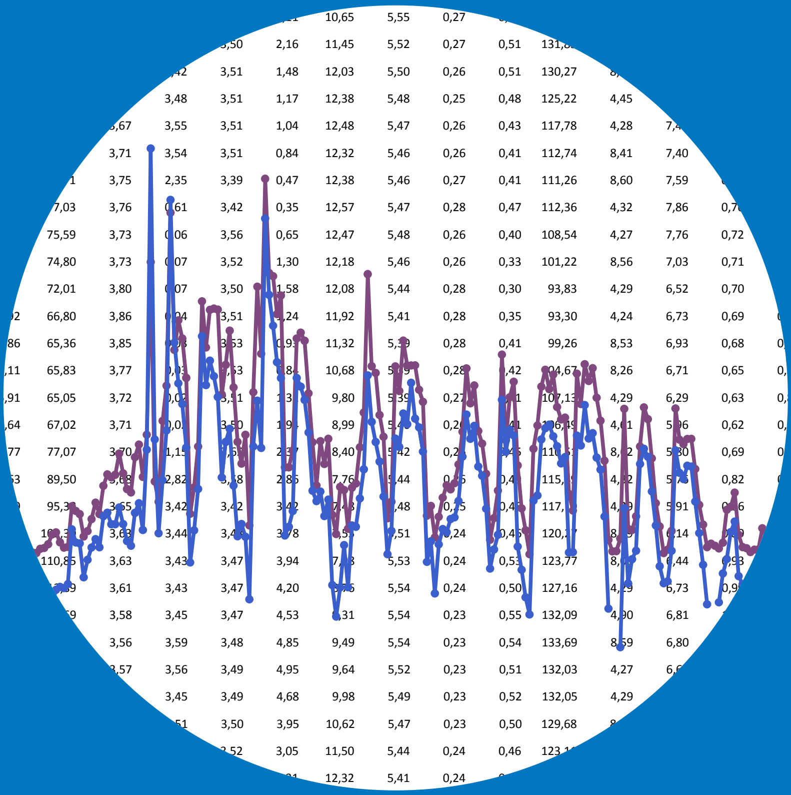
Department of Built Environment

Data-derived soft sensors in biological wastewater treatment

With application of multivariate statistical methods

Henri Haimi



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 24/2016

Data-derived soft sensors in biological wastewater treatment

With application of multivariate statistical methods

Henri Haimi

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Engineering, at a public examination held at the lecture hall R1 of the school on 26 February 2016 at 12.

Aalto University
School of Engineering
Department of Built Environment
Water and Environmental Engineering

Supervising professor
Prof. Riku Vahala, Aalto University, Finland

Thesis advisors
Dr. Michela Mulas, Aalto University, Finland / Federal University of Campina Grande, Brazil
Dr. Francesco Corona, Aalto University, Finland / Federal University of Ceará, Brazil

Preliminary examiners
Prof. Krist Gernaey, DTU Technical University of Denmark, Denmark
Prof. Daniel Aguado García, Universitat Politècnica de València, Spain

Opponent
Prof. Emer. Gustaf Olsson, Lund University, Sweden

Aalto University publication series
DOCTORAL DISSERTATIONS 24/2016

ISBN 978-952-60-6649-3 (printed)
ISBN 978-952-60-6650-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-6650-9

Unigrafia Oy
Helsinki 2016

Finland

NORDIC ECOLABEL
441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Henri Haimi

Name of the doctoral dissertation
Data-derived soft sensors in biological wastewater treatment: With application of multivariate statistical methods

Publisher School of Engineering

Unit Department of Built Environment

Series Aalto University publication series DOCTORAL DISSERTATIONS 24/2016

Field of research Water and Environmental Engineering

Manuscript submitted 7 August 2015 **Date of the defence** 26 February 2016

Permission to publish granted (date) 25 January 2016 **Language** English

□ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

The increased awareness about the ecological status of the waterbodies and, on the other hand, the advances in the treatment technology have acted as driving forces behind the gradual tightening of wastewater purification requirements. Achieving the stringent treatment targets of wastewater treatment plants cost-efficiently is crucially dependent on the high-grade monitoring and control of the process units. Those, in turn, necessitate reliable real-time information about the primary process variables. In spite of the considerable developments of on-line sensors, demanding conditions in biological treatment processes sometimes give rise to an insufficient performance of instruments.

The main motivation for this thesis was to design software tools that enable more efficient and safer treatment process operation by complementing conventional instrumentation. Since modern facilities are amply instrumented and there are plenty of accessible historical data, data-derived approaches were used in the studies. The data were processed together with predictive models providing virtual instruments often referred to as soft sensors.

In this thesis, the models at the core of the soft sensors are based on multivariate statistics. In particular, principal component analysis with its variants and least-squares-based regression methods were employed in the soft sensor development. The moving-window techniques were applied so as to adapt the models to time-varying wastewater treatment processes. Both linear and nonlinear regression methods were explored.

The technical studies of the thesis concern a large-scale municipal wastewater treatment plant. An array of soft sensors for the on-line prediction of nitrate concentrations was developed to support the operation of the biological post-filtration unit. Then, a system that enables the complementary use of the soft sensor estimates and the corresponding hardware instrument measurements was designed. The soft sensors were found to model nitrate concentrations accurately and, especially when integrated with the proposed switching system, to allow for a more secure control of the unit. In addition, a soft sensor for detecting process and instrument anomalies in the activated sludge process was investigated. The presented anomaly detection system motivates a more efficient use of sensors in the process control.

It was demonstrated that soft sensors were applicable to the considered tasks and that they have strong potential for providing support to the operations of treatment facilities. The employed multivariate techniques proved to be capable of extracting easily understandable and practicable information from the high-dimensional data.

Keywords data-derived modelling, multivariate statistics, process monitoring, soft sensors, wastewater treatment

ISBN (printed) 978-952-60-6649-3 **ISBN (pdf)** 978-952-60-6650-9

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2016

Pages 211 **urn** http://urn.fi/URN:ISBN:978-952-60-6650-9

Aalto-yliopisto

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tiivistelmä

Tekijä
Henri Haimi

Väitöskirjan nimi
Datapohjaiset virtuaaliset anturit biologisessa jätevedenkäsittelyssä: tilastollisten monimuuttujamenetelmien sovelluksia

Julkaisija Insinööritieteiden korkeakoulu

Yksikkö Rakennetun ympäristön laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 24/2016

Tutkimusala Vesi- ja ympäristötekniikka

Käsikirjoituksen pvm 07.08.2015 **Väitöspäivä** 26.02.2016

Julkaisuluvan myöntämispäivä 25.01.2016 **Kieli** Englanti

□ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Jäteveden käsittelyvaatimukset ovat kiristyneet asteittain, mihin on vaikuttanut lisääntynyt vesistöjen ekologista tilaa koskeva valveutuneisuus ja, toisaalta, puhdistustekniikoiden kehittyminen. Korkealaatuinen prosessinvalvonta ja -ohjaus ovat olennaisia edellytyksiä tiukentuvien käsittelytavoitteiden saavuttamiseen kustannustehokkaasti. Niihin puolestaan tarvitaan luotettavaa reaaliaikaista informaatiota tärkeistä prosessimuuttujista. Vaikka ajantasaisesti muuttujia analysoivat instrumentit ovat kehittyneet merkittävästi, ei niiden suorituskyky biologisten käsittelyprosessien vaativissa olosuhteissa ole aina riittävä.

Väitöskirjan päätavoite oli suunnitella perinteistä instrumentaatiota täydentäviä ohjelmistotyökaluja, jotka mahdollistavat käsittelyprosessien tehokkaamman ja riskittömämmän operoinnin. Koska nykyaikaiset laitokset ovat hyvin instrumentoituja ja niiden historiallista käyttödataa on runsaasti saatavilla, tutkimuksissa käytettiin datapohjaisia lähestymistapoja. Tutkielmassa kehitettiin usein termillä ″soft sensor″ kutsuttavia virtuaalisia antureita, joissa hyödynnetään dataa prosessoivia ennustavia malleja.

Tässä työssä virtuaalisissa antureissa käytetyt mallit perustuvat tilastollisiin monimuuttujamenetelmiin. Niiden suunnittelussa hyödynnettiin erityisesti pääkomponenttianalyysiä muunnoksineen sekä pienimpään neliösummaan perustuvia regressiomenetelmiä. Liikkuvan ikkunan tekniikoita sovellettiin mallien mukauttamiseen prosessien vaihteleviin olosuhteisiin. Työssä tarkasteltiin sekä lineaarisia että epälineaarisia regressiomenetelmiä.

Väitöstyön tekniset tutkimukset koskevat suurta yhdyskuntajätevedenpuhdistamoa. Biologisen jälkisuodatusyksikön operointia tukemaan kehitettiin nitraattikonsentraatioita ajantasaisesti ennustavia virtuaalisia antureita. Tutkimuksen jatkeena suunniteltiin mallien tuottamien estimaattien ja niitä vastaavien instrumenttimittausten toisiaan täydentävän käytön mahdollistava vaihtamisjärjestelmä. Virtuaalisten anturien osoitettiin mallintavan nitraattikonsentraatioita tarkasti ja, etenkin integroituina vaihtamisjärjestelmään, lisäävän yksikön ohjauksen käyttövarmuutta. Lisäksi prosessi- ja instrumenttianomalioiden valvontaan soveltuvaa virtuaalista anturia tutkittiin aktiivilieteprosessissa. Ehdotettu järjestelmä anomalian havaitsemiseen edesauttaa mittausten tehokasta käyttöä prosessinohjauksessa.

Virtuaalisten anturien demonstroitiin soveltuvan tarkasteltuihin käyttökohteisiin ja omaavan varteenotettavaa potentiaalia käsittelylaitosten operoinnin tukemiseen. Käytettyjen monimuuttujatekniikoiden osoitettiin pystyvän suodattamaan helposti ymmärrettävää ja käyttökelpoista informaatiota monidimensionaalisesta datasta.

Avainsanat datapohjainen mallinnus, tilastolliset monimuuttujamenetelmät, prosessinvalvonta, virtuaaliset anturit, jätevedenkäsittely

ISBN (painettu) 978-952-60-6649-3 **ISBN (pdf)** 978-952-60-6650-9

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2016

Sivumäärä 211 **urn** http://urn.fi/URN:ISBN:978-952-60-6650-9

Preface

This dissertation summarizes the research that I have been carrying out when working in the Water and Environmental Engineering research group at the Aalto University School of Engineering during 2010–2015. The most significant financial supporter of my doctoral research was Maa- ja vesitekniikan tuki ry. that funded projects which produced several scientific articles the thesis covers. Other financiers include the Helsinki Region Environmental Services HSY, the Aalto University School of Engineering and the Aalto University Department of Civil and Environmental Engineering. I am extremely thankful to all of them.

I want to express my gratitude Prof. Riku Vahala who has supervised this work. Riku has provided me the opportunity and the resources required to prepare this thesis. I am very fortunate and thankful of having had two excellent instructors who have taught me what academic research in practice is. Dr. Michela Mulas has shared much of her remarkable expertise in the modelling and control of WWTPs with me. Michela has always been available when any help with the research has been needed and kept me on the track when obstacles have occurred on my path towards the defense. Dr. Francesco Corona has instructed me about the data-derived modelling techniques used in the thesis. In particular, collaborating with Francesco has demonstrated me how the competent research work is realized and how the solid scientific articles are prepared.

A significant part of the work reported in this thesis has been conducted in cooperation with the Helsinki Region Environmental Services, particularly with the staff of the Viikinmäki WWTP, who I would like to sincerely thank. Mari Heinonen and Tommi Fred provided the opportunities for our research team to investigate novel solutions to support the critical plant operations. Laura Sundell and Paula Lindell helped in collecting the required operational data and in learning precisely the process configurations and the instrumentation. Anna Kuokkanen reviewed the plant description given in this dissertation. I also want to acknowledge Mari, Laura and Paula for being co-authors of articles included in the thesis.

I am grateful to Prof. Krist V. Gernaey and Prof. Daniel Aguado García for pre-examining this thesis and for providing valuable feedback. I feel honored that Prof. Emer. Gustaf Olsson agreed to act as my opponent in the public defense of the dissertation. Prof. Olsson is greatly appreciated for his pioneering research achievements in the field of the WWTP automation. He has also been of notable help by introducing me to other participants in the specialist conferences I have attended, which I am thankful for. Actually, I would like to express my gratitude the whole ICA community of the International Water Association for the enormous source of inspiration. Additionally, I acknowledge Prof. Stefano Marsili-Libelli who has made me feel welcome to the global research collective and who also has co-authored an article attached to this dissertation.

I have enjoyed very much my time in the Water and Environmental Engineering group where the atmosphere has always been encouraging and positive. Particularly, I want to thank Ari Järvinen, Aino Peltola and Antti Louhio for taking care of numerous practicalities that have made my work easier – or at times even possible. A large number of doctoral students, professors, post docs, master students and other employees have worked in the group during the preparation of my thesis and I have shared plenty of good moments with many of them. Above all, I have had many helpful discussions about the PhD research with the other doctoral students and the post docs under the water and wastewater engineering discipline. All the co-workers are too many to name, but especially I would like to mention Matti Keto and Maryam 'Roza' Yazdani who I have had pleasure to become friends with in recent years. Thank you colleagues.

Eventually, I am most thankful to my parents, other relatives and friends who have supported me during this PhD project. In particular, my girlfriend Erja has always encouraged me when I have faced challenges with this dissertation and hard times with the research work in general.

Espoo, January 26, 2016,

Henri Haimi

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Haimi H., Mulas M., Corona F., Vahala R. Data-Derived Soft sensors for Biological Wastewater Treatment Plants: An overview. *Environmental Modelling & Software*, 47:88–107, June 2013.

II Corona F., Mulas M., Haimi H., Sundell L., Heinonen M., Vahala R. Monitoring nitrate concentrations in the denitrifying post-filtration unit of a municipal wastewater treatment plant. *Journal of Process Control*, 23(2):158–170, February 2013.

III Haimi H., Corona F., Mulas M., Sundell L., Heinonen M., Vahala R. Shall we use hardware sensor measurements or soft sensor estimates? Case study in a full-scale WWTP. *Environmental Modelling & Software*, 72:215–229, October 2015.

IV Haimi H., Mulas M., Corona F., Marsili-Libelli S., Lindell P., Heinonen M., Vahala R. Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant. *Engineering Applications of Artificial Intelligence*, revised and submitted, December 2015.

Author’s Contribution

Publication I: “Data-Derived Soft sensors for Biological Wastewater Treatment Plants: An overview”

The author explored, summarized and analyzed the presented case studies. He contributed to the general description of the soft sensor design framework and was mainly responsible for writing the article.

Publication II: “Monitoring nitrate concentrations in the denitrifying post-filtration unit of a municipal wastewater treatment plant”

The author participated particularly in innovating and developing the sample selection procedure. He also contributed to designing the soft sensors and writing the article.

Publication III: “Shall we use hardware sensor measurements or soft sensor estimates? Case study in a full-scale WWTP”

The author was mainly responsible for developing the proposed switching system. He also tested the system’s performance, analyzed the results and wrote the article.

Publication IV: “Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant”

The author had the main responsibility for selecting the methods and for designing and testing the proposed anomaly detection systems. He analyzed the results and wrote the article manuscript.

List of Abbreviations

AdMSPCA	Adaptive Multiscale Principal Component Analysis
AF	Anaerobic Filter
AMWPCA	Adaptive Moving-window Principal Component Analysis
ANFIS	Adaptive Network-based Fuzzy Inference System
ANN	Artificial Neural Network
APCA	Adaptive Principal Component Analysis
ASM	Activated Sludge Model
ASP	Activated Sludge Process
BOD	Biochemical Oxygen Demand
BSM	Benchmark Simulation Model
COD	Chemical Oxygen Demand
CPV	Cumulative Percent Variance
DO	Dissolved Oxygen
DSVI	Diluted Sludge Volume Index
EBPR	Enhanced Biological Phosphorus Removal
EFPCA	Expert-driven Functional Principal Component Analysis
FCM	Fuzzy *C*-Means
FFNN	Feedforward Neural Network
FPCR	Fuzzy Principal Component Regression
FPLS	Fuzzy Partial Least Squares
GK	Gustafson-Kessel algorithm
GLSR	Generalized Least Squares Regression
IPLS	Interval Partial Least Squares
k-NN LLR	*k* Nearest Neighbor Local Linear Regression
KPCA	Kernel Principal Component Analysis
KPLS	Kernel Partial Least Squares
LAMDA	Learning Algorithm for Multivariable Data Analysis

LARS	Least Angle Regression
LASSO	Least Absolute Shrinkage and Selection Operator
LLR	Local Linear Regression
LOO	Leave-One-Out
LTP	Lagoon Treatment Process
MBR	Membrane Bioreactor
MLCA	Multilevel Component Analysis
MLSS	Mixed Liquor Suspended Solids
MPCA	Multiway Principal Component Analysis
MPLS	Multiway Partial Least Squares
MSPCA	Multiscale Principal Component Analysis
MSPLS	Multiscale Partial Least Squares
MWPCA	Moving-window Principal Component Analysis
NNPLS	Neural Network Partial Least Squares
OLSR	Ordinary Least Squares Regression
OP	Orthophosphate-Phosphorus
PARAFAC	Parallel Factor Analysis
PC	Principal Component
PCA	Principal Component Analysis
PCM	Possibilistic *C*-Means
PCR	Principal Component Regression
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
QPLS	Quadratic Partial Least Squares
RAPCA	Reflection-based Algorithm for Principal Components Analysis
RAPLS	Robust Adaptive Partial Least Squares
RF	Random Forest
RMSE	Root Mean Squared Error
SBR	Sequencing Batch Reactor
SOM	Self-Organizing Map
SS	Suspended Solids
SHARON	Single reactor system for High activity Ammonium Removal Over Nitrite
SVI	Sludge Volume Index
TMP	Transmembrane Pressure
TN	Total Nitrogen
TOC	Total Organic Carbon
TOD	Total Oxygen Demand
TP	Total Phosphorus
WWTP	Wastewater Treatment Plant

1. Introduction

1.1 Background

During recent decades, awareness and concern about the negative impact of eutrophication in the quality of water bodies has increased (Ansari et al., 2010; HELCOM, 2013; Ansari and Gill, 2014). Eutrophication is the excessive enrichment of surface waters with nutrients which promotes the high production of especially, algae and cyanobacteria and results in the depletion of oxygen in the water and decomposition of aquatic flora and fauna. The attention drawn to the nutrients in municipal wastewater that cause eutrophication in the receiving watercourses and, on the other hand, the advances in the best available techniques (BAT) have been driving forces to more stringent wastewater treatment requirements and regulations (Olsson et al., 2005; Olsson, 2012). In wastewater treatment plants (WWTPs), the tightening treatment regulations lead towards the addition of new process units, for instance for tertiary treatment purposes, and towards the renewal of the existing units. A typical example the process unit renewal in municipal wastewater treatment is the update of the ammonium removal process towards total nitrogen removal. This is usually achieved through the conversion of the biological reactor from a single aerated tank to a sequence of anoxic and aerobic zones.

The subsequent increase in operational and management investments, mostly associated with energy consumption and chemical dosing, stimulates modern WWTPs to face the challenges of improving effluent quality, while guaranteeing efficient and safe operations and optimizing costs. Achieving these goals is crucially dependent on the high-grade monitoring and control of the process units (Olsson et al., 2005, 2014). For an effective exploitation of advanced monitoring, control and optimization strategies in WWTPs, the real-time availability of primary process indicators is invaluable. Additionally, flexible and controllable actuators are required (Olsson and Newell, 1999).

On-line instrumentation provides the plant operators and automation systems with information on the process variables. These data are also stored in the data acquisition systems of the plant for later analysis and deployment purposes. Already in early publications, the challenges of the reliable field measurements due to the harsh conditions in biological wastewater treatment processes have been highlighted and discussed (Molvar et al., 1976; Olsson, 1977). The typical problems of on-line instrumentation included solids deposition, slime build-up and precipitation, which gave rise to poor performance and to the frequent need for maintenance of the instrumentation. Thereafter, considerable developments in the on-line instrumentation in WWTPs have occurred, providing more real-time information on the evolving process conditions (Gernaey et al., 1998; Vanrolleghem and Lee, 2003; Vanrolleghem et al., 2006; Madsen et al., 2011; Campisano et al., 2013; Olsson et al., 2014). In spite of these developments, the instruments still tend to get fouled in the biological treatment processes, resulting in an inadequate performance of sensors (Olsson, 2012). In addition, all the field instruments are also subjected to down-time, for instance, due to their maintenance. Also, some of the relevant process variables are typically analyzed only in a laboratory, since reliable and moderately-priced real-time instruments are not available. This leads to considerably time-delayed responses unsuitable for the control of dynamic processes. For instance, a survey conducted in Finnish WWTPs indicates that use of on-line measurements of the organic matter content is rather exceptional and in none of the investigated plants were those measurements being used in control loops (Haimi et al., 2009, 2010).

In the future, the introduction of new objectives and regulations may further increase the process control requirements and, hence, the necessity for high-quality real-time information about a growing number of variables. Such objectives may relate, for instance, to priority pollutants in wastewater or greenhouse gases produced in wastewater treatment (Poch et al., 2014). In addition, plant-wide control that merges the control of process units of WWTPs (Olsson and Jeppsson, 2006) and integrated control of the urban sewer-WWTP systems (Benedetti et al., 2013) have been discussed and investigated. When these control perspectives become the dominant reality in WWTPs, for example, by introducing multi-criteria optimization, the importance of reliable measurements of the primary variables increases even more so, in order to avoid defective control actions on plant-wide or system-wide scale.

One solution to the real-time measurement challenges relies on the plentiful historical process data collected in the modern-day WWTPs. Historical data can be utilized together with mathematical modelling algorithms for designing software tools that enhance the usability of the hardware instruments. Data-derived modelling approaches can be used, for instance, to estimate in real-time those variables that are crucial to an efficient process operation; however, their hardware measurements are not sufficiently reliable and in some plants they do not exist at all (Lin et al., 2007; Kadlec et al., 2009; Budka et al., 2014). Another common purpose for data-derived models is for monitoring processes or instruments and, thus providing operators with timely information about malfunctions or chancing process states (Venkatasubramanian et al., 2003; Qin, 2011; Ge et al., 2013). Such computer programs are often called software sensors or soft sensors.

1.2 Objectives and scope of the research

The main motivation for this thesis was to investigate the possibility of supporting the operation of a WWTP by introducing soft sensors. The technical studies of the thesis concern the Viikinmäki WWTP in Finland. It is a large-scale municipal facility where total nitrogen and phoshorus are removed efficiently in a sequence of treatment process units. Both the observations presented in the literature and the practical explorations of the operational data of the plant indicate that the quality of on-line measurements is occasionally inadequate. This diminishes the operational efficiency when such measurements are used in process control and, furthermore, does not motivate towards the inclusion of instruments in control loops. This problem was approached by examining the possibility of using modelling techniques for providing practicable information that complements the data produced by hardware instrumentation. Since the WWTP was well instrumented and, therefore, there was plenty of accessible historical data stored in the data acquisition system, data-derived approaches were used in the studies. The following research questions (RQs) are considered under this objective:

1. What are the dominant trends in data-derived soft sensor applications proposed for wastewater treatment systems and which modelling techniques have been successfully used for soft sensor design?

2. How can historical process data and established real-time measurements be used to provide a tool for supporting efficient process control that conventionally relies on vulnerable field instruments?

3. How can hardware measurements and soft sensor estimates determining the same variable be used in a complementary manner to provide the most reliable information for process monitoring and control?

4. What kind of software tool that has been designed using historical operational data can efficiently detect and diagnose anomalies based on unseen on-line measurements under dynamic process conditions?

As a starting point, Publication I answers RQ 1 by overviewing and analysing a large number of studies where data-derived soft sensors have been presented for biological wastewater treatment processes. Additionally, research gaps were identified at this stage and they guided the choice of the technical applications that were later investigated. Publication II addresses RQ 2 by evaluating the performance of an array of soft sensors designed for the on-line prediction of pollutant concentrations in a biological tertiary treatment unit. Publication III focuses on the development of a switching system that enables the selection between measurements or corresponding estimates in the tertiary treatment process and, thus, provides an answer to RQ 3. Publication IV addresses RQ 4 by investigating the applicability of methodologies for detecting and isolating process and instrument anomalies in a full-scale activated sludge process.

This thesis aims to provide a theoretical backgrounds and practical examples of developing data-derived soft sensors. The soft sensor performances are often studied by using data from pilot-scale processes or by employing simulation platforms. However, the case studies presented in this thesis concern the processes and data of a large-scale WWTP. One of the motivations of the work as a matter of fact is to inspire researchers, consultants and operators to consider implementations of soft-sensing applications in real-life treatment facilities.

Only data-derived modelling techniques are considered for the practical soft sensor development in this thesis. The employed methods in the data-derived modelling techniques belong to multivariate statistics. The feasibility of methods affiliated with the other modelling families is not examined. The presented case studies concern specific process units in a well-monitored municipal WWTP. Only the process units of the wastewater treatment line are investigated and, therefore, the sludge treatment units of WWTPs or sewer networks are not explored in this thesis. The proposed methodologies can without doubt be adapted to a large number of WWTPs, as adequate historical data records are available in present-day facilities equipped with plentiful on-line instruments.

1.3 Outline of the thesis

This dissertation is divided into seven chapters. After this introduction to the thesis, biological wastewater treatment phenomena and process units are briefly presented in Chapter 2, with specific consideration to the WWTP that was considered in the experimental studies. Chapter 3 provides a brief description of soft sensors followed by a general framework for designing data-derived soft sensors. Next, Chapter 4 summarizes case studies where data-derived soft sensors based on multivariate statistical methods have been proposed for biological treatment applications. Chapter 5 describes those multivariate statistical techniques that are employed in the technical studies presented in this thesis. Thereafter, the main results and findings of Publications I–IV are presented and discussed in Chapter 6. Finally, the conclusions arising from the study are set out in Chapter 7.

