014.
- [185] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.
- [186] M. J. Lesot, M. Rifqi, and H. Benhadda, "Similarity measures for binary and numerical data: A survey," *International Journal of Knowledge Engineering and Soft Data Paradigms*, vol. 1, no. 1, pp. 63–84, 2008.
- [187] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [188] H.-T. Lin, C.-J. Lin, and R. C. Weng, "A note on Platt's probabilistic outputs for support vector machines," *Machine Learning*, vol. 68, pp. 267–276, 2007.
- [189] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.
- [190] E. Fix and J. L. Hodges, "Discriminatory analysis. Nonparametric discrimination: Consistency properties," *USAF School of Aviation Medicine, Randolph Field, Texas*, 1951.
- [191] J. Xu, J. Wang, I. Izadi, and T. Chen, "Performance assessment and design for univariate alarm systems based on FAR, MAR, and AAD," *IEEE Transactions on Automation Science and Engineering*, vol. 9, no. 2, pp. 296–307, 2012.
- [192] F. Yang and C. Guo, "Survey on advanced alarm strategies based on multivariate analysis," in *2017 6th International Symposium on Advanced Control of Industrial Processes, AdCONIP 2017*, pp. 612–617, IEEE, 2017.
- [193] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [194] M. Ardakani, M. Askarian, A. Shokry, G. Escudero, M. Graells, and A. Espuña, "Optimal features selection for designing a fault diagnosis system," *Computer Aided Chemical Engineering*, vol. 38, pp. 1111–1116, 2016.
- [195] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [196] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLoS ONE*, vol. 9, no. 2, p. e87357, 2014.

- [197] A. Kraskov, H. Stögbauer, and P. Grassberger, “Estimating mutual information,” *Physical Review E - Statistical Physics*, vol. 69, no. 6, pp. 1–16, 2004.
- [198] E. Parzen, “On estimation of a probability density function and mode,” *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 2007.
- [199] G. H. John and P. Langley, “Estimating continuous distributions in Bayesian classifiers,” *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI1995)*, 1995.
- [200] M. Wand, M. Jones, and M. Jones, *Kernel Smoothing*. Chapman and Hall/CRC, 1994.
- [201] A. Rohatgi, “Webplotdigitizer: <https://automeris.io/WebPlotDigitizer>,” 2011.
- [202] M. A. A. S. Choudhury, S. L. Shah, N. F. Thornhill, M. S. Choudhary, S. L. Shah, and N. F. Thornhill, *Diagnosis of Process Nonlinearities and Valve Stiction*. Heidelberg: Advances in Industrial Control, 2008.
- [203] M. Sajidman and H. B. Kuntze, “Integration of fuzzy control and model based concepts for disturbed industrial plants with large dead-times,” *IEEE International Conference on Fuzzy Systems*, vol. 2, pp. 1007–1013, 1997.
- [204] I. M. Cecílio, J. R. Ottewill, J. Pretlove, and N. F. Thornhill, “Nearest neighbors method for detecting transient disturbances in process and electromechanical systems,” *Journal of Process Control*, vol. 24, no. 9, pp. 1382–1393, 2014.
- [205] N. Li, M. Hollender, A. Cohen, M. Chioua, M. Lucke, and R. Bauer, “Unravelling events and alarms with data analytics tools,” *ABB Review*, vol. 3, pp. 36–43, 2019.
- [206] T. Gamer, M. Hoernicke, B. Kloepper, R. Bauer, and A. J. Isaksson, “The autonomous industrial plant - Future of process engineering, operations and maintenance,” *IFAC-PapersOnLine*, vol. 52, no. 1, pp. 454–460, 2019.
- [207] B. S. Lindner, *Improving the Performance of Causality Analysis Techniques for Automated Fault Diagnosis in Mineral Processing Plants*. Diss. Stellenbosch University, 2019.
- [208] L. Cao, “Practical method for determining the minimum embedding dimension of a scalar time series,” *Physica D*, vol. 110, no. 1-2, pp. 43–50, 1997.
- [209] K. Li, L. Xie, X. Ge, and X. Wang, “Application of DTF method in disturbance propagation analysis of complex chemical process,” in *Proceedings of the 11th IEEE International Conference on Networking, Sensing and Control, ICNSC 2014*, pp. 685–689, 2014.
- [210] D. Mønster, R. Fusaroli, K. Tylén, A. Roepstorff, and J. F. Sherson, “Inferring causality from noisy time series data,” in *Proceedings of COMPLEXIS 2016*, pp. 1–9, 2016.