Aalto-yliopisto
Insinööritieteiden
korkeakoulu

Sanni Eerikäinen

Developing process control and operation of Finnish municipal wastewater treatment plants with data analytics: examples from process industries and international utilities

Master's Thesis submitted for examination for the degree of Master of Science in Technology.

In Espoo 24.7.2020

Supervisor: Professor Anna Mikola
Advisor: D.Sc. (Tech.) Henri Haimi

Aalto-yliopisto
Insinööritieteiden korkeakoulu

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

Diplomityön tiivistelmä

Tekijä Sanni Eerikäinen	
Työn nimi Jätevedenpuhdistamojen prosessinohjauksen ja operoinnin kehittäminen data-analytiikan avulla: esimerkkejä teollisuudesta ja kansainvälisiltä puhdistamoilta	
Maisteriohjelma Vesi- ja ympäristötekniikka	**Koodi** ENG29
Työn valvoja Anna Mikola	
Työn ohjaaja(t) Henri Haimi	
Päivämäärä 24.7.2020 **Sivumäärä** 78+6	**Kieli** Englanti

Tiivistelmä

Prosessinohjaus- ja automaatiojärjestelmillä on keskeinen rooli modernien jätevedenpuhdistamojen operoinnissa. Prosessi- ja laitetietoa paremmin hyödyntämällä prosessia voidaan ohjata entistä tehokkaammin ja luotettavammin. Kiertotalous, ilmastonmuutos ja infrastruktuurin ikääntyminen korostavat entisestään tarvetta ymmärtää ja ohjata myös eri osaprosessien välisiä vuorovaikutuksia.

Tässä työssä tarkastellaan tarpeita, esteitä, kannustimia ja mahdollisuuksia kehittää jätevedenpuhdistamojen ohjausta ja operointia data-analytiikan avulla. Eri sidosryhmien näkemyksiä kartoitetaan haastatteluilla, joiden tuloksia käsitellään temaattisen analyysin kautta. Löydösten perusteella potentiaalisia ratkaisuja kartoitetaan suomalaisten ja kansainvälisten puhdistamojen sekä prosessiteollisuuden jo käyttämistä sovelluksista.

Löydökset osoittavat, että monilla puhdistamoilla tarvitaan nykyistä merkittävästi kattavampia menetelmiä instrumentoinnin, laitteiston ja ohjauksen laadunvarmistukseen, ennen kuin edistyneempien prosessinohjausmenetelmien käyttöönotto on mahdollista. Operoinnin toimintavarmuutta ja luotettavuutta voitaisiin kehittää monin tavoin hyödyntämällä jo kerättyä prosessi- ja laitetietoa.

Työssä esitellään yhteensä 14 esimerkkiä puhdistamoilla ja prosessiteollisuudessa käytössä olevista prosessinohjaus- ja laadunvarmistusmenetelmistä. Osalla ratkaisuista arvioidaan sellaisenaan olevan laajaa sovelluspotentiaalia suomalaisilla jätevedenpuhdistamoilla. Useiden ratkaisujen käyttöönottoa voitaisiin edistää pilotoinnilla tai jatkotutkimuksella potentiaalisten hyötyjen ja kustannusten arvioimiseksi. Jo kerättyä prosessi- ja laitetietoa hyödyntävien ratkaisujen kysynnän odotetaan tulevaisuudessa lisääntyvän, kun puhdistamojen operointi keskittyy ja paineet kustannus- ja energiatehokkuudelle kasvavat.

Avainsanat jätevedenpuhdistus, prosessinohjaus, operointi, data-analytiikka

Aalto University
School of Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Sanni Eerikäinen

Title of thesis Developing process control and operation of Finnish municipal wastewater treatment plants with data analytics: examples from process industries and international utilities

Master programme Water and Environmental Engineering **Code** ENG29

Thesis supervisor Anna Mikola

Thesis advisor(s) Henri Haimi

Date 24.7.2020 **Number of pages** 78+6 **Language** English

Abstract

Instrumentation, control and automation are central for operation of municipal wastewater treatment plants. Treatment performance can be further improved and secured by processing and analyzing the collected process and equipment data. New challenges from resource efficiency, climate change and aging infrastructure increase the demand for understanding and controlling plant-wide interactions.

This study aims to review what needs, barriers, incentives and opportunities Finnish wastewater treatment plants have for developing current process control and operation systems with data analytics. The study is conducted through interviews, thematic analysis and case studies of real-life applications in process industries and international utilities.

Results indicate that for many utilities, additional measures for quality assurance of instruments, equipment and controllers are necessary before advanced control strategies can be applied. Readily available data could be used to improve the operational reliability of the process.

14 case studies of advanced data processing, analysis and visualization methods used in Finnish and international wastewater treatment plants as well as Finnish process industries are reviewed. Examples include process optimization and quality assurance solutions that have proven benefits in operational use.

Applicability of these solutions for identified development needs is initially evaluated. Some of the examples are estimated to have direct potential for application in Finnish WWTPs. For other case studies, further piloting or research efforts to assess the feasibility and cost-benefits for WWTPs are suggested. As plant operation becomes more centralized and outsourced in the future, need for applying data analytics is expected to increase.

Keywords wastewater treatment, process control, operation and maintenance, data analytics

Forewords

This work reviews the development needs, barriers, incentives and opportunities of process control and operation in Finnish municipal wastewater treatment plants. Several professionals are interviewed and case studies from other countries and industries reviewed.

Idea for the topic originated already in Autumn 2018, when I did some elective studies on automation and electrical engineering. These courses made me realize how little us environmental engineers are taught of process control and automation – the systems actually operating the process we should be able to design. With all the digitalization hype seeming to have little impact to practice, I wanted to understand where the practitioners see the needs for development.

During this project, I have had the pleasure and privilege to meet, discuss and exchange ideas with over 30 professionals of wastewater treatment, process instrumentation, control and automation in Finland and several other countries. I want to express my sincerest gratitude to all of you for contributing to this work with your time and expertise. Each discussion has further enlarged my understanding and perspective to the issues at hand. Especially the site visits to nine WWTPs were very educational and memorable, not the least for all the kind and helpful people I met.

Most importantly, I would like to thank the six Finnish WWTPs that have supported and guided the work in the project steering committee. I want to thank also my advisor Henri for continuous guidance, supervisor Anna and also Riku for the comments and advice on the article. I would like to thank also Finnish Water Services Development Fund and the Finnish foundation Maa- ja vesitekniikan tuki ry for supporting the work.

Despite the exceptional circumstances in Spring 2020, I have enjoyed my time writing this thesis, learning so much and discovering new things to learn of. Warmest thanks to my family for the support and a superior workplace at our summer cottage. Lastly, thanks to Petro and all my wonderful friends: not that much for helping with the thesis (you didn't, really), but for making life outside it so much more fun.

In Espoo 24.7.2020

Sanni Eerikäinen

Table of Content

Symbols

K_c	gain factor
T_d	derivative time
T_i	integrative time
X^2	chi-square
c	control signal
e	control error
x_k	feature
ξ	slack variable

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Networks
ASP	Activated Sludge Process
DaaS	Data as a Service
DCS	Distributed Control Systems
DO	Dissolved Oxygen
ERP	Enterprise Resource Planning
FBD	Function Block Diagram
HMI	Human-Machine Interface
ICA	Instrumentation, Control and Automation
IoT	Internet of Things
IP	Internet Protocol
LIMS	Laboratory Information Management System
MES	Manufacturing Execution Systems
MLP	Multi-Layer Perceptron
MPC	Model Predictive Control
PCA	Principal Component Analysis
PID	Proportional, Integral and Derivative
PLC	Programmable Logic Controller
RTU	Remote Terminal Unit
SCADA	Supervisory Control and Data Acquisition
SPC	Statistical Process Control
SS	Suspended Solids
SVM	Support Vector Machine
WWTP	Wastewater Treatment Plant

Figures

Tables

1 Introduction

1.1 Background

Municipal wastewater treatment plants (WWTPs) protect the public health and environment from wastewater pollution. While understanding and requirements for pollutants removal continue to develop, also concerns for greenhouse gas emissions of WWTPs have arised (Monteith *et al.,* 2005). Plant performance is increasingly evaluated in terms of not only treatment result, but also energy and resource efficiency.

With also circular economy gaining more traction (Geissdoerfer *et al.,* 2017), recovery of nutrients, heat and clean water from the wastewater is progressing. For WWTPs, resource recovery will further highlight the need to understand and control plant-wide interactions in a more integrated manner (Solon *et al.,* 2019).

Meanwhile, the operating environment of WWTPs is changing. In Finland, climate change is expected to increase the variation of influent flows due to intensified periods of heavy rains, floods and drought (Vienonen *et al.,* 2012). Moreover, influent temperature variation and power outages during storm events are expected to cause additional challenges for process safety and efficiency.

Instrumentation, Control and Automation (ICA) systems are essential for control and operation of modern WWTPs to meet the effluent quality criteria, while minimizing the operational and capital expenses (Olsson, 2012). ICA systems have been utilized for systematic control and monitoring of the dynamics and disturbances in the treatment process since 1970s.

After that, digitalization has drastically transformed the world. Among other developments, it is estimated that volume of data in the world is doubled every two years (IDC, 2011). Simultaneously, computational capacity and tools for extracting meaningful information from incomprehensible amounts of data have been developed. Quantity of data has increased also at WWTPs. Olsson (2012) states that despite this development, collected data is not fully utilized for improving the process knowledge and operation strategies, *i.e.* data analytics. Instead, tacit knowledge and experience of process operation often leaves with the operators.

It seems that in process industries, collected data is more effectively utilized in process control and operation. Non-linear multi-variable control strategies, such as Model Predictive Control (MPC), have been and are increasingly applied in oil & gas, chemical, pulp & paper and food processing sectors (Jämsä-Jounela, 2007). Same industries have adopted data-based quality assurance tools, such as control loop performance monitoring (Bauer *et al.,* 2016).

Researchers have proposed similar applications also for municipal WWTPs. Yet, a review by Corominas *et al.* (2018) demonstrated that only 9 % of data analysis solutions studied for WWTPs were applied or validated in full-scale applications. This result seems to indicate a major gap between research and practice.

In the field of software engineering, the barriers in technology diffusion and acceptance are widely addressed. For better understanding of user perceptions and organizational factors, qualitative study methods are often applied (Dybå *et al.,* 2011). Barriers of advanced process control (APC) methods have been qualitatively studied also in *e.g.* power generation industry in the US (Smuts & Hussey, 2011).

This study aims to review the barriers, needs, incentives and opportunities of data analytics for process control and operation of Finnish municipal WWTPs. Practical barriers and future expectations are studied through interviews with different stakeholders involved in the design, operation and maintenance of ICA systems at municipal WWTPs. Interview contents are assessed with thematic analysis to conclude a broader view of practitioner perspectives.

Following the findings of the thematic analysis, several case studies of full-scale applications are reviewed to identify potential tools and practices in the future. 14 case studies are surveyed and compiled from both Finnish and international WWTPs and Finnish process industries. These case studies do not provide an exhaustive list of best practices, but concise examples of potential approaches for tackling current and future needs in process control and operation of Finnish WWTPs.

11 years ago, a state-of-the-art survey by Haimi *et al.* (2009) reviewed ICA technologies applied in Finnish WWTPs. All 24 respondents mildly or strongly agreed that ICA will become more important at WWTPs in the near future. After that, development of ICA in Finnish WWTPs has not been broadly reviewed, but a digitalization strategy for water utilities was published in February 2020 by Finnish Water Utilities Association (Ikäheimo, 2020). This study aims to also complement and update the information of these publications.

Results from the first part of the study, thematic analysis of the interviews, is published in the Water Science & Technology journal (Eerikäinen *et al.,* 2020). These materials and results are presented as a part of this thesis with reference to the paper accordingly.

1.2 Objectives

Objective of the study was outlined to the following research questions:

Question 1. What development incentives, needs and barriers do Finnish WWTPs have for improving the process control and operation with data analytics now and in the future?

Question 2. What solutions for developing process control and operation with data analytics are used in other sectors or countries, that might be feasible for Finnish WWTPs?

> *Sub-question 2.1.* What solutions for process control and operation with data analytics are used at WWTPs in other countries?
>
> *Sub-question 2.2*. What solutions for process control and operation with data analytics are used in process industries?

In this study, process industry refers to all those fields of industry, where the primary process is operated as continuous or batch-process. These include *e.g.* pulp and paper, minerals, food processing, oil and gas and chemical industries.

1.3 Limitations

Focus of the study was limited into topics with most foreseen potential for application in Finnish municipal WWTPs. Some aspects beyond these limitations are considered in the Chapter 6.

ICA in municipal wastewater treatment. Focus of the study was selected as wastewater treatment instead of drinking water treatment, due to generally higher complexity of the process. It was considered that *e.g.* energy saving potential from optimizing the control is higher in wastewater treatment. Industrial wastewater treatment was not discussed due to the large variation of process types and practices.

Control and operation of WWTPs. In terms of process control efficiency, also operation practices, including maintenance, were considered highly relevant and included in the study. Sewage network data and operation practices were considered whereas they are directly related to control, operation or maintenance of the WWTP.

Readily available process and equipment data. Focus of the study was in applications utilizing the data readily available at the plant. Also data from process instrumentation and equipment, *e.g.* electricity consumption and valve positions, was considered relevant.

Data processing, analysis and visualization. Description of case studies was focused on how the collected data is processed, analyzed and utilized to support the process control and operation. Issues related to different measurement technologies were not thoroughly discussed due to large site-specific variation. Similarly, issues related to data transfer, communication protocols and cyber security were not thoroughly discussed.

1.4 Structure of the thesis

Introduction to the structure and components of process control system is given in Chapter 2. In Chapter 3, some data analysis and information extraction methods are reviewed as presented in the literature. Chapter 4 describes the materials and methods used in the study. Results are introduced in Chapter 5 and further discussed in Chapter 6. Final conclusions are drawn in Chapter 7.

2 Process instrumentation, control and automation

In this chapter, common structure, components and factors affecting the performance of process control systems in municipal wastewater treatment are presented.

