

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	ix
List of Figures	xiv
List of Tables	xvii
1 Introduction	1
1.1 Objectives of the research	4
1.2 Outline of the thesis	5
1.3 Contributions	6
1.4 Statement of Originality	7
1.5 Publications	7
1.5.1 Publications in scientific Journals and Books	8
1.5.2 Works presented in conferences at national or international level	9
2 Biofilm in drinking water distribution systems	13
2.1 Biofilm overview	14
2.2 Biofilm development in DWDSs	15
2.3 Problems associated with biofilm development in DWDSs	17
2.3.1 Health risk	17
2.3.2 Aesthetic deterioration of water	19
2.3.3 Proliferation of higher organisms	20
2.3.4 Disinfectant decay	22
2.3.5 Biocorrosion	23
2.3.6 Operational problems	26
2.4 Biofilm control in DWDSs	27
3 Current approaches to study biofilm development in DWDSs	32
3.1 Biofilm growth devices	32
3.2 <i>In situ</i> biofilm sampling	36
3.2.1 Pipe cut-out sampling	37

3.2.2	Pipe device sampling	38
3.3	Microbial quantification in DWDS biofilm	39
3.4	Heterotrophic plate count (HPC)	42
4	Case Studies	47
4.1	Selection of case studies	48
4.2	Case study 1. Drinking water distribution system of Thessaloniki, Greece	49
4.2.1	Sampling protocol	52
4.2.2	Descriptive analysis	56
4.3	Case study 2. Pennine Water Group pilot distribution system in Sheffield, United Kingdom	57
4.3.1	Pennine Water Group's experimental facility and operating con- ditions	62
4.3.2	Biofilm sampling	65
5	Getting and pre-processing data	70
5.1	Data collection	71
5.2	Data pre-processing	74
5.2.1	Data unification	75
5.2.1.1	Variables design	75
5.2.1.1.1	Physical characteristics of the system	75
5.2.1.1.2	Hydraulic characteristics of the system	77
5.2.1.1.3	Sampling and incubation	80
5.2.1.1.4	Physico-chemical characteristics of water	84
5.2.1.1.5	Biofilm	86
5.2.2	Data cleansing	87
5.2.2.1	Variables cleansing	87
5.2.2.1.1	Variability reduction	87
5.2.2.1.2	Non-significant/informative variables identifi- cation	88
5.2.2.1.3	Removal of variables with high percentage of missing values	89
5.2.2.2	Cases cleansing	90
5.2.2.2.1	Inconsistent data identification	90
5.2.2.2.2	Removal of cases with high percentage of miss- ing values	90
5.2.2.2.3	Outlier detection	90
5.2.2.3	Clean data set	92
5.2.3	Data set reconstruction	94
5.2.3.1	Imputation of missing values	96
5.2.3.2	Complete data set	97
6	Data set: Exploratory Data Analysis	99
6.1	Descriptive data analysis	99
6.1.1	Target attribute	100
6.1.2	Categorical attributes	100
6.1.3	Continuous attributes	105
6.2	Exploratory data analysis	108

6.2.1	Categorical attributes	108
6.2.2	Continuous attributes	110
6.2.2.1	Data set clustering	111
7	Model development	118
7.1	Regression Trees	118
7.2	Regression Tree implementation	120
7.2.1	Testing the Regression Tree model	122
7.3	Random Forests	125
7.4	Random Forests implementation	126
7.4.1	Testing the Random Forest model	128
7.5	Conclusions	129
8	From pipe to network	131
8.1	Multi-agent systems	131
8.2	Discriminant Analysis via Label Propagation	133
8.3	Graph Theory Measurements to Assess the Importance of the Edges	134
8.3.1	Edge betweenness centrality	134
8.4	Case Study	135
8.5	Further application: Biofilm susceptibility as criteria for rehabilitation actions in DWDSs	138
8.6	Conclusions	139
9	Conclusions and Future Work	143
9.1	Merits of the new approach	144
9.2	Practical implications	145
9.3	Future perspectives	147
9.4	Final conclusions	149
A	Compiled variables with less than the 15% of data	150
B	Extract of the first 50 elements of the synthetic database	152
C	Ensemble of naïve Bayesian approaches for the study of biofilm development in drinking water distribution systems	154
C.1	Naïve Bayesian approaches	154
C.1.1	Augmented Bayesian Classifiers	155
C.1.2	A combined approach: bagging naïve bayes	156
C.1.3	A hybrid approach: Bagging leafs of naïve Bayesian trees	157
C.1.4	Summary of the results and conclusions	158
D	Modelling the Biofilm Development Process within pipes with Multi-agent systems	160
D.1	Modelling the Biofilm Development Process	160
E	Presentation of the web page sections	163
E.1	Presentation of the web page sections	163

Bibliography

166