2. Biological wastewater treatment

This Chapter begins with brief descriptions of the main phenomena that biological wastewater treatment relies on. A particular consideration is assigned to the activated sludge process. Thereafter, the treatment process of the Viikinmäki WWTP is described as it was at the time of the technical studies represented in Chapter 6 of this thesis. The objective of the plant description is also to provide an overview of the process units and real-time measurements of a typical modern WWTP designed for total nitrogen removal. However, an exception to the conventional wastewater treatment line concerns a rather uncommon tertiary treatment unit.

2.1 Introduction to biological wastewater treatment

Today's municipal wastewater treatment aims at reducing concentrations of nitrogen, phosphorus, organic matter (determined, for instance, as Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD) or Total Organic Carbon (TOC)) and Suspended Solids (SS) in influent wastewater. The reason for removing these groups of compounds relates to their negative effect on the qualities of water bodies where urban wastewater is carried. The removal targets are achieved by using several process units, including biological, chemical and physical treatment methods. The typical wastewater treatment units have evolved historically (Angelakis and Rose, 2014) with the major driving forces in the evolution of the process units in WWTPs during the recent decades being the tightening of treatment regulations and changes in the catchment areas of the plants (Dominguez and Gujer, 2006; Neumann et al., 2015).

Usually, the core of the wastewater treatment line is a biological reactor, such as an Activated Sludge Process (ASP, Jenkins and Wanner, 2014). In the ASP, a high concentration of activated sludge consisting mainly of bacteria and protozoa is recycled in cascaded zones under different Dissolved Oxygen (DO) conditions. Primary and secondary clarifiers are applied for separation and thickening of sludge. To maintain sufficient microbiological population in the bioreactors, the thickened sludge from the secondary clarifiers is recirculated into the bioreactor and, additionally, internal sludge recirculation is used in many ASP configurations aimed at total nitrogen removal. Other biological process units are also used for corresponding wastewater treatment purposes, but the ASP is the most popular one in today's plants (Jenkins and Wanner, 2014).

Organic matter removal in biological treatment units is achieved by bacterial conversion to gaseous end-products, such as carbon dioxide in aerobic conditions (Metcalf & Eddy, 2003). Organic matter is also converted into bacterial cells and is removed from the treatment process with the excess sludge. The removal of the suspended particles is usually realized with physical processes such as gravitational settling. The settling process is often facilitated by adding chemicals for coagulating and flocculating the solids. The larger flocs allow for the more efficient separation of solids from liquid by settling for instance in a primary clarifier.

Nowadays, the ASP is especially used for nitrogen removal purposes through the employment of nitrification and denitrification processes. Nitrification is a two-step process taking place in aerobic conditions. Ammonium in the wastewater is, first, converted to nitrite in the presence of ammonium-oxidizing bacteria, which are typically chemoautotrophs such as *Nitrosomonas*. According to Metcalf & Eddy (2003), an approximate equation for this reaction is

$$55\,\mathrm{NH_4^+} + 76\,\mathrm{O_2} + 109\,\mathrm{HCO_3^-} \rightarrow \mathrm{C_5H_7O_2N} + 54\,\mathrm{NO_2^-} + 57\,\mathrm{H_2O} + 104\,\mathrm{H_2CO_3}$$

In the second step of nitrification, nitrite is converted to nitrate by the use of nitrite-oxidizing chemoautotrophic bacteria such as *Nitrobacter*. An approximate equation for this reaction is

$$400\,\mathrm{NO_2^-} + \mathrm{NH_4^+} + 4\,\mathrm{H_2CO_3} + \mathrm{HCO_3^-} + 195\,\mathrm{O_2} \rightarrow \mathrm{C_5H_7O_2N} + 3\,\mathrm{H_2O} + 400\,\mathrm{NO_3^-}$$

The denitrification process, on the other hand, converts nitrate into gaseous nitrogen and is realized by means of nitrate-reducing bacteria, typically in anoxic conditions. There are several genera of bacteria capable of nitrate reduction, being primarily heterotrophs. Denitrification is a two-step process, the first one being the conversion of nitrate to nitrite. In the second step, nitric oxide, nitrous oxide and nitrogen gas are produced (Metcalf & Eddy, 2003), with the reactions of the whole denitrification process being

$$NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$$

Phosphorus removal is generally achieved using chemical precipitation, usually, with ferric or aluminum salts. In addition, bacteria in wastewater treatment processes consume phosphorus in the wastewater and phosphorus is thus incorporated into the bacterial biomass. Process configurations targeting for Enhanced Biological Phosphorus Removal (EBPR, Oehmen et al., 2007) also exist. In the EBPR process, polyphosphate-accumulating organisms that accrue large quantities of polyphosphate within their cells are selectively enriched in the bacterial community, while sludge containing excess phosphorus is removed from the process. In the EBPR process, an additional anaerobic process stage and a more sophisticated sludge recirculation scheme are required in contrast with an ASP that aims at total nitrogen removal and chemical phosphorus precipitation.

2.2 Description of the Viikinmäki WWTP

The Viikinmäki WWTP with its over 800 000 population equivalent is the largest municipal plant in the Nordic countries. The plant is built inside bedrock and it treats wastewater sewered from the Helsinki metropolitan area. An average influent flow rate is approximately 250 000 m^3/d of which about 85% is domestic and 15% industrial wastewater. The wastewater treatment line of the plant comprises bar screening, grit removal, pre-aeration, primary sedimentation, activated sludge process, secondary sedimentation and denitrifying post-filtration. The sludge treatment is achieved with mesophilic digesters and subsequent sludge dewatering systems. The biogas from the sludge digestion is utilized for electricity and heat production, which covers about 50% and 100% of their demand in the plant, respectively. Yearly averages of total nitrogen removal of approximately 90%, total phosphorus removal of 95% and Biochemical Oxygen Demand (BOD_7) removal of 95% are achieved in the Viikinmäki WWTP. The wastewater treatment line of the plant with the primary online measurements is depicted in Figure 2.1. not all the measurements, for instance for the flow rates, are included in the figure. The primary treatment units, secondary treatment unit (ASP), tertiary treatment unit and sludge treatment units of the plant are briefly described in the following.

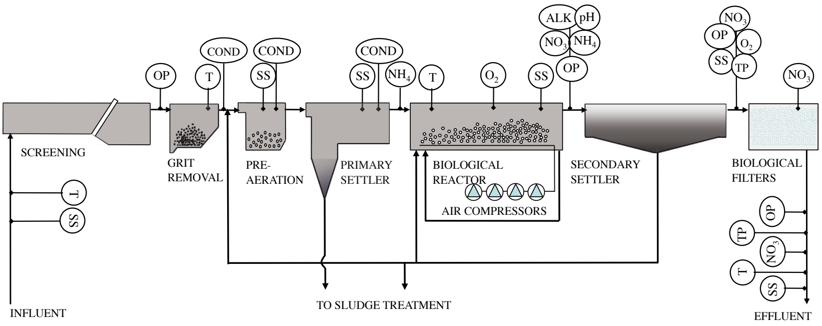


Figure 2.1. Simplified layout for the wastewater treatment line of the Viikinmäki WWTP.

Primary treatment units

Bar screens are mechanical filters that consist of a series of vertical steel bars. The aim of the bar screening is to remove coarse solid waste that would disturb the operation of the process units downstream. The on-line measurements of SS and temperature are performed before the screening and of orthophosphate-phosphorus (OP) after the process unit.

Grit removal is realized in rectangular basins that are aerated from the other long side. Grit is removed from the bottom of the basins with scrapers whereas grease and oil are scraped from the surface of the water in zones where turbulence is dampened. Ferrous sulphate and, when needed, lime are added at the beginning of the grit removal basins. Ferrous sulphate is diluted in water and dosed for precipitating the soluble phosphorus that wastewater contains. The ferrous ions (Fe^{2+}) are oxidized into ferric ions (Fe^{3+}), which effectively precipitate the phosphate ions (PO_4^{3-}) corresponding for the vast majority of the soluble phosphorus in the influent wastewater. The precipitated solid phosphorus is then removed from wastewater together with other solids in the subsequent process units. Lime is added in wastewater to maintain the sufficient alkalinity level particularly for the bacteria in the activated sludge process that is located further in the treatment line. The purpose of grit removal is to decrease wear and need for maintenance on actuators, pipelines and instrumentation downstream in the treatment line. Further, the removal of grease and oil prevents disturbances in the biological process and worsening of the sludge settling properties. A temperature measurement is installed in the grid removal basin and a conductivity measurement after the unit.

After grit removal, wastewater is fed into a rectangular pre-aeration basin. Excess sludge from the secondary sedimentation and reject water from sludge treatment are mixed with wastewater before the pre-aeration unit. The purpose of pre-aeration is to equalize the quality and flow rate of wastewater entering the primary sedimentation lines, to ensure sufficient mixing of chemicals and to increase DO concentration. The on-line measurements of SS and conductivity are conducted in the pre-aeration phase.

Primary sedimentation consists of seven treatment lines, each of which are divided into two basins. The settled sludge is collected with scrapers at the bottom of the basins into sludge pockets located at the beginning of the tanks, which are also used to thicken the excess sludge. From the sludge pockets, the thickened sludge is further pumped to the sludge treatment. The scrapers are also used for removing surface sludge in the basins to the surface sludge wells. The treated wastewater is collected from the surface into overflow chutes and carried to the sequential process units. If the sequential biological process units are partly by-passed due to the too high influent flow rate, polymer and polyaluminum chloride are added in the primary sedimentation in order to increase the efficiency of mechanical and chemical treatment. Apart from conductivity, the on-line measurements performed in the primary sedimentation units are listed as the influent measurements to the bioreactor in Table 2.1.

Activated sludge process

The activated sludge process consists of a bioreactor and secondary sedimentation units, which are schematically represented in Figure 2.2. At the time of the investigation, the ASP was divided into eight treatment lines (the ninth ASP line has been introduced recently).

Each line begins with a non-aerated mixing zone where pre-settled wastewater, return sludge from two secondary sedimentation basins and internal recycle sludge flow are fed and which are mixed mechanically with agitators. Next, the reactors are split in six cascaded zones, with the anoxic zones enabling denitrification located near the input. The anoxic volume defined by the number of the non-aerated zones is flexible. It depends on the aeration mode, which is controlled in such a way that

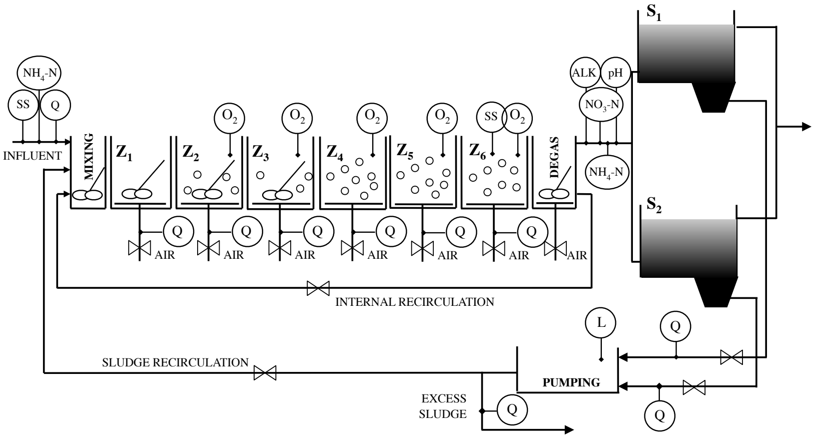


Figure 2.2. Simplified layout for a single ASP line and location of on-line measurements.

the effluent ammonium-nitrogen (NH_4-N) concentration is within the set target range while using the minimum required aerated volume. Time-delays are also included in the aeration mode control scheme in order to increase the stability of the control. In practice, Zone 1 is never aerated and is mixed mechanically. Zones 2 and Zone 3 are equipped with agitators and are either aerated or non-aerated (and mechanically mixed) depending on the aeration mode in use. In contrast, Zones 4–6 are always aerated. The aeration of the zones is realized with air compressors that blow air through disc-shaped fine bubble diffusers located at the bottom of the basins. In addition to nitrogen removal, the amount of organic matter in wastewater is reduced both in the aerobic and anoxic zones of ASP. Some ferrous sulphate is added to the degassing zone located after the last aerobic zone in order to complete the phosphorus removal of the targeted level. The bioreactor including its influent and effluent is amply monitored with the on-line measurements shown in Figure 2.2 and collected with the TAGs and the units of measurement in Table 2.1.

The operational objective of the secondary sedimentation process is to separate activated sludge from the wastewater and return an appropriate amount of the settled sludge into the bioreactor. The mixed liquor flows to the secondary sedimentation basins from the degassing zones by gravity. The treatment lines are divided into two rectangular basins for each ASP line. The mixed liquor is carried in the middle of the basins from where it flows to both ends. The sludge is collected with scrapers at the bottom of the basins to sludge pockets located in the middle of the tanks. From the sludge pockets, the thickened sludge is carried to the return sludge pump-

Table 2.1. Process variables in the ASP.

TAG	Description	Unit
BI-NH_4	Bioreactor influent ammonium-nitrogen	mg/l
BI-SS	Bioreactor influent suspended solids	mg/l
BI-Q	Bioreactor influent wastewater flow rate	m^3/s
$Z2$-$O2$	Dissolved oxygen in zone 2	mg/l
$Z3$-$O2$	Dissolved oxygen in zone 3	mg/l
$Z4$-$O2$	Dissolved oxygen in zone 4	mg/l
$Z5$-$O2$	Dissolved oxygen in zone 5	mg/l
$Z6$-$O2$	Dissolved oxygen in zone 6	mg/l
$Z6$-SS	Mixed liquor suspended solids in zone 6	g/l
BE-NH_4	Bioreactor effluent ammonium-nitrogen	mg/l
BE-NO_3	Bioreactor effluent nitrate-nitrogen	mg/l
BE-pH	Bioreactor effluent pH	–
BE-ALK	Bioreactor effluent alkalinity	mmol/l
QA	Internal recycle flow rate	m^3/s
$S1$-QR	Return sludge flow rate from settler 1	dm^3/s
S2-QR	Return sludge flow rate from settler 2	dm^3/s
QW	Excess sludge flow rate	dm^3/s

ing station and from there pumped further to the ASP's mixing zone and part of the sludge to the pre-aeration as excess sludge. Surface sludge is removed by using separate scrapers. The clarified wastewater is collected from the surface to overflow chutes at both ends of the basins and carried further to the post-filtration process.

Denitrifying post-filtration process

The operational objective of the post-filtration treatment unit is to remove nitrate-nitrogen (NO_3-N) contained in wastewater after the activated sludge treatment by means of denitrification. The post-filtration unit receives wastewater from the secondary sedimentation. It consists of ten Biostyr filters arranged in parallel, as depicted in Figure 2.3a. The wastewater is equally distributed to ten filter cells and methanol diluted in water is independently added to each line to provide an external carbon source to enhance denitrification (see, Figure 2.3b). The methanol flow rate in each line is manipulated by a feedback loop that controls the NO_3-N concentration in the cell outlet. The NO_3-N concentration in the cell is measured *in situ* by means of an optical instrument.

Inside the cell, wastewater flows upwards through floating polystyrene support media, on which biomass is attached. Due to the biomass attachment, periodic backwashes are needed. The cells are usually backwashed one at a time using the effluent wastewater with a counter-current air

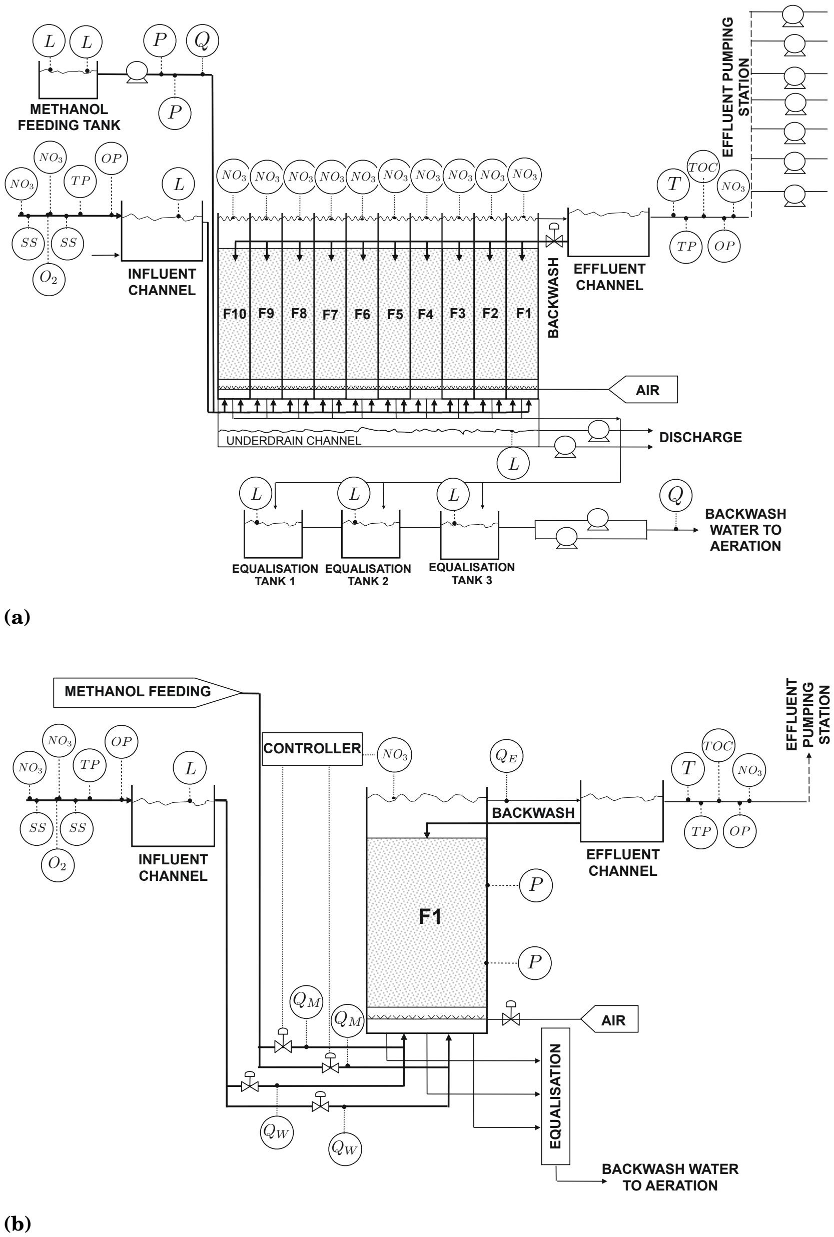


Figure 2.3. Schematic representation of the post-denitrification filtration unit (a) with a highlight on one filter (b).

Table 2.2. Process variables in the post-filtration unit.

TAG	Description	Unit
PI-NO_3-N(1)	Post-filtration influent nitrate-nitrogen (sensor 1)	mg/l
PI-NO_3-N(2)	Post-filtration influent nitrate-nitrogen (sensor 2)	mg/l
PI-SS(1)	Post-filtration influent suspended solids (sensor 1)	mg/l
PI-SS(2)	Post-filtration influent suspended solids (sensor 2)	mg/l
PI-O_2	Post-filtration influent dissolved oxygen	mg/l
PI-OP	Post-filtration influent orthophosphate-phosphorus	mg/l
PI-TP	Post-filtration influent total phosphorus	mg/l
Fi-QWW	*i-th* Filter backwashing water flow rate	m^3/s
Fi-QWA	*i-th* Filter backwashing air flow rate	m^3/s
Fi-QW(1)	*i-th* Filter wastewater flow rate (line 1)	m^3/s
Fi-QW(2)	*i-th* Filter wastewater flow rate (line 2)	m^3/s
Fi-QM(1)	*i-th* Filter methanol flow rate (line 1)	m^3/h
Fi-QM(2)	*i-th* Filter methanol flow rate (line 2)	m^3/h
Fi-P(1)	*i-th* Filter pressure at the bottom	kPa
Fi-P(2)	*i-th* Filter pressure at the top	kPa
Fi-NO_3-N	*i-th* Filter effluent nitrate-nitrogen	mg/l
Fi-HL	*i-th* Filter head loss	m
Fi-CR	*i-th* Filter clogging rate	%
Fi-HRU	*i-th* Filter hour in use	0-1
Fi-ITW	*i-th* Filter intermediate time of backwash	0-1
PE-NO_3-N	Post-filtration effluent nitrate-nitrogen	mg/l
PE-TOC	Post-filtration effluent total organic carbon	mg/l
PE-OP	Post-filtration effluent orthophosphate-phosphorus	mg/l
PE-TP	Post-filtration effluent total phosphorus	mg/l
PE-T	Post-filtration effluent temperature	° C

flow. The backwash water is pumped to the pre-aeration unit. After filtration, the treated wastewater is discharged into the effluent channel where streams from Filter 10 to 1 are collected. The unit is well monitored with the on-line measurements illustrated in Figure 2.3 and listed with the TAGs and the units of measurement in Table 2.2.

Sludge treatment

The excess sludge from primary and secondary sedimentation is treated by anaerobic digestion, which is a bacterial process carried out in the absence of oxygen. In the Viikinmäki WWTP, four mesophilic digesters are used for the sludge treatment. Digestion is realized as a two-step process where the sludge goes through two digesters. In the sludge digestion, complex proteins and sugars are broken down to form simpler compounds. The process reduces the total mass of solids, destroys pathogens and produces biogas, which is collected. Sludge in the digesters is mixed with mechanical agitators and recycle pumping. Anti-foaming chemical is dosed in the digesters when needed. The total solids concentration of the influent to the digestion process is measured in real-time.

After the anaerobic digestion, the sludge is pumped to intermediate storage basins where the digestion process is stopped by means of aeration and from there pumped further to dewatering which is realized with centrifuges. Polyelectrolytes are used to enhance the dewatering of the sludge. The polyelectrolyte feed is controlled by the on-line measured total solids load into the centrifuges with a feedforward strategy.

The dewatered sludge is processed into soil by composting. The biogas produced in digestion is used for producing electricity and heat at the plant's power station, which consists of gas power engines, boilers and an organic Rankine cycle system. The nitrogen-rich reject water from sludge dewatering is fed first into equalization basins and, then, to sedimentation basins. The sludge separated in the reject water sedimentation process is pumped back into the intermediate storage basins and the clarified reject water back into the pre-aeration unit of the wastewater treatment line.

3. Data-derived soft sensors

Software sensors or soft sensors are virtual instruments that can be used for similar purposes as their hardware counterparts. They are computer programs and a model at their core will process information produced typically by hardware instruments. On the basis of their internal model, soft sensors are often divided into two main classes, phenomenological (also called mechanistic, first-principles-driven, theoretical, deterministic or white-box) and data-derived (also called data-driven, data-based, empirical, process-history-based or black-box). Phenomenological soft sensors are based on first-principle process models, whereas data-derived soft sensors are built around process models derived from historical data. Section 3.1 introduces the main modelling classes with emphasis on their use in the wastewater treatment systems. The principles of the data-derived soft sensors and their typical applications are also described. Section 3.2 provides a general framework for designing data-derived soft sensors. The design steps and a number of potential techniques that can be applied to them are briefly introduced and discussed.

3.1 Introduction to soft sensors

This section first discusses the phenomenological models specifically developed for wastewater treatment and their potential for real-time soft sensor applications in the field of industry. Then, the basic concepts of data-derived modelling and its employment in soft-sensing with a focus on the wastewater treatment sector are provided.

In wastewater treatment, the most commonly used first principle models for the biological treatment processes belong to the Activated Sludge Model family (ASM, Henze et al., 2000). The clarifying and thickening processes taking place in settlers are often described using the Takács model (Takács et al., 1991) in engineering practice. Further, a number of alternative one-dimensional settling models have been proposed by researches (Li and Stenstrom, 2014), including also reactive settler models (Gernaey et al., 2006). As for phenomenological models for sludge treatment, the Anaerobic Digestion Model is the most famous contribution (Batstone et al., 2002).

Phenomenological models are capable of describing both linear and nonlinear phenomena and of providing information on the internal states of the process. The detailed phenomenological modelling approach has proven to be efficient, for example, in wastewater treatment process design, renovation, employee training, optimization of the plant operation, and in understanding the behaviour of the system and interactions of the components (Gernaey et al., 2004; Hauduc et al., 2009; Phillips et al., 2009; Brjdanovic et al., 2015). However, there are major challenges in using the first-principle models for real-time applications. Characterizing the organic matter and determining the rate constants for the volatile fatty acid uptake is challenging, expensive and time-consuming and, yet, fundamental to the successful calibration of the model (Dochain and Vanrolleghem, 2001; Petersen et al., 2003; Hauduc et al., 2011; Choubert et al., 2013). The models are calibrated for certain operational conditions, often for dry weather circumstances, which diminishes their on-line use under the varying process states (Gernaey et al., 2004). The theoretical limitations concerning several phenomenological activated sludge models have also been recently reported by Hauduc et al. (2013). Moreover, the high-dimensionality of detailed phenomenological models results in enormous computational requirements and ill-conditioned problems due to the interaction between fast and slow dynamics (Dochain and Vanrolleghem, 2001).

The large amount of process data are routinely measured in real-time and collected in modern-day WWTPs. The on-line measurements and the stored historical operational data permit data-driven modelling as an interesting alternative for soft sensor design (Haimi et al., 2013c). Today, data-derived soft sensors are becoming more common in the wastewater treatment sector, even though they are still not as widespread as, for instance, in the process industry where soft sensors are extensively exploited and have shown great potential (see e.g. Fortuna et al., 2007; Kadlec et al., 2009; Kadlec, 2009; Slišković et al., 2011a). Data-derived soft sensors have also been developed for a considerable number of applications in mineral processing (González, 2010), building systems (Li et al., 2011), bioprocessing (Luttmann et al., 2012), electronics (Liukkonen et al., 2012), the pharmaceutical industry (Gernaey et al., 2012), in the manufacture of bio-therapeutics (Mandenius and Gustavsson, 2014) and in the steel industry (Kano and Nakagawa, 2014).