2.1 Process control system

ICA refers to the integrated practice of monitoring and controlling of industrial processes and systems (Olsson *et al.,* 2014). The objectives of process control in WWTPs can be summarized as to keep the plant running, satisfy effluent requirements, minimize the costs and integrate the plant operation (Olsson *et al.,* 2005).

Process control system consists of various layers that can be classified in several ways. Figure 1 presents an automation pyramid commonly used for illustration. In several fields of industries, Manufacturing Execution Systems (MES) are often used for plant planning and scheduling. On top of MES, additional Enterprise Resource Planning (ERP) system can be utilized for overall management of different business operations (Modrák & Mandulák, 2009). In case of WWTPs, MES can be considered analogic to plant-wide control, *i.e.* expanding the control to interactions between unit processes (Olsson *et al.,* 2005). Similarly, ERP can be considered analogic to system-wide control, taking into account also *e.g.* water production, consumers, network or discharge water body.

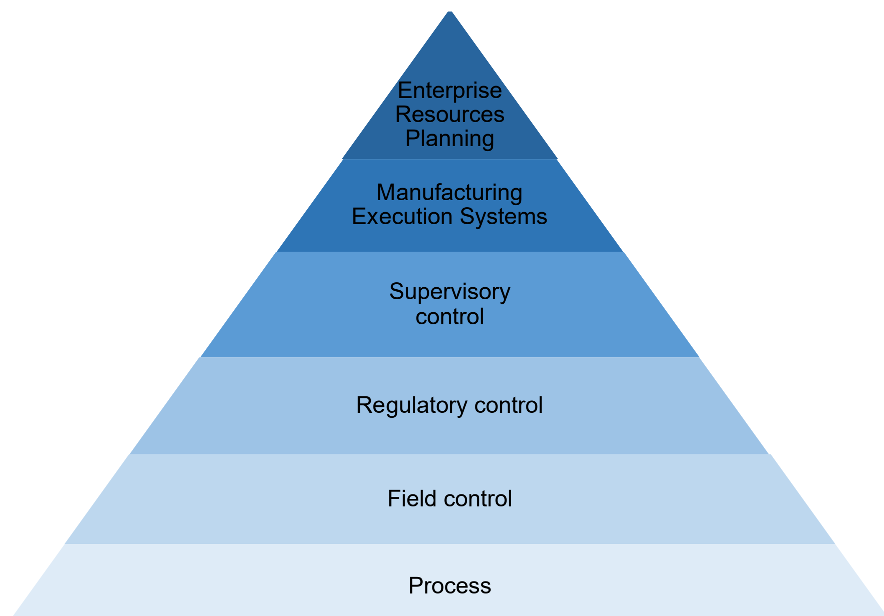

Figure 1. Conventional illustration of a process control system structure.

Dynamic processes are affected by external and internal disturbances. Regulatory control is implemented with several control loops (Agachi & Cristea, 2014), such as the example in Figure 2. Sensors measure the value of a process parameter and convert it into electrical signal (Anderson, 1997). Based on the deviation of the measured variable from the setpoint, the controller sends a control signal to the actuator. Actuators operate the devices that implement the actual change of energy or mass flow in the process. Term actuator is sometimes used also to refer to the combination of the actuator and actuating device.

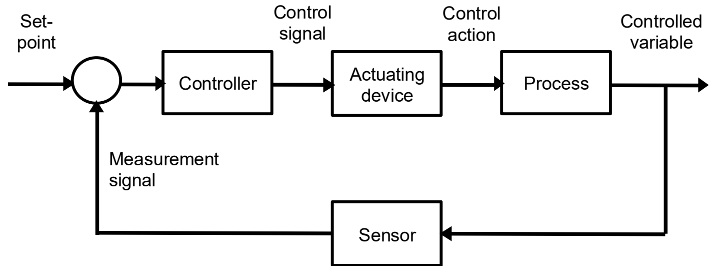

Figure 2. Block diagram of a simple feedback control loop.

In Chapter 2.3, process control layers are reviewed reflecting the automation pyramid. In place of MES and ERP, some aspects of plant-wide and system-wide control are discussed. First, some commonly applied control strategies are presented.

2.2 Process control strategies

Control strategy is outlined based on the control objective, controlled variables, manipulated variables and restrictions for other process variables (Olsson *et al.,* 2005). Based on this information, the control strategy and algorithms for implementation are selected. Control strategy is dependent also on the type of control signal applied. For this purpose, most commonly used on-off and PID control signals are first described.

2.2.1 Control signal

Controller can produce either a continuous or discontinuous control signal, which has a major impact for performance of the selected control strategy. Different control signals as described by Agachi & Cristea (2014) are briefly presented below.

Most simple controller is an on-off controller, that defines only two discrete control states for the actuating device. *E.g.* schedule based on-off control may be cost-effective and robust option for applications without strict performance requirements. More accurate control can be achieved with a PID controller, that can combine any of the proportional (P), integral (I) and derivative (D) control actions.

In proportional (P) control, control signal *c(t)* is proportional to the control error *e(t)*, *i.e.* deviation of the input signal from the desired setpoint multiplied by the gain factor K_c:

$$c(t)=K_c e(t)+c_0 \tag{1}$$

In steady state (*e(t)* = 0), control action equals to c_0. P term enables straightforward adjustment of manipulated variable to optimize the value of controlled variable. However, P term alone results into an offset from the setpoint.

Integral (I) term is proportional to the integral of the control error *e(t)*. Combined with the proportional term, integral term takes into account the past error values:

$$c(t)=K_c\left[e(t)+\frac{1}{T_i}\int e(t)dt\right] \tag{2}$$

Where T_i is the integral time. Combined PI control reduces the oscillation of the control signal and eliminates the steady-state error, *i.e.* makes it more accurate compared to single P control. Due to its stable nature, it is widely applied in the industry (Ketonen & Marttinen, 2001).

Derivative (D) term is based on the predicted future value of the control error *e(t)*. Derivative term takes into account the changing rate of *e(t)*, thus accelerating the control response time:

$$c(t)=K_c\left[e(t)+\frac{1}{T_i}\int e(t)dt+\ T_d\frac{de(t)}{dt}\right] \tag{3}$$

Where T_d is the derivative time. As a result, control response is significantly faster than with PI control, but also more vulnerable to noise in the measurement signal. In practice, the derivate term is often applied with a filter (Harju & Marttinen, 2000).

In the following section, most common control strategies are discussed as presented by Seborg *et al*. (2014).

2.2.2 Feedback control

In feedback control, the control signal is directly based on the difference between the measurement signal and the setpoint. Controller sends a control signal to the actuating device, which alters the manipulated variable accordingly. Typical feedback control loop is presented in Figure 2.

As a result, corrective action occurs as soon as the deviation of controlled variable from the setpoint is measured. Control action takes place regardless of the source or type of disturbance. However, complete elimination of setpoint deviation is not possible with feedback control, since disturbances can only be responded to once already occurred. In processes with long time delays, steady state might be difficult to reach (Seborg *et al.,* 2014).

Feedback control with on/off or PID controller is the most widely applied control method in WWTPs (Vanrolleghem, 2003).

2.2.3 Feedforward control

Feedforward control enables sending a correcting control signal to the actuator before a deviation from the setpoint has occurred. This is possible, if the disturbance entering the control system can be measured online. After that, feedback controller can correct the manipulated variable to maintain the controlled variable at the setpoint. Simplified block diagram of feedforward control is presented in Figure 3.

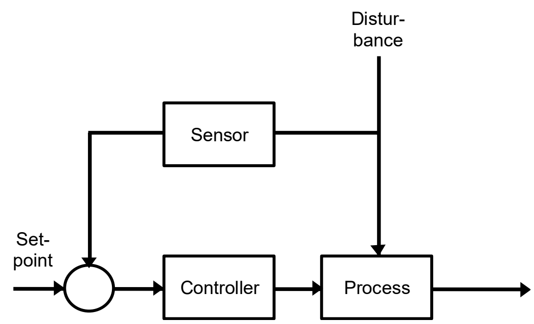

Figure 3. Simplified feedforward control loop.

Performance of feedforward control loop depends on the utilized process model, *i.e.* how the measured disturbance is predicted to impact the controlled and manipulated variables. Process model may be a steady-state or dynamic model. Feedforward is also commonly applied as ratio control, which aims to maintain a defined stoichiometric ratio between two variables, *e.g.* flow rates. In practice, a feedback loop is often added to compensate for uncertainties in the process model (Seborg *et al.,* 2014).

In wastewater treatment, feedforward control is applied as *e.g.* ratio control of sludge recycle flow rate to influent flow rate (Vanrolleghem, 2003).

2.2.4 Cascade control

In cascade control, two or more feedback loops are connected to each other in such a way that the master controller, *i.e.* primary loop, determines a setpoint for the slave controller, *i.e.* secondary loop. Simplified block diagram of cascade control is presented in Figure 4.

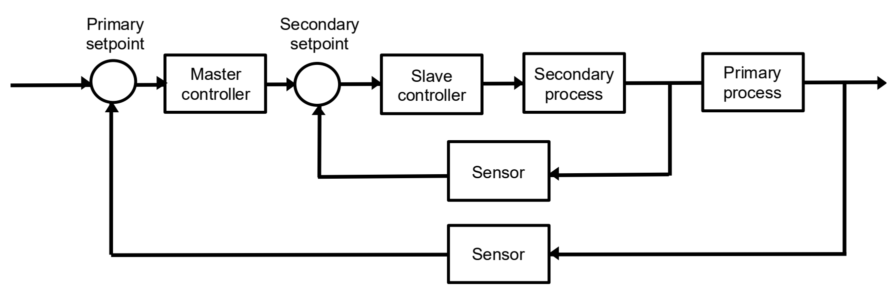

Figure 4. Simplified cascade control loop.

Cascade control enables filtering out disturbances that affect the manipulated variable. Secondary loop provides a faster response to potential disturbances. To function properly, the secondary loop should be significantly faster than the master loop (Seborg *et al.,* 2014). Cascade control is commonly applied in WWTPs for aeration, in which the setpoint for airflow in secondary loop is often determined with DO measurement in primary loop (Vanrolleghem, 2003).

2.2.5 Enhanced single loop control

In addition to cascade control, several control strategies combining one or more feedback and feedforward loops have been developed. Some of these strategies are shortly presented in Table 1 as described by Seborg *et al*. (2014).

Table 1. Summary of enhanced control strategies built on conventional PID control.

Time-delay compensation *(e.g. Smith predictor)*	Major time delays in feedback loops, *i.e.* deadtime between the control action and observed process state, may be eliminated with a Smith predictor. Essentially, a process model is added to the control system as an inner loop to yield two estimates for the controlled variable: one in the absence of disturbances, and one in the absence of both disturbances and deadtime. Comparing these to the actual process variable provides the controller the predicted variable with disturbances but deadtime left out from the loop. As a model-based control strategy, predictor may become unstable if process dynamics are changed and the controller is not properly tuned. Improvement in comparison to conventional feedback control is possible only if model parameters are within ± 30 % of actual values.
Inferential control	If controlled variables are difficult or costly to measure online, their value can be inferred and controlled with a secondary measurement, *i.e.* a soft sensor. In *e.g.* several chemical reactions, composition of a mixture may be measured only with lab samples but can be indirectly estimated from temperature and pressure measurements.
Selectors	In control problems with fewer manipulated than controlled variables, selectors may be applied to adjust a single controlled variable with multiple manipulated variables. In case of *e.g.* multiple temperature measurements, a median selector can employ their median as the controller input, instead of a single measurement. Another type of selectors is override systems, in which a second controller may overtake the first controller in certain condition. Selectors are used also for safety applications, such as level control of tanks.
Nonlinear control modifications *(e.g. gain scheduling)*	Conventional linear control systems may be enhanced for non-linear applications by *e.g.* making the controller gain a function of the control error, using logarithmic transformations as controlled variable or automatically adjusting the controller settings based on a scheduling variable. Most common method for this is gain scheduling, in which the controller gain is adjusted according to a controlled variable or some other slowly changing process variable. Depending on the application, adjustment may be continuous or divided into regions.

2.2.6 Model predictive control

Model predictive control (MPC) refers to a variety of multivariable control algorithms utilizing a process model for control. A linear or non-linear process model is used to calculate predicted future outputs. The residuals between the actual and predicted outputs are used as an input for the controller, which calculates new set-points for regulatory control loops based on the defined optimization objective, which may be *e.g.* maximizing a profit function within the defined upper and lower constraints (Seborg *et al.,* 2014).

A block diagram of model-predictive control is presented in Figure 5. Based on the process model, a sequence of control actions within the control horizon is defined to reach the setpoints in an optimal manner during a certain prediction horizon. Control is based on a receding horizon approach, meaning that from a sequence of defined control moves, only the first move is implemented and after that, the sequence is re-calculated with updated measurement data (Seborg *et al.,* 2014).

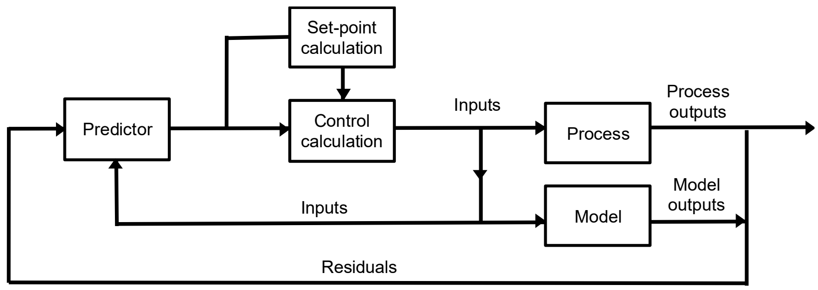

Figure 5. Model-predictive control.

MPC has been applied in thousands of multivariable control problems since 1970s, especially in the oil refining and petrochemical industries (Jämsä-Jounela, 2011). Full-scale studies have been implemented also for WWTPs, see *e.g.* Mulas *et al.* (2016).