A data-derived soft sensor is conventionally described as an input-output process model. The model inputs typically consist of *secondary variables* that are easy to measure reliably, with reasonable costs (often stated as *easy-to-measure* variables). The inputs are in the form of measurements of the plant and, sometimes, numerically encoded expert knowledge. The model outputs consist of information associated with those *primary variables* whose reliable measurement is challenging or high-priced (often stated as *difficult-to-measure* variables). The information that the input and output variables contain is modelled empirically in the soft sensor. A data-derived soft sensor model with its inputs (six variables in this example) and output (one variable) is sketched in Figure 3.1.

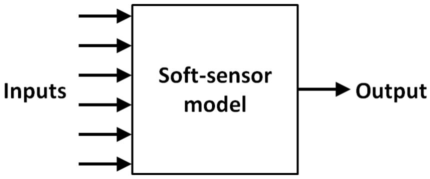


Figure 3.1. Data-derived soft sensor described as an input-output model.

The range of tasks that can be performed by data-derived soft sensors is broad. The original application area of soft sensors is the *on-line prediction* of process variables that can only be measured at low sampling rates or off-line. Another motivation for designing soft sensors for on-line prediction tasks is the need for a back-up system for on-line hardware measurements, which are crucial for the safe and successful operation of a system. The input-output relationship is encoded in the historical data, which are used to calibrate the soft sensor model. The calibrated model is used for reconstructing the input-output relationship and it estimates the output variables once new inputs are available. This type of soft sensors addresses a supervised learning problem in the form of regression or classification. Other typical application areas are related to *monitoring the state of the process* and to *monitoring the state of the instrumentation*. In these cases, the outputs are information on the operation of the process and the instruments, in the form of diagnostics and status characterization. Soft sensors of this kind usually address an unsupervised learning problem in the form of dimensionality reduction or clustering.

3.2 Framework for designing data-derived soft sensors

This section addresses the practical steps to be undertaken in the design of data-derived soft sensors. Also, the most commonly used data-derived modelling techniques are briefly discussed. An overview of the design procedure is given in Figure 3.2. The procedure consists of several independent steps: *data acquisition*, *data pre-processing*, *model design* and *model maintenance*. Soft sensor development is an iterative process where choices made during the design procedure often need to be reconsidered before the soft sensor is ready for deployment. A framework for developing data-derived soft sensors has also recently been presented by Budka et al. (2014).

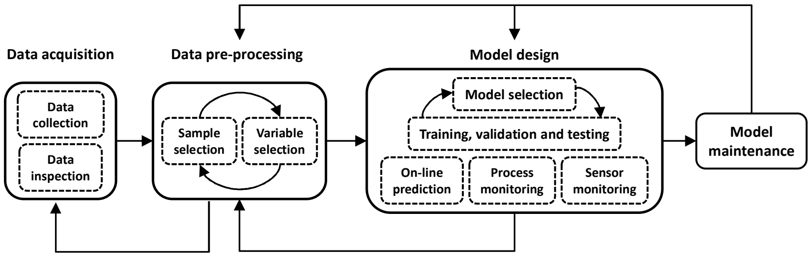


Figure 3.2. Overview of the design steps for data-derived soft sensors. Figure adopted from Publication I.

3.2.1 Data acquisition

The historical process and laboratory data are routinely stored in the data acquisition system of WWTPs and are easily retrieved. *Data collection* and subsequent *data inspection* are the first steps in soft sensor development. During the initial inspection, a preliminary exploration of the measurements is performed in order to obtain an overview of the prominent structures in the data and to identify the presence of obvious problems (*e.g.*, locked measurements, missing and drifting data and measurements outside the operating range of the instruments). Periods of instrument calibrations and process unit maintenance are also annotated together with a selection of representative operations. The data inspection typically requires a large amount of manual work and expertise in the underlying processes. It typically includes an extensive investigation of time series, scatter plots and histograms of data.

3.2.2 Data pre-processing

The remarkable characteristics of the data acquired in wastewater treatment facilities are redundancy and, possibly, insignificance. Disturbances that corrupt the measurements are also sometimes present. Often, the amount and quality of the data together with their high dimensionality can be a limiting factor for soft sensor development. Therefore, it is necessary to prepare the data before they are processed by the soft sensor's model. Process understanding and *a priori* knowledge is required in this phase (Kadlec and Gabrys, 2009). Such knowledge can be supported and complemented by many statistical techniques for *variable selection* and *sample selection*. Applicable methods for the data pre-processing have also been discussed by Slišković et al. (2011b).

Variable selection
The choice of the input variables is a crucial stage. Variable selection consists of choosing those secondary variables that are the most informative for the process being modelled, as well as those that provide the highest generalisability. This step is fundamental because models are built from a finite number of observations and having a model with too many inputs, may lead to over-fitting and give rise to a large computational burden. In addition, describing a process in terms of a few selected variables allows one to retain interpretability.

The most commonly used techniques for variable selection are often categorized as *filtering*, *wrapping* and *embedded* methods (Guyon and Elisseeff, 2003; Lu et al., 2014). Embedded methods perform the variable selection inside the soft sensor's model and they are overviewed in Subsection 3.2.3. Filters and wrappers select variables and subsets of variables by ranking them on the basis of their significance for the task. The wrappers combine two independent elements: (i) a *relevance criterion*: to score an input variable or a group of input variables according to its informative power; and (ii) a *search procedure*: for finding from among all the available variables the subset that optimizes the chosen criterion. The criterion is often based either on statistical dependence or model accuracy measures. Given a criterion, the simplest strategy for variable selection consists first of scoring all inputs and ranking them accordingly. A threshold for the score is then set and only the variables that score more than the threshold are retained.

In many situations, the computational applicability of criterion-search schemes is limited by the too large number of potential inputs. An alternative approach to reduce the dimensionality of the problem consists of producing a small number of combinations of the original variables. New variables may be constructed on the basis of process knowledge (*e.g.*, using balance calculations and averaging) or they can be derived using statistical projection methods for dimensionality reduction (Lee and Verleysen, 2007). Linear multivariate techniques such as Principal Component Analysis (PCA, Jolliffe, 2002) and Partial Least Squares (PLS, Wold et al., 2001) are the most common methods for learning a low-dimensional representation of a set of data. PCA and PLS techniques are described in Sections 5.1 and 5.3, respectively. The conventional PCA and PLS methods are commonly used for continuous processes, but when addressing batch processes, the multiway nature of data is more appropriately processed with the multiway variants of the techniques (Smilde et al., 2004).

Sample selection

When real data are analyzed, it is common that some observations are different from the majority. Such observations are often called outliers. They may be due to data acquisition mistakes or they correspond to exceptional process circumstances. In general, one can distinguish between two main types of outliers: i) *obvious outliers* are observations that violate physical or technological limitations and ii) *non-obvious outliers* are observations that do not violate any threshold but still fall outside of typical ranges. Sample selection consists of discarding or pinpointing outlying observations, since they are not necessarily representative of normal operations and because their use may impair the performance of the soft sensor model (Rosen et al., 2003). Alternatively, sample selection consists of choosing only those observations that are truly representative of the normal operation of the processes and instruments being modelled.

Multivariate statistics such as PCA and PLS, coupled with a model residual analysis are frequently used in sample selection (Robinson et al., 2005). However, these classical models are sensitive to outliers in the data and to overcome such a limitation, their robust extensions should be used instead (Rousseeuw and Hubert, 2011). Again, the classical approaches are most conveniently used for continuous processes, whereas the multiway variants are more suitable for batch processes (Hubert et al., 2012).

A conceptually different approach to sample selection is the application of *clustering* and *classification* methods (Hastie et al., 2009). Clustering and classification techniques aim at reducing the amount of data by grouping the observations into subsets, either clusters or classes, that consist of observations similar to each other but different from observations in other subsets. Clustering is essentially an unsupervised learning problem, whereas classification is an analogous supervised problem based on pre-labelled data. The most popular techniques for clustering are based on the k-means algorithm (Hartigan and Wong, 1979) and on fuzzy c-means (FCM, Bezdek et al., 1984). Classification, on the other hand, is typically performed using methods such as Linear Discriminant Analysis (Fisher, 1936), Artificial Neural Networks (Haykin, 1999) and Support Vector Machines (Cristianini and Shawe-Taylor, 2000).

3.2.3 Model design

Model design is a critical step in soft sensor development. In particular, the model structure defines the specific application task and the selection of the model parameters determines the generalization ability of the soft sensor. However, a consistent approach to the task does not exist. Instead, the model structure and parameters are often selected in an *ad hoc* manner for each soft sensor (Kadlec et al., 2009). This is, firstly, due to the fact that model design depends on the task at hand and, secondly, it is often subjected to the developer's past experience and personal preference.

Despite the lack of a consistent approach to model design, two main tasks can be recognized: i) *model structure selection* and ii) *model training, validation and testing*. The common practice suggests starting with simple model types, assessing their performances and then gradually increasing complexity, as long as significant improvements are observed. Furthermore, it is important that the models are not only accurate, but also computationally efficient, interpretable and with a low maintenance cost. In the following, the most popular model structures are briefly overviewed and then the optimization of their parameters discussed.

Model structure selection

i. *Models for on-line prediction*: Such models address the problem of reconstructing the functionality existing between the inputs that ideally are low-cost measurements with a good analytical performance and the outputs that often are difficult to measure precisely or with reasonable costs. Usually, the inputs and the outputs take continuously varying real values and their relationships can be modelled as a regression problem. Less common is the case where the outputs take categorical values. In this case, their relationship can be modelled as a classification problem (Duda et al., 2000).

The simplest regression techniques assume the existence of a linear input-output relationship and they fit a linear model to reconstruct it. Most of the commonly used techniques belong to *multivariate statistics* (Anderson, 2003), which examine relationships among multiple variables at the same time. Ordinary Least Squares Regression (OLSR, Ryan, 2008), also known as Multiple Linear Regression, is based on the ordinary least squares approximation of the linear model, which is simple and sometimes accurate. The accuracy and interpretability of OLSR can be improved by shrinking the regression coefficients, which is achieved with models that also perform an *embedded variable selection* scheme (Hastie et al., 2009). The most commonly used subset selection methods are Best Subset Selection, Forward and Backward Stepwise Selection, Forward Stagewise Regression, whereas popular shrinking methods are Ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO) and Least Angle Regression (LARS). In situations with a large number of inputs, multivariate statistical methods combining linear projection and linear regression can be used to reduce the dimensionality of the modelling problem, at the price of interpretability. The most commonly used in this category of methods are Principal Component Regression (PCR) and PLS regression (PLSR).

A number of variants of the classical methods have been proposed in order to improve the models' performances in different applications. Adaptive and recursive extensions of PCR and PLSR (see Kadlec et al., 2011) can be used for capturing the dynamic nature of process data. Typically in these methods, the model is updated or re-constructed when a new data sample or a block of new samples are available. The recursive methods update the data matrices by including new data according to certain weights, which in the course of time are exponentially decreasing so that old data are increasingly disregarded in comparisons with the new data (Li et al., 2000). The Multiscale PCA (MSPCA, Bakshi, 1998) is a method for separating the data in different time scales and, essentially, it is a combination of multiresolution analysis (Strang and Nguyen, 1997) and PCA.

In the Adaptive Multiscale PCA (AdMSPCA, Lennox and Rosen, 2002), the models at each scale are made adaptive. Nonlinear kernel extensions like KPCA and KPLS (Rosipal and Trejo, 2001) can be used in the presence of nonlinearities. In these methods, kernel functions, for instance polynomial kernels, are used in the nonlinear mapping of the original variables and, then, the originally linear operations of PCA or PLS are performed (Schölkopf et al., 1998). Apart from the kernel extensions, several other nonlinear generalizations applicable to PCA and PLS have also been presented (see Jolliffe, 2002; de Leeuw, 2014, for references).

Other nonlinear methods do not necessarily rely on any assumption on the input-output relationships. Nevertheless, they are widespread among researchers and practitioners. Methods based on supervised *artificial neural networks* (ANN, Haykin, 1999) and *neuro-fuzzy systems* (Fullér, 2000) are among the most popular.

An ANN is a network of artificial neurons arranged in layers and connected to each other. The neurons nonlinearly transform the incoming signals using an activation function and, then, they distribute the result to the other neurons. The input-output relationship is encoded in the connection weights, which are adapted to minimize the error between the network outputs and the targets. In particular, Feedforward Neural Networks (FFNN, Haykin, 1999) have been popular in the wastewater treatment sector.

Neuro-fuzzy systems combine the features of ANNs with the human-like reasoning style of fuzzy systems, aiming at complementary techniques and enhanced performance compared with the individual methodologies. Typically in neuro-fuzzy systems, the first layer corresponds to input variables, the middle layers encode fuzzy IF-THEN rules and the last layer corresponds to the output variables. The advantages of neuro-fuzzy systems include the ability of the ANN learning algorithms to learn both fuzzy sets and fuzzy rules, as well as the potential to use *a priori* knowledge. In particular, an Adaptive Network-based Fuzzy Inference System (ANFIS, Jang, 1993) has been popular among soft sensor designers in the wastewater treatment industry.

ii. *Models for process and sensor monitoring*: Such models address the problem of detecting, identifying and diagnosing normal and abnormal behaviours in the processes and in the field instruments, using the easy-to-measure process variables as inputs. Usually, no prior information about the outputs is available and it must be extracted from data using dimensionality reduction and clustering approaches. In the less common case where the output information is available, either as real or categorical values, the input-output relationship can be modelled as a regression or a classification problem, respectively.

In order to detect the occurrence of any variation having an exceptional cause, univariate statistical control charts have been traditionally used to monitor a small number of process variables. Examining one variable at a time, as though they were independent, however, makes interpretation and diagnosis extremely difficult in environments where a large number of variables continuously vary relative to one another. However, when the number of variables is large, one often finds that they are also highly dependent on one another.

In the monitoring of the continuous processes and the hardware sensors, the conventional and adaptive PCA and PLS methods are popular for reducing the dimensionality of the variables of interest. An established technique to isolate the variable(s) responsible for the detected anomalies is to study their contributions to the model residual statistics. Another traditional monitoring approach with PCA is to use low-dimensional scatter plots defined by the most significant principal components, which include most of the information of the original variables. In such a way, the transitions in the process or in the relationships between the supervised sensors can be observed. When using PLS for process monitoring, the output variables are usually difficult, or impossible, to determine in real-time and they indicate the presence of anomalous situations. In wastewater treatment, such variables are, for instance, indicators of the sludge settling properties determined by field and laboratory experiments.

In the batch process, an additional dimension to the data structure is addressed by the batch, the other dimensions representing time and the variables. The Multiway PCA (MPCA) and PLS (MPLS) are commonly used for dimensionality reduction when multiway data are considered, for instance in the cases of the batch processes (Smilde et al., 2004). The multiway extensions first unfold a three-dimensional data structure into a two-dimensional structure and, then, PCA or PLS is executed. The multiway methods are sometimes called Unfold or Unfolding methods. Multilevel Component Analysis (MLCA, Timmerman, 2006) is an extension of PCA that is useful if the variation in the data occurs on different levels simultaneously. For monitoring the batch processes, MLCA enables separate interpretations of the transitions both within the batches and between the batches in low-dimensional subspaces.

The Self-Organizing Map (SOM, Kohonen, 2001) is the most common unsupervised ANN method and it has also been used in many wastewater treatment applications. In the model training, the neurons adapt themselves to the relationships within a set of input signals and the SOM outputs a low-dimensional (usually two-dimensional) representation of the patters encoded in the training data. In this representation, clusters corresponding to the characteristic features of the data are formed onto a topographic map that provides an interpretation of the input information.

Clustering methods (Everitt et al., 2011) are applied for monitoring using the process variables or new variables created for instance by PCA approaches as the model inputs. The observations among the training data are grouped in the clusters based on their similarity. In the conventional clustering approaches based, for example, on the k-means algorithm, each observation belongs to one of the clusters. Instead, in fuzzy clustering methods, each observation belongs to all the clusters to some extent, represented by their fuzzy memberships. When introducing unseen data, their discrete properties can be observed by monitoring their transition between the clusters defined in the model training step.

Model training, validation and testing

Most of the model types discussed in this subsection are characterized by a number of basic parameters and a number of meta-parameters that define their structure and optimize it in terms of its generalization performances. The basic parameters of the models are, for instance, the regression coefficients of linear regression methods, the connection weights of neural and neuro-fuzzy systems, the loading components in multivariate statistical methods like PCA and PLS. The meta-parameters are, on the other hand, the number of components to be retained in methods like PCA and PLS, the regularization parameter in linear shrinkage methods like LARS and LASSO, the number of neurons and layers in the ANN and neuro-fuzzy systems, and the number of clusters, among others. Before a model is able to operate on new unseen observations, it has to be trained to estimate its basic parameters and it has to be validated to optimize its meta-parameters. Model validation is a highly important step in soft sensor development, in which the designer estimates how well the model will perform on the unseen data.

Ideally, if enough data were available, the soft sensor developer would set aside a validation set and use it to assess a model whose basic parameters are calibrated on a training set, for the different values of its meta-parameters. After finding the optimal set of meta-parameters, the developer would then calibrate the model to set its basic parameters, using all the available training data. The resulting model is eventually assessed on an independent testing data set.

However, it may be difficult to obtain a sufficient amount of historical data for training the model according to the aforementioned procedure. In such a situation, the soft sensor developer has to rely on error estimation techniques, like the simple and widely used *cross-validation* (Hastie et al., 2009). K-fold cross-validation uses part of the training data to calibrate the model and a different part to validate it. The procedure consists of, firstly, splitting the data in K roughly equal-size parts, secondly, to set aside the k-th part and calibrating the model to the other K-1 parts and, thirdly, to calculate a measure of model accuracy over the k-th part. After repeating the procedure for all the K parts, the accuracies are combined to give an average performance of the model, for a specific set of meta-parameters. The model whose meta-parameters have the best generalization accuracy is finally trained and assessed against a testing data set.

3.2.4 Model maintenance

After the successful design and implementation, it is not uncommon to observe a degradation of the performance of a data-derived soft sensor. Such degradation is often due to changes in the process and instrumental characteristics or operating conditions. In the wastewater treatment applications, the reason for this may be, for example, variations in the influent wastewater composition, temperature and flow rate, instruments recalibrations or operational changes inside the plant. To overcome such limitations, soft sensors should be regularly maintained and updated as the system characteristics change, but their manual and repeated redesign should be avoided due to the heavy workload.

Many of the soft sensors currently found in full-scale environments do not provide any automated mechanisms for their maintenance. A suggested improvement for regular soft sensor maintenance is to statistically analyze the residuals between the soft sensor estimates and the hardware instrument measurements if they are available and, then, to perform the maintenance only when the analysis indicates that it is required (Abonyi et al., 2014). However, this approach also necessitates a lot of manual work.

To automatically cope with the changes in process characteristics and operating conditions, a number of data-derived approaches have been designed and are available for the soft sensor developer (Kadlec et al., 2011). The majority of these approaches are inherently encoded in the adaptive and recursive versions of multivariate statistical methods like PCA and PLS. Recently, Kaneko et al. (2014) proposed the use of multivariate statistical process control to select an adequate adaptive PLS-based soft sensor out of models employing three different adaptation techniques: moving-window, just-in-time and time difference. Algorithms for designing adaptive soft sensors based on just-in-time models have been reviewed by Saptoro (2014). An approach related to the neuro-fuzzy methods also providing adaptation possibilities is local learning (Atkeson et al., 1997). An adaptive soft sensor developed in this framework was published in Kadlec and Gabrys (2008). An alternative for the adaptive maintenance approaches was presented by Chen et al. (2015) who developed a Kalman filter (Kalman, 1960) based model mismatch index and a procedure for updating soft-sensor PLS model on-line when significant degradation occurs. In addition, Fujiwara et al. (2009) and Zhu et al. (2011) have discussed the development of maintenance-free soft sensors for on-line prediction using local linear regression methods.

4. Applications of multivariate statistics -based soft sensors in biological wastewater treatment

In this Chapter, an overview of case studies where soft sensors have been proposed for biological wastewater purification is given. Since the experimental research in this thesis concerns soft sensor development using multivariate statistics, only the works where those methods have been exploited are considered. However, other data-derived modelling families have also been used in soft sensor design. Such families are, for example, artificial neural networks (Capodaglio et al., 1991; Bongards, 2001; Çinar, 2005; Ráduly et al., 2007; Lee et al., 2008; Rustum et al., 2008; Aguado et al., 2009; Dellana and West, 2009; Dürrenmatt and Gujer, 2012; Bagheri et al., 2015) and neuro-fuzzy systems (Tay and Zhang, 1999; Civelekoglu et al., 2007; Fernandez et al., 2009; Huang et al., 2010; Wan et al., 2011; Dzakpasu et al., 2015). In addition, hybrid methods that combine two or more modelling approaches have been proposed for soft sensor design in WWTPs (Côté et al., 1995; Cohen et al., 1997; Choi and Park, 2001; Lee et al., 2005; Kim et al., 2009; Rustum, 2009; Liu et al., 2014).

In the following, the case studies that concern both municipal and industrial wastewater treatment applications are considered and the investigations of full-scale, pilot-scale and laboratory-scale processes as well as simulated processes are included. In particular, the presented studies are divided according to the applications of the soft sensors: on-line prediction (Section 4.1); process monitoring and fault detection (Section 4.2); and instrument monitoring and fault detection (Section 4.3). The case studies are arranged according to methods and processes in the following order: classical methods for continuous applications; the extensions of the classical methods for continuous applications; and any methods for batch applications.

4.1 On-line prediction

A common application for the data-derived soft sensors in wastewater treatment systems involves predicting the primary process variables. Multivariate methods, especially PLS and its extensions, are one of the typical techniques used for the on-line prediction tasks. Nitrogen and COD concentrations along with the variables describing the sludge settling properties have been popular predicted outputs. However, a wide range of primary variables has been approximated, depending, for instance, on the process type and on the treatment regulations. The prediction applications in the overviewed publications are summarized in Table 4.1, where studies are arranged in the same order as they appear in the text.

Conventional multivariate techniques have been found adequate for soft sensor design in several studies. In an early application, Aarnio and Minkkinen (1986) used PLS to estimate the Total Phosphorus (TP) and COD concentrations and turbidity as indicators of the effluent quality in a municipal ASP. The authors found the methodology feasible for recognizing the reasons for sludge bulking episodes, which diminish the quality of the treated effluent. Blom (1996) designed a PLS model for estimating the influent TP concentration in a municipal WWTP and considered the model accuracy based on daily laboratory analyses to be acceptable. Teppola et al. (1999b) used OLSR, PCR and PLS models for predicting Diluted Sludge Volume Index (DSVI) and COD removal in the ASP of a paper mill. They also investigated updating the static models with a Kalman filter which remarkably improved the predictions of DSVI, but did not notably increase the performance of the COD prediction. Jansson et al. (2002) examined soft sensors in a municipal ASP for estimating phosphate-phosphorus (PO_4-P) and TP concentrations, the data of which were acquired with sampling campaigns and laboratory analyses. The authors found PLS models to be the most precise of the tested methods and the model accuracies to improve by including past observations using the finite impulse response filter (Mitra and Kaiser, 1993). The soft sensor estimates could have been used in the precipitation chemical dosage control improving the system in use at the time of the study. Amaral et al. (2013) explored the prediction of Sludge Volume Index (SVI) and Mixed Liquor Suspended Solids (MLSS) in a laboratory-scale ASP in the presence of intentionally caused sludge settling disturbances. PCA and decision trees

Table 4.1. Prediction applications of soft sensors in the reviewed publications. M = municipal, I = industrial, L = laboratory-scale, P = pilot-scale, S = simulated.

Publication	**Method(s)**	**Appl.**	**Process(es)**	**Predicted variable(s)**
Aarnio and Minkkinen (1986)	PLS	M	ASP	TP, COD, turbidity
Blom (1996)	PLS	M	WWTP	TP
Teppola et al. (1999b)	OLSR, PCR, PLS	I	ASP	DSVI, COD
Jansson et al. (2002)	OLSR, PCR, PLS	M	ASP	PO_4-P and TP
Amaral et al. (2013)	PLS	L	ASP	SVI, MLSS
Galinha et al. (2012)	PLS	P	MBR	TMP, COD, NO_3-N, TN, TP, MLSS
Chen et al. (2014)	IPLS	I	ASP	COD
Platikanov et al. (2014)	PLS	M	ASP	NO_3-N, TOC
Yoo et al. (2004)	PLS, FPLS, QPLS	S, I	BSM1, ASP	NH_4-N, NO_3-N, SVI, cyanide, COD
Lee et al. (2007)	PLS, APLS, RAPLS	I	AF	TOD, CH_4 production
Woo et al. (2009)	PLS, NNPLS, KPLS	I	ASP	COD, TN, cyanide
Dürrenmatt and Gujer (2012)	GLSR, FFNN, SOM, RF	M	ASP	COD, NH_4-N
Sulthana et al. (2014)	FCPR	M	LTP	COD, BOD
Aguado et al. (2006)	PCR, PLS, MPLS, FFNN	P	SBR	PO_4-P

(Quinlan, 1986) were used to cluster samples according to four operating conditions: pinpoint flocs formation, filamentous bulking, viscous bulking and normal conditions. The accuracies of the individual models for each cluster were found to be considerably better than the accuracies of the models that covered all the samples.

In a few case studies, spectral data have been used as inputs into PLS models. Galinha et al. (2012) investigated PLS-based prediction in a pilot-scale Membrane Bioreactor (MBR). In MBRs, suspended solids are separated with membranes instead of the secondary settlers of the conventional ASPs and, therefore, considerably larger MLSS concentrations can be used. In the study, the spectroscopy data was compressed using Parallel Factor Analysis (PARAFAC, Bro, 1997) and, subsequently, its most descriptive components were applied as PLS inputs along with conventional process variables. The transmembrane pressure (TMP) indicating the membrane fouling and the permeate COD concentrations were precisely estimated, but the nutrient and MLSS approximations were less satisfying. PLS-based estimations applying spectral data have also been proposed by Chen et al. (2014) in an airport WWTP application with Interval PLS (IPLS, Andersen and Bro, 2010) and by Platikanov et al. (2014) in a municipal WWTP application with the classical method.