Performance of MPC is dependent on the process model accuracy. Control performance should be regularly monitored to avoid degradation arising from changing process dynamics, instrumentation and disturbances (Santín *et al.*, 2017). MPC algorithms can be further refined with *e.g.* neuro-fuzzy techniques, which have been studied also for air supply control in WWTPs (Bernardelli *et al.,* 2020).

Neuro-fuzzy techniques and various other mathematical methods utilized in non-linear process control applications are briefly presented in Chapter 3.

2.3 Process control layers

In the following section, process control layers and their common components in municipal wastewater treatment are discussed.

2.3.1 Wastewater treatment process

Primary objective of modern municipal wastewater treatment process is to protect the public health and environment by removing suspended solids, organic matter, pathogens and nutrients from the influent wastewater (Tchobanoglous *et al.* 1991). Additionally, resources such as nutrients, energy and clean water might be recovered. Specific treatment requirements depend on the environmental permit and other objectives of the plant and may include e.*g.* removal of micropollutants (von Sperling, 2007).

Treatment process is composed of both physical, chemical and biological unit processes. Unit processes are commonly classified to preliminary, primary, secondary and tertiary treatment steps. In water reuse applications, additional advanced process steps may be necessary. Sludge generated during the process is collected and treated. Example treatment steps and unit processes are presented in Figure 6.

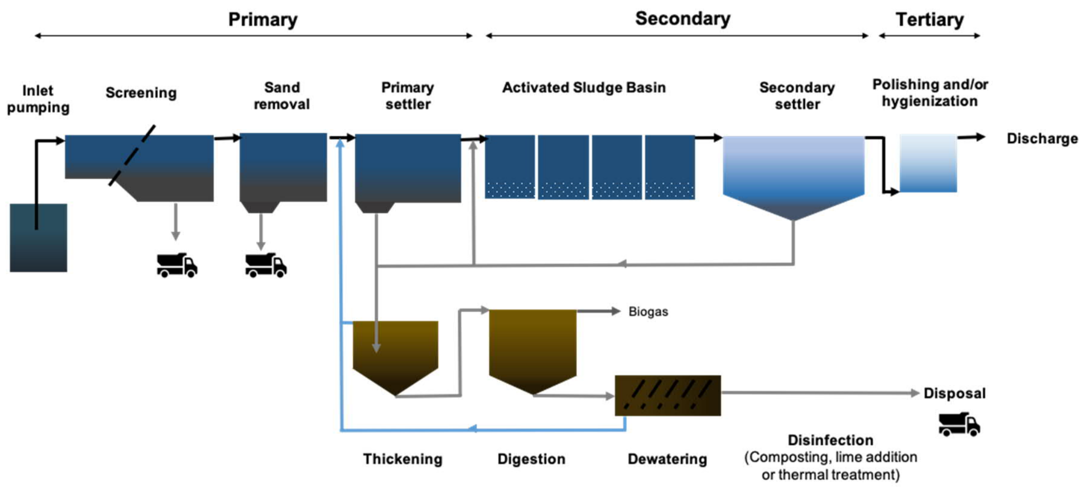

Figure 6. Simplified example of wastewater and sludge treatment process.

In terms of process control, municipal wastewater treatment process has some distinguishable characteristics. These are described by Olsson *et al.* (2005) as follows:

- Extremely large disturbances in both raw water quality and quantity. Disturbances arise both from daily, weekly and annual variation in influent rates but also from stormwater infiltration and *e.g.* illegal spills of unknown substances to sewerage.
- Complex biological and chemical reactions. Dynamic operation and control might also result into variation in microbial communities, their composition and behavior.
- Various time spans between different process responses. Process dynamics are affected by *e.g.* adjusting of valves (seconds), changes in dissolved oxygen concentration (minutes), changes of flow rate from primary pump (20–40 minutes), microbial variations (days to weeks) and seasonal changes (year).

2.3.2 Instrumentation

Instrumentation refers to both on-line sensors and in-situ analyzers, generating a close-to real-time measurement signal from process parameters. *Smart instrument* typically means a sensor with a built-in micro-processor, that enables computational capacity and a communication interface (Bailey & Wright, 2003).

In 2009, Haimi *et al.* reviewed online measurements used at Finnish WWTPs. Online sensors used in the studied WWTPs are presented in Figure 7. Over 50 % of these utilities reported measurements of dissolved oxygen (DO), temperature, suspended solids, pH, influent flow rate, air pressure, level, air flow rate, ammonium, nitrate and phosphate. For control purposes, only DO, influent flow rate, air pressure and level were used by the majority.

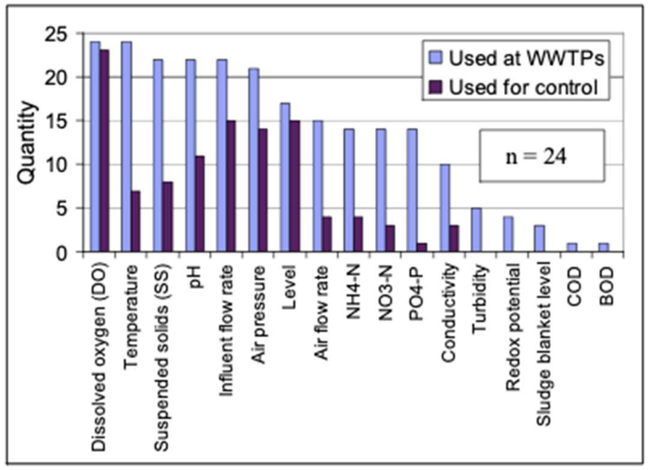

Figure 7. Online sensors used at Finnish WWTPs (Haimi *et al.,* 2009).

In 2005, Olsson stated that reliable sensors even for online in-situ nutrient monitoring are available in the market. Application of online sensors has become easier, as knowledge and standards related to operation, maintenance and measurement uncertainty have improved.

Measurement uncertainties and operating ranges vary between measurement technologies and products. In addition, sensor performance can be compromised for several reasons. Faults may arise from *e.g.* measurement noise, faulty calibration, erroneous gain or sticking to a fixed value (Rosén *et al.,* 2008).

Some conventional methods for validating raw monitoring data are described by Olsson *et al.* (2005) as gap filling, range check, outlier detection, running variance check, drift detection and cross validation. Additionally, soft sensors may be used for fault detection or back-up of physical sensors during maintenance breaks or failures (Kadlec *et al.*, 2009).

Soft sensors are usually a software or function that computes a value for the desired variable based on other variables measured from the process (Kadlec *et al.*, 2009). Soft sensors can be either model-derived, commonly based on a first principle model of the process, or data-derived, commonly based on data from the actual process conditions.

2.3.3 Actuating devices

Actuating devices manipulate the energy and mass flows in the system. Common actuating devices include valves, pumps and aeration devices, *i.e.* compressors and aerators. Actuating devices are controlled by the actuator, which is either pneumatic, hydraulic or electric (Agachi & Cristea, 2014). In this study, actuator characteristics and performance are not thoroughly discussed.

Olsson (2012) considers that controllability of actuating devices is often a limiting factor for applying more efficient control strategies. Challenges include *e.g.* poor dimensioning, non-linear behavior and lack of continuous control mechanisms.

Valves

Valves are the most common final control elements in process control. Valve limits the flow of gas or liquid with a plug, ball, diagram or baffle. Common types of control valves include global valve, angle valve, butterfly valve, ball control valve and gate valve (Agachi & Cristea, 2014).

Article of Beall (2010) describes common shortcomings of control valve systems. These include inaccurate movement, position and windup of valve stem, shaft or actuator. Beall describes a study of over 5 000 control loops, from which 30 % were negatively impacted by poor control valve performance.

Pumps

Various types of both dynamic and displacement pumps are applied throughout the treatment process for manipulating the flow of wastewater, sludge and chemicals. Centrifugal pumps are the most common type of pump applied (Creaco *et al.,* 2019).

Pump motor may be operated either with a constant speed (on/off) or variable speed, adding a variable speed drive (VSD, alternating current drive or variable frequency drive) to the system. VSD enables the pump speed to be adjusted by varying the input frequency and voltage of the motor. Compared to using a throttling valve, VSD control can improve the energy efficiency and stability of the process (Al-Khalifah & McMillan, 2012). Impact of the pump control method on power consumption is illustrated in Figure 8.

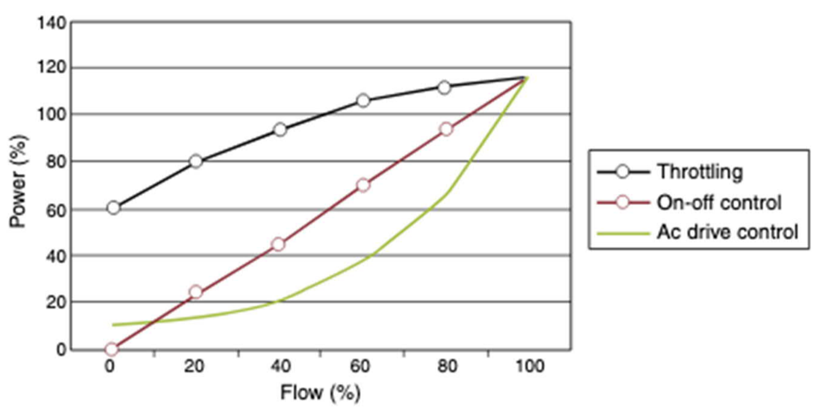

Figure 8. Typical power consumption per (nominal) flow rate with throttling, on-off or alternating current drive control (Tolvanen, 2007).

Operating range of the pump is limited by both the maximum capacity and minimum flow rate required to overcome the stiction head, *i.e.* to maintain the flow direction positive. Running the pump at too low flow rate may result into cavitation, impeller and shaft failures. In addition, pumps may face operational problems such as plugging or pump seal failure (Creaco *et al.,* 2019).

To increase the operating range and reliability, pumps are often installed in parallel or in series. Control performance of the pump should then be reviewed from the perspective of the pump system as a whole (Bloch, 2011).

Aeration devices

Typical aeration system in an activated sludge process has compressors supplying air into a pressurized piping system, from which airflow to the diffusers, submerged to the bottom of the aeration tank, is controlled with valves. Compressors with low pressure rise are often also referred to as blowers. In Finland, aeration compressors are estimated to correspond to 45 – 75 % of the total energy consumption of WWTPs (Motiva, 2019).

Åmand *et al.* (2013) reviewed typical implementations of aeration control. Centrifugal blowers, such as turbo blowers, are described common in large installations, and positive displacement blowers more common in small installations. Similar to pumps, blowers are often operated in parallel and equipped with VSDs to enable flexible air supply for varying process conditions.

Aerators can be categorized as coarse-bubble diffusers, fine-bubble diffusers and surface aerators. Fine-bubble diffusers are the most common solution in large municipal WWTPs (Rosso *et al.,* 2008). Fine-bubble diffusers are described to have the highest aeration efficiency, which however declines by time due to fouling and weakened oxygen transfer efficiency. Schematic of a fine-bubble diffuser is presented in Figure 9. Fine-bubble diffusers are described to require continuous cleaning and easily suffer from clogging or changes in orifice characteristics. In addition, diffuser layout, density and age have been studied to affect their performance (Groves *et al.,* 1992).

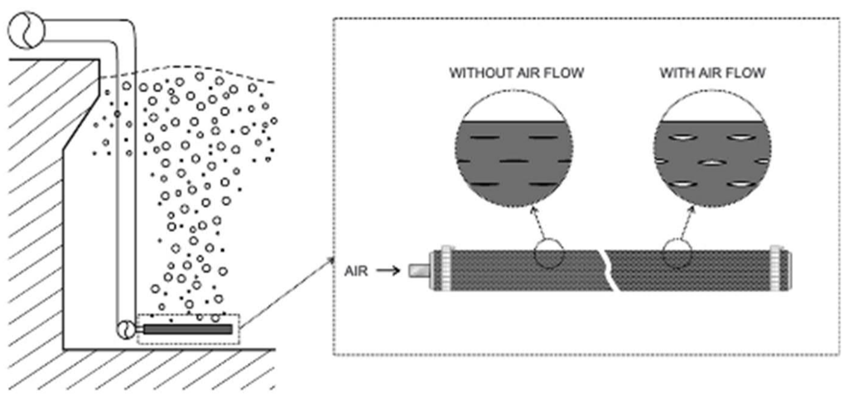

Figure 9. Schematic of fine-bubble diffusers (Kaliman *et al.,* 2008).

Equipment maintenance and condition monitoring

In the long run, equipment performance is highly dependent on the maintenance (Hegg *et al.,*1979). Maintenance often forms the largest single expense of the equipment (Stamboliska *et al.*, 2014). Maintenance strategy may be based on faults (reactive), schedule (preventive) or condition (predictive/proactive).

Hashemian (2010) discusses that based on the data from US Department of Defense, only 11 % of industrial equipment is most effectively maintained on a time-basis. For 89 %, maintenance should be based on observed condition.

Three main categories of condition-based maintenance are described as per the source of the condition data: existing process data (*e.g.* pressure sensors), secondary sensors (*e.g.* accelerometers measuring motor vibration) or separate test signals, that are injected to the equipment to test their response (Hashemian, 2010).

In low-speed machinery operating below 600 rpm, such as compressors and wind turbines, occurring faults are often more gradual and more difficult to detect than with high speed. For better fault detection and prevention, condition data can be further analyzed with visual and statistical methods (Stamboliska *et al.*, 2014).

2.3.4 Regulatory control

Controllers implement the regulatory control and communicate between the instrumentation and central control room. Traditionally, configuration and control programs are dynamically downloaded to the controller from the central computer station (Bailey & Wright, 2003). In case of a Distributed Control Systems (DCS), controllers may communicate to each other on a peer-to-peer basis through a fieldbus network.

Controller hardware

Selection of controller hardware depends on the overall process control system and control strategy. For on-off control, a simple on-off relay can be used as the controller. Remote Terminal Units (RTU) are common in geographically large control systems, such as the sewage network. For more versatile local control, such as the plant environment, programmable logic controllers (PLC) are common (Bailey & Wright, 2003).

PLC is based on a ladder logic that originates from the physical wiring of electrical circuits. Standard programming languages include ladder diagram, function block diagram (FBD), structured text and sequential function chart (Bailey & Wright, 2003). Example illustrations are given in Figure 10.

Nowadays PLCs come with a variety of programming languages and are utilized also for more sophisticated control logics. PLCs have become popular due to their inexpensiveness, flexibility, easy design and installation, physical compactness and easy diagnostics (Bailey & Wright, 2003).

Figure 10. Examples of simple PLC program for one pump control action implemented with ladder diagram (left) and FBD (Bolton, 2015).

In DCS, control functions are autonomously implemented by microprocessors situated near the controlled instrument or process (Bailey & Wright, 2003). Controllers are connected to each other through a fieldbus network that enables high speed real-time communication. While PLC-based systems are event driven, DCS is process state driven. In practice, elements from both approaches are often mixed and today both technologies share most of the functionalities.

Failure of a single PLC controller can induce problems in the governed control loop (Metcalf & Eddy, 1979). This risk can be mitigated with a net of PLCs, similar to DCS approach, enabling the use of duplicate controllers and operator terminals. Additionally, the use of personal or industrial computers and smart instruments as controllers can improve the control system reliability in case of failures.

Controller tuning

Controller tuning is the practice of adjusting the control parameters to ensure best possible response of the controller (Buckbee, 2009). Several tuning methods exist, such as manual tuning (trial and error), Ziegler–Nichols, Cohen-Coon and Åström-Hagglund. Buckbee argues that most effective methods always involve some form of a process model, *e.g.* integral, first-order or second-order model. Today, also autotuning methods exist.

Tuning may include process experiments to verify the process response to different types of test signals, *e.g.* step change, sine wave, pulse, pulse queue or random pulse queue. From these, step change is the easiest and most commonly applied (Harju & Marttinen, 2000).

McMillan (2014) states that PID controllers are used in 99.9 % of control loops in the industry, including as a part of MPC applications. PID control behavior is largely influenced by the tuning of the control parameters, *i.e.* gain factor K_c, integral time T_i and derivative time T_d, and also the frequency of the control signal.

According to Buckbee (2009), tuning should be considered whenever changes occur in the process, instrumentation, control devices, raw material or operating procedures. Changes in process dynamics may be identified by monitoring the control loop performance from *e.g.* percentage of time in normal mode, control loop oscillation and integral absolute error, among others. Loop performance monitoring may be applied as a periodical audit or a continuous service (Ketonen & Marttinen, 2001).

2.3.5 Supervisory control

Supervisory control and data acquisition (SCADA) refers to the overall system combining telemetry and data acquisition for collecting information, transferring it to central site, carrying out pre-defined control tasks and displaying the information on operator screens (Bailey & Wright, 2003). Operator display is also called as human-machine interface (HMI). In addition, SCADA software tasks include alarms, process monitoring and reporting. Example configuration is illustrated in Figure 11.

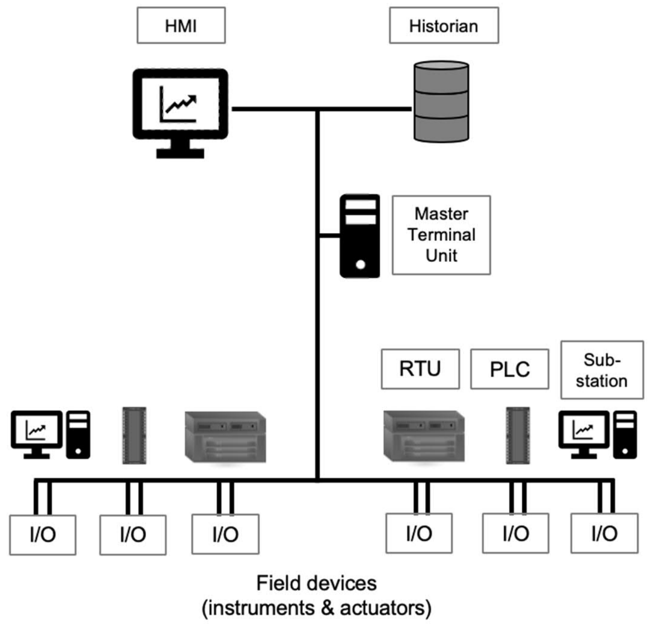

Figure 11. Example SCADA configuration.

Monitoring data in SCADA may be utilized also for online prediction, fault detection and diagnosis, reporting and documentation. Olsson *et al*. (2005) predict that automated control tuning and performance reporting become a standard in WWTPs. These tools can be combined to SCADA systems either as a separate module or an integrated functionality.

Rieger *et al.* (2012) discuss that the impact of human operators and organizations for control performance is often underestimated. In addition to operators managing the supervisory control, other stakeholders such as utility management, owners and authorities, influence the objectives and incentives for treatment performance, that consequently determine also the goals and strategies for control and operation.

2.3.6 Plant-wide control

Traditionally, each unit process in a WWTP is considered as an independent part of the control system. Different time scales are often decoupled by assuming interactions between time scales as constant (Olsson *et al.,* 2005). It is however clear that there are several couplings and interactions between different unit processes. In plantwide control, these couplings are accounted for. Drivers for plant-wide control can be summarized to complex interactions in plant hydraulics, recycle flows, sequential units and optimal resource allocation.

In plant-wide control, unit processes may be locally controlled, but *e.g.* set-points are adjusted by the integrated plant control system, taking into account the interaction of different unit processes in different time scales. System boundaries can be further enlarged by including the wastewater producer, receiving water body or clean effluent consumers to the control system, which can be referred to as system-wide control.

Olsson (2005) describes that in municipal WWTPs, it is common to integrate the control of sewage network with plant control. Some benefits from integrated control are described as prediction of plant influent flow rate, extension of plant capacity to minimize bypasses and extension of sewer system capacity to minimize combined sewer overflows. Common sub-objectives and variables for integrated network and plant control are described in Figure 12.

	Partial aim	**Measurements**	**Control handles**
Sewer system	Minimise upstream overflow	Rain	Pumping stations
	Utilise basins for most polluted water	Levels	Adjustable weirs
		Flow rates	Basins
Wastewater treatment plant	To treat as much wastewater as possible during and after rainfall	Flow rates (inlet, outlet, return sludge, recycles)	Return sludge pumping (control of sludge blanket in sec. sedimentation tanks)
	Reduce hydraulic load and sludge load in secondary sedimentation tanks	Suspended solids (aeration tanks and return sludge)	ATS control (sedimentation in aeration tanks)
		Sludge blanket	Primary pumping (bypass before biological section or the total plant)

Figure 12. Sub-objectives, measurements and control handles for system-wide control (Olsson & Jeppsson, 2006).

3 Data analysis and extraction of information

Process control systems in municipal WWTPs are described to be “data rich but information poor” (Olsson *et al.,* 2005). Existing tools are often not fully utilized for extracting useful information, such as patterns, from the collected data. Olsson *et al.* (2014) summarize the incentives for data analysis and information extraction as data validation, detection of process state anomalies and fault detection and diagnosis.

In the following sections, data analysis methods are presented following the introduction of Runkler . Main categories and example algorithms are presented. In addition, some existing application areas and examples are reviewed as found in the literature.

3.1 Terminology

Data analysis, or *data mining*, refers to any type of extraction of any type of knowledge from the data. Common data analysis phases are summarized in Figure 13. *Data analytics* implies to the use of data analysis to consequently support decision making (Runkler, 2012).

Figure 13. Phases of data analysis.

A range of different expressions are associated with data analysis methods and applications. Exact definitions seem to quickly evolve and vary depending on the context and industry. For purposes of clarification, some definitions are introduced.

Big data refers to “extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations” (Oxford University Press, 2019). These data sets may be analyzed with a range of data analysis methods. According to Ingildsen & Olsson (2016), *smart water* refers to the systematic and integrated approach of measuring, analyzing and decision making in the water utility. This might involve using a public or private internet protocol (IP) network of physical devices with built-in computational capacity and communication interface, forming an *Internet of Things* (IoT). In an industry setting, this may also be referred to as *industrial internet* (Patel & Patel, 2016).

In addition, smart water may include the use of *intelligent control*, referring to control applications utilizing *artificial intelligence* (AI). AI can be defined as a range of computer programs or machines that utilize data analysis for simulating or mimicking the human intelligence – a definition that varies over time, as the definition of human intelligence evolves (Antsaklis, 1997). A *narrow AI* refers to an algorithm designed for only a limited or specific task, instead of all human abilities (DeepAI, 2020). One family of AI methods is *machine learning*, which can be seen to mimic the human ability of learning from experience (Antsaklis, 1997). In practice, machine learning refers to automatic adjustment and calibration of parameters with training data.

3.2 Data analysis methods

In the following sections, data analysis methods are presented following the introduction of Runkler (2012). Main categories and some example algorithms are presented. In addition, some existing applications in wastewater treatment are reviewed as found in the literature.

3.2.1 Data preprocessing

In most cases, raw measurement data suffers from some errors and measurement noise. Raw data might also need to be scaled, transformed or merged, *e.g.* if it is collected from several potentially heterogeneous sources (Runkler, 2012).

Errors in the data may be deterministic or stochastic. Deterministic errors arise from *e.g.* wrong scaling, wrong calibration or sensor drift, and can be corrected if the error systematic is known. Stochastic errors originate from *e.g.* measurement errors or packet losses in data transmission, resulting into random outliers or missing data.

Identified outliers, invalid and missing data can be individually handled by listing the invalid values, replacing the values with *e.g.* NaN, or correcting, estimating or removing them (Runkler, 2012). Another common way to handle the errors is filtering the whole data series.

Filtering

The goal of filtering is to remove outliers and measurement noise. Common filtering methods apply a statistical measure, such as mean or median, over a moving window. With same window size, moving mean and moving median result into similar noise reduction, but moving median suppresses the outliers much more efficiently. This can be observed from example filtering outputs in Figure 14.

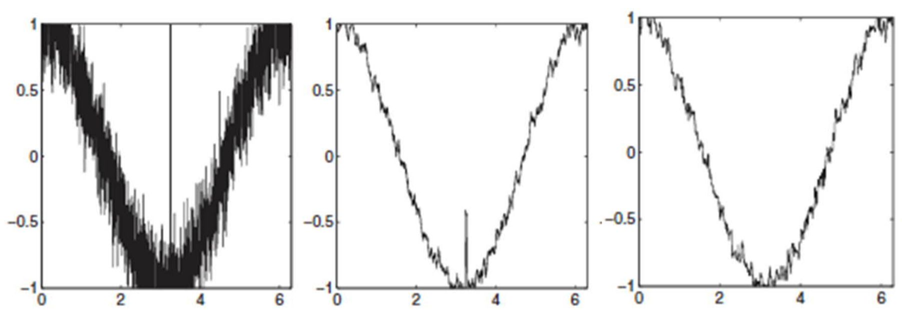

Figure 14. Original (left), moving mean (middle) and moving median filtered data (right) (Runkler, 2012).

Another common filtering method is an exponential filter, which is a type of discrete linear filter. In exponential filter, past observations are weighted exponentially to decrease their weight over time, in opposite to equal weights in moving average. Exponential filter is defined as

$$y_k = y_{k-1} + \eta(x_{k-1} - y_{k-1}), \qquad k=1, \ldots, n-1 \tag{4}$$

where y_k is the filter output, x_k is the filter error and η a coefficient for weighting the previous filter input. Value of η defines the level of noise reduction and has to be carefully selected to achieve sufficient filtering result without losing the essential characteristics in the data. This is demonstrated in Figure 15.

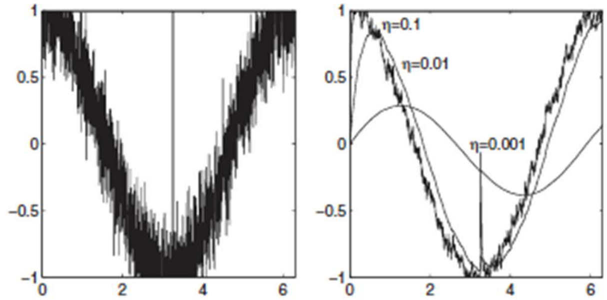

Figure 15. Original (left) and exponentially filtered data (right) with different values of η (Runkler, 2012).

Exponential filters have been used at Rya WWTP in Sweden to estimate flow rates based on changes in tank level (Lumley, 2002). Exponential filter eliminates the impact of measurement noise in the level signal, which can then be used to estimate the flow rate from change of tank level by time, when the tank area is known. In addition, exponentially filtered signals are applied at Rya WWTP for estimating mixed liquor sludge age, sludge mass in the sludge blanket and anaerobic digester detention times.

3.2.2 Data visualization

Visualization of data is an effective method to analyze, document and communicate data characteristics. One- or two-dimensional data sets are often presented as simple diagrams, scatter diagrams, histograms or spectrums. For visualization of high-dimensional data, a projection must be made. Several linear and non-linear projection methods exist, one of them being Principal Component Analysis (PCA) described by Runkler (2012).

Principal Component Analysis

The objective of PCA is to make a linear projection of multi-dimensional intercorrelated data, capturing the data structure and variance as much as possible while dimensions are reduced. Remaining dimensions are also called principal components, that represent the maximum variance in the data but are not correlated with each other.

PCA is performed in several steps by calculating a covariance matrix for the data, for which an eigenvalue decomposition is performed. As a result, PCA yields coordinate axes, *i.e.* projection, and transformation matrix, that maximize the variance. In addition, PCA yields the variance values.

Results of PCA may be interpreted and visualized on several ways depending on the analysis purpose. In the examples of Figure 16 (Ringnér, 2008), graph a) plots the data based on two categories (black and red) and values of two variables in the x- and y-axis. In graph b), two uncorrelated principal components, PC1 and PC2, capturing the maximum variance of the data are identified. In graph c), data is projected to PC1 all together and separately. Graph d) describes how much of the variation is covered by each principal component. Graphs e) and f) present a projection to both PC1 and PC2 with the color indicating a classification to different data categories.