A number of multivariate variants have been proposed for overcoming the challenges of the nonlinear and time-evolving nature of WWTPs. Yoo et al. (2004) predicted process variables in ASPs with the conventional PLS, Quadratic PLS (QPLS, Baffi et al., 1999) and Fuzzy PLS (FPLS, Bang et al., 2003) using the Benchmark Simulation Model 1 (BSM1, Gernaey et al., 2014) and an industrial ASP as the test environments. BSM1 is a virtual test platform for the performance assessment of control and monitoring strategies for the ASP. In the case of the BSM1 application, the FPLS model predicted most precisely the outputs, which were the effluent NH_4-N and NO_3-N concentrations. In the real-plant application, the outputs were SVI and the reductions of cyanide and COD. The PLS and QPLS showed a slightly better prediction performance than the FPLS, but the prediction of SVI was challenging due to the influent disturbances. Lee et al. (2007) proposed Robust Adaptive PLS (RAPLS) for prediction in an Anaerobic Filter (AF) treating industrial wastewater. In the AFs, wastewater is carried through a bed of medium on which anaerobic bacteria grow. The effluent Total Oxygen Demand (TOD) and the production rate of methane were successfully estimated with the presented technique.

Woo et al. (2009) predicted the effluent COD, Total Nitrogen (TN) and cyanide concentrations in an ASP treating cokes wastewater. While the conventional PLS was not capable of modelling the outputs satisfactorily, Neural Network PLS (NNPLS, Qin and McAvoy, 1992) achieved improved estimation accuracies. The best prediction performance was, however, obtained by a KPLS model. The models were further compared using the Bayesian Information Criterion (Leonard and Hsu, 1999), which confirmed the ability of KPLS to outperform the other investigated methods when considering both accuracy and complexity. Dürrenmatt and Gujer (2012) compared Generalized Least Squares Regression (GLSR, Kariya and Kurata, 2004), ANN and Random Forest (RF, Breiman, 2001) methods for approximating COD and NH_4-N concentrations in a municipal ASP. Even though GLSR estimates were not the most accurate, the authors considered the transparency of the GLSR models to be a significant advantage, for instance, compared with the opaqueness of the ANNs. Therefore, they considered the interpretability to justify the selection of GLSR for soft sensor development. Sulthana et al. (2014) proposed Fuzzy PCR (FCPR) for approximating the COD and BOD reductions in a municipal Lagoon Treatment Process (LTP), where earthen basins are used as reactors and surface aerators are exploited to provide the required oxygen and mixing. The authors found the output estimations to be precise using daily measurements as the inputs, but a larger training data set would have been needed for coping with a wider range of process conditions.

Most of the treatment processes considered in the reviewed studies were continuous, though batch processes for purifying wastewaters were also considered. Sequencing Batch Reactors (SBRs) containing activated sludge are operated according to a different number of phases, such as filling, aeration, mixing, settling and decanting. The phases and their lengths depend on the treatment targets. A study by Aguado et al. (2006) compared the performances of several approaches based on PCR, PLS and ANNs for predicting the PO_4-P concentration profile in a pilot-scale SBR treating synthetic sewage and operated for the EBPR purpose. The authors found the batchwise unfolding MPLS models outperformed the other techniques used for the estimation task.

4.2 Process monitoring and fault detection

Classical multivariate techniques and their extensions have been widely used in the development of soft sensors aimed at monitoring the biological WWTPs. Pioneering work on the applications of multivariate statistics and developing adequate methods for monitoring the ASPs in the pulp and paper industry was done by Mujunen (1999) and Teppola (1999) and for monitoring the municipal ASPs by Rosen (2001) and Lennox (2002). Supervising the batch reactors for wastewater treatment using multivariate techniques has been extensively investigated by Aguado (2005) and Villez (2007). The reviewed studies that aimed at monitoring treatment processes are summarized in Table 4.2.

Table 4.2. Process monitoring and process fault detection applications of soft sensors in the reviewed publications. M = municipal, I = industrial, L = laboratory-scale, P = pilot-scale, S = simulated.

Publication	Method(s)	Appl.	Process
Rosen and Olsson (1998)	PCA, PLS	M	ASP
Teppola et al. (1999a)	PCA + FCM	I	ASP
Tomita et al. (2002)	PCA	S	ASM1
Olsson et al. (2003)	PCA	M	ASP
Miettinen et al. (2004)	PCA, PARAFAC	M	LTP
Dias et al. (2008)	PCA	L	ASP
Moon et al. (2009)	PCA + K-means	M	ASP
Teppola et al. (1997)	PLS + autocorrelation function	I	ASP
Teppola et al. (1998)	PLS + FCM	I	ASP
Mujunen et al. (1998)	PLS	I	ASP
Teppola and Minkkinen (1999)	PLS + FCM/PCM	I	ASP
Rosen and Yuan (2001)	APCA + FCM	S	BSM1
Rosen et al. (2002)	APCA + FCM	S	BSM1
Rosen et al. (2003)	PCA, APCA	M	ASP
Aguado and Rosen (2008)	PCA, APCA, MPCA	S	BSM1_LT
Lee et al. (2008)	PCA, APCA	M	ASP
Rosen and Lennox (2001)	PCA, APCA, MSPCA	M	ASP
Mirin and Wahab (2014)	PCA, MSPCA	M	ASP
Lennox and Rosen (2002)	APCA, AdMSPCA	M	ASP
Yoo et al. (2003)	FPCR, PCR	I	ASP
Maere et al. (2012)	PCA + GK, EFPCA + GK	L	MBR
Aguado et al. (2007b)	MPLS	L	SBR
Aguado et al. (2007a)	MPCA	L	SBR
Villez et al. (2008)	MPCA + LAMDA	P	SBR

The conventional PCA technique has been used for process-state identification especially in earlier applications. Rosen and Olsson (1998) applied PCA for monitoring the operational states in a municipal ASP. After building a PCA model using the samples representing the normal operating conditions, they monitored deviations from the normal process behaviour with the two-dimensional score plot. Teppola et al. (1999a) proposed combined approaches of PCA and FCM clustering for monitoring and visualizing process states and seasonal fluctuations in an ASP treating wastewater from a paper mill. In the pulp and paper industry, the main treatment goal is to oxidize the organic compounds in the wastewater instead of targeting nutrient removal. Actually, nitrogen and phosphorus usually need to be added in order to maintain a sufficient balance between organic matter and nutrients for the microbiological population. In this work, the researchers were able to detect and isolate an excess addition of phosphorus into the influent wastewater. Tomita et al. (2002) applied a PCA model for the analysis and disturbance detection in a simulated ASP. They found three groups of process variables that successfully characterized the system behaviour. Olsson et al. (2003) employed PCA for monitoring a municipal ASP as an example of using information technology for decision support purposes. They identified separate clusters that describe the operational states and applied a score plot for monitoring shifting of the new data between the clusters. Miettinen et al. (2004) investigated the use of PCA and PARAFAC for characterizing the operation of a municipal multistage LTP. The results established that the operational states and the ponds where particular reactions occurred could be identified using both methods, which complemented each other well. Dias et al. (2008) employed PCA for monitoring perturbations in a laboratory-scale ASP and used spectral data as the model inputs. The authors found the application to be successful in monitoring the process states and the wastewater quality. Moon et al. (2009) published a methodology for the identification of the operational states in a municipal ASP by means of PCA and K-means clustering. They identified five operational groups and demonstrated that the proposed operational map could visually provide information on the dynamic trends of the process states.

Moreover, conventional PLS has been employed for process monitoring. Teppola et al. (1997) used PLS combined with an auto-correlation function (Massart et al., 1988) for modelling DSVI and COD, nitrogen and phosphorus reductions in the ASP of a paper mill in order to detect process shifts. Whereas the model was able to explain DVSI well, it did not yield a good performance in relation to the reductions due to scarce data representing the daily values of the variables. Nevertheless, the researchers concluded that the disturbance was successfully isolated in almost in every case. In another soft sensor application, Teppola et al. (1998) combined PLS with FCM clustering aiming at novel monitoring tools for the ASP of a paper mill. The methodology was successfully applied in monitoring DSVI, with seasonal variations, for instance, being recognized in the process. Mujunen et al. (1998) applied PLS for monitoring three ASPs treating wastewater from the pulp and paper industry. Sludge settling properties determined with SVI and DSVI were monitored and used for identifying the process states associated with poor treatment efficiency caused by bulking sludge. The authors found that the sampling frequency of the 2–3 day composite samples and the daily mean values of on-line measurements were too low for modelling the peak values successfully. Rosen and Olsson (1998) demonstrated the use of PLS for monitoring the operational states of a municipal WWTP when considering the effluent turbidity as the model output. The process variables associated with the disturbances were isolated by analyzing the variables' contributions to the model residuals. Teppola and Minkkinen (1999) combined PLS with FCM and Possibilistic *C*-means (PCM, Krishnapuram and Keller, 1993) clustering methods for monitoring the ASP of a paper mill. The authors demonstrated the use of the combined methodologies for real-time process monitoring.

Adaptive PCA (APCA) extensions have been proposed for WWTP monitoring in a number of publications for improved adjustment to the dynamic conditions. Rosen and Yuan (2001) used APCA and FCM clustering for monitoring and control set point definition in a case study where a simulated step-feed ASP was used as a test bench. In a step-feed ASP, the wastewater is fed at several stages in the bioreactor. The influent data for a preliminary version of BSM1 modified with an extreme NH_4-N load disturbance was applied in the simulations. The authors defined five operational states with the procedure: normal operation, storm with sewer flush-out, storm, rain and high NH_4-N load. Supervisory control strategies determining the set-points of the manipulated variables were successfully applied in the occurrence of the specific operational states. Later, Rosen et al. (2002) utilized the same process-state estimation approach in an investigation of a predictive supervisory controller during extreme events. Rosen et al. (2003) focused on the challenges of multivariate monitoring in wastewater treatment. The authors did not find the conventional PCA to be adequate for dealing with the process dynamics of a municipal ASP, but an adaptive scaling of the model parameters improved the monitoring performance considerably. Aguado and Rosen

(2008) investigated efficient monitoring tools for municipal ASPs. The dynamic influent data of the BSM1 long-term model (BSM1_LT, Rosen et al., 2004) that was modified to include a realistic set of process disturbances was used in the study. The researchers used PCA approaches and FCM clustering for monitoring the operational states and for tracing the most likely disturbance causes. They also found the performance of the monitoring tool to improve significantly when an APCA method was used instead of the classical PCA. An example of a practical implementation was provided by Lee et al. (2008) who applied APCA for the real-time remote monitoring of small-scale municipal ASPs. They noticed that the APCA models overcame the problem of evolving dynamics and reduced the number of false alarms significantly. The models were used to provide the plant operators with an early warning of an anomalous process behaviour.

Other types of extensions to the conventional multivariate techniques have also been presented for inspecting the process units. Rosen and Lennox (2001) proposed Multiscale PCA (MSPCA), a combination of PCA and multiresolution analysis (Strang and Nguyen, 1997), which decomposes measurement signals into several time scales, for monitoring a municipal ASP. Both MSPCA and APCA techniques had the potential to overcome the challenges created by time-varying process conditions, but the MSPCA was shown to provide more information about the process disturbances. Mirin and Wahab (2014) recently used the MSPCA approach to monitor a municipal step-feed ASP. The authors indicated the employment of MSPCA to reduce the number of false alarms in comparison with the conventional PCA method. Lennox and Rosen (2002) continued developing the MSPCA to be more adequate for the field of operation and proposed Adaptive Multiscale PCA (AdMSPCA). The authors compared the performances of the AdMSPCA and APCA techniques and indicated AdMSPCA showed the ability to adapt to a much broader range of process changes. Yoo et al. (2003) presented a Fuzzy Principal Component Regression (FPCR) method for adaptive monitoring of an industrial ASP and demonstrated the technique was able to distinguish between a large process change and a short disturbance. Maere et al. (2012) investigated the use of three different PCA approaches combined with the Gustafson-Kessel fuzzy clustering algorithm (GK, Gustafson and Kessel, 1979) for monitoring membrane fouling in a laboratory-scale MBR. The authors found the Expert-driven Functional PCA (EFPCA, Ramsay and Silverman, 2005) to be the most suited for the monitoring task out of the tested PCA approaches. However, they concluded that a larger training data set would have been needed for the correct classification of all samples and that a full-scale study would widen the validity of the proposed monitoring approach.

Multiway methods have been popular in the process monitoring of batch processes. Aguado et al. (2007b) used a MPLS to correlate several online variables with the phosphorus removal efficiency in a laboratory-scale SBR aiming at EBPR. They identified conductivity as a suitable variable for monitoring the process upsets associated with a negative effect on the phosphorus removal efficiency. Aguado et al. (2007a) compared different multivariate techniques for fault detection and diagnosis in a laboratory-scale SBR operated for an EBPR purpose. The authors found the MPCA methodology to be straightforward and consistent in the monitoring and diagnosis of process abnormalities. Villez et al. (2008) applied a combination of MPCA and Learning Algorithm for Multivariable Data Analysis (LAMDA, Aguilar-Martin and López de Mántaras, 1982) clustering for an analysis of a pilot-scale SBR. They showed the combined methodology provided an efficient and robust tool for screening and interpreting data from a batch process.

4.3 Instrument monitoring and fault detection

Another soft sensor application type used in wastewater treatment is the monitoring and fault detection of hardware instrumentation. These types of soft sensors also provide a decision support system for the maintenance of hardware sensors and analyzers. PCA and PLS approaches have been applied in instrument monitoring and for identifying reasons for sensor anomalies, such as bias, drift, complete failure and precision degradation. The sensor monitoring applications in the reviewed publications are summarized in Table 4.3.

Classical PCA methods have been considered in a few investigations. Yoo et al. (2008) proposed PCA for sensor anomaly identification and sensor reconstruction. The case study concerned a laboratory-scale Single reactor system for High activity Ammonium Removal Over Nitrite (SHARON) process that performs partial nitrification and treats wastewater streams containing high nitrogen concentrations. In particular, two realistic fault scenarios concerning a DO sensor were successfully tested.

Table 4.3. Instrument monitoring and instrument fault detection applications of soft sensors in the reviewed publications. M = municipal, I = industrial, L = laboratory-scale, P = pilot-scale, S = simulated.

Publication	Method(s)	Appl.	Process
Yoo et al. (2008)	PCA	L	SHARON
Alferes et al. (2013)	PCA	M	ASP
Tao et al. (2013)	PCA	P	SBR
Lee et al. (2004)	PCA, APCA	S	BSM1
Lee et al. (2006)	PCA, APCA	S	BSM1
Baggiani and Marsili-Libelli (2009)	APCA	M	ASP
Lee et al. (2009)	PLS, MSPLS	I	AF

Alferes et al. (2013) used PCA and the model residual statistics for monitoring sensors at the inlet of a municipal WWTP. As the model inputs, they used eight on-line measured variables that consisted of four pairs of redundant measurements. In a provided example, the other turbidity sensor was defective and the anomaly was correctly detected and isolated. Tao et al. (2013) investigated the underlying sensor anomalies in a pilot-scale SBR. Four on-line measured process variables were applied as the inputs to the PCA model. The researchers illustrated that the fault causes could be successfully identified by the methodology that combined the information provided by the loadings and scores of the model.

Advanced multivariate variants have been studied for instrument monitoring purposes. Lee et al. (2004) proposed a PCA-based sensor anomaly detection and isolation application using the BSM1 protocol as a test environment. They used a time-lagged APCA model to identify faults in seven sensors. Two case studies were tested: the precision degradation of the influent NH_4-N sensor and the bias in the influent flow rate sensor. In particular, a reconstruction method based on the ratio between two kinds of modelling residuals was applied for identification. The APCA model detected and isolated defective sensors clearly and consistently in both of the cases. Lee et al. (2006) applied APCA for sensor fault detection in the BSM1 platform. The studied anomaly scenarios included the influent NH_4-N sensor corrupted by a drifting fault, the influent flow rate sensor corrupted by the bias fault and the NO_3-N sensor in the ASP corrupted by precision degradation. The authors indicated that the proposed approach performed well for sensor anomaly detection and in identifying the failing sensors efficiently. However, a limitation of the proposed technique concerned its inability to identify the malfunctioning sensors that cause process transitions, *i.e.* the situations when a faulty instrument is connected to a control loop. Baggiani and Marsili-Libelli (2009) studied APCA models for real-time fault detection and isolation in an ASP treating municipal wastewater and septic tank discharges. In particular, the measurements of one NH_4-N and two NO_3-N sensors were used as the model inputs. The abnormalities were classified into three categories: sensor faults, spikes and process anomalies. The researchers demonstrated that the models were capable of detecting all the investigated faults. In addition, the sensors responsible for the faults were isolated by studying the contributions of input variables to the model residuals. Lee et al. (2009) proposed a Multiscale PLS (MSPLS) algorithm combining PLS and wavelet analysis (Strang and Nguyen, 1997) for sensor fault detection. As a case study, the methodology was used for anomaly detection in an AF process treating petrochemical industry wastewater. The fault detection ability of the MSPLS approach was found to be good and, moreover, it was shown to properly diagnose the detected sensor failures and to provide scale-level information about the fault characteristics.

5. Multivariate methods used in soft sensor development

Multivariate statistical analysis concerns analyzing relationships among data in high dimensions. In particular, the multivariate statistical techniques deal with data comprising multiple variables simultaneously, in contrast with the univariate statistical techniques that handle the observations of only one variable at time. Multivariate statistics is also considered an important discipline in chemometrics, which is the science of extracting information from chemical systems by data-driven means (Wold, 1995). In the wastewater treatment sector, research where multivariate techniques have been used under the domain of chemometrics has also been conducted (*e.g.*, Teppola, 1999; Rosen, 2001; Haimi and Hurme, 2004; Haimi, 2006).

A number of classical multivariate techniques are available for researchers to convert high-dimensional data into easily interpretable and actionable information. Novel variants of the classical methods are also being developed continuously in order to find adequate solutions for different applications and problems. In this chapter, the multivariate statistical methods and their extensions that are applied in the experimental research presented in Chapter 6 of this thesis are described: Principal Component Analysis, Ordinary Least Squares Regression, Partial Least Squares Regression and k-Nearest Neighbor Local Linear Regression.

5.1 Principal Component Analysis

5.1.1 Conventional PCA

Principal Component Analysis (PCA, Jolliffe, 2002) is a multivariate statistical technique for learning a low-dimensional representation of a set of data. PCA extracts the dominant patterns in the data by eliminating information redundancy due to variables cross-correlation. PCA searches in the original data space for new directions that are maximally independent in a linear sense, hence uncorrelated. In such a way, PCA identifies the principal directions in which the data varies. Before performing PCA, data are often autoscaled *i.e.* centered around the sample mean and scaled with respect to their unit variance in order to give the variables the same influence in the PCA model. However, in every case autoscaling is not appropriate before performing PCA and, in addition, there are several other options for centering and scaling data (Bro and Smilde, 2003; van den Berg et al., 2006).

Let $\mathbf{X}$ indicate a $K \times D$ data matrix with the K observations each comprising D variables. Each of the K observations $\mathbf{x}(k) = [x_1(k), \ldots, x_d(k), \ldots, x_D(k)]'$ at time k represents a point in the D-dimensional data space. PCA factorizes the $K \times D$ data matrix $\mathbf{X}$ using eigenvalue decomposition, to obtain

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{5.1}$$

where $\mathbf{T}$ is a $K \times S$ score matrix, $\mathbf{P}$ is a $D \times S$ loading matrix and $\mathbf{E}$ is a $K \times D$ residual matrix. S is the number retained Principal Components (PCs) and each of the K measurements at time k is modelled as a S-dimensional point $\mathbf{t}(k) = \mathbf{x}(k)\mathbf{P}$. The scores are understood as the new coordinates of the point in a (sub)space whose directions are defined by the set of loadings $\{\mathbf{p}_1, \ldots, \mathbf{p}_s, \ldots, \mathbf{p}_S\}$, or PCs, which are eigenvectors of the covariance matrix $\mathbf{X}'\mathbf{X}$. Typically, most of the variation in the data can be explained by retaining a small number of PCs compared with the original dimension of the data matrix $\mathbf{X}$ (*i.e.* S « D). The discarded PCs are associated with the smallest eigenvalues λ and they represent the directions with the least variance among the data.

The example in Figure 5.1 demonstrates the projection of data from the original space into the principal subspace. The observations are shown in the original space defined by three variables, X_1, X_2 and X_3 (Figure 5.1a). A plane defined by two PCs (PC 1 and PC 2) is depicted in the original space (Figure 5.1b). The direction of PC 1 is associated with the largest variation among the observations. PC 2 is orthogonal to PC 1 and, within that constraint, it describes the largest possible variation left in the data.

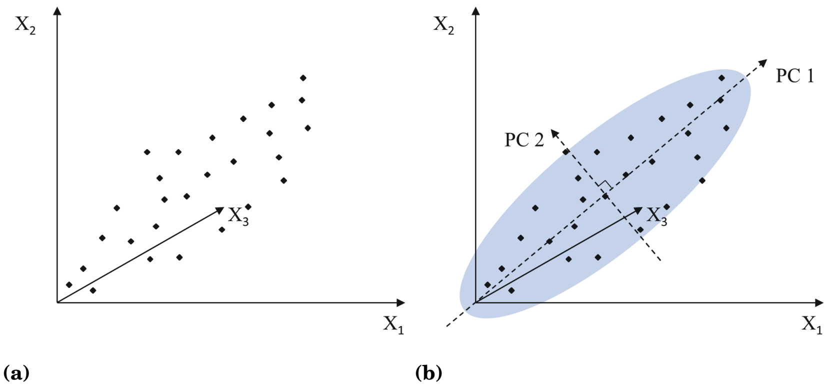


Figure 5.1. Observations in an original three-dimensional space (a) and in a principal subspace formed by two PCs (b).

Methods for selecting a subset of PCs

A large number of methods have been presented for selecting a sufficient subset of PCs, including *heuristic* and *statistical* approaches (Jackson, 1993; Valle et al., 1999; Jolliffe, 2002). The heuristic methods are experience-based techniques, for instance graphical procedures. Their justification is that they have often been found to work sufficiently for selecting the number of retained PCs in practice. The statistical approaches base the choice of an adequate subset of PCs on significance tests. For example, in a wide range of cross-validation methods (Hastie et al., 2009; Arlot and Celisse, 2010) the estimated values of a data set that has not been used for building the PCA model are compared statistically with the actual values. Some of the common methods for determining the number of retained PCs are collected in Table 5.1. Only the methods that are used for selecting a sufficient subset of PCs or other latent variables in the experimental studies presented in Chapter 6 are described below. However, references for the detailed procedures of other relevant approaches to assessing the number of retained PCs are provided in Table 5.1.

One of the most applied methods for selecting the number of the retained PCs is *Cumulative Percent Variance* (CPV). The eigenvalues associated with PCs correspond to the variance explained by each dth PC, $\{1, ...d, ..., D\}$. Therefore, the cumulative variance that is explained by the first d PCs is:

$$CPV(d) = \frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{D} \lambda_i} \times 100\% \tag{5.2}$$

Jolliffe (2002) has stated that a sufficient cutoff for CPV in the selection of

Table 5.1. Popular heuristic and statistical methods for selecting the subset of retained Principal Components in a PCA model.

Method	**Reference(s)**
Heuristic:	
Cumulative Percent Variance	Jolliffe (2002)
Cut-off limit for value of λ	Kaiser (1960); Jolliffe (1972)
Maximum eigengap	Davis and Kahan (1970)
Scree diagram	Cattell (1966)
Log-eigenvalue diagram	Farmer (1971)
Statistical:	
Cross-validation	Wold (1978); Diana and Tommasi (2010)
Bootstrapping	Diaconis and Efron (1983)
Variance of reconstruction error	Qin and Dunia (2000)
Partial correlations	Velicer (1976)

the number of the retained PCs, corresponding for the dimensionality S of the subspace, is often in the range 70–90%. However, the adequate minimum CPV is application-specific and dependent on the subjective evaluation of the modeller.

The *eigengap* technique (Davis and Kahan, 1970) can be used for selecting an appropriate number of PCs for the models. When the eigenvalues are sorted in descending order $\lambda_1 \geq, ..., \geq \lambda_d \geq, ..., \geq \lambda_D$, the eigengap is defined as $\mu_d = \lambda_d - \lambda_{(d+1)}$, with $d = \{1, ...D-1\}$. The index d of the eigenvalue associated with the largest eigengap defines the dimensionality S of the projection subspace *i.e.* $S = \arg\max \mu_d$.

Leave-One-Out (LOO) is a standard resampling method used in a *cross-validation* procedure (Stone, 1974). One sample in the training data set is excluded at the time and the value of the excluded sample is estimated using the learned model. The procedure is repeated for each sample, in other words as many times as there are observations K in the training data set. The cross-validation accuracies are then measured in terms of an appropriate statistic that describes the deviations between the measured and estimated values. LOO cross-validation is performed for models with d retained PCs (or other latent variables) and d resulting in the smallest value of the error statistic defines the dimension S.

Statistics for monitoring the variability in data

The Hotelling's T^2 statistic and the Q statistic (Jackson and Mudholkar, 1979) and their confidence limits, T^2_{lim} and Q_{lim} (Atkinson et al., 2004; Nomikos and MacGregor, 1995, respectively), are often employed, for instance, in process monitoring. The T^2 statistic measures the (normalized) Mahalanobis distance (Mahalanobis, 1936) of the projected observation $\mathbf{t}(k)$ from the origin of the principal component subspace:

$$T^2(k) = \mathbf{t}(k)\mathbf{\Lambda}^{-1}\mathbf{t}(k) \tag{5.3}$$

where $\mathbf{\Lambda}^{-1}$ denotes a diagonal matrix with the inverse of the eigenvalues associated with the retained PCs. An example of T^2 in a space defined by 2 PCs is shown with a red double-headed arrow in Figure 5.2a. The Q statistic measures the (orthogonal) distance of an observation $\mathbf{x}(k)$ from its reconstruction $\hat{\mathbf{x}}(k) = \mathbf{t}(k)\mathbf{P}'$ on the principal component subspace (indicated with a red double-headed arrow in Figure 5.2b):

$$Q(k) = \sum_{d=1}^{D}(x_d(k) - \hat{x}_d(k))^2 \tag{5.4}$$

The confidence limits T^2_{lim} and Q_{lim} are calculated for a certain confidence level $z = \{0, ..., 1\}$. Typically, the values given for z range between 0.95 and 0.99, a low z value providing a stricter T^2_{lim} and Q_{lim} than a large z value. The appropriate confidence level is case-specific and it is set by the user of the PCA technique.