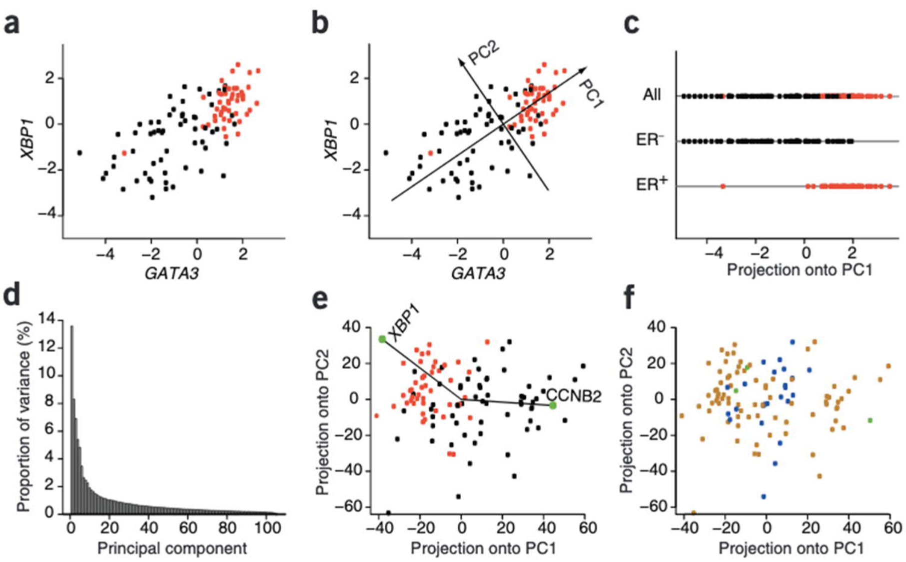

Figure 16. Example visualizations of PCA results (Ringnér, 2008).

In essence, PCA enables highly robust linear mapping with low computational effort. It is a suitable tool for various purposes of analyzing data structure, variation and patterns. Several extensions of PCA have been developed, including also non-linear and adaptive alternatives. For complex non-linear data, also additional methods such as multidimensional scaling, Sammon mapping and auto-associators exist (Runkler, 2012).

In wastewater treatment, PCA can be applied for *e.g.* purposes of fault detection and diagnosis. It can be used to identify principal causes for process variations and faults (Garcia-Alvarez, 2009) or develop soft sensors for instrumentation quality assurance (Haimi *et al.,* 2013).

3.2.3 Correlation

Correlation is a quantification of the relationship between two features. A correlation between features x and y does not imply causality but may indicate one or more of the following scenarios: i) x causes y, ii) y causes x, iii) z causes x and y, iv) correlation is a coincidence (Runkler, 2012).

Linear correlation may be analyzed with a covariance matrix C of a data set $X \subset \mathbb{R}^p$, where each matrix element c_{ij} indicates the covariance between features x^i and x^j, where $i,j = 1,\ldots,p$,

$$c_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k^{(i)} - \overline{x}^{(i)}\right)\left(x_k^{(j)} - \overline{x}^{(j)}\right) \tag{5}$$

If the features are multiplied by a constant factor, the covariance will increase. Dividing the covariance with the product of standard deviations of the features a further value called *Pearson correlation*. In comparison to covariance, it provides a more descriptive measure for correlation, since acknowledging the standard deviations eliminates the impact of constant scaling.

In addition to linear correlation, non-linear correlation can be identified. One method for this is chi-square test, presented below as described by McHugh (2013). Several other methods have been developed, such as Spearman's and Kendall's rank correlations.

Chi-square test for independence

Chi-square test enables quantifying of non-linear correlation between features, *i.e.* to determine if the features are independent or not (McHugh, 2013). Independence of the features is set as the null hypothesis and chi-square test may reject or fail to reject it. Chi-square is produced as

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{6}$$

where O_i are the observed and E_i are the estimated frequencies of features as per the null hypothesis. Significance of their deviance, *i.e.* if it may arise from random sampling error or indicate a correlation, can be estimated from a chi-square table that indicates the value of X^2 for a certain probability. Assigned probability depends also on the degree of freedom defined by the number of features and parameters, which can be expressed as (no of parameters - 1)*(no of features - 1). If the value of X^2 exceeds the indicated value, null hypothesis can be rejected and a correlation between the features concluded.

Chi-square test is applicable to only classified data and a proper result can be expected only from large sample sets. It is a common method for statistical analysis and has been applied for *e.g.* research in variation of virus types in hospital wastewater (Prado *et al.,* 2011) and detection of intrusion attempts to information and control systems, *i.e.* cyber security (Ye & Chen, 2001).

3.2.4 Regression

Regression provides an estimate for functional dependency of features. Whereas correlation quantifies the relationship of two features, regression models a value for the dependent feature, based on the value of the feature used as input (Runkler, 2012). Simplest form of linear regression can be presented as

$$x_k^{(i)} \approx a * x_k^{(j)} + b \tag{7}$$

where x_k are the features, for which the relation is estimated with parameters *a* and *b* by minimizing the selected error function, *e.g.* average quadratic regression error.

Forecasting is a specific regression task, in which the future values of a times series are assumed to be generated by a deterministic process. The process is described as a state machine with input, state and output. If the number of possible states is limited, the system is called a finite state machine. In a recurrent model, possible inputs and their expected outputs are known. If future inputs are not known, an auto-regressive model can be developed.

There are various linear and non-linear regression methods. For example, partial least squares regression has been applied also in WWTP research (Haimi *et al.,* 2013). Another popular class of non-linear regression is universal approximators, from which one example is multi-layer perceptrons (MLP), also called as feedforward neural networks. Some general characteristics of artificial neural networks (ANN) are presented below as described by Santín *et al.* (2017).

Artificial neural networks

ANN consists of input layer, one or more hidden layers and an output layer of process parameters. Each node, *i.e.* neuron, in the hidden layer is a simple processor that computes a single output from a multitude of inputs. A range of different network topologies have been developed for different purposes. A deep neural network refers to ANN with multiple layers. Structure of a simple feedforward neural network, with only one hidden layer and all nodes connected, is presented in Figure 17.

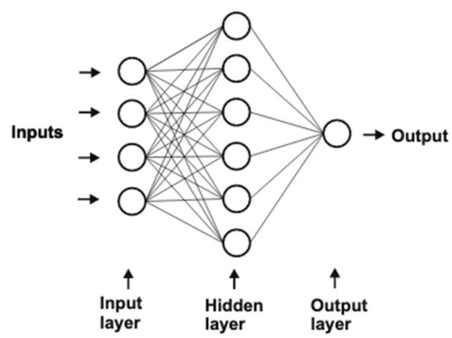

Figure 17. Structure of a simple ANN algorithm (Santín *et al.*, 2017).

The output of each node is calculated with a non-linear activation function, for which sigmoid functions are commonly used. Sigmoid functions are s-shaped functions that can be used to map a large (finite or infinite) range of inputs, into a limited range of outputs. In the training phase, the weighting of input vectors to reach the certain output vector is estimated, after which the output for any combination of input parameters can be predicted.

In wastewater treatment, ANN algorithms have been studied and applied for prediction of *e.g.* plant performance, effluent quality and biogas production (Mjalli *et al.*, 2007; Nasr *et al.,* 2012; Yetilmezsoy *et al.,* 2013).

3.2.5 Classification

Classification is a type of supervised learning method. In the training phase, data is labeled to assign objects into classes. After that, several classifier types can be used to assign additional objects to classes. Runkler (2012) distinguishes four types of classifiers:

- Probabilistic, such as naive Bayes classifier, in which the classification is defined by the direct probability of an object with specific features to belong to a certain class. Deterministic classifier is straight-forward but assumes statistical independence between the features, which is often not true for real-life data.
- Linear, in which one or more linear lines are computed to the projection as class borders. Linear classifier is related to PCA and is efficient with correlated features. Though, it assumes a Gaussian distribution of the features and is therefore not applicable if the class borders are non-linear.
- Prototype based, such as nearest neighbor classifier, in which the class of an object is assigned to be the same as in a labeled vector with most similar features, *i.e.* nearest neighbor. For higher accuracy, instead of the class of one near neighbor, most frequent class of several nearest neighbors can be selected.
- Hierarchical, such as decision trees, in which the features are ranked and evaluated in an order of importance to assign the class. Decision trees can be useful for data with several features of varying significance. As a disadvantage, complex non-linear class borders are not efficiently accounted for.

Further example of support vector machine, a type of linear classifier, is described next (Runkler, 2012).

Support vector machine

Support vector machine (SVM) is based on linear class borders and a margin around the discriminant line, plane or hyperplane, that yields the most accurate classification. It is possible that two or more classes partly overlap each other. For two overlapping classes, it is required that

$$w * x_k^T + b \geq +1 - \xi_k \qquad \text{if } y_k = 1 \tag{8}$$

$$w * x_k^T + b \leq -1 - \xi_k \qquad \text{if } y_k = 2 \tag{9}$$

where x_k is the feature, y_k is the class, w is a coefficient parameter, b the margin parameter and ξ the slack variable that accounts for the overlapping. If multiple solutions exist, optimal solution is defined by minimal value of the cost function J

$$J = \frac{1}{2} ||w||^2 + \gamma * \sum_{k=1}^{n} \xi_k \qquad \gamma > 0 \tag{10}$$

Parameters w and b are found by minimizing the cost function using quadratic programming. With no overlapping or close adjacency of classes, the slack variable can be eliminated.

In the wastewater sector, SVM has been applied for *e.g.* prediction of sewer condition grade (Mashford *et al.,* 2011). In this case, four different models were developed with different sets of input variables, such as material, age, diameter, soil characteristics and change of angle. Best-performing model indicated a 91 % prediction accuracy in classification.

3.2.6 Clustering

Distinct to classification, clustering is an unsupervised learning method for identifying structures and categorizing unlabeled objects. This is particularly useful in cases where there are no pre-defined labels, or they are difficult to obtain in advance. Clustering may be done with either sequential or partitional methods. In sequential clustering, all objects are first interpreted as individual clusters. These clusters are gradually merged into larger clusters with most similar or least dissimilar cluster, which can be defined from *e.g.* the minimum distance (Runkler, 2012).

In proportional clustering, a model, prototype or characteristics are utilized. Range of methods include *e.g.* linear, heuristic, fuzzy and noise clustering. Clustering with fuzzy logic, applied in *e.g.* fuzzy control, is presented as an example as described by Santín *et al.* (2017).

Fuzzy logic

Fuzzy logic is based on intuitive understanding and experience of the process, instead of a mathematical model. In fuzzy logic, input parameters have continuous values instead of discrete values of 0 or 1. Simplified block diagram for a fuzzy logic controller is presented in Figure 18.

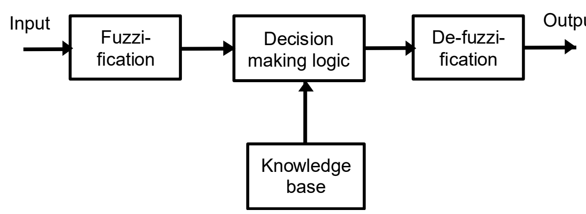

Figure 18. Fuzzy logic control.

In the fuzzification stage, *i.e.* fuzzy clustering, discrete input values are decomposed into fuzzy sets. Fuzzy sets are defined with membership functions that describe the grade at which a certain input belongs to the category, as illustrated in the Figure 19.

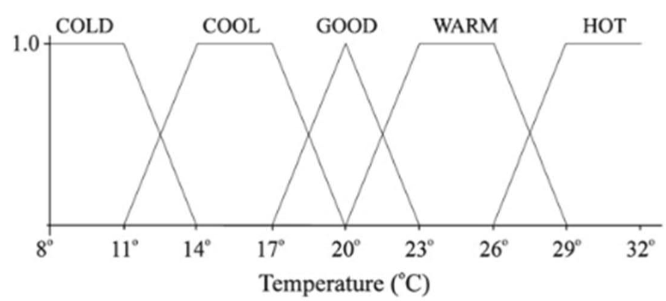

Figure 19. Example of membership functions for room temperature (Santín *et al.*, 2017).

After fuzzification, fuzzy sets are interpreted with Boolean logic operations, *e.g.* if-then rules. Output of the Boolean operations is similarly a fuzzy statement, which is de-fuzzified back to yield a discrete value for implementing the control action.

Application of fuzzy logic for WWTP control has been widely studied, *e.g.* (Boger, 1992; Carrasco *et al.,* 2004; Traore *et al.,* 2005). Performance depends on the extent of applied human knowledge for implementing the control, for which downsides may include challenging knowledge base development, tuning, testing and verification of optimal performance (Haimi *et al.,* 2009).

4 Study methods

4.1 Interviews

To study practitioner perspectives, interviews were selected as the primary data collection method. Objective of the interviews was to review what barriers, needs and incentives are seen for developing the process control and operation in WWTPs by utilizing data analytics. A range of stakeholders involved in the process of data acquisition, processing and analysis at WWTPs were invited.

In total, 19 interviews were held with approximately 2–3 hours length each. Interviews included personnel of WWTPs and consulting, automation and technology companies. 10 of the interviews were held with personnel of Finnish WWTPs: most of them chief operators of the plant, but also automation engineers, process engineers and managing directors. It should be noted that the smallest scale of WWTPs is not represented in the interviews. It was assumed that current sample of WWTPs already provides sufficient variance in plant size and represents a target group with most potential and resources for development.

In addition, nine representatives from Finnish consulting, automation and technology companies were interviewed. Interviewees from the companies included senior specialists or leaders of water business unit. Invited organizations included the companies with most references and experience of municipal wastewater sector. Interviewed organizations are presented in Table 2. Average flows of WWTPs are only indicative values. Based on the range of plant and organization types interviewed, these perceptions were presumed to provide a good overview of the municipal wastewater sector in Finland.

Table 2. Organizations interviewed for the thematic analysis. CAS = Conventional Activated Sludge, MBBR = Moving Bed Biofilm Reactor, MBR = Membrane BioReactor.