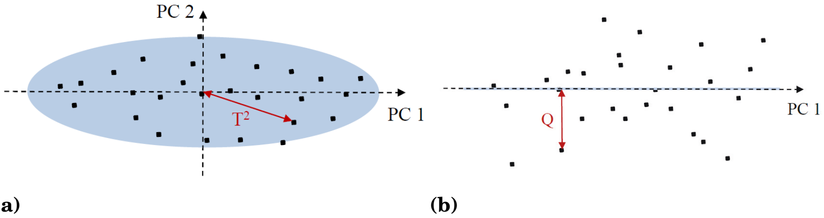


Figure 5.2. T^2 measures the distance from the origin of the principal subspace to the projected observation; the plane formed by PC 1 and PC2 viewed from the top (a). Q measures the distance of an observation to its reconstruction on the subspace; the plane formed by PC 1 and PC2 viewed from the side (b).

Squared Mahalanobis distances of normally distributed scores are approximately χ^2_S-distributed. Hence, the cut-off value T^2_{lim} for the score distances can be determined as $\chi^2_{S,z}$, where z refers to the set confidence level (Atkinson et al., 2004).

The distribution of orthogonal distances is not known exactly. However, it can be well approximated by a scaled chi-squared distribution with orthogonal distances to the power of 2/3 approximately normally distributed with estimated mean μ and variance σ^2 of the Q statistic. The threshold value Q_{lim} is then determined as $(\hat{\mu} + \hat{\sigma}_z)^3$ with z denoting the $(z \times 100)\%$ quantile of the Gaussian distribution (Nomikos and MacGregor, 1995).

An alternative technique is to use a single statistic that integrates the information provided by the T^2 and Q statistics. The weighted combination of the T^2 and Q statistics, J statistic (Raich and Çinar, 1996) is formulated as

$$J(k) = \lambda T^2(k) + (1 - \lambda)Q(k) \tag{5.5}$$

where λ is a parameter that gives weight for the observations inside the principal component subspace over the observations outside the principal component subspace ($0 \leq \lambda \leq 1$). Confidence limit J_{lim} is determined as a weighted combination of T^2_{lim} and Q_{lim} using the same value of λ as for calculating the J statistic. Other combined indices of T^2 and Q have also been presented, for instance, by Yue and Qin (2001).

The variables' contributions to the T^2 and Q statistics (MacGregor et al., 1994) can be investigated, for instance, when a T^2_{lim} and Q_{lim} violation takes place in the monitored process. The contributions along the dth PC to the T^2 statistic are calculated as

$$\mathbf{c}(k) = \mathbf{x}(k)\mathtt{diag}(\mathbf{p}_d) \tag{5.6}$$

Particularly, $\mathtt{diag}(\mathbf{p}_d)$ denotes the diagonal matrix of the column vector $\mathbf{p}_d$ and $\mathbf{x}(k)$ denotes the vector of original data at time k. The contributions for a PCA model with d PCs to the Q statistic are calculated as follows

$$\mathbf{e}(k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k) \tag{5.7}$$

where $\hat{\mathbf{x}}(k)$ denotes its reconstruction using a model with d PCs. The reconstruction $\hat{\mathbf{x}}(k)$ is determined as follows: $\hat{\mathbf{x}}(k) = \mathbf{t}(k)\mathbf{p}_d$.

Alternatives for the conventional contribution analysis that can also be used in the modular soft sensor design procedure presented in the thesis have been proposed. One of them is reconstruction-based contribution method that is based the reconstruction of a monitoring statistic, typically T^2 or Q, along the direction of a variable (Alcala and Qin, 2009).

5.1.2 Robust PCA

The conventional PCA method is strongly affected by anomalous observations and the estimation of the center of the data (usually, the sample mean) that is employed for the centering of the data is affected by outliers that deviate significantly from the other members of the data set. Robust PCA methods try to address such shortcomings by applying robust statistic estimators for location (the center of the data) and scale (the entries in the covariance matrix) that are more resistant to outliers. A large number of robust PCA extensions have been presented in the literature and references for these methods are available, for instance, in the review papers by Frosch Møller et al. (2005); Rousseeuw et al. (2006); Rousseeuw and Hubert (2011); Bro and Smilde (2014).

One of the robust PCA variants is called the Reflection-based Algorithm for Principal Components Analysis (RAPCA, Hubert et al., 2002). The general concept of PCA, such as score and loading matrices that were introduced in Subsection 5.1.1, apply also for this robust PCA variant. RAPCA is a dimension reduction method based on projection pursuit (Li and Chen, 1985). In RAPCA, the spatial L^1-median (Daszykowski et al., 2007) is used as the robust center of the data, around which the data are centered instead of the sample mean. The L^1-median $\hat{\mu}^R$ is defined as the point θ in the original data space that minimizes the sum of Euclidean distances to all the observations $\mathbf{x}(k)$ with $k = (1, ..., k, ..., K)$:

$$\hat{\mu}^R = \arg\min_{\theta} \sum_{k=1}^{K} \parallel \mathbf{x}(k) - \theta) \parallel \tag{5.8}$$

where $\parallel ... \parallel$ represents the L^1 norm (Galpin and Hawkins, 1987). The robust scale in the RAPCA procedure is measured using the Q_n estimator (Rousseeuw and Croux, 1993), which is essentially the first quartile of all pairwise distances between two data samples, i and j. For any univariate data set $(z_1, ..., z_n)$, the Q_n estimate is defined as

$$Q_n = 2.2219 \cdot d \cdot \{|z_i - z_j|; i < j\} \tag{5.9}$$

where d is a small sample correction factor that approaches 1 for increasing n.

The robust principal components are constructed starting from the direction $\mathbf{p}_1$ (or eigenvector, in analogy with the conventional PCA) such that the scale (the robust equivalent of standard deviation) of the robustly centered observations $\mathbf{x}(k)$ projected onto $\mathbf{p}_1$ is maximal. In analogy with the conventional PCA, the squared robust scale of the projections onto the first eigenvector represents the first eigenvalue of the RAPCA model. Once $\mathbf{p}_1$ is found, the observations $\mathbf{x}(k)$ are projected onto the orthogonal complement of $\mathbf{p}_1$. The next direction $\mathbf{p}_2$ is found in this orthogonal complement by looking for the direction that maximizes the second robust scale of the observations projected onto $\mathbf{p}_2$. At each step, the working dimensionality is reduced by one and the procedure can be continued until all the sources of variation have been accounted for.

5.1.3 Moving-window PCA

A major limitation of PCA-based monitoring in many industrial applications is that once the model has been built, it is time-invariant while the processes are time-varying. When such models are used, false interpretations on the instrumental conditions and process operations might result. This is because a PCA model describes the process conditions represented by the training period and is applicable to testing only in corresponding conditions. However, if the conditions change considerably during the testing period, the trained model is no longer valid. PCA methods based on moving-windows have been proposed for monitoring tasks when processes with considerable dynamic behaviour are considered in order to overcome some of the deficiencies of the static PCA approach (Ku et al., 1995; Baggiani and Marsili-Libelli, 2009; Kadlec et al., 2011).

In the moving-window approach (Kruger and Xie, 2012), historical data from a time period defined by the *window-length* L are used for building PCA models. New PCA models are built at the time intervals of a *shift-size* Z. In such a manner, a window shifts along time and a new model is trained at each step by including the newest data (of the size Z) and excluding the oldest ones (of the size Z). In addition, the unseen testing data sets associated with each PCA model are of the size Z. This procedure is depicted for n PCA models in Figure 5.3.

The moving-window models can be categorized as *sample-wise* and *block-wise* models (Kadlec et al., 2011). In the sample-wise techniques, Z corresponds to each data sample coming in, *i.e.* the PCA model is recalculated after every new sample. When the process operating conditions change abruptly, sample-wise moving-window models are efficient in monitoring (Choi et al., 2006). As for the block-wise moving-window

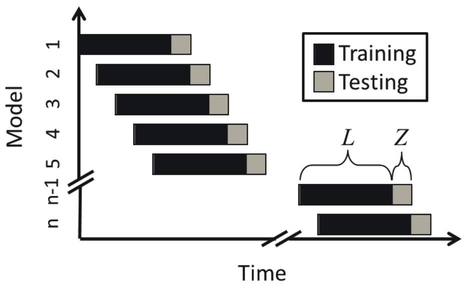


Figure 5.3. Moving-window monitoring procedure using fixed window-lengths L and fixed shift-sizes Z with $1 \ldots n$ PCA models. Figure adopted from Publication IV.

techniques, Z corresponds for a certain number of samples or samples of a certain time period after which the PCA model is recalculated. The advantages of the block-wise moving-window techniques include a low computational cost in comparison with the sample-wise techniques. The block-wise techniques also reduce the risk of recalculating the model based on an anomalous observation (Choi et al., 2006). Conventionally, in the moving-window applications each model covers the same window-length L, and also the shift-size Z is fixed (Kadlec et al., 2011).

Even though the moving-window PCA extension provides considerable advantages over the static PCA approach in the monitoring of time evolving processes, one of its limitations is the fixed window-length. This is due to the fact that rapidness of the process transitions varies. In general, if the process changes rapidly, the window-length should be shortened and when the changes are slow, a large window-length should be preferred (He and Yang, 2008). For this reason, adaptive window-lengths have been considered (Kadlec et al., 2011). Two of the approaches where window-lengths adapt are described in the following.

In the approach presented by He and Yang (2008), L for each model $\{1, ..., n, ..., N\}$ in the moving-window procedure is defined as:

$$\begin{aligned} L(n) =& L_{min} + (L_{max} - L_{min}) \exp\Bigg\{ - \Bigg(\alpha \frac{\|\Delta \mathbf{b}(n-1)\|}{\|\Delta \mathbf{b}_0\|} \\ & + \beta \frac{\|\Delta \mathbf{R}(n-1)\|}{\|\Delta \mathbf{R}_0\|} \Bigg)^{\gamma} \Bigg\} \end{aligned} \tag{5.10}$$

where L_{min} and L_{max} are minimum and maximum window-lengths, respectively. $\|\Delta \mathrm{b}(n-1)\|$ is the Euclidean vector norm (Deza and Deza, 2014) of the difference between the previous two consecutive $1 \times D$ mean vectors, $\mathbf{b}(n-1)$ and $\mathbf{b}(n-2)$, calculated from training data. Correspondingly, $\|\Delta\mathbf{R}(n-1)\|$ is the Euclidean matrix norm (Deza and Deza, 2014) of the difference between the two consecutive $D \times D$ correlation matrices, $\mathbf{R}(n-1)$ and $\mathbf{R}(n-2)$. $\|\Delta\mathbf{b}_0\|$ and $\|\Delta\mathbf{R}_0\|$ represent the Euclidean vector norm of difference between two consecutive mean vectors and the Euclidean matrix norm of the difference between two consecutive correlation matrices in reference conditions, respectively. They are calculated correspondingly as $\|\Delta\mathbf{b}(n-1)\|$ and $\|\Delta\mathbf{R}(n-1)\|$, using two sets of reference data that are associated with normal process conditions without anomalous observations. Three parameters are used for tuning the function; α and β are weights given for $\|\Delta\mathbf{b}(n-1)\|/\|\Delta\mathbf{b}_0\|$ and $\|\Delta\mathbf{R}(n-1)\|/\|\Delta\mathbf{R}_0\|$, respectively, and γ is an exponential parameter that affects the sensitivity of L to the process change.

With the approach of Ayech et al. (2012), the window-lengths are determined accordingly:

$$L(n) = L_{max} - (L_{max} - L_{min})[1 - \exp(-\delta(\|\Delta\mathbf{R}_{ref}(n-1)\|))] \tag{5.11}$$

where $\|\Delta\mathbf{R}_{ref}(n-1)\|$ is the Euclidean matrix norm of the difference between $\mathbf{R}(n-1)$ and $\mathbf{R}_{ref}$. Otherwise $\|\Delta\mathbf{R}_{ref}(n-1)\|$ is calculated like $\|\Delta\mathbf{R}(n-1)\|$, but instead of using the second previous correlation matrix in its calculation, $\mathbf{R}_{ref}$ representing the correlation matrix of a reference data set is utilized. The parameter δ controls the sensitivity of the change in L.

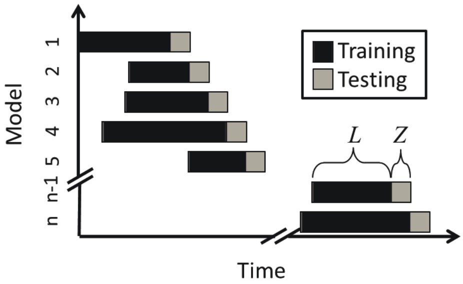


Figure 5.4. Moving-window monitoring procedure using adaptive window-lengths L and fixed shift-sizes Z with $1 \ldots n$ PCA models. Figure adopted from Publication IV.

The monitoring procedure with a moving-window PCA approach, where window-lengths adapt, follows the same principles as the approach with a fixed window-length. The procedure using an adaptive moving-window PCA technique for n models is visualized in Figure 5.4.

5.2 Ordinary Least Squares Regression

Regression models are input-output models where the relationships between the (explanatory) input variables and the (dependent) output variables are estimated statistically. Let $\mathbf{X}$ indicate an input data matrix with the K observations from D variables and y indicate an output column matrix with the K observations from one variable.

Ordinary Least Squares Regression (OLSR, Björk, 1996; Ryan, 2008) regression is a classical method that learns a reconstruction of the functionality between the D-dimensional input observations $\mathbf{x}(k)$ and the output observations $y(k)$. This is done by estimating the regression vector $\boldsymbol{\beta} = [\beta_1, ..., \beta_D]'$ that parameterizes the linear model

$$y(k) = \boldsymbol{\beta}'\mathbf{x}(k) + r(k) \tag{5.12}$$

where $r(k)$ is the residual additive noise. The least-biased estimation of $\boldsymbol{\beta}$ is based on a criterion that minimizes the residual sum of squares by globally fitting a D-dimensional hyperplane over the training data.

Once $\boldsymbol{\beta}$ is defined using the training data, the calibrated OLSR model is usually tested using an unseen data set. After satisfying testing performance, the model can be used, *e.g.*, for estimating new output $\hat{y}(k)$ values by introducing new input observations $\mathbf{x}(k)$ when they are available.

5.3 Partial Least Squares Regression

Partial Least Squares Regression (PLSR, Wold, 1975; Wold et al., 2001) is a global linear regression method that learns a model between the $K \times D$ input matrix $\mathbf{X}$ and the $K \times E$ output matrix $\mathbf{Y}$. Specifically, E stands for the number of output variables and it often, though not necessarily, is equal to one. Instead of using the original input variables, the fitting in PLSR is an iterative procedure and it is based on original variables' decomposition in latent variables. The variable's decomposition aims at maximizing simultaneously the variance in the inputs and the covariance between the inputs and the outputs. The PLSR model is a parametric model that has an additional meta-parameter, the number S of latent variables to be retained. The basic equations of PLSR are as follows:

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{5.13}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}' + \mathbf{F} \tag{5.14}$$

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{F} \tag{5.15}$$

where Equation 5.13 that concerns the decomposition of the input matrix **X** similarly as in PCA, Equation 5.1. **U** and **Q** are the $K \times S$ score matrix and $E \times S$ loading matrix for the outputs **Y**, respectively, and **F** is a $K \times E$ residual matrix. **B** is a $D \times E$ matrix of regression coefficients that describes the relation between the latent variables of **X** and **Y**.

Similarly as in PCA, the subset of retained latent variables needs to be selected when using PLSR. For testing the predictive significance of the different number of retained latent variables in a PLSR model, cross-validation methods have become standard (Wold et al., 2001). For instance, LOO cross-validation is a potential technique for this task.

5.4 Local Linear Regression based on *k*-Nearest Neighbours

Local Linear Regression (LLR) techniques (*e.g.* Stone, 1977; Cleveland and Devlin, 1988) are nonlinear regression methods that are based on the same principle as OLSR. The main difference between the LLR and OLSR methods is the approach used in the estimation of the regression coefficients $\boldsymbol{\beta}$. The LLR methods fit linear models by locally weighted least squares instead of global fitting, which is performed, for instance, in the OLSR method. In other words, the local fitting procedure is executed only in the neighbourhood of the input observation $\mathbf{x}(k)$ for which the output $y(k)$ is to be predicted. Therefore, LLR has the feature of simplicity of traditional linear regression and, in spite of such, it can overcome the drawback of low model accuracy with which linear regression is often associated. For a random input test observation $\mathbf{x}(k)$, LLR estimates its output as $\hat{y}(k) = \hat{\boldsymbol{\beta}}'\mathbf{x}(k)$ by fitting a hyperplane over the local neighbourhood $\mathcal{J}_{\mathbf{x}(k)}$ of $\mathbf{x}(k)$:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{\mathbf{x} \in \mathcal{J}_{\mathbf{x}(k)}} (y(k) - \boldsymbol{\beta}'\mathbf{x}(k))^2 \tag{5.16}$$

The definition of the neighbourhood and the number of neighbours are crucial in LLR. Several strategies for the definition of locality are available when LLR is used in soft sensor development (Zhu et al., 2011). In k-Nearest Neighbour LLR (k-NN LLR, Stone, 1977), the neighbourhood of the input observation $\mathrm{x}(k)$ is defined by the K of its neighbours, according to a predefined metric. Conventionally, the Euclidean distance is used for determining the nearness of the observations. The size K is the meta-parameter of k-NN LLR models. Usually, K is either fixed beforehand or determined by using cross-validation.

6. Results and discussion

This Chapter summarizes the main results and discusses the findings of Publications I–IV. In Section 6.1, the status and the trends of the data-derived soft sensors presented for biological WWTPs are analyzed. Section 6.2 addresses soft sensors designed for on-line prediction in biological post-filters. A novel system enabling the complimentary use of the aforementioned soft sensors and the corresponding hardware instruments is discussed in Section 6.3. Finally, a soft sensor designed for anomaly detection purpose in an ASP is presented in Section 6.4. More comprehensive information on the works is provided in the original publications.

6.1 Status of data-derived soft sensors in biological wastewater treatment

The status of data-derived soft sensors proposed for biological wastewater treatment processes was investigated by reviewing about 100 case studies that were available in literature. The works were published between 1986 and 2012 and they covered processes from laboratory-scale to full-scale operations for both municipal and industrial wastewater treatment. The specific foci of the research were:

- Applications of the data-derived soft sensors;
- Data-derived methods used in soft sensor design;
- Wastewater treatment systems in data-derived soft sensor proposals.

6.1.1 Soft sensor applications

Data-derived soft sensor applications were divided in the review into the following categories:

- On-line prediction of the primary process variables;
- Process monitoring and fault detection;
- Instrument monitoring and fault detection.

The review showed that on-line prediction has been the most widely used application of the data-driven soft sensors in biological wastewater treatment. This is demonstrated in Figure 6.1, where the shares of the applications have been divided into the studies published in 2005 or earlier and the studies published between 2005 and 2012.

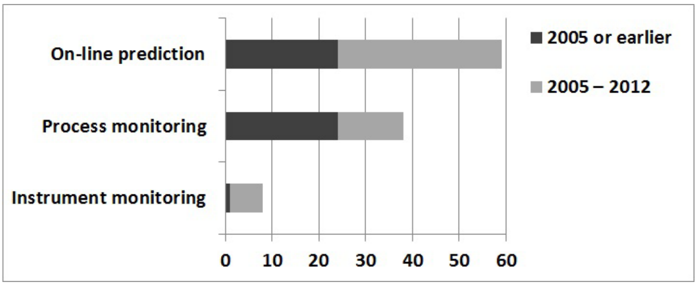


Figure 6.1. Amounts of the soft sensor applications in the reviewed publications. The case studies are divided into the works published in 2005 or earlier and the works published in 2005–2012. Figure adopted from Publication I.

The on-line predictions are typically employed for enhancing process monitoring and control in WWTPs by complementing the conventional approach that relying on the information obtained from the hardware instrumentation. Especially during the past decades, the treatment regulations typically applying to organic matter removal have most importantly affected the operation and process configurations of WWTPs. However, the variables describing the content of organic matter have been challenging to measure in real-time and information about them was mostly provided by the laboratory analyses of daily or grab samples. For these reasons, the content of organic matter has been the most commonly predicted output that enable monitoring its dynamic behaviour. Nitrogen compounds have been the estimated primary variables in an extensive number of soft sensor applications, especially in more recent studies. This finding is associated with the trend that present-day municipal treatment processes are being designed, in particular, for nutrient removal.

Process monitoring has been another popular application among soft sensor developers, especially in the earlier publications (Figure 6.1). This is partly explained by a few researches having actively published studies in the process-monitoring field at the turn of the millennium, whereas soft sensors for on-line prediction have been developed by a more diverse group of researchers over the course of the investigated time span. Any specific process variable is not typically estimated in the process monitoring applications. Instead, the tools such as model residual indexes or positions on operational maps that provide information about the status of the process are often the model outputs. In addition, data-derived models for sensor and analyzer monitoring have been proposed recently, but the rather small number of existing publications suggests that the research area is still emerging in the wastewater treatment sector. The scarce number of publications indicates that data-derived instrument monitoring should be more intensively addressed by the scientific community. The quantity of sensors has increased in the plants due to the evolution of the process configurations, while the reliable information produced by the instruments has become more crucial in conjunction with the introduction of more advanced control schemes. This provides further motivation for future research in the domain of instrument anomaly detection.

6.1.2 Methods for soft sensor design

A number of data-derived techniques for developing soft sensors exists. Based on the frequencies of their reported use in the reviewed literature, the following main modelling families were recognized:

- Principal component analysis;
- Partial least squares;
- Supervised artificial neural networks;
- Self-organizing maps;
- Neuro-fuzzy systems.

Most commonly, different multivariate and ANN methodologies have been used for the development of the data-derived soft sensors in biological wastewater treatment. The distribution of the main method families is presented in Figure 6.2.

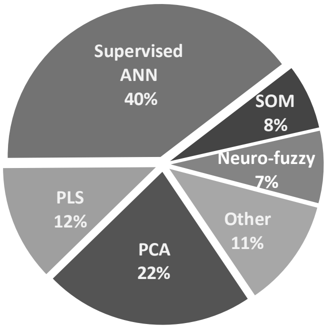


Figure 6.2. The distribution of the main method families employed for the data-derived soft sensor design in the reviewed publications. Figure adopted from Publication I.

Among the main modelling families, different techniques and extensions of the conventional methodologies were deployed depending *e.g.* on the considered application. Supervised ANN techniques were the most popular methods for predicting the primary variables, the other popular methods being the ANFIS and PLS approaches. PCA techniques were employed in the majority of the studies targeting monitoring of treatment processes and instruments.

Due to the dynamic and nonlinear nature of the wastewater treatment processes, the conventional multivariate techniques were often found to be unsatisfying for soft sensor development as such. Therefore, a number of adaptive and nonlinear PCA and PLS extensions were proposed and were demonstrated to be more feasible for soft sensor design, especially in the more recent studies. In particular, the adaptive approaches were shown to overcome the difficulties associated with the changing process conditions. As for the nonlinear extensions, researchers have established them to be adequate for on-line prediction tasks, but in all the case studies the performance of nonlinear methods was not found to be superior to linear methods. Researchers have also indicated that multiway extensions are useful, in particular, for monitoring and analysis of SBRs and have exhibited the multiscale approaches for extracting the features of treatment processes in different time-scales. In addition, the PCA methods were popular pre-processing techniques applied in many soft sensors, which indicates the potential of PCA techniques in the compression of information contained in the high-dimensional data.

As discussed in Subsection 3.2.3, the choice of the methodologies not only depends on the considered application, but also on other issues such as personal preferences and backgrounds. Regional traditions and trends for employing different method families were investigated based on the affiliation of the first authors of the papers. In Europe, the multivariate techniques have most often been applied in soft sensor design. In addition, SOM has been widely used by European researchers compared with others. The Asian and North American research communities have preferred the supervised ANNs. Additionally, ANFIS accounts for a more significant share of the selected methodologies in Asia in comparison with the other regions. The reviewed papers from other continents are so few that any representative analysis could not be done.

6.1.3 Investigated wastewater treatment systems

The survey covered data-derived soft sensor investigations in different types of biological wastewater treatment systems, which were divided into the following categories:

- Municipal;
- Industrial;
- Pilot-scale and laboratory-scale;
- Simulated.

The distribution between the different wastewater treatment systems in the case studies is shown in Figure 6.3. The soft sensors designed for the full-scale municipal and industrial treatments systems accounted for the majority of the explored works. Pilot- and laboratory-scale processes were used in a significant share of the investigations, most of them being SBRs. Only a few full-scale batch processes were included while the continuous treatment processes were typically full scale. The simulated processes proved to be popular in the reviewed case studies, where the BSM platforms were by far the most popular virtual test environments.

While pilot- and laboratory-scale processes and simulated protocols provide valuable opportunities for development of the modelling methodologies and control systems, the tests on full-scale processes are typically more challenging due to the unforeseen features of real-life conditions and data. Therefore, full-scale experiments provide irreplaceable platforms

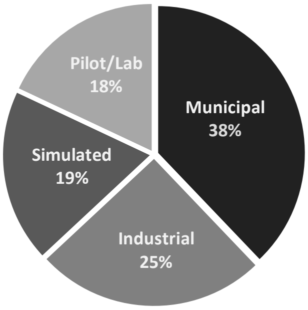


Figure 6.3. The distribution of the wastewater treatment systems for which the data-derived soft sensors were designed in the reviewed publications. Figure adopted from Publication I.

for designing and piloting soft sensors when aiming at their practical implementations in WWTPs. However, initial studies about the soft sensor performances using simulated processes, particularly when integrated with a realistic influent disturbance generator (Gernaey et al., 2011; Martin and Vanrolleghem, 2014) can definitely support full-scale experiments and executions.

6.2 Soft sensors for on-line prediction in biological post-filtration

The objective of the study was to develop an array of soft sensors that estimate in real-time primary process variables in the denitrifying post-filtration unit of the Viikinmäki WWTP. The process unit is described in Section 2.2. The goal was to provide a back-up system for the hardware sensors employed in process control and, in that manner, to ensure the cost-efficient operation of the unit. The initial work on designing the soft sensors is presented in Mulas et al. (2012), whereas a complete application is available in Publication II.

6.2.1 Development of soft sensors

The investigation addresses a denitrifying post-filtration unit that consists of ten parallel filter cells where methanol is dosed as a carbon source to enhance nitrogen removal. Methanol flow to each filter is manipulated by a feedback loop controlling the NO_3-N concentration in the outlet of the cell. Therefore, reliable analyses of NO_3-N contents are essential because of the economic and environmental implications due to incorrect methanol dosing. To sustain the correct dosing, an array of soft sensors that predict on-line the NO_3-N concentrations was developed. To design the soft sensors, process measurements were collected for three years of operations recorded as hourly averages, in which form the long-term data are stored in the plant. The first-year data were used for model training and data from the second and third years for testing the models' performances.

Sample selection
The sample selection was approached as a problem of outlier detection where the goal was to discard anomalous observations as dissimilar from the global behaviour of the data, initially described in Haimi et al. (2011). RAPCA models with the subset of retained PCs selected using the CPV approach were considered for this task. For the filter-specific input sample selection, the T^2 and Q statistics were employed and the samples that respected T^2_{lim} and Q_{lim} were maintained. For the output sample selection, a RAPCA model was built for the matrix that contained NO_3-N concentrations of each filter and the J statistic with the weight coefficient λ set to be equal to 0.1 was employed. The data where J did not exceed J_{lim} were retained.

Variable selection
The selection of the variables to be used for the models among all the candidate inputs was the next task. Since the design of the filters is ideally similar, the subset of input variables was logically seen to be corresponding to all the filters. Variable selection was pragmatically approached starting from the physical understanding of the nitrogen removal process. The aim was for regression models that have only a few representative variables but are accurate and intelligible. Based on the described approach and the target, six soft sensor inputs x^R_{ij} shown in Table 6.1 were selected to estimate the output $\hat{y}_i$ in Filter i. The symbol j refers to the sequence number of the input.

Model selection
Simplicity was one of the main requirements for allowing an easy implementation of the soft sensors in the plant's control system. Mainly for this reason, linear regression methods OLSR and PLSR were considered for the models at their core. However, it is also well known that many subprocesses within a WWTP show a nonlinear behaviour and that nonlinear models might give more accurate approximations. For that reason, a nonlinear k-NN LLR method was also considered.

Table 6.1. Input and output variables used for the regression soft sensor models.

Symbol	TAG	Description	Unit
x_{i1}^R	$PI\text{-}NO_3\text{-}N$	Post-filtration influent nitrate-nitrogen	mg/l
x_{i2}^R	$PE\text{-}NO_3\text{-}N$	Post-filtration effluent nitrate-nitrogen	mg/l
x_{i3}^R	$PE\text{-}T$	Post-filtration effluent temperature	° C
x_{i4}^R	$Fi\text{-}HL$	*i-th* Filter head loss	m
x_{i5}^R	$Fi\text{-}QW$	*i-th* Filter wastewater flow rate	m^3/s
x_{i6}^R	$Fi\text{-}QM$	*i-th* Filter methanol flow rate	m^3/h
$\hat{y}_i$	$Fi\text{-}NO_3\text{-}N$	*i-th* Filter effluent nitrate-nitrogen	mg/l

6.2.2 Performance of soft sensors

The samples of the training period were used for calibrating the regression models and for optimizing the number of latent variables for the PLSR models with the LOO cross-validation technique. The accuracies of the soft sensors were tested using Root Mean Squared Errors (RMSE) between the NO_3-N measurements and estimates during the testing period as the criterion. RMSE is defined as follows:

$$RMSE = \left(\frac{1}{K} \sum_{k=1}^{K} (y(k) - \hat{y}(k))^2 \right)^{1/2} \tag{6.1}$$

where $y(k)$ and $\hat{y}(k)$ denote the measurement at time k and its reconstruction, respectively. The experimental results summarized in Table 6.2 showed that the overall accuracy achieved by the soft sensors in terms of RMSE was about 0.2 mg/l, which is comparable with the nominal resolution of the hardware sensors. Interestingly, only minor improvements compared with the linear modelling approaches were achieved with the k-NN LLR models. It is worth noting that OLSR and PLSR always achieved identical accuracies due to the fact that the number S of latent variables determined for the PLSR models was equal to the dimensionality D of the original input space used by OLSR.

Even though the amply instrumented unit does not suffer from lack of on-line data in normal operational situations, there are periods of sensor malfunctions and downtime. For instance, about one week of measurements of continuous process operation in Filter 7 were inaccurate due to failing (Figure 6.4b). At the same time and under similar operating conditions, the dependence between the measurements and estimates using both the OLSR and the k-NN LLR in the Filter 3 was strong (Figure 6.4a). The NO_3-N estimates in Filter 7 were available during the investigated

Table 6.2. Estimation accuracies of the soft sensors using the different modelling techniques in terms of RMSE (mg/l).

	Filter 1	Filter 2	Filter 3	Filter 4	Filter 5
OLSR	0.40	0.18	0.15	0.16	0.19
PLSR	0.40	0.18	0.15	0.16	0.19
k-NN LLR	0.37	0.20	0.15	0.15	0.18

	Filter 6	Filter 7	Filter 8	Filter 9	Filter 10
OLSR	0.18	0.25	0.25	0.26	0.31
PLSR	0.18	0.25	0.25	0.26	0.31
k-NN LLR	0.16	0.26	0.23	0.24	0.28

period and, based on the prediction accuracy in the other examined filter, they could have been used for process monitoring and for methanol dosage control instead of the erroneous hardware sensor measurements. By using estimates in this manner when available, process monitoring and control would not suffer from out-of-order measurements or lack of data, for instance, during instrumentation maintenance. That is of prime importance, since the methanol flow rates are manipulated by feedback loops controlling the NO_3-N concentrations and methanol responds to a major share of the chemical costs in the plant (Sundell, 2008). Since the estimation models were able to reconstruct the dynamics in the faulty filters and to accurately estimate the output concentrations, they also provide a useful tool for validating the existing instrumentation.

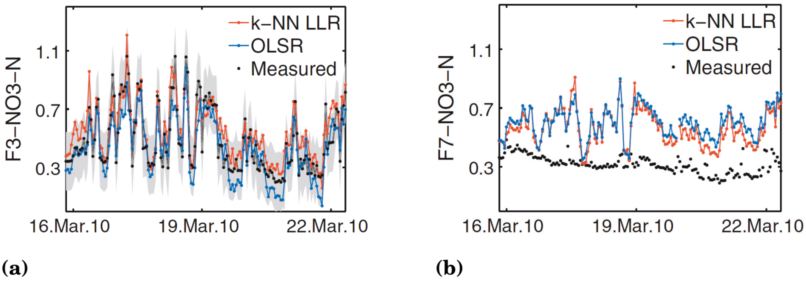


Figure 6.4. Measured and estimated NO_3-N concentrations in Filter 3 (c) and Filter 7 (d). Figure adopted from Publication II.

Unlike in many studies presented in Chapter 4, the performances of linear and nonlinear modelling methods were not found to differ much from each other. This demonstrates that the choice of method should be application-specific. One reason for relatively similar performance might relate to the fairly short retention time of about 25 min of the post-filtration unit. Most of the studies showing benefits of the nonlinear approaches concern ASPs with retention times of several hours. In this case, simpler linear methods would be selected due to the model selection criterion preferring computationally light and transparent models.

From a wider perspective, the study further demonstrates the potential benefits for monitoring and supervision of WWTPs through the use of data-derived soft sensors. Such devices can be used as inexpensive back-up systems for conventional analytical instrumentation. Even though this research considered real-time estimation in specific process units, denitrifying post-filters, which has not been investigated previously with multivariate techniques, a corresponding soft sensor modelling approach can be applied to other units in WWTPs. The only major requirements for this are the availabilities of an adequate historical operation data set and on-line measurements.

6.3 Switching system allowing the complementary use of measurements and estimates

The soft sensors described in Section 6.2 were further investigated. In particular, the aim was to develop a system for automatic selection of whether to use a soft sensor estimate or a hardware measurement at any time individually in each filter. Usually, the presented switching systems for choosing between estimates and measurements simply prefer measurements whenever they are available. Otherwise, estimates are to be employed as back-ups (*e.g.*, Bhuyan, 2011). However, the target of the system development was to enable a complementary use of measurements and estimates providing the best available information for the process operation. The preliminary version of the switching system is presented in Haimi et al. (2013b) and, later, it has been finalized as described at length in Publication III.

6.3.1 Development of the switching system

The reasoning employed in the sample selection step of the soft sensor design (Subsection 6.2.1) was adapted to rank the samples. Specifically, the ranking of the input data was connected to the assessment of soft sensor estimates, while the ranking of the output data concerned the hardware measurements. Two independent rankings were combined into a filter-specific *preferability index* that suggested whether to use measurements or estimates, given the current situation.

Ranking of the soft sensor estimates

In the soft sensor ranking procedure, the quality of the input data used for the regression models was considered instead of the actual estimates. The presented approach to ranking the quality of the estimates considering each Filter i at any time k consisted of the following conditions:

1. Existence of all the soft sensor input signals x_{ij}^{PS};
2. Availability of the T^2 and Q statistics from the model of Filter i;
3. Respects of the statistics' confidence limits T^2_{lim} and Q_{lim}.

A soft sensor estimate in Filter i that satisfied its three (two, one or none) conditions could be denoted as 3E (2E, 1E or 0E, respectively). As the procedure was based on a filter-specific approach, the labels given to the estimates in each filter at the same time were different.

Ranking of the hardware sensor measurements

The reasoning of the output sample selection procedure could also have been applied in the ranking of the NO_3-N measurements. However, a limitation of using a similar logic was derived from the fact that the approach considered the global quality of the signals from ten filter cells and, therefore, an anomalous measurement signal in one or a few filter(s) could have led to a violation of J_{lim}.

For this reason, the output data were further examined by using an analysis of the variables' contributions to T^2 and Q along with their statistical control limits (*e.g.*, Oakland, 2003), determined independently for each Filter i. The upper and lower control limits for the contributions to the statistics were defined using the observations that represented normal operating conditions (see *e.g.*, Westerhuis et al., 2000; Choi and Lee, 2005). The contribution analysis with the control limits provided an additional condition for the ranking of the measurements:

1. Existence of the hardware sensor measurement x_i^{PH} in Filter *i*;
2. Availability of the J statistic from the model concerning all filters;
3. Respect of the statistic's confidence limit J_{lim};
4. Respects of the contribution control limits of T^2 and Q of Filter *i*.

A hardware sensor measurement $Fi\text{-}NO_3\text{-}N$ in Filter i that satisfied its four (three, two, one or none) conditions could be denoted as 4M (3M, 2M, 1M or 0M) at any time k. The first condition concerned the existence of a filter-specific NO_3-N measurement whereas the measurements from all the filters were needed for continuing the assessment by computing J. The label 3M was given in two different cases considered as representing the same severity level: (i) when J_{lim} was violated, but both the control limits were respected, or (ii) when J_{lim} was respected, but one or both of the control limits were violated.

Switching map

The above-described rankings were employed for deciding whether it is preferable to use estimates or measurements. The possible combinations of the given E and M labels are depicted in the switching map in Figure 6.5. Each combination was quantified in such a way that the suitability of one option over the other was determined numerically with the preferability index P combining the rankings :

$$P = \frac{l(\mathtt{M}) - l(\mathtt{E})}{\sqrt{l(\mathtt{M})^2 + l(\mathtt{E})^2}} \tag{6.2}$$

where $l(\mathtt{M})$ is the distance of an observation on the map from the origin (0E,0M), along the y-axis and $l(\mathtt{E})$ is the corresponding distance along the x-axis. If $P > 0$, the measurement is to be preferred, whereas the estimates are favored if $P < 0$. When $P = 0$, the quality of both options is ranked as equal. The larger the absolute value of P is, the more obvious the choice is. The P values are used to dye the map as shown in the colorbar.

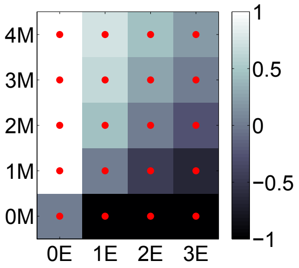


Figure 6.5. Switching map showing the combinations of soft sensor estimates and hardware sensor measurements denoted with different rankings. Figure adopted from Publication III.

6.3.2 Use of the switching system

The shares of hits in the different positions on the switching maps, considering three years of filter-specific operations, are depicted in Figure 6.6. The sizes of the red dots represent the proportional shares of the hits in each position. Most of the maps are relatively similar, indicating consistency in the ranks. However, some variation between maps can be observed. For instance, in Filter 1 there are more hits in the position `3E,0M` compared with the other filters, expressing a situation when the measurement signal was not available, but the corresponding soft sensor estimate was of a good quality. In Filter 7, the ratio of the hits in the positions `3E,4M` and `3E,3M` is smaller in comparison with the other filters, indicating more violations of the control limits when J_{lim} was often globally respected.

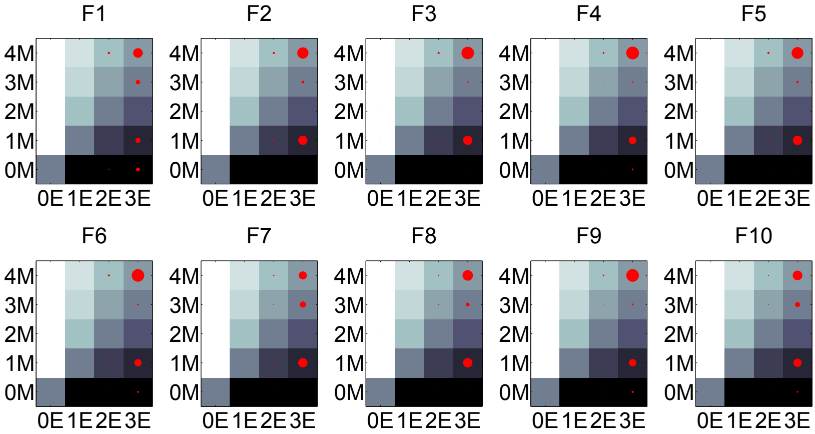


Figure 6.6. Shares of the hits in different positions on the switching maps for the individual filters indicated as the sizes of the red dots. Figure adopted from Publication III.

Percentages of the preferred options according to the proposed switching procedure and the situations when neither the measurement nor the estimate were available over three years are presented in Table 6.3. The hardware measurements were preferred in the largest share of the occasions in all filters (38.4–54.6%), but the use of the estimates was also favored in many situations (38.7–39.1%). When both measurements and estimates were of equal quality (6.0–22.4%), the one to be used could have be chosen based on the previous occasion when one of them was preferred ($P \neq 0$).

The proposed switching system is a novel tool tested for a tertiary treatment unit in a large municipal WWTP. The main benefit of the system

Table 6.3. Percentages of the preferred choices for measurement or estimation, when both were of equal quality and when neither of them were available, for all the filters over three years.

Filter	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Measurement	45.1	50.5	54.3	54.6	52.1	54.3	38.4	46.6	53.4	42.7
Estimate	38.7	39.1	39.0	38.9	38.8	38.9	39.1	39.0	38.7	39.0
Both equal	16.0	10.4	6.7	6.0	9.1	6.5	22.4	14.3	7.8	18.3
Neither available	0.2	0.0	0.0	0.5	0.0	0.3	0.1	0.1	0.1	0.0

is in enabling the complementary use of measurements and estimates based on their quality rankings, always preferring the best information available. For instance, when the NO_3-N measurements in a filter cell evolve from a normal concentration range to an abnormal range and, at the same time, the NO_3-N estimates show consistent behaviour, the values of the preferability index typically suggest employing the estimates in the process operation. Usually in WWTPs, the additional carbon source is fed into the anoxic part of an ASP, but in the plant considered, methanol is dosed in the post-filtration unit that finalizes the nitrogen removal of the targeted level. Therefore, the proposed switching system that prefers more valid information about the NO_3-N concentrations would support the efficient operation of the unit and secure achieving the treatment goals.

Typically, soft sensor estimates are not used in conjunction with hardware measurements. Instead, they provide the on-line approximations of variables that are otherwise only analyzed in a laboratory or are used as back-ups only when real-time instruments suffer from down-time. The novelty of the presented system results from combining the rankings based on well-established techniques such as PCA, model residual statistics, contribution analyses, and control limits in a simple and understandable manner that enhances the process operation. The switching system was developed using the data of a rather uncommon treatment unit, but its basic principles can be adapted also to other applications involving data-derived real-time estimates and hardware measurements that are alternatives for each other. However, because the output ranking in this case concerns measurements from ten filters, in a case dealing with only one process unit and one output signal, some adaptation is required. In such a case, for example, univariate control charts with confidence limits could be used to rank the quality of the outputs.

6.4 Soft sensor for anomaly detection in an activated sludge process

An anomaly detection and isolation system was developed for the activated sludge process in Viikinmäki WWTP. The considered ASP is described in Section 2.2. The aim of designing the soft sensor tool was to provide the operators with an early warning of process and instrument abnormalities. This would also motivate a more efficient use of the hardware sensors in the process operation. A preliminary version the soft sensor design and discussion on the performance of the system are presented in Haimi et al. (2013a) and are available in completed and detailed forms especially in Publication IV.

6.4.1 Development of soft sensors

Sample selection

The process data considered in the study concern one ASP line. The collected data covered two years of process operations recorded as hourly average values. Sample selection considered discarding only the obvious outliers that violated the technological limitations of hardware instruments. Such observations were considered the measurements that exceeded the instrument range or that were associated with unfeasible zero-values.

Variable selection

From all the acquired data, the variables selected for anomaly monitoring are collected in Table 6.4. The primary criterion in variable selection was their potentiality to be employed in the future advanced control schemes. The only investigated sensor that at the time of the research used in the aeration control was $BE\text{-}NH_4$. However, it was also included because the initial data inspection showed frequent unexpected peaks. The presence of problems in the measurement reliability of all the selected variables was detected in the data inspection step. Therefore, an adequate anomaly detection system would increase the feasibility of the investigated variables for process control purposes. DO sensors were not considered because they were already successfully used in the aeration control and the data inspection did not reveal any signs of unreliability.

Table 6.4. Process variables considered for the anomaly detection study in the ASP.

Symbol	TAG	Description	Unit
x_1	$BI\text{-}NH_4$	Bioreactor influent ammonium-nitrogen	mg/l
x_2	$BI\text{-}SS$	Bioreactor influent suspended solids	mg/l
x_3	$BI\text{-}Q$	Bioreactor influent wastewater flow rate	m^3/s
x_4	$Z6\text{-}SS$	Mixed liquor suspended solids in zone 6	g/l
x_5	$BE\text{-}NH_4$	Bioreactor effluent ammonium-nitrogen	mg/l
x_6	$BE\text{-}NO_3$	Bioreactor effluent nitrate-nitrogen	mg/l
x_7	$BE\text{-}pH$	Bioreactor effluent pH	–

Model selection and parametrization

PCA-based models were considered for the soft sensor development and the T^2 and Q statistics were employed in the anomaly detection procedure. In the preliminary studies, the stationary PCA was found to be inadequate for the considered application due to the excessive number of alarms in the model testing. Thus, adaptive PCA methods were used that aimed at better adjustment to the time-evolving processes. Specifically, moving-window approaches with fixed window-lengths MWPCA and two variants with adaptive window-lengths (He and Yang, 2008; Ayech et al., 2012, denoted AMWPCA_1 and AMWPCA_2 hereafter, see Subsection 5.1.3) were explored.

Several parameters had to be adjusted for the anomaly monitoring procedure. A separate data set was used for the selection of the window-length calculation parameters for the AMWPCA approaches and the shift-sizes for all the investigated approaches.

To calculate the window-lengths L adaptively, the minimum window-length L_{min} was set at 24 h (1 day) and the maximum window-length L_{max} at 168 h (1 week). The chosen values associate with the diurnal and weekly trends typical for the influent flow rate and concentrations in municipal WWTPs (Henze et al., 2008). In the selection of the parameters and the shift-sizes of the AMWPCA methods, sufficient responses in the window-lengths to changes among the relationship of the process variables were targeted. On the other hand, the practicable anomaly detection sensitivity of the monitoring systems was required. Such parameters were chosen that satisfied both the criteria. For the tested MWPCA method, window-length and shift-size were selected to approximately correspond to the average levels of those in the AMWPCA techniques.

Anomaly monitoring procedure

A 30-day period at the beginning of the acquired data set was, first, cleaned from the samples exceeding T^2_{lim} or Q_{lim} and, then, used for defining the reference values for the AMWPCA approaches. In the on-line monitoring, the PCA models were built with a standardized training data set $\mathbf{X}_{trn}$ defined by the window-length and the thresholds T^2_{lim} and Q_{lim} were calculated for every model. The eigengap technique was applied for selecting a subset of retained PCs individually for each model. Finally, T^2_{lim} and Q_{lim} were used in monitoring a testing data set $\mathbf{X}_{tst}$ comprised of samples of a time span defined by the shift-size. The variables' contributions to T^2 and Q were examined for isolating the fault source in the cases of T^2_{lim} and Q_{lim} violations.

If the required proportion P of the samples in $\mathbf{X}_{trn}$ was not available due to the discarding procedure in the sample selection step, the model was not considered representative and the previous valid model was maintained. The required P value varied depending on L, the criterion being stricter for shorter windows in order to have sufficiently samples for building descriptive PCA models.

A detailed description of the anomaly monitoring algorithm is available in Publication IV. Simplifying, the main steps of the monitoring procedure for the AMWPCA approaches are:

1. Calculating the references $\|\mathbf{\Delta b}_0\|$ and $\|\mathbf{\Delta R}_0\|$ (AMWPCA_1), or $\mathbf{R}_{ref}$ (AMWPCA_2) using $\mathbf{X}_{ref}$ off-line;
2. Calculating the window-length L using $\mathbf{X}_{trn}$ at the intervals set by Z;
3. Calculating the PCA model and the statistics' confidence limits T^2_{lim} and Q_{lim} at the intervals set by Z;
4. Calculating T^2 and Q for each incoming sample of $\mathbf{X}_{tst}$;
5. If $T^2 > T^2_{lim}$ and/or $Q > Q_{lim}$, calculating the variables' contributions to T^2 and/or Q for isolating the anomaly source.

The step 1 is performed only once off-line, whereas the other steps consider on-line anomaly monitoring. For the anomaly detection procedure using MWPCA, only the steps $3 - 5$ described above are considered.

6.4.2 Performance of soft sensors

The shares of normal and anomalous observations among the testing data using the different approaches are collected in Table 6.5. The slight difference between the monitoring performances of AMWPCA_1 and AMWPCA_2 is explained by their significantly dissimilar average window-lengths, 101.6 h and 58.3 h, respectively. A corresponding divergence occurred also in their variabilities, which can be seen in Figure 6.7 where their window-lengths are depicted. A different choice of the function parameters would naturally impact the monitoring performances of the approaches. The AMWPCA_1 technique was found to be widely adjustable whereas the tuning capacity of AMWPCA_2 was indicated to be more limited. The anomaly monitoring performance of MWPCA, defined in terms of the total number of the detected anomalies, corresponded with the AMWPCA_2 approach. In the MWPCA method, the selection of window-length was indicated to affect considerably the anomaly detection sensitivity, the models with smaller windows being stricter.

Table 6.5. Shares of normal and anomalous samples in testing data using the different moving-window PCA methods.

Method	Normal, %	Anomalous, %
AMWPCA_1	76.7	23.3
AMWPCA_2	78.9	22.1
MWPCA	78.9	22.1

The variables most frequently found to be responsible for anomalies were the same for all the approaches when investigating the largest contributions during the threshold violations. $BI\text{-}NH4$ was isolated most often as the fault source, followed by $BI\text{-}SS$. All the approaches suggested $BI\text{-}Q$ and $Z6\text{-}SS$ as causing the smallest number of anomalies. The average number of retained PCs ranged between 2.25 and 2.30 using the different approaches. On average, the models reconstructed 72.3–75.5% of the total variation with different approaches.

The experimental results showed that when adequate window-lengths were defined, drifts and peaks in measurements as well as process anomalies can be detected. Moreover, the correct isolation of the variables causing the anomalies was demonstrated. The results indicated that the AMWPCA_1 approach successfully modified the window-lengths according to the changes taking place among the relationship of the considered

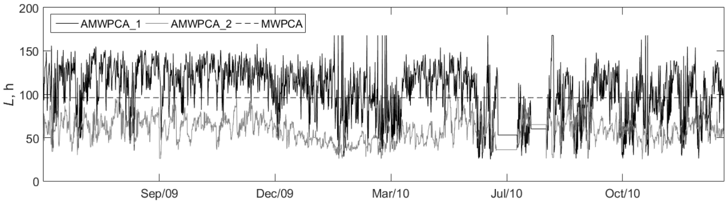


Figure 6.7. Time series of the window-lengths of the investigated anomaly monitoring approaches. Figure adopted from Publication IV.

process variables. For the techniques with adapting window-lengths, the tuning of the parameters of the window-length definition equations and of the shift-sizes specifying the model recalculation intervals proved to be the critical factors for the anomaly monitoring performances.

One of the motivations for the investigation from a scientific perspective was to search for adequate techniques for instrument monitoring in WWTPs, the area being scarcely researched earlier as concluded in Subsection 6.1.1. This goal was approached by using not only well-established moving-window techniques but also their recently proposed adaptive window-length variants for exploring their suitability for the considered application. Also the number of retained PCs was defined specifically for each model, instead of using the same number of PCs throughout the monitoring as is often the case in moving-window applications.