	Organization type	**Average flow (m^3/d)**	**Process type**
1	WWTP	84 000	CAS
2	WWTP	250 000	CAS
3	WWTP	20 000	CAS
4	WWTP	18 000	CAS
5	WWTP	42 000	CAS
6	WWTP	22 000	CAS
7	WWTP	2 000	MBBR
8	WWTP	5 000	CAS
9	WWTP	6 000	CAS
10	WWTP	50 000	CAS + MBR
11	Design & consulting		
12	Design & consulting		
13	Design & consulting		
14	Process automation		
15	Process automation		
16	Process automation		
17	Technology		
18	Technology		
19	Technology		

Interviews were implemented as semi-structured interviews. Semi-structured interview has pre-determined questions, but interviewer can explore additional themes in a conversational manner (Chism *et al.*, 2008). To avoid possible conflicts of interest between different interviewees, interviewed organizations were not published.

Interview questions for WWTPs were focused on current and future process control and operation strategies, tools and challenges. Questions for companies discussed views of development needs, opportunities and differences as compared to other sectors they work with. Interviews were recorded and summarized transcripts were prepared afterwards. Question lists for WWTPs and companies can be found in Appendices 2 and 3. Some interviewees sent additional material, which was included in the analysis.

4.2 Thematic analysis

After the interviews, collected material was processed with thematic analysis. Thematic analysis is a transparent method of identifying, analyzing and reporting patterns in qualitative data (Braun & Clarke, 2006). Quotations from the research data, *e.g.* interview transcripts, are systematically coded with the meanings, ideas or views that they reflect. From these codes, patterns are identified to construct generalized findings, *i.e.* themes.

The phases of the analysis process are presented in Figure 20. The analysis process was carried out with ATLAS.ti software (Scientific Software Development GmbH, v 8.4.4.). Following the instructions given by Friese *et al.* (2018), analysis phases are described next.

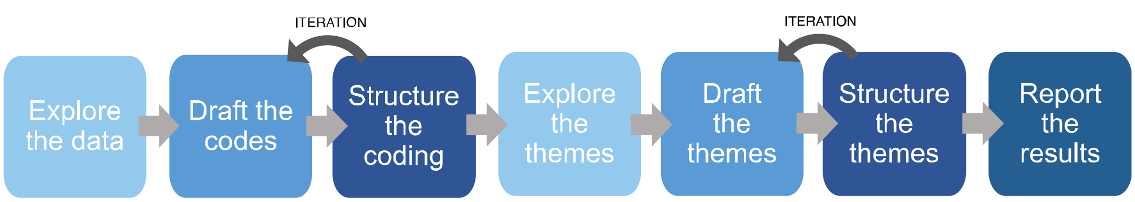

Figure 20. Thematic analysis process (Eerikäinen *et al.,* 2020).

Phase 1: Explore the data

First, collected data is reviewed and initial remarks written down. Material is continuously reviewed already during data collection and transcription of recordings. Inconsistencies, faults and missing parts are revised and corrected.

Phase 2: Draft the codes

After familiarizing with the data, coding process is started. A code is allocated for a quotation from the material to summarize the meaning of individual comments. Same code can be assigned to several quotations reflecting similar views. New codes are continuously added, code list regularly reviewed, and codes adjusted to structure them in a logical manner.

Phase 3: Structure the coding

Based on the initial long list of codes, a coding framework is prepared to develop further structure in the coding. This framework is updated and complemented throughout the coding phase, being an iterative process between phases 2 and 3.

The framework is adjusted to conclude grounded codes with sufficient detail, meaning that original codes are either split into smaller codes describing the data in more detail, or merged into larger codes, giving more generalized and grounded findings.

Phase 4: Explore the themes

Once the coding framework is ready, analysis is focused to a broader level. Linkages and relations between different codes are assessed and codes are further grouped into potential themes.

Phase 5: Draft the themes

After the initial themes are built, their validity is reviewed through both included codes and quotations linked to them. This is done to ensure that the conclusions arising from the themes are well grounded on the quotations from the material. Some recoding of original data can be done to complement previously missed information, that is now identified as relevant.

Phase 6: Structure the themes

After reviewing the themes, linked codes and related quotations, the theme is written out into a coherent analysis of the findings and their relations. Each theme is given a name and a short description of its essential aspects, including their sub-themes.

Phase 7: Report the results

Once all identified themes are reviewed and defined, a coherent report of all themes is prepared. Quantitative data can be beneficial to include to the analysis to reflect and verify the findings.

In addition to the thematic analysis, some quantitative data of current control strategies in Finnish WWTPs was compiled from the interviews. This data was seen as beneficial background information for discussing further the needs and barriers in process control.

4.3 Case studies

Case studies can be used to investigate complex real-life issues, such as human interaction with technology (Runeson & Höst, 2009). In software engineering, case studies are applied to reveal how software development, operation and maintenance are carried out in practice by the individuals, groups and organizations, that are influenced by social and political environment.

Case studies for this work were developed by means of case study design, data collection, analyzing the data by classification and reporting the results (Runeson & Höst, 2009). These steps are further described below.

4.3.1 Data collection

To assess and demonstrate potential benefits and limitations of developing ICA systems and data analysis in practice, case studies of various real-life applications were identified. Examples were sought from both Finnish WWTPs, international WWTPs and process industries. This was done with the purpose to identify solutions of different maturity level and potential for both research, piloting and direct implementation at Finnish WWTPs.

Potential solutions were searched from literature, online materials and interviews. In addition to interviews assessed in the thematic analysis, several discussions about the case studies were held with both solution developers and users, whereas available.

4.3.2 Descriptions

Targeted information of case studies consisted of process description, control and/or monitoring objectives, achieved benefits and pre-requisites for implementation.

Initial case study descriptions were presented to the project steering committee for initial evaluation of their feasibility in Finnish WWTPs. Based on the received comments, collected information was further complemented.

Since the thesis is public, some of the case study information was left out from the published version and shared only for the steering committee WWTPs. This included further information of costs, pricing models and reference clients of the described examples.

4.3.3 Classification

Case studies were presented based on the application area, *i.e.* Finnish WWTPs, international WWTPs and process industries. A classification was made according to the primary process control layer involved as was illustrated in Figure 1. In addition, case studies were classified according to the primary type of data analysis, as described by research and advisory company Gartner (Banerjee *et al.*, 2013). This classification is presented in Table 3.

Table 3. Classification of data analysis types in case studies.

Diagnostic	**Proactive**
Active and informative *e.g. root cause analysis*	Active and operative *e.g. automatic control*
Descriptive	**Predictive**
Passive and informative *e.g. data visualization*	Passive and operative *e.g. forecasts*

5 Results

Results from the thematic analysis (Chapter 5.2) are published in Water Science & Technology journal (Eerikäinen *et al.,* 2020). The chapter discusses the results as presented in the paper with some additional information related to Finnish WWTPs.

5.1 Control strategies

In total, 10 WWTPs were interviewed and their process control strategies briefly reviewed. Since interviewed WWTPs represent various different process types and configurations, an exhaustive description of control strategies for each unit process was not considered favorable. For several primary or tertiary treatment steps, *e.g.* sand removal or disc filtration, control strategy is largely dependent on the selected technology or variation between plants is small. Applied control algorithms were either feedback, feedforward or cascade control. Therefore, review of control strategies was focused to assess which online parameters are used in control of wastewater and sludge flows, aeration and chemicals.

Control methods of major unit processes in interviewed utilities are described in Table 4. Most commonly mentioned methods for each unit process is highlighted in green. In addition to online measurements, strategies based on manual, constant or scheduled control were identified. For influent pumping, constant level means operating the pump based on the water level in the influent tank. Aeration airflow refers to the airflow demand, which defines the valve operation. Aeration air pressure refers to the pressure level in the piping system that defines the compressor operation.

Table 4. Control methods for main unit processes in interviewed WWTPs.

Control method	Influent pumping	Primary sludge pumping	Aeration airflow	Aeration air pressure	RAS	Secondary sludge pumping	Sludge feed to centrifuge
% of flow	0	2	0	0	8	3	0
TS	0	3	0	0	0	2	5
Constant level	9	1	0	0	0	0	0
Constant pressure	0	0	0	9	0	0	0
DO	0	0	9	0	0	0	0
Constant flow	1	0	0	0	0	1	3
Manual	0	1	0	0	0	0	2
Scheduled	0	0	0	0	0	2	0
NH4	0	0	1	0	0	0	0

From the results, it can be observed that for almost all plants, influent flow pumping aims to maintain a constant level in the influent tank. Both flow, level and TS measurements are utilized for sludge removal from the clarifiers. For aeration, DO cascade control with constant pressure level in the piping system and constant sludge age is dominant. Sludge dewatering is controlled with TS in larger plants, whereas most of the manual or flow-based control took place in smaller plants.

Control variables for chemical dosing are presented in Table 5. For simplicity, various types of precipitation chemicals and polymers have been combined to their respective groups. Results indicate that majority of all chemicals is currently dosed based on flow measurements, *i.e.* solely wastewater quantity. Some of these values are simplified from the reality: for example, methanol dosing in one WWTP was based on a combination of flow rate, nitrate and DO concentrations. Manual adjustments by the operator were mentioned also in many cases where automatic control was constant.

Table 5. Control methods for chemical dosing in interviewed WWTPs.

Control method	Precipitation chemicals	Polymers	Lime	Soda	Methanol
Flow	10	8	3	2	0
Constant or step-wise constant	3	1	1	1	0
Manual	3	0	2	0	0
Nitrate / ammonium / phosphate	1	0	0	0	2
TS / TS-tons	1	2	0	0	0
pH	1	0	0	1	0

5.2 Thematic analysis

Based on the interviews, 198 codes were created with an average groundedness of 2.6, indicating from how many quotations the code consists of. All created codes are presented in Appendix 1. Codes identified as significant for the research question were organized into themes consisting of 65 codes, with an average of 5.3 quotations linked to each code. This indicates that the selected codes are well grounded on the interview findings. Main themes and sub-themes used in thematic analysis are presented in Table 6.

Table 6. Analysis themes (Eerikäinen *et al.,* 2020).

Main theme	**Sub-theme**	**Definition**
Physical systems	Actuators	Process equipment, *e.g.* pumps, compressors, valves, motors, centrifuges
	Instrumentation	Sensors and analyzers in the process and its surroundings
	Maintenance	Maintenance & condition monitoring of process equipment and sensors
Digital systems	Data analysis	Tasks related to data processing, analysis and visualization
	Digital tools	Tools used for data collection, processing, analysis and visualization
Process control		Process control logics and strategies, including use of controllers
Operating environment		Non-technical issues, *e.g.* personnel, management & purchasing, cooperation with contractors & providers etc.

Perceived barriers and future development were categorized also by the source, *i.e.* the interviewee group presenting these views. Topics per interviewee group are presented in Figure 21. For WWTP staff, a variety of barriers for optimizing the process control, operation and maintenance with data analytics originate already from the limitations in the physical systems.

Consultants shared a similar view with WWTPs but emphasized more the barriers in maintenance work and less in the equipment and instrumentation. Technology providers emphasized barriers in digital tools, such as integration of software, while automation companies highlighted the barriers in actual analysis phase, such as required maintenance and updating if process models are used.

When discussing future development, very few interviewees emphasized the development of equipment and instrumentation. Most expectations were focused on increasingly developed digital systems and process control strategies shifting towards more predictive and automated control. Maintenance practices were expected to develop mostly along with new digital tools.

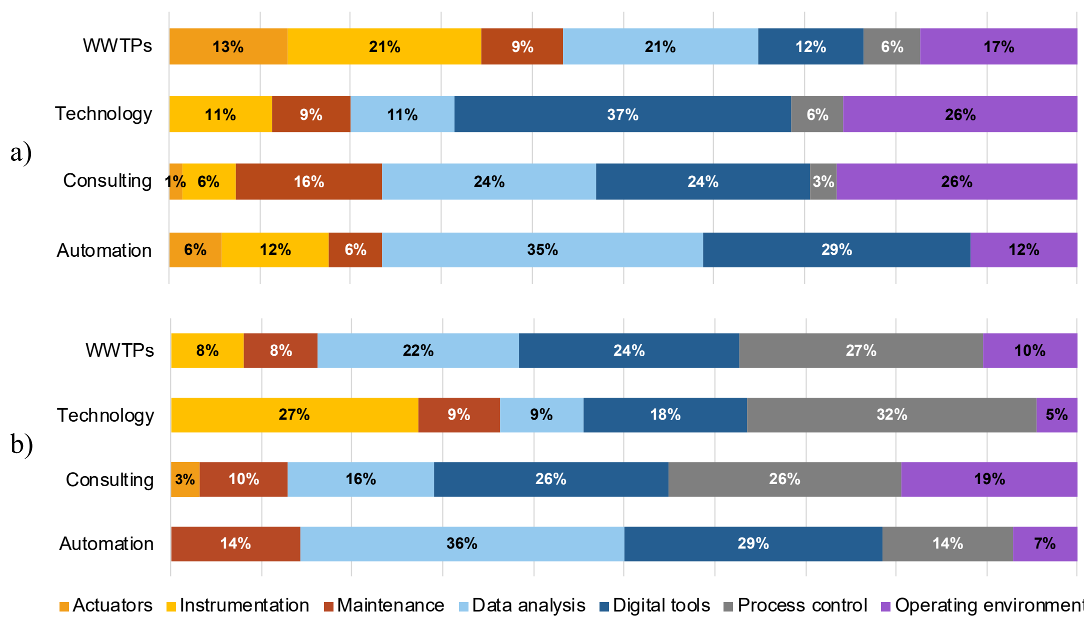

Figure 21. Division of perceived a) barriers and b) future development considered by different interviewees (Eerikäinen *et al.,* 2020).

Further on, interviewed WWTPs were divided into three size categories described in Table 7. The division is not based on any general classification but only for the purpose of distinguishing differences between the plant types interviewed in this study. Results for perceived barriers and future development are presented in Figure 22.

Table 7. Interviewed WWTP size categories.

Category	**Average flow rate**	**No of WWTPs**
Small	0 – 10 000 m^3/d	3
Medium	10 000 – 30 000 m^3/d	3
Large	30 000 – 250 000 m^3/d	4

From the barriers, it is clear that small plants suffer the most from problems with the process equipment. Role of instrumentation and maintenance is more highlighted in the larger utilities. Digital systems play a significant role in all size categories, but operating environment is highlighted in smaller ones, referring to *e.g.* personnel resources, incentives and decision making.

For future development, process control is emphasized more in smaller utilities, whereas medium and large utilities expect most development to happen in digital tools. This could be interpreted to indicate a higher need for automatic control in smaller utilities, while larger utilities have more interest for new analysis tools to understand and optimize the process with their own expertise.

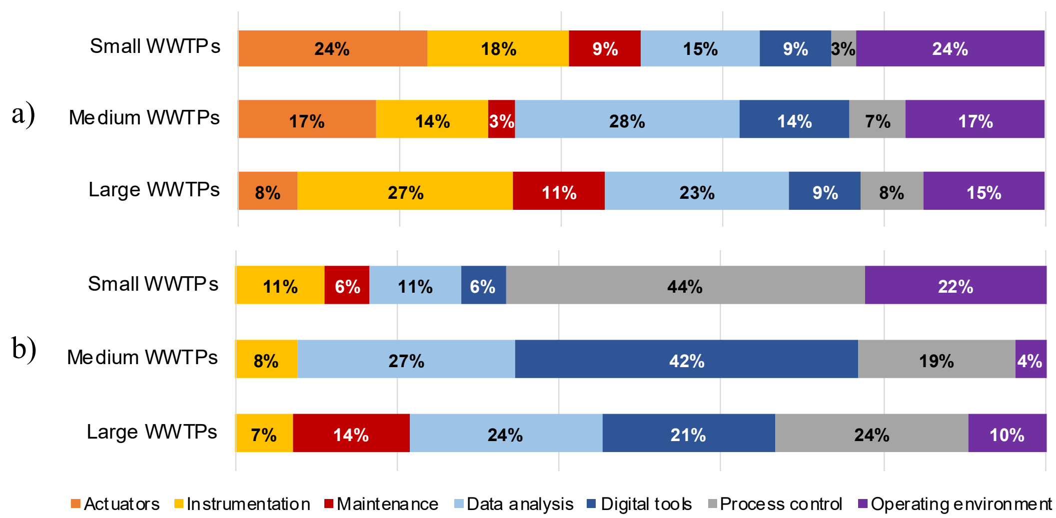

Figure 22. Division of perceived a) barriers and b) future development considered by differently sized WWTPs.

Overall themes and their key content were compiled into figures and underlying views described in further detail. Theme figures conclude frequently mentioned views from all interviewees related to the respective theme: descriptions of current practices (now), limitations for more optimal use (barriers), identified development needs (needs) and how this theme is expected to develop in the future (future). The first theme of physical systems, consisting of the sub-themes of instrumentation, process equipment and maintenance, is presented in Figure 23.

Theme 1: Physical systems

Limited resources are prioritized for basic operation and maintenance tasks.

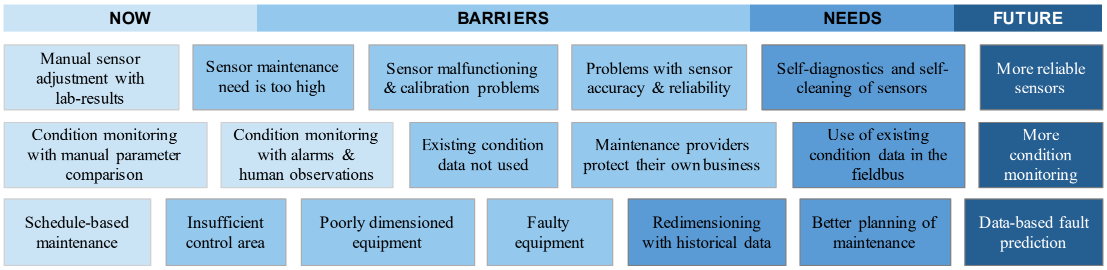

Figure 23. Theme 1: Physical systems (Eerikäinen *et al.,* 2020).

Maintenance need of instrumentation appeared as the key barrier for several control applications, contributing also to a lack of personnel resources, data quality issues and unstable control loops. Consequently, the potential benefit from data analytics is reduced. Condition monitoring of sensors and equipment is often based on alarm limits and human senses with some exceptions. For example, sludge dewatering centrifuges commonly have vibration and temperature measurements and some utilities monitor also the temperature of critical bearings. Although it is recognized that sensor signals often contain relevant condition data, extracting the information from the fieldbus network is deemed difficult and laborious by the personnel.

Functionalities of instrumentation, such as self-cleaning and self-calibration, have largely developed and according to the providers, will continue to do so. Intelligent sensors, *i.e.* sensors with built-in data processing capabilities, have gained some ground at WWTPs, but providers still consider that wider application is hindered by the lack of requirements in the purchasing criteria.

Optimizing of control is sometimes also prevented by the physical limits of equipment: for example, the minimum frequency of aeration compressor is too high for night-time wastewater flow, or capacity of polymer dosing equipment is not sufficient for large flows. Poorly dimensioned equipment was found to be more common in smaller WWTPs, in which the actual flowrates might derive easier from the original design values. Some companies considered that re-dimensioning of the equipment based on historical data could often be economically feasible**.** Calculating the payback time for replacing the equipment, *e.g.* changing one of the three aeration compressors to a smaller one, could justify the investment.

Currently, some data-driven fault detection and fault prediction tools for process equipment are offered directly by the technology providers. In these cases, condition data is collected and analyzed by the individual provider. For now, this service has been utilized only for the most critical equipment, *e.g.* aeration compressors. Fault prediction is expected to further develop in the future, but there is no clear indication yet whether WWTPs will expand the outsourcing of equipment condition data collection and analysis to several different providers, in a distributed manner, or aim to centralize the data collection and analysis to either the automation or the maintenance software.

Theme 2: Digital systems

Data analysis is limited by lack of quality assurance and customized tools.

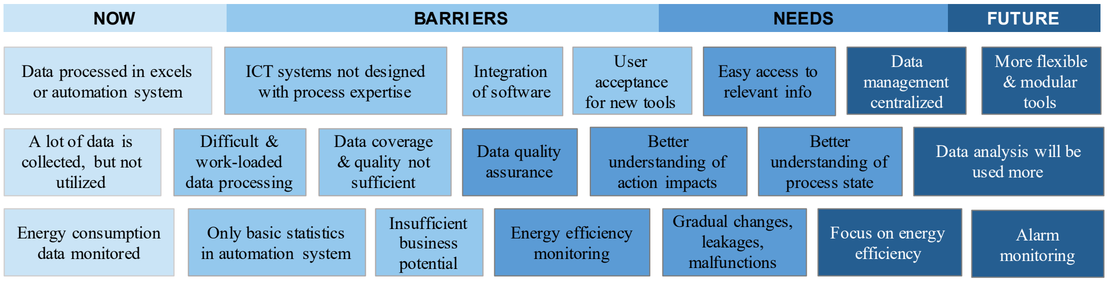

Figure 24. Theme 2: Digital systems (Eerikäinen *et al.,* 2020).

The second theme of digital systems, including the sub-themes of data analysis and digital tools, is presented in Figure 24. There are few tools for automatically monitoring or verifying the data quality, or for easy marking of faulty data. Lack of quality assurance makes data processing excessively laborious, further hindering its daily use for analysis. The automation contractor often designs and updates the automation and reporting dashboards and functions based directly on the client's requests. Involving more process expertise in the design of digital tools for WWTPs was considered beneficial by several interviewees. For now, the involvement of consultants as the source of process expertise has not been seen economically feasible.

Almost all interviewed utilities planned to analyze their process data more in the future, but for most utilities clear targets were not yet defined. The most frequently mentioned goal was monitoring energy efficiency, for which the current reporting and analysis tools in the automation system were considered as insufficient. Automation companies saw that a lot more analysis could already be done with the current software, but plant personnel has no time or expertise to adjust the system to their needs. Defining of common Key Performance Indicators (KPIs) for energy efficiency to enable utility benchmarking was considered beneficial by some WWTPs and companies.

An increasing interest was mentioned for identifying gradual changes in equipment performance, such as degradation of pumps or fouling of sensors. There are also development plans for basic statistics tools to distinguish trends in alarm sources and types. Some potential was seen in both predictive and descriptive analytics for overall supervision of the process state and verifying the impacts of operator actions.

In addition to current tools used in the automation software, some utilities are planning to purchase or have already purchased additional third-party software for further data analysis and visualization, not only for use at the plant but also within all units of the company. It is not yet clear how the overall data infrastructure will develop and who will provide the next generation of tools used in overall process monitoring. Automation providers probably have an advantage over competitors to provide simple add-ins to the already familiar software. However, some interviewees deem the viewpoint of automation companies to utility operations as too narrow for tools that would require a thorough understanding of process behavior.

Theme 3: Process control

Process control suffers from instability and requires predictability.

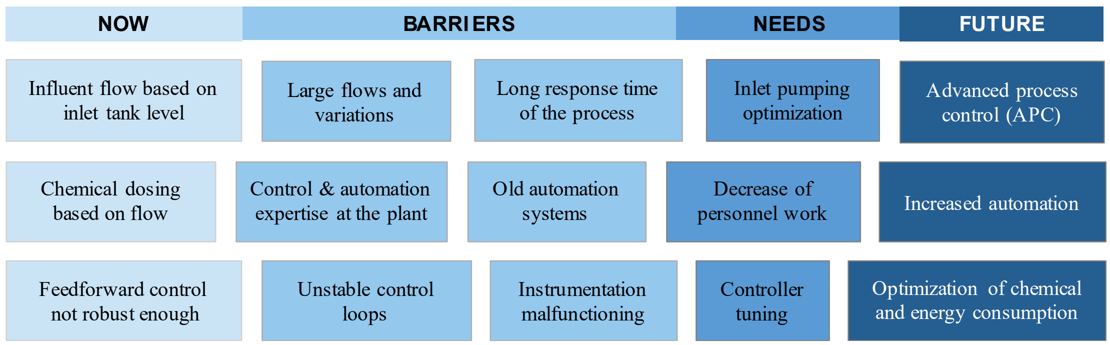

Figure 25. Theme 3: Process control (Eerikäinen *et al*., 2020).

Key findings related to process control are presented in Figure 25. Several optimized control strategies that interviewed WWTPs had tested, *e.g.* ammonium-based feedforward control of aeration, were eventually disregarded due to the increased need for sensor maintenance or instability of the control loop in varying plant conditions. In some cases, instability could be also partly due to poor controller tuning, an issue which various companies considered a major barrier for process optimization. Tuning of controllers was considered to be often inadequate already in the start-up phase and to decay throughout the years if the plant has no dedicated staff with sufficient expertise for the task.

Most of the barriers mentioned for applying more optimal control strategies are directly or indirectly influenced by the large variation of wastewater flowrate and quality. Integration of plant operation with sewage network operation was rarely mentioned, with a few exceptions of individual development projects to improve access of plant personnel to network data. Several companies mentioned the detached network operation as a major development barrier. Most WWTPs considered that stabilizing the influent flowrate is prevented by the lack of a dedicated equalization basin, but a few had also considered equalization of the influent flowrate with predictive data analytics.

Despite the strong focus on energy efficiency, optimization of chemicals was also deemed relevant now and in the future. Many utilities already had some development ideas and plans for reducing chemical consumption by using new parameters for dosage control, *e.g.* alkalinity, and either purchasing a new instrument for the purpose or using a linear correlation value from an already measured parameter, *i.e.* simple soft sensor.

Highly complex or black-box control systems did not seem to attract a lot of interest among practitioners. Instead, operator support systems to help the operator to identify and prevent disturbances or decide suitable control actions, especially in the complex parts of the process like aeration, would be adopted more easily.

A lot of future expectations involved more predictive control. This could indicate some interest in integrating the network and plant operation and moving towards a system-wide control. The level of automation was expected to increase, which was seen to change the operator duties in the future.

Theme 4: Operating environment
New generation and operation model will push demand.

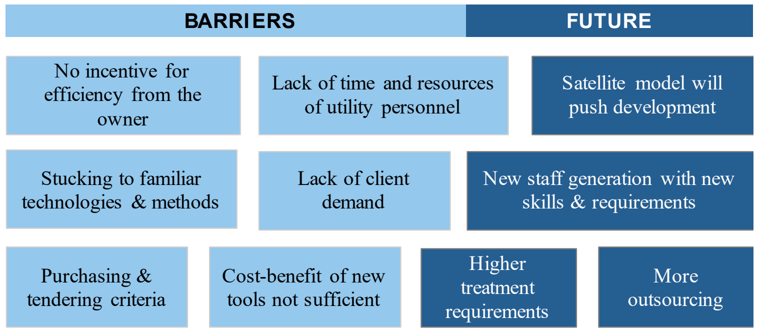

Figure 26. Theme 4: Operating environment (Eerikäinen *et al*., 2020).

Key findings related to the operating environment are presented in Figure 26. Lack of utility personnel time and resources was considered a major barrier by all interviewees. Interviewed WWTP staff mostly felt that aside of their obligatory daily tasks, they simply don't have enough time for additional process optimization or major development tasks.

Several companies and some WWTP personnel also mentioned the lack of incentives from the owner, *i.e.* municipality, to improve the cost-efficiency of plant operation. Consequently, interviewed technology and service providers saw rather small business potential in expert services and tools for the sole purpose of process optimization. Low quality criteria in tenders, for example in the purchasing of instrumentation, was considered to hinder the application of more reliable solutions. Also, insufficient emphasis on life cycle costs, in terms of duration and maintenance need, in the purchasing criteria was highlighted by the providers.

In addition to cost-efficiency, greenhouse gas emissions and overall environmental impact of WWTPs are receiving increased attention especially in larger utilities. Some WWTPs mentioned that small cost savings of a certain control strategy, that resulted in better performance but required more operator work, would not in all cases be a sufficient motive for implementation. However, the reduction of negative environmental impacts, in addition to the cost savings, was seen to justify the investment of personnel time and resources. These examples were mentioned only by WWTPs with concrete development targets and incentives defined by the company board and management.