The starting point being instrument monitoring, the scope of the work was later extended to also include monitoring process anomalies that, in the worst case, can cause significant operational problems in biological treatment units such as ASPs. An example of this is a sudden drop of pH in the ASP that gives rise to a long-term diminishing of nitrification due to the decreased growth-rate of the ammonium-oxidizing bacteria. Such a change in a process variable, pH in this case, can be rapidly detected with the investigated techniques, as demonstrated in the original publication and an early warning for the process operators can be provided. Therefore, the use of a soft sensor for anomaly monitoring could prevent severe problems in the treatment units.

As for the practical perspective, the proposed techniques could be installed as an inexpensive software tool for monitoring sensor and process abnormalities. This would also increase the potential of sensors to be used in advanced control schemes, such as the model predictive controller proposed for the considered ASP (Mulas et al., 2015). However, more studies for optimizing the model parameters should be performed before the system is implemented. In addition, assessing the false alarm rate of the monitoring system would be beneficial in order to facilitate its successful adoption. The presented algorithm could easily be extended to include more sensors and adapted for other process units in WWTPs that are equipped with abundant on-line instrumentation.

7. Conclusions

This thesis has been concerned with the development of data-derived soft sensors for biological wastewater treatment systems by means of multivariate statistical techniques. The main motivation for this was to design software tools that enable more efficient and safer process operation by complementing the information produced by the conventional instrumentation. As today's treatment plants are amply instrumented, extensive historical process data are stored in their data acquisition systems. The historical data contain plentiful information on the process operation and provide invaluable material for the data-derived soft sensor design. In point of fact, the development of the soft sensors in the thesis was realized by utilizing the historical process data and mathematical algorithms, coupled with the process understanding also needed for the task. On the other hand, the high dimensionality of the measured data provided a motivation for using multivariate techniques instead of univariate methods. The employed multivariate statistical techniques include principal component analysis with its extensions and least squares based regression methods.

The thesis presents a framework for designing data-derived soft sensors that is general in nature. The trends in soft sensor development in the wastewater sector and the research gaps were analyzed based on an extensive number of case studies presented in the literature. An array of soft sensors for on-line prediction was developed for supporting the cost-efficient operation of the biological post-filtration unit of a large-scale municipal WWTP. The performance of the soft sensors was demonstrated to be good, with accuracies comparable with the hardware sensors. The on-line estimation study was extended by developing a switching system that enables the complementary use of the aforementioned soft sensor estimates and the corresponding measurements of the existing hardware instruments. The system was based on real-time ranking of the qualities of the measurement and prediction data. Finally, a soft sensor for detecting process and instrument anomalies in the activated sludge process of the same WWTP was designed. Adaptive methods based on moving-window techniques were successfully used in this application for coping with the time-evolving process conditions.

The results of the thesis give rise to both theoretical and practical implications. The most important ones are summarized in the following.

- Novel approaches to using multivariate techniques for developing data-derived soft sensors for biological wastewater treatment applications were provided. Both classical multivariate methods and their extensions were employed in an innovative manner in the different steps of the soft sensor design based on the requirements set by the acquired operational data and the tasks at hand. Some of these techniques have not been used in wastewater treatment applications earlier and, hence, the thesis presents new approaches for the development of software tools and provides inspiration for future studies. These techniques include robust, nonlinear and adaptive multivariate methods, the feasibility of which for the field of operation was investigated. Different options for defining the window-lengths of the adaptive models were also examined. In addition, the thesis addresses an on-line estimation task in a rather uncommon treatment unit, the denitrifying post-filtration, which has not earlier been considered in corresponding research.
- The possibility of using both hardware instrument measurements and soft sensor estimates in a complimentary way in the process operation was demonstrated by means of the proposed switching system. Based on earlier literature, that is not a typical manner of exploiting soft sensors. Contradicting, soft sensor estimates usually serve as back-up options for existing instruments when they suffer down-time or they are used when hardware sensors are not applied at all in measuring the process variables. This proposal, in particular, opens new insights for the scientific community, not only when dealing with wastewater treatment, but also with relevance for other industrial sectors.
- The presented results provide novel examples about the design of soft sensors based on the real-life data of a large municipal WWTP. Imple-

mentation of these kinds of software aiming at improved information for control purposes would allow for safer process operation under the tight treatment regulations and demanding economic constraints. The designed tools can be used, for instance, to secure an appropriate chemical dosing and for providing an early warning about process and online instrument malfunctioning. Therefore, the adaptation of presented frameworks and ideas provides plant operators and engineering companies involved in the wastewater sector with new perspectives and flexibly designable tools that complement the traditionally used automation and instrumentation in the industry. Additionally, corresponding soft sensors can be adapted to other fields of industry when sufficiently operational data are available.

The findings of this thesis and the experience accrued during the doctoral work provide several insights for further development and future research.

Soft sensor development is a time-consuming task. In particular, the pre-processing of the data acquired from the plant requires a considerable effort. Therefore, developing frameworks for automated pre-processing would provide considerable benefits in comparison with the typical manual approach. Also, the model maintenance step is of high importance for the successful soft sensor implementation under time-evolving process conditions. More practical experiences and reports of the soft sensor maintenance, for instance, by means of model adaptation would be beneficial for future soft sensor developers.

Particularly in the wastewater treatment sector, research focusing on hardware instrument monitoring has been limited. The treatment units and their sequence in the treatment lines have become more complex and this progression is likely to continue into the future. Thus, process control needs are growing more varied, which makes the role of reliable real-time information about primary variables even more crucial. For this reason, more attention should be paid to developing of practicable techniques for sensor monitoring. This would motivate, for instance, a more intensive use of the sensors close to the inlet of WWTPs in feedforward control and, hence, support the adaptation of operation practices to the changing process conditions in advance.

A switching system for using soft sensor estimates and instrument measurements in a consistent and complementary manner was proposed in this thesis. The economic evaluation of the switching system operated practically with closed control loops in comparison with the traditional approaches, where only estimates or only measurements are employed, would be useful. Additionally, it would be beneficial to investigate alternatives and improvements to the presented switching system.

More practical experience of the use of soft sensors in WWTPs would be needed, especially as the literature suggests that only a small share of the proposed soft sensors have actually been implemented in the plants. The reasons for the scarce number of reported soft sensor implementations might be diverse. Kordon (2012) has recognized, for instance, the following issues as limiting the application of intelligent systems in industry: wrong expectations of the final users from the technology; lack of professional marketing of the developed systems; the proposals looking too academic or not understandable; and underestimating the maintenance and support needs. All of these are things worth considering and discussion when researchers develop soft sensors for wastewater treatment practitioners. In fact, many implemented soft sensors in WWTPs have been designed by people who actually work in the facilities and, hence, understand the properties of the software tool and how it responds to the assigned operational problem (Lumley, 2002; Cecil, 2004; Äijälä and Lumley, 2006; Cecil and Kozlowska, 2010), although university researchers (Bongards, 2001; González and López García, 2006; Lee et al., 2008) and experts from engineering companies (Boger, 1992; Cohen et al., 1997) have also reported practical implementations. In any case, close co-operation between the researchers and the plant operators is needed for successful software development. An increased number of references of soft sensor implementations in WWTPs would, without doubt, raise interest in applying these tools in the wastewater industry.

The data-derived soft sensors employing multivariate statistical techniques proved to be capable of extracting easily understandable and practicable information that the high-dimensional data contain. The use of such soft sensors allows for the more intensive use of real-time measured data and, consequently, for more advanced process operations that satisfy the treatment targets of the wastewater treatment facilities cost-efficiently.

References

Aarnio, P., Minkkinen, P., 1986. Application of partial least-squares modelling in the optimization of a waste-water treatment plant. Anal. Chim. Acta 191, 457–460.

Abonyi, J., Farsang, B., Kulcsar, T., 2014. Data-driven development and maintenance of soft-sensors, in: Proc. of IEEE the 12th International Symposium on Applied Machine Intelligence and Informatics, Herl'any, Slovakia. pp. 239–244.

Aguado, D., 2005. Application of multivariate statistical methods for modelling and monitoring a sequencing batch reactor for wastewater treatment. Ph.D. thesis. Universitat Politècnica de València, Department of Hydraulic Engineering and Environment. Valencia, Spain. In Spanish.

Aguado, D., Ferrer, A., Ferrer, J., Seco, A., 2007a. Multivariate SPC of a sequencing batch reactor for wastewater treatment. Chemometr. Intell. Lab. 85, 82–93.

Aguado, D., Ferrer, A., Seco, A., Ferrer, J., 2006. Comparison of different predictive models for nutrient estimation in a sequencing batch reactor for wastewater treatment. Chemometr. Intell. Lab. 84, 75–81.

Aguado, D., Ribes, J., Montoya, T., Ferrer, J., Seco, A., 2009. A methodology for sequencing batch reactor identification with artificial neural networks: A case study. Comput. Chem. Eng. 33, 465–472.

Aguado, D., Rosen, C., 2008. Multivariate statistical monitoring of continuous wastewater treatment plants. Eng. Appl. Artif. Intel. 21, 1080–1091.

Aguado, D., Zarzo, M., Seco, A., Ferrer, A., 2007b. Process understanding of a wastewater batch reactor with block-wise PLS. Environmetrics 18, 551–560.

Aguilar-Martin, J., López de Mántaras, R., 1982. The process of classification and learning the meaning of linguistic descriptors of concepts, in: Gupta, M., Sanchez, E. (Eds.), Approximate Reasoning in Decision Analysis. North-Holland Publishing Company, New York, pp. 165–175.

Äijälä, G., Lumley, D., 2006. Integrated soft sensor model for flow control. Water Sci. Technol. 53, 473–482.

Alcala, C., Qin, S., 2009. Reconstruction-based contribution for process monitoring. Automatica 45, 1593–1600.

Alferes, J., Tik, S., Copp, J., Vanrolleghem, P.A., 2013. Advanced monitoring of water systems using *in situ* measurement stations: data validation and fault detection. Water Sci. Technol. 68, 1022–1030.

Amaral, A.L., Mesquita, D.P., Ferreira, E.C., 2013. Automatic identification of activated sludge disturbances and assessment of operational parameters. Chemosphere 91, 705–710.

Andersen, C.M., Bro, R., 2010. Variable selection in regression – a tutorial. J. Chemometr. 24, 728–737.

Anderson, T., 2003. An Introduction to Multivariate Statistical Analysis (3rd edition). Wiley Series in Probability and Statistics, John Wiley & Sons.

Angelakis, A.N., Rose, J.B., 2014. Evolution of Sanitation and Wastewater Technologies through the Centuries. IWA Publishing.

Ansari, A.A., Gill, S.S., 2014. Eutrophication: Causes, Consequences and Control, Volume 2. Springer.

Ansari, A.A., Gill, S.S., Lanza, G.R., Rast, W., 2010. Eutrophication: Causes, Consequences and Control. Springer.

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Statist. Surv. 4, 40–79.

Atkeson, C., Moore, A., Schaal, S., 1997. Locally weighted learning. Artif. Intell. Rev. 11, 11–73.

Atkinson, A.C., Riani, M., Cerioli, A., 2004. Exploring multivariate data with the forward search. Springer Series in Statistics, Springer.

Ayech, N., Chakour, C., Harkat, M.F., 2012. New adaptive moving window PCA for process monitoring fault detection, in: Proc. of the 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Mexico City, Mexico. pp. 606–611.

Baffi, G., Martin, E., Morris, A., 1999. Non-linear projection to latent structures revisited: the quadratic PLS algorithm. Comput. Chem. Eng. 23, 395–411.

Baggiani, F., Marsili-Libelli, S., 2009. Real-time fault detection and isolation in biological wastewater treatment plants. Water Sci. Technol. 60, 2949–2961.

Bagheri, M., Mirbagheri, S., Ehteshami, M., Bagheri, Z., 2015. Modeling of a sequencing batch reactor treating municipal wastewater using multi-layer perceptron and radial basis function artificial neural networks. Process Saf. Environ. 93, 111–123.

Bakshi, B.R., 1998. Multiscale PCA with application to multivariate statistical process monitoring. AIChE J. 40, 1596–1610.

Bang, Y., Yoo, C., Lee, I.B., 2003. Nonlinear PLS modeling with fuzzy inference system. Chemometr. Intell. Lab. 64, 137–155.

Batstone, D., Keller, J., Angelidaki, I., Kalyuzhny, S., Pavlostathis, S., Rozzi, A., Sanders, W., Siegrist, H., Vavilin, V., 2002. Anaerobic Digestion Model No. 1 (ADM1). Scientific and Technical Reports, No. 13, IWA Publishing.

Benedetti, L., Langeveld, J., Comeau, A., Corominas, L., Daigger, G., Martin, C., Mikkelsen, P.S., Vezzaro, L., Weijers, S., Vanrolleghem, P.A., 2013. Modelling and monitoring of integrated urban wastewater systems: review on status and perspectives. Water Sci. Technol. 68, 1203–1215.

Bezdek, J., Ehrlich, R., Full, W., 1984. FCM: The fuzzy *c*-means clustering algorithm. Comput. Geosci. 10, 191–203.

Bhuyan, M., 2011. Intelligent Instrumentation: Principles and Applications. CRC Press Taylor & Francis Group.

Björk, Å., 1996. Numerical Methods for Least Squares Problems. Other Titles in Applied Mathematics, SIAM.

Blom, H.A., 1996. Indirect measurement of key water quality parameters in sewage treatment plants. J. Chemometr. 10, 697–706.

Boger, Z., 1992. Application of neural networks to water and wastewater treatment plant operation. ISA T. 31, 25–33.

Bongards, M., 2001. Improving the efficiency of a wastewater treatment plant by fuzzy control and neural networks. Water Sci. Technol. 43, 189–196.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brjdanovic, D., Meijer, S.C.F., Lopez-Vazquez, C.M., Hooijmans, C.M., van Loosdrecht, M.C.M., 2015. Applications of Activated Sludge Models. IWA Publishing.

Bro, R., 1997. PARAFAC: Tutorial and applications. Chemometr. Intell. Lab. 38, 149–171.

Bro, R., Smilde, A.K., 2003. Centering and scaling in component analysis. J. Chemometr. 17, 16–33.

Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6, 2812–2831.

Budka, M., Eastwood, M., Gabrys, B., Kadlec, P., Salvador, M.M., Schwan, S., Tsakonas, A., Žliobaitė, I., 2014. From sensor readings to predictions: On the process of developing practical soft sensors, in: Blockeel, H., van Leeuwen, M., Vinciotti, V. (Eds.), Lect. Notes Comput. Sc.. Springer-Verlag. volume 8819, pp. 49–60.

Campisano, A., Cabot Ple, J., Muschalla, D., Pleau, M., Vanrolleghem, P.A., 2013. Potential and limitations of modern equipment for real time control of urban wastewater systems. Urban Water J. 10, 300–311.

Capodaglio, A.G., Jones, H.V., Novotny, V., Feng, X., 1991. Sludge bulking analysis and forecasting: Application of system identification and artificial neural computing technologies. Water Res. 25, 1217–1224.

Cattell, R.B., 1966. The scree test for the number of factors. Multiv. Behav. Res. 1, 245–276.

Çinar, Ö., 2005. New tool for evaluation of performance of wastewater treatment plant: Artificial neural network. Process Biochem. 40, 2980–2984.

Cecil, D., 2004. A software nitrate sensor based on ammonium and redox signals. Water Sci. Technol. 48, 259–265.

Cecil, D., Kozlowska, M., 2010. Software sensors are a real alternative to true sensors. Environ. Modell. Softw. 25, 622–625.

Chen, B., Wu, H., Li, S.F.Y., 2014. Development of variable pathlength UV-vis spectroscopy combined with partial-least-squares regression for wastewater chemical oxygen demand (COD) monitoring. Talanta 120, 325–330.

Chen, K., Castillo, I., Chiang, L.H., Yu, J., 2015. Soft sensor model maintenance: A case study in industrial processes, in: Proc. of the 9th IFAC International Symposium on Advanced Control of Chemical Processes, Whistler, Canada. pp. 427–432.

Choi, D.J., Park, H., 2001. A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process. Water Res. 35, 3959–3967.

Choi, S.W., Lee, I.B., 2005. Multiblock PLS-based localized process diagnosis. J. Process Contr. 15, 295–306.

Choi, S.W., Martin, E.B., Morris, A.J., Lee, I.B., 2006. Adaptive multivariate statistical process control for monitoring time-varying processes. Ind. Eng. Chem. Res. 45, 3108–3118.

Choubert, J.M., Rieger, L., Shaw, A., Copp, J., Spérandio, M., Sørensen, K., Rönner-Holm, S., Morgenroth, E., Melcer, H., Gillot, S., 2013. Rethinking wastewater characterisation methods for activated sludge systems – a position paper. Water Sci. Technol. 67, 2363–2373.

Civelekoglu, G., Perendeci, A., Yigit, N.O., Kitis, M., 2007. Modeling carbon and nitrogen removal in an industrial wastewater treatment plant using an adaptive network-based fuzzy inference system. Clean – Soil, Air, Water 35, 617–625.

Cleveland, W., Devlin, S., 1988. Locally weighted regression: An approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83, 596–610.

Cohen, A., Janssen, G., Brewster, S.D., Seeley, R., Boogert, A.A., Graham, A.A., Mardani, M.R., Clarke, N., Kasabov, N.K., 1997. Application of computational intelligence for on-line control of a sequencing batch reactor (SBR) at Morrinsville sewage treatment plant. Water Sci. Technol. 35, 63–71.

Côté, M., Grandjean, B.P.A., Lessard, P., Thibault, J., 1995. Dynamic modelling of the activated sludge process: improving prediction using neural networks. Water Res. 29, 995–1004.

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press.

Daszykowski, M., Kaczmarek, K., Vander Heyden, Y., Walczak, B., 2007. Robust statistics in data analysis – A review: Basic concepts. Chemometr. Intell. Lab. 85, 203–219.

Davis, C., Kahan, W.M., 1970. The rotation of eigenvectors by a perturbation, iii. SIAM J. Numer Anal. 7, 1–46.

de Leeuw, J., 2014. History of nonlinear principal component analysis, in: Blasius, J., Greenacre, M. (Eds.), Visualization and Verbalization of Data. Chapman and Hall / CRC, pp. 45–60.

Dellana, S.A., West, D., 2009. Predictive modeling for wastewater applications: Linear and nonlinear approaches. Environ. Modell. Softw. 24, 96–106.

Deza, M.M., Deza, E., 2014. Encyclopedia of Distances (3rd edition). Springer-Verlag.

Diaconis, P., Efron, B., 1983. Computer-intensive methods in statistics. Sci. Am. 248, 116–130.

Diana, G., Tommasi, C., 2010. Cross-validation methods in principal component analysis: A comparison. Stat. Methods Appl. 11, 71–82.

Dias, A.M.A., Moita, I., Páscoa, R., Alves, M.M., Lopes, J.A., Ferreira, E.C., 2008. Activated sludge process monitoring through *in situ* near-infrared spectral analysis. Water Sci. Technol. 57, 1643–1650.

Dochain, D., Vanrolleghem, P.A., 2001. Dynamical Modelling and Estimation in Wastewater Treatment Processes. IWA Publishing.

Dominguez, D., Gujer, W., 2006. Evolution of a wastewater treatment plant challenges traditional design concepts. Water Res. 40, 1389–1396.

Duda, R., Hart, P., Stork, D., 2000. Pattern Classification (2nd edition). John Wiley & Sons.

Dürrenmatt, D., Gujer, W., 2012. Data-driven modeling approaches to support wastewater treatment plant operation. Environ. Modell. Softw. 30, 47–56.

Dzakpasu, M., Scholz, M., McCarthy, V., Jordan, S., Sani, A., 2015. Adaptive neuro-fuzzy inference system for real-time monitoring of integrated-constructed wetlands. Water Sci. Technol. 71, 22–30.

Everitt, B.S., Landau, S., Leese, M., Stahl, D., 2011. Cluster Analysis (5th edition). Wiley's series in probability and statistics, John Wiley & Sons.

Farmer, S.A., 1971. An investigation into the results of principal component analysis of data derived from random numbers. J. Roy. Stat. Soc. D-Sta. 20, 63–72.

Fernandez, F.J., Seco, A., Ferrer, J., Rodrigo, M.A., 2009. Use of neurofuzzy networks to improve wastewater flow-rate forecasting. Environ. Modell. Softw. 24, 686–693.

Fisher, R., 1936. The use of multiple measurements in taxonomic problems. Annals Eugen. 7, 179–188.

Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M.G., 2007. Soft Sensors for Monitoring and Control of Industrial Processes. Advances in Industrial Control, Springer.

Frosch Møller, S., von Frese, J., Bro, R., 2005. Robust methods for multivariate data analysis. J. Chemometrics 19, 549–563.

Fujiwara, K., Kano, M., Hasebe, S., Takinami, A., 2009. Soft-sensor development using correlation-based just-in-time modeling. AIChE J. 55, 1754–1765.

Fullér, R., 2000. Introduction to Neuro-Fuzzy Systems. volume 2 of *Advances in Intelligent and Soft Computing*. Springer.

Galinha, C.F., Carvalho, G., Portugal, C.A.M., Guglielmi, G., Reis, M.A.M., Crespo, J.G., 2012. Multivariate statistically-based modelling of a membrane bioreactor for wastewater treatment using 2D fluorescence monitoring data. Water Res. 46, 3623–3636.

Galpin, J., Hawkins, D., 1987. Methods of L_1 estimation of a covariance matrix. Comput. Stat. Data An. 5, 305–319.

Ge, Z., Song, Z., Gao, F., 2013. Review of recent research on data-based process monitoring. Ind. Eng. Chem. Res. 52, 3543–3562.

Gernaey, K., Vanderhasselt, A., Bogaert, H., Vanrolleghem, P., Verstraete, W., 1998. Sensors to monitor biological nitrogen removal and activated sludge settling. J. Microbiol. Meth. 32, 193–204.

Gernaey, K.V., Cervera-Padrell, A.E., Woodley, J.M., 2012. A perspective on PSE in pharmaceutical process development and innovation. Comput. Chem. Eng. 42, 15–29.

Gernaey, K.V., Flores-Alsina, X., Rosen, C., Benedetti, L., Jeppsson, U., 2011. Dynamic influent pollutant disturbance scenario generation using a phenomenological modelling approach. Environ. Modell. Softw. 26, 1255–1267.

Gernaey, K.V., Jeppsson, U., Batstone, D.J., Ingildsen, P., 2006. Sensors to monitor biological nitrogen removal and activated sludge settling. Water Sci. Technol. 53, 159–167.

Gernaey, K.V., Jeppsson, U., Vanrolleghem, P.A., Copp, J.B., 2014. Benchmarking of Control Strategies for Wastewater Treatment Plants. Scientific and Technical Reports, No. 23, IWA Publishing.

Gernaey, K.V., van Loosdrecht, M.C.M., Henze, M., Lind, M., Jørgensen, S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation: state of the art. Environ. Modell. Softw. 19, 763–783.

González, G.D., 2010. Soft sensing, in: Sbárbaro, D., del Villar, R. (Eds.), Advanced Control and Supervision of Mineral Processing Plants. Springer. Advances in Industrial Control, pp. 143–212.

González, I.M., López García, H., 2006. End-point detection of the aerobic phase in a biological reactor using SOM and clustering algorithms. Eng. Appl. Artif. Intel. 19, 19–28.

Gustafson, E., Kessel, W., 1979. Fuzzy clustering with a fuzzy covariance matrix, in: Proc. of the IEEE Conference on Decision and Control, San Diego, USA.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Haimi, H., 2006. An intermittently aerated carrier media process for wastewater treatment. Licentiate thesis. Helsinki University of Technology, Department of Chemical Engineering. In Finnish.

Haimi, H., Hurme, M., 2004. Use of chemometrics in the research of activated sludge plants. Plant Design Report Series 77. Helsinki University of Technology, Department of Chemical Engineering. In Finnish.

Haimi, H., Mulas, M., Corona, F., Marsili-Libelli, S., Lindell, P., Heinonen, M., Vahala, R., 2013a. Data derived sensor fault detection in the activated sludge process of the Viikinmäki wastewater treatment plant, in: Proc. of the 11th IWA Conference on Instrumentation, Control and Automation, Narbonne, France.

Haimi, H., Mulas, M., Corona, F., Sundell, L., Heinonen, M., Vahala, R., 2011. Outlier detection for the denitrifying post-filtration unit of a municipal wastewater plant: The Viikinmäki case, in: Proc. of Watermatex 2011, the 8th IWA Symposium on Systems Analysis and Integrated Assessment, San Sebastián, Spain. pp. 793–800.

Haimi, H., Mulas, M., Corona, F., Sundell, L., Heinonen, M., Vahala, R., 2013b. Shall we use hardware sensor measurements or soft-sensor estimates? Case study in a full-scale WWTP, in: Proc. of the 11th IWA Conference on Instrumentation, Control and Automation, Narbonne, France.

Haimi, H., Mulas, M., Sahlstedt, K., Vahala, R., 2009. Advanced operation and control methods of municipal wastewater treatment processes in Finland. Technical Report. Helsinki University of Technology. Water and Wastewater Engineering Publications.

Haimi, H., Mulas, M., Vahala, R., 2010. Process automation in wastewater treatment plants: the Finnish experience. E-WAter, 2010/03, 1–17.

Haimi, H., Mulas, M., Vahala, R., Corona, F., 2013c. Soft-sensors in wastewater treatment: The benefits of the data-driven approach, in: Proc. of the Automaatio XX, Helsinki, Finland.

Hartigan, J., Wong, M., 1979. A k-means clustering algorithm. Appl. Stat. 28, 100–108.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd edition). Springer Series in Statistics, Springer.

Hauduc, H., Gillot, S., Rieger, L., Ohtsuki, T., Shaw, A., Takács, I., Winkler, S., 2009. Activated sludge modelling in practice: An international survey. Water Sci. Technol. 60, 1943–1951.

Hauduc, H., Rieger, L., Oehmen, A., van Loosdrecht, M.C.M., Comeau, Y., Héduit, A., Vanrolleghem, P.A., Gillot, S., 2013. Critical review of activated sludge modeling: State of process knowledge, modeling concepts, and limitations. Biotechnol. Bioeng. 110, 24–46.

Hauduc, H., Rieger, L., Ohtsuki, T., Shaw, A., Takács, I., Winkler, S., Héduit, A., Vanrolleghem, P.A., Gillot, S., 2011. Activated sludge modelling: Development and potential use of a practical applications database. Water Sci. Technol. 63, 2164–2182.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation (2nd edition. Prentice Hall.

He, X.B., Yang, Y.P., 2008. Variable MWPCA for adaptive process monitoring. Ind. Eng. Chem. Res. 47, 419–427.

HELCOM, 2013. Approaches and methods for eutrophication target setting in the Baltic Sea region. volume 133 of *Baltic Sea Environment Proceedings*. Helsinki Commission.

Henze, M., Gujer, W., Mino, T., van Loosedrecht, M., 2000. Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. Scientific and Technical Reports, No. 9, IWA Publishing.

Henze, M., van Loosdrecht, M.C.M., Ekama, G.A., Brdjanovic, D., 2008. Biological Wastewater Treatment – Principles, Modelling and Design. IWA Publishing.

Huang, M.Z., Wan, J.Q., Ma, Y.W., Li, W.J., Sun, X.F., Wan, Y., 2010. A fast predicting neural fuzzy model for on-line estimation of nutrient dynamics in an anoxic/oxic process. Bioresource Technol. 101, 1642–1651.

Hubert, M., Rousseeuw, P.J., Verboven, S., 2002. A fast method for robust principal components with applications to chemometrics. Chemometr. Intell. Lab. 60, 101–111.

Hubert, M., Van Kerckhoven, J., Verdonck, T., 2012. Robust PARAFAC for incomplete data. J. Chemometr. 26, 290–298.

Jackson, D.A., 1993. Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. Ecology 74, 2204–2214.

Jackson, J.E., Mudholkar, G.S., 1979. Control procedures for residual associated with principal component analysis. Technometrics 21, 341–349.

Jang, J.S.R., 1993. ANFIS: Adaptive-network-based fuzzy inference system. IEEE T. Syst. Man Cy. 23, 665–685.

Jansson, Å., Röttorp, J., Rahmberg, M., 2002. Development of a software sensor for phosphorus in municipal wastewater. J. Chemometr. 16, 542–547.

Jenkins, D., Wanner, J., 2014. Activated Sludge – 100 Years and Counting. IWA Publishing.

Jolliffe, I.T., 1972. Discarding variables in a principal component analysis. i: Artificial data. J. Roy. Stat. Soc. C-App. 21, 160–173.

Jolliffe, I.T., 2002. Principal Component Analysis (2nd edition). Springer Series in Statistics, Springer.

Kadlec, P., 2009. On robust and adaptive soft sensors. Ph.D. thesis. Bournemouth University, School of Design, Engineering & Computing. Bournemouth, UK.

Kadlec, P., Gabrys, B., 2008. Soft sensor based on adaptive local learning, in: Proc. of the 15th International Conference On Neural Information Processing, Auckland, New Zealand. pp. 1172–1179.

Kadlec, P., Gabrys, B., 2009. Soft sensors: Where are we and what are the current and future challenges?, in: Proc. of the 2nd IFAC International Conference on Intelligent Control Systems and Signal Processing, Istanbul, Turkey.

Kadlec, P., Gabrys, B., Grbić, R., 2011. Review of adaptation mechanisms for data-driven soft sensors. Comput. Chem. Eng. 35, 1–24.

Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven soft sensors in the process industry. Comput. Chem. Eng. 33, 795–814.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educ. Psychol. Meas. 2, 141–151.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Trans. ASME, Ser. D: J. Basic Eng. 82, 35–45.

Kaneko, H., Okada, T., Funatsu, K., 2014. Selective use of adaptive soft sensors based on process state. Ind. Eng. Chem. Res. 53, 15962–15968.

Kano, M., Nakagawa, Y., 2014. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. Comput. Chem. Eng. 34, 12–24.

Kariya, T., Kurata, H., 2004. Generalized Least Squares. Wiley Series in Probability and Statistics, John Wiley & Sons.

Kim, M.H., Kim, Y.S., Prabu, A.A., Yoo, C.K., 2009. A systematic approach to data-driven modeling and soft sensing in a full-scale plant. Water Sci. Technol. 60, 363–370.

Kohonen, T., 2001. Self-Organizing Maps. volume 30 of *Springer Series in Information Sciences*. Springer.

Kordon, A., 2012. Applying intelligent systems in industry: A realistic overview, in: Proc. of the 6th IEEE International Conference Intelligent Systems, Sofia, Bulgaria.

Krishnapuram, R., Keller, J., 1993. A possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. 1, 98–110.

Kruger, U., Xie, L., 2012. Advances in Statistical Monitoring of Complex Multivariate Processes: With Applications in Industrial Process Control. Statistics in Practice, John Wiley & Sons.

Ku, W., Storer, R.H., Georgakis, C., 1995. Disturbance detection and isolation by dynamic principal component analysis. Chemometr. Intell. Lab. 30, 179–196.

Lee, C., Choi, S.W., Lee, I.B., 2004. Sensor fault identification based on time-lagged PCA in dynamic processes. Chemometr. Intell. Lab. 70, 165–178.

Lee, C., Choi, S.W., Lee, I.B., 2006. Sensor fault diagnosis in a wastewater treatment process. Water Sci. Technol. 53, 251–257.

Lee, D.S., Vanrolleghem, P.A., Park, J.M., 2005. Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. J. Biotechnol. 115, 317–328.

Lee, H.W., Lee, M.W., Park, J.M., 2007. Robust adaptive partial least squares modeling of a full-scale industrial wastewater treatment process. Ind. Eng. Chem. Res. 46, 955–964.

Lee, H.W., Lee, M.W., Park, J.M., 2009. Multi-scale extension of PLS algorithm for advanced on-line process monitoring. Chemometr. Intell. Lab. 98, 201–212.

Lee, J., Verleysen, M., 2007. Nonlinear Dimensionality Reduction. Information Science and Statistics, Springer.

Lee, M.W., Hong, S.H., Choi, H., Kim, J.H., Lee, D.S., Park, J.M., 2008. Real-time remote monitoring of small-scaled biological wastewater treatment plants by a multivariate statistical process control and neural network-based software sensors. Process Biochem. 43, 1107–1113.

Lennox, J., Rosen, C., 2002. Adaptive multiscale principal components analysis for online monitoring of wastewater treatment. Water Sci. Technol. 45, 227–235.

Lennox, J.A., 2002. Multivariate subspaces for fault detection and isolation: with application to the wastewater treatment process. Ph.D. thesis. The University of Queensland, School of Engineering. Brisbane, Australia.

Leonard, T., Hsu, J.S.J., 1999. Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Li, B., Stenstrom, M.K., 2014. Research advances and challenges in onedimensional modeling of secondary settling tanks – A critical review. Water Res. 65, 40–63.

Li, G., Chen, Z., 1985. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. J. Am. Stat. Assoc. 80, 759–766.

Li, H., Yu, D., Braun, J.E., 2011. A review of virtual sensing technology and application in building systems. HVAC&R. Res. 17, 619–645.

Li, W., Yue, H.H., Valle-Cervantes, S., Qin, S.J., 2000. Recursive PCA for adaptive process monitoring. J. Process Contr. 10, 471–486.

Lin, B., Recke, B., Knudsen, J.K.H., Jørgensen, S.B., 2007. A systematic approach for soft sensor development. Comput. Chem. Eng. 31, 419–425.

Liu, Y., Chen, J., Sun, Z., Li, Y., Huang, D., 2014. A probabilistic self-validating soft-sensor with application to wastewater treatment. Comput. Chem. Eng. 71, 263–280.

Liukkonen, M., Havia, E., Hiltunen, Y., 2012. Computational intelligence in mass soldering of electronics – A survey. Expert Syst. Appl. 39, 9928–9937.

Lu, B., Castillo, I., Chiang, L., Edgar, T.F., 2014. Industrial PLS model variable selection using moving window variable importance in projection. Chemometr. Intell. Lab. 135, 90–109.

Lumley, D., 2002. On-line instrument confirmation: how can we check that our instruments are working? Water Sci. Technol. 53, 469–476.

Luttmann, R., Bracewell, D.G., Cornelissen, G., Gernaey, K.V., Glassey, J., Hass, V.C., Kaiser, C., Preusse, C., Striedner, G., Mandenius, C.F., 2012. Soft sensors in bioprocessing: A status report and recommendations. Biotechnol. J. 7, 1040–1048.

MacGregor, J.F., Jaeckle, C., Kiparissides, C., Koutoudi, M., 1994. Process monitoring and diagnosis by multiblock PLS methods. AIChE J. 40, 826–838.

Madsen, M., Holm-Nielsen, J.B., Esbensen, K.H., 2011. Monitoring of anaerobic digestion processes: A review perspective. Renew. Sust. Energ. Rev. 15, 3141–3155.

Maere, T., Villez, K., Marsili-Libelli, S., Naessens, W., Nopens, I., 2012. Membrane bioreactor fouling behaviour assessment through principal component analysis and fuzzy clustering. Water Res. 46, 6132–6142.

Mahalanobis, P.S., 1936. On the generalised distance in statistics. Proc. Nat. Inst. Sci. India 2, 49–55.

Mandenius, C.F., Gustavsson, R., 2014. Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics. J. Chem. Technol. Biotechnol. 90, 215–227.

Martin, C., Vanrolleghem, P.A., 2014. Analysing, completing, and generating influent data for WWTP modelling: A critical review. Environ. Modell. Softw. 60, 188–201.

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kaufman, L., 1988. Chemometrics: A Textbook. Elsevier Science.

Metcalf & Eddy, 2003. Wastewater Engineering – Treatment and Reuse (4th edition). McGraw-Hill Inc. Revised by G. Tchobanoglous, F.L. Burton and H.D. Stensel.

Miettinen, T., Hurse, T.J., Connor, M.A., Reinikainen, S.P., Minkkinen, 2004. Multivariate monitoring of a biological wastewater treatment process: a case study at Melbourne Water's Western Treatment Plant. Chemometr. Intell. Lab. 73, 131–138.

Mirin, S.N.S., Wahab, N.A., 2014. Fault detection and monitoring using multiscale principal component analysis at a sewage treatment plant. Jurnal Teknologi 70, 87–92.

Mitra, S.K., Kaiser, J.F., 1993. Handbook for Digital Signal Processing. John Wiley & Sons.

Molvar, A.E., Roesler, J.F., Babcock, R.H., 1976. Instrumentation and automation experiences in wastewater-treatment facilities. volume EPA-600/2-76-198 of *Environmental protection technology*. U.S. Environmental Protection Agency, Office of Research and Development, Municipal Environmental Research Laboratory.

Moon, T.S., Kim, Y.J., Kim, J.R., Cha, J.H., Kim, D.H., Kim, C.W., 2009. Identification of process operating state with operational map in municipal wastewater treatment plant. J. Environ. Manage. 90, 772–778.

Mujunen, S.P., 1999. Multivariate monitoring of wastewater treatment processes in pulp and paper industry. Ph.D. thesis. Lappeenranta University of Technology, Department of Chemical Technology. Lappeenranta, Finland.

Mujunen, S.P., Minkkinen, P., Teppola, P., Wirkkala, R.S., 1998. Modeling of activated sludge plants treatment efficiency with PLSR: a process analytical case study. Chemometr. Intell. Lab. 41, 83–94.

Mulas, M., Corona, F., Haimi, H., Sundell, L., Heinonen, M., Vahala, R., 2012. Nitrate estimation in a denitrifying post-filtration unit of a municipal wastewater plant: The Viikinmäki case. Water Sci. Technol. 65, 1521–1529.

Mulas, M., Tronci, S., Corona, F., Haimi, H., Lindell, P., Heinonen, M., Vahala, R., Baratti, R., 2015. Predictive control of an activated sludge process: An application to the Viikinmäki wastewater treatment plant. J. Process Contr. 35, 89–100.

Neumann, M.B., Rieckermann, J., Hug, T., Gujer, W., 2015. Adaptation in hindsight: Dynamics and drivers shaping urban wastewater systems. J. Environ. Manage. 151, 404–415.

Nomikos, P., MacGregor, J.F., 1995. Multivariate SPC charts for monitoring batch processes. Technometrics 37, 41–59.

Oakland, J.S., 2003. Statistical Process Control (5th edition). Butterworth-Heinemann.

Oehmen, A., Lemos, P.C., Carvalho, G., Yuan, Z., Keller, J., Blackall, L.L., Reis, M.A.M., 2007. Advances in enhanced biological phosphorus removal: From micro to macro scale. Water Res. 41, 2271–2300.

Olsson, G., 1977. State of the art in sewage treatment control. American Inst. of Chemical Engineers, Symp Series 159, 52–76.

Olsson, G., 2012. ICA and me – a subjective review. Water Res. 46, 1585–1626.

Olsson, G., Carlsson, B., Comas, J., Copp, J., Gernaey, K.V., Ingildsen, P., Jeppsson, U., Kim, C., Rieger, L., Rodríguez-Roda, I., Steyer, J.P., Takács, I., Vanrolleghem, P.A., Vargas, A., Yuan, Z., Åmand, L., 2014. Instrumentation, control and automation in wastewater – from London 1973 to Narbonne 2013. Water Sci. Technol. 69, 1373–1385.

Olsson, G., Jeppsson, U., 2006. Plant-wide control: dream, necessity or reality? Water Sci. Technol. 53, 121–129.

Olsson, G., Newell, B., 1999. Wastewater Treatment Systems: Modelling, Diagnosis and Control. IWA Publishing.

Olsson, G., Newell, B., Rosen, C., Ingildsen, P., 2003. Application of information technology to decision support in treatment plant operation. Water Sci. Technol. 47, 35–42.

Olsson, G., Nielsen, M., Yuan, Z., Lynggaard-Jensen, A., Steyer, J.P., 2005. Instrumentation, Control and Automation in Wastewater Systems. Scientific and Technical Reports, No. 15, IWA Publishing.

Petersen, B., Gernaey, K., Henze, M., Vanrolleghem, P.A., 2003. Calibration of activated sludge models: a critical review of experimental designs, in: Agathos, S.N., Reineke, W. (Eds.), Biotechnology for the Environment: Wastewater Treatment and Modeling, Waste Gas Handling. Kluwer Academic Publishers.

Phillips, H.M., Sahlstedt, K.E., Frank, K., Bratby, J., Brennan, W., Rogowski, S., Pier, D., Anderson, W., Mulas, M., Copp, J.B., Shirodkar, N., 2009. Wastewater treatment modelling in practice: a collaborative discussion of the state of the art. Water Sci. Technol. 59, 695–704.

Platikanov, S., Rodriguez-Mozaz, S., Huerta, B., Barceló, D., Cros, J., Batle, M., Poch, G., Tauler, R., 2014. Chemometrics quality assessment of wastewater treatment plant effluents using physicochemical parameters and UV absorption measurements. J. Environ. Manage. 140, 33–44.

Poch, M., Comas, J., Porro, J., Baserba L. Corominas, M.G., Pijuan, M., 2014. Where are we in wastewater treatment plants data management? A review and a proposal, in: Proc. of the 7th International Congress on Environmental Modelling and Software, iEMSs 2014, San Diego, USA. pp. 1450–1455.

Qin, S., McAvoy, T., 1992. Nonlinear PLS modeling using neural networks. Comput. Chem. Eng. 16, 379–391.

Qin, S.J., 2011. Survey on data-driven industrial process monitoring and diagnosis. Annu. Rev. Control 36, 220–234.

Qin, S.J., Dunia, R., 2000. Determining the number of principal components for best reconstruction. J. Process Contr. 10, 245–250.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.

Ráduly, B., Gernaey, K.V., Capodaglio, A.G., Mikkelsen, P.S., Henze, M., 2007. Artificial neural networks for rapid WWTP performance evaluation: Methodology and case study. Environ. Modell. Softw. 22, 1208–1216.

Raich, A., Çinar, A., 1996. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. AIChE J. 42, 995–1009.

Ramsay, J., Silverman, B.W., 2005. Functional Data Analysis (2nd edition). Springer Series in Statistics, Springer.

Robinson, R.B., Cox, C.D., Odom, K., 2005. Identifying outliers in correlated water quality data. J. Environ. Eng. 131, 651–657.

Rosen, C., 2001. A Chemometric Approach to Process Monitoring and Control – with Applications to Wastewater Treatment Operation. Ph.D. thesis. Lund University, Department of Industrial Electrical Engineering and Automation. Lund, Sweden.

Rosen, C., Jeppsson, U., Vanrolleghem, P.A., 2004. Towards a common benchmark for long-term process control and monitoring performance evaluation. Water Sci. Technol. 50, 41–49.

Rosen, C., Larsson, M., Jeppsson, U., Yuan, Z., 2002. A framework for extreme-event control in wastewater treatment. Water Sci. Technol. 45, 299–308.

Rosen, C., Lennox, J.A., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. Water Res. 35, 3402–3410.

Rosen, C., Olsson, G., 1998. Disturbance detection in wastewater treatment plants. Water Sci. Technol. 37, 197–205.

Rosen, C., Röttorp, J., Jeppsson, U., 2003. Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation. Water Sci. Technol. 47, 171–179.

Rosen, C., Yuan, Z., 2001. Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering. Water Sci. Technol. 43, 147–156.

Rosipal, R., Trejo, L.J., 2001. Kernel partial least squares regression in reproducing kernel hilbert space. J. Mach. Learning Res. 2, 97–123.

Rousseeuw, P.J., Croux, C., 1993. Alternatives to the median absolute deviation. J. A. Stat. Assoc. 88, 1273–1283.

Rousseeuw, P.J., Debruyne, M., Engelen, S., Hubert, M., 2006. Robustness and outlier detection in chemometrics. Crit. Rev. Anal. Chem. 36, 221–242.

Rousseeuw, P.J., Hubert, M., 2011. Robust statistics for outlier detection. WIREs Data Mining Knowl. Discov. 1, 73–79.

Rustum, R., 2009. Modelling Activated Sludge Wastewater Treatment Plants Using Artificial Intelligence Techniques (Fuzzy Logic and Neural Networks). Ph.D. thesis. Heriot-Watt University, School of the Built Environment. Edinburgh, Scotland.

Rustum, R., Adeloye, A.J., Scholz, M., 2008. Applying Kohonen self-organizing map as a software sensor to predict biochemical oxygen demand. Water Environ. Res. 80, 32–40.

Ryan, T.P., 2008. Modern Regression Methods (2nd edition). Wiley Series in Probability and Statistics, John Wiley & Sons.

Saptoro, A., 2014. State of the art in the development of adaptive soft sensors based on just-in-time models. Procedia Chem. 9, 226–234.

Schölkopf, B., Smola, A., Müller, K., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. 10, 1299–1319.

Slišković, D., Grbić, R., Ž. Hocenski, 2011a. Methods for plant data-based process modeling in soft-sensor development. Automatika 52, 306–318.

Slišković, D., Grbić, R., Ž. Hocenski, 2011b. Online data preprocessing in the adaptive process model building based on plant data. Tehn. Vjesn. 18, 41–50.

Smilde, A., Bro, R., Geladi, P., 2004. Multi-way Analysis: Applications in the Chemical Sciences. John Wiley & Sons.

Stone, C., 1977. Consistent nonparametric regression. Ann. Stat. 5, 595–620.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. Roy. Stat. Soc. B 36, 111–147.

Strang, G., Nguyen, T., 1997. Wavelets and Filter Banks (2nd edition). Wellesley-Cambridge Press.

Sulthana, A., Latha, K.C., Imran, M., Rathan, R., Sridhar, R., Balasubramanian, S., 2014. Non-linear modeling using fuzzy principal component regression for Vidyaranyapuram sewage treatment plant, Mysore – India. Water Sci. Technol. 70, 1040–1047.

Sundell, L., 2008. Improvement of chemical dosing in a wastewater treatment plant. Master's thesis. Helsinki University of Technology, Department of Chemical Engineering. Espoo, Finland. In Finnish.

Takács, I., Patry, G., Nolasco, D., 1991. A dynamic model of the clarification-thickening process. Water Res. 25, 1263–1271.

Tao, E.P., Shen, W.H., Liu, T.L., Chen, X.Q., 2013. Fault diagnosis based on PCA for sensors of laboratorial wastewater treatment process. Chemometr. Intell. Lab. 128, 49–55.

Tay, J.H., Zhang, X., 1999. Neural fuzzy modeling of anaerobic biological wastewater treatment systems. Systems J. Environ. Eng., ASCE 125, 1149–1159.

Teppola, P., 1999. Multivariate process monitoring of sequential process data – A chemometric approach. Ph.D. thesis. Lappeenranta University of Technology, Department of Chemical Technology. Lappeenranta, Finland.

Teppola, P., Minkkinen, P., 1999. Possibilistic and fuzzy C-means clustering for process monitoring in an activated sludge waste-water treatment plant. J. Chemometr. 13, 445–459.

Teppola, P., Mujunen, S.P., Minkkinen, P., 1997. Partial least squares modeling of an activated sludge plant: A case study. Chemometr. Intell. Lab. 38, 197–208.

Teppola, P., Mujunen, S.P., Minkkinen, P., 1998. A combined approach of partial least squares and fuzzy c-means clustering for the monitoring of an activated-sludge waste-water treatment plant. Chemometr. Intell. Lab. 41, 95–103.

Teppola, P., Mujunen, S.P., Minkkinen, P., 1999a. Adaptive Fuzzy C-Means clustering in process monitoring. Chemometr. Intell. Lab. 45, 23–38.

Teppola, P., Mujunen, S.P., Minkkinen, P., 1999b. Kalman filter for updating the coefficients of regression models. A case study from an activated sludge waste-water treatment plant. Chemometr. Intell. Lab. 45, 371–384.

Timmerman, M., 2006. Multilevel component analysis. Brit. J. Math. Stat. Psy. 59, 301–320.

Tomita, R.K., Park, S.W., Sotomayor, O.A.Z., 2002. Analysis of activated sludge process using multivariate statistical tools – a PCA approach. Chem. Eng. J. 90, 283–290.

Valle, S., Li, W., Qin, S.J., 1999. Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods. Ind. Eng. Chem. Res. 38, 4389–4401.

van den Berg, R.A., Hoefsloot, H.C.J., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 7, 1–15.

Vanrolleghem, P.A., Clercq, B.D., Clercq, J.D., Devisscher, M., Kinnear, D.J., Nopens, I., 2006. New measurement techniques for secondary settlers: a review. Water Sci. Technol. 53, 419–429.

Vanrolleghem, P.A., Lee, D.S., 2003. On-line monitoring equipment for wastewater treatment processes: state of the art. Water Sci. Technol. 47, 1–34.

Velicer, W.F., 1976. Determining the number of components from the matrix of partial correlations. Psychometrika 41, 321–327.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., Yin, K., 2003. A review of process fault detection and diagnosis: Part III: Process history based methods. Comput. Chem. Eng. 27, 327–346.

Villez, K., 2007. Multivariate and qualitative data analysis for monitoring, diagnosis and control of sequencing batch reactors for wastewater treatment. Ph.D. thesis. Ghent University, Faculty of Bioscience Engineering. Ghent, Belgium.

Villez, K., Ruiz, M., Sin, G., Colomer, J., Rosen, C., Vanrolleghem, P.A., 2008. Combining multiway principal component analysis (MPCA) and clustering for efficient data mining of historical data sets of SBR processes. Water Sci. Technol. 57, 1659–1666.

Wan, J., Huang, M., Ma, Y., Guo, W., Wang, Y., Zhang, H., Li, W., Sun, X., 2011. Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based fuzzy inference system. Appl. Soft Comput. 11, 3238–3246.

Westerhuis, J.A., Gurden, S.P., Smilde, A.K., 2000. Generalized contribution plots in multivariate statistical process monitoring. Chemometr. Intell. Lab. 51, 95–114.

Wold, H., 1975. Soft modeling by latent variables: the nonlinear iterative partial least squares (NIPALS) approach, in: Gani, J. (Ed.), Perspectives in Probability and Statistics, Papers in Honor of M.S. Barlett. Academic Press, pp. 117–142.

Wold, S., 1978. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 20, 397–405.

Wold, S., 1995. Chemometrics; what do we mean with it, and what do we want from it? Chemometr. Intell. Lab. 30, 109–115.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometr. Intell. Lab. 58, 109–130.

Woo, S.H., Jeon, C.O., Yun, Y.S., Choi, H., Lee, C.S., Lee, D.S., 2009. On-line estimation of key process variables based on kernel partial least squares in an industrial cokes wastewater treatment plant. J. Hazard. Mater. 161, 538–544.

Yoo, C.K., Bang, Y.H., Lee, I.B., Vanrolleghem, P.A., Rosen, C., 2004. Application of fuzzy partial least squares (FPLS) modeling nonlinear biological processes. Korean J. Chem. Eng. 21, 1087–1097.

Yoo, C.K., Vanrolleghem, P.A., Lee, I.B., 2003. Nonlinear modeling and adaptive monitoring with fuzzy and multivariate statistical methods in biological wastewater treatment plants. J. Biotechnol. 105, 135–163.

Yoo, C.K., Villez, K., Van Hulle, S.W.H., Vanrolleghem, P.A., 2008. Enhanced process monitoring for wastewater treatment systems. Environmetrics 19, 602–617.

Yue, H.H., Qin, S.J., 2001. Reconstruction-based fault identification using a combined index. Ind. Eng. Chem. Res. 40, 4403–4414.

Zhu, Z., Corona, F., Lendasse, A., Baratti, R., Romagnoli, J.A., 2011. Local linear regression for soft-sensor design with application to an industrial deethanizer, in: Proc. of the18th IFAC World Congress, Milano, Italy. pp. 6212–6217.

Aalto-DD 24/2016

9 789526 066493

ISBN 978-952-60-6649-3 (printed)
ISBN 978-952-60-6650-9 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Engineering
Department of Built Environment
www.aalto.fi