All interviewees shared the view that utility operations will be more centralized and outsourced in the future. A satellite model, *i.e.* centralizing operation of several small plants to one larger one, was expected to gradually take place and push the development towards more centralized data management, increased process automation and higher level of outsourcing. These changes were seen to provide further motivation for using computer-based data analysis to quickly extract relevant information from the vast amount of process data, reducing the need for site visits and manual inspections. Another highlighted factor was the new generation of personnel, who are expected to bring in new skills and demand better performance and usability from the digital tools.

5.3 Case studies

In total, 14 case studies were identified and described as one-page summaries in Chapters 5.3.1 – 5.3.3. Descriptions are based on first-hand information from the users and developers and in some cases, a thesis work, study or other written report of the solution.

Classification of case studies based on their primary analysis method is presented in Table 8. Examples have been introduced for each analysis type, descriptive analysis being the most common.

Table 8. Case studies classified according to type of primary analysis method.

Diagnostic	**Proactive**
5.3.2.5 Aeration control performance monitoring 5.3.3.1 Process diagnostics and root cause analysis 5.3.3.4 Control loop performance monitoring	5.3.1.3 Sludge dewatering optimization 5.3.2.6 UV254-based ozone control 5.3.2.4 Integrated network and plant control 5.3.3.2 Model predictive control of paste thickener
Descriptive	**Predictive**
5.3.1.1 Energy efficiency monitoring 5.3.1.2 Auto-adjustment of online sensors 5.3.2.1 Plant-wide flow model 5.3.2.2 Process data analysis platform 5.3.2.3 Process data visualization and dashboards	5.3.3.3 Neural MPC operator support system for crude oil distillation 5.3.3.5 Predictive maintenance with artificial neural network

In addition, case studies were classified based on the primarily concerned process control layer as illustrated in Figure 27. In this illustration, conventional terms for the control system layers of ERP and MES have been adjusted to correspond the WWTP control system as system-wide and plant-wide control. It can be observed that several case studies are discussed for each control system layer, indicating a wide variance in types of applications. It should be noted that there is also large variance in the order of magnitude of the solutions.

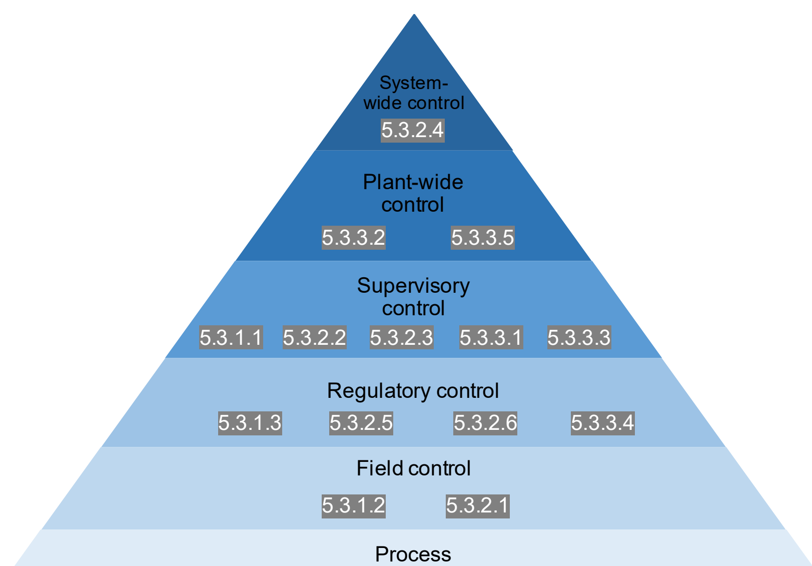

Figure 27. Case studies classified by the primarily concerned process control layer.

5.3.1 Finnish WWTPs

5.3.1.1 Energy efficiency monitoring

Location	Kariniemi WWTP, Lahti, Finland
Developer	Lahti Aqua and Insta Automation Oy
Type	Descriptive analysis for supervisory control
Application	Kariniemi WWTP uses an energy efficiency monitoring system based on Grafana software. Grafana is an open-source data analytics platform with emphasis on visualization, customized dashboards and metrics. The monitoring system was implemented as a thesis work in cooperation with the automation provider. Energy consumption data of the whole plant is processed in Grafana, which is used in parallel with Insta Wahti process monitoring software. Data is collected from 33 electrical distribution centers and 99 individual devices.
Benefits	Software is used by process engineers and plant operators for process equipment and efficiency monitoring. Implementation of the software has simplified evaluation of energy efficiency, estimation of payback times for energy saving actions, such as ammonium-based aeration control, and led to additional actions, such as replacement of supply air fan.
Pre-requisites	Cloud-based software is used in browser and can be accessed from any device by defined users. Data is collected with one-minute interval and categorized by treatment trains and 16 different unit processes. Comparison of collected data to total electricity consumption showed a difference of 2.2 %, indicating that the data gives a realistic overview of energy consumption.
Feasibility for Finnish WWTPs	Similar solutions can be applied at any Finnish WWTP for overall energy efficiency monitoring. As planned at Kariniemi WWTP, also process data and operating time of technical equipment could be brought to the software for comparison of the energy efficiency of different control strategies.

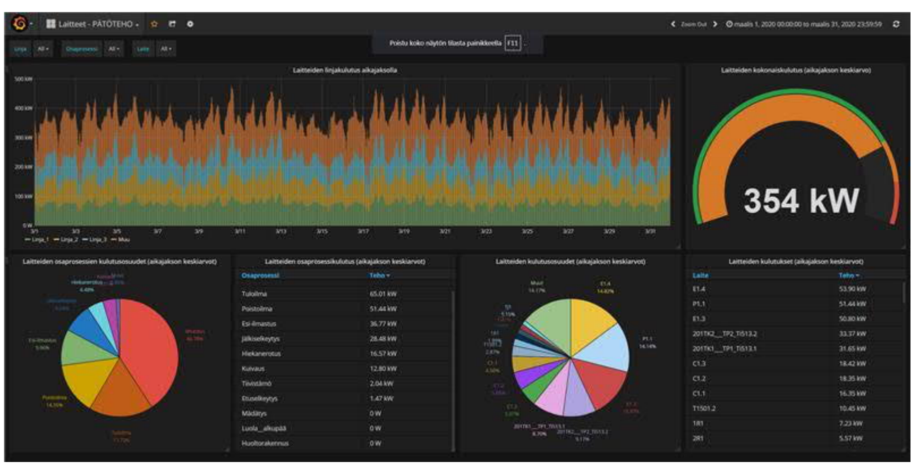

Figure 28. Example of a dashboard view in the program.

5.3.1.2 Auto-adjustment of online instruments

Location	Viikinmäki WWTP, Helsinki, Finland
Developer	Helsinki Region Environmental Services (HSY)
Type	Descriptive analysis for field control
Application	Online sensors without auto-cleaning suffer from gradual fouling that degrades their performance. With auto-adjustment, online sensor data is automatically compared with latest laboratory results using statistical methods. First, laboratory results are automatically updated to ValmetDNA from the LIMS with a macro function, while automatically checked to be within reasonable limits. Each sensor has their own (linear break point) calibration curve. Calibration curves are updated with an autocalibration loop using a combination of recursive least square and Kalman filter methods. The user may define individually for each sensor, how many previous laboratory results are used for the adjustment. After that, a separate supervision loop calculates the deviation of online values from the calibrated values with two different statistical maps. If the deviation violates a defined limit, a red dot will appear next to the parameter in the automation user interface. If *e.g.* maintenance affects the process or sensor behavior, auto-adjustment can be reset to start over.
Benefits	Auto-adjustment automatically detects gradual deviation between lab samples and online signal, indicating a need for maintenance. Most of the time, adjusted data is closer to lab results than online signal. Auto-adjusted values can also replace the physical sensors during maintenance or failure.
Pre-requisites	Auto-adjustment was implemented in 2011 as a thesis work. Both loops have been done with ValmetDNA block programming libraries. Extensive knowledge of the statistical methods is not required. Regular follow-up of the values and occasional resets may be necessary.
Feasibility for Finnish WWTPs	Similar system could be directly applied to ValmetDNA at any Finnish WWTP, and most probably any other DCS, and utilized for quality assurance for any parameters with both online and lab measurements.

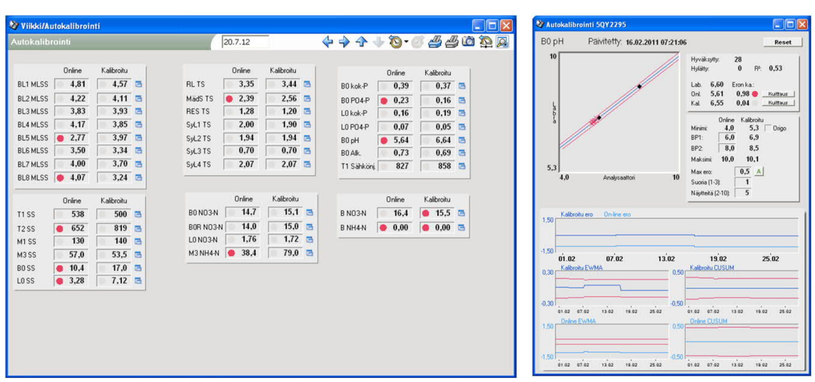

Figure 29. Overview of auto-adjustments (left) and info page for individual sensor (right) in the DCS (Jokelainen, 2011).

5.3.1.3 Sludge dewatering optimization

Location	Viinikanlahti WWTP, Tampere, Finland
Developer	Valmet Automation Oy
Type	Proactive analysis for supervisory control
Application	In Viinikanlahti, sludge treatment process is optimized to achieve savings in polymer consumption, energy use, dry cake transportation and to reduce solids concentration in centrate water. Online solids measurements were installed before the thickener, digester and centrifuge, and to centrate and to dry cake (Figure 30). For one of three centrifuges, an MPC module is used. MPC aims to optimize the centrate and dry cake quality, while minimizing the energy and chemical use. Solids measurements in incoming sludge, centrate water and dry cake are used as inputs for a multi-variable algorithm. As the output, optimal setpoints for sludge feed flow, polymer dosing and centrifuge torque are defined.
Benefits	After first year of implementation, control of primary sludge pump with total solids content (instead of flow) had reduced the pumping need with 35%, equaling to 5 000 €/a savings in energy use, and reduced the sludge treatment need by 32%. In dewatering, polymer usage decreased by 40%, saving 49 000 €/a. Solids content in the centrate decreased by 50%. Estimated savings in dry cake transportation were 80 000 €/a.
Pre-requisites	MPC algorithm requires several accurate online solids measurements in different process stages. Sensors have to be regularly maintained to ensure stable control performance. Changing process parameters from cubic meters to kilos takes time for operators to get used to.
Feasibility for Finnish WWTPs	The solution is readily applicable to Finnish WWTPs. New sensors increase the need for instrumentation maintenance, but potential savings are significant and improved stability of sludge treatment process might compensate for the increase in workload.

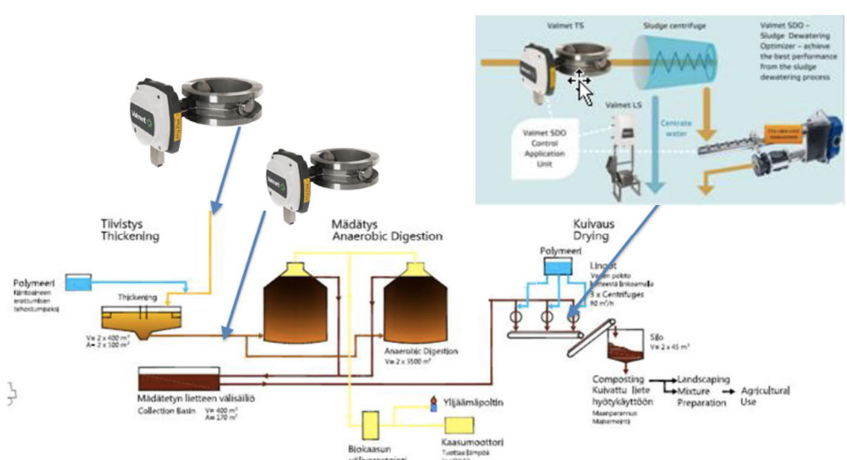

Figure 30. Measurement locations. (© Tampereen vesi)

5.3.2 International WWTPs

5.3.2.1 Plant-wide flow model

Location	Rya WWTP, Göteborg, Sweden
Developer	Gryaab AB
Type	Descriptive analysis for field control
Application	Plant-wide flow model enables quality assurance for conventional flow meters, and flow estimates for locations unsuitable for physical meters. Rya WWTP has built an online plant-wide flow model with simple IN/OUT models for each unit process, see Figure 31. Flow estimates are derived from online sensors and where not available, from arithmetic calculations, soft sensors and level measurements. For pumping stations, flow estimate comes from pump lift height and frequency/power. For one spot, weir and level measurements are used to derive a flow estimate. Deviation in estimates sets an alarm for the operator to investigate the cause.
Benefits	Values from the flow model are used to control internal plant circulations, bypass flows and different unit processes. All flows are monitored from a plant-wide view in the DCS. The system has pinpointed faulty values in extreme flow conditions, which would have otherwise gone unnoticed and deteriorated process performance. Improved quality of also mass flows and mass balances used in control reduces process disturbances.
Pre-requisites	When the plant configuration is updated, new modules can be added to the model. Pump characteristics used in soft sensors might need to be readjusted, *i.e.* if the pump or its parts are replaced.
Pricing model	At Rya WWTP, implementation of the flow model required no additional investments in instrumentation. Estimated investment (2005) for implementation was 35 000 €, consisting of only staff input: process engineer (85 h), engineer trainee (430 h) and DCS programming (345 h).
Feasibility for Finnish WWTPs	A similar flow monitoring system could be applied at any Finnish WWTP for instrumentation quality assurance or additional flow estimates in tricky locations. Depending on the plant configuration and instruments, existing online sensors might already suffice. Soft sensors can be added for quality assurance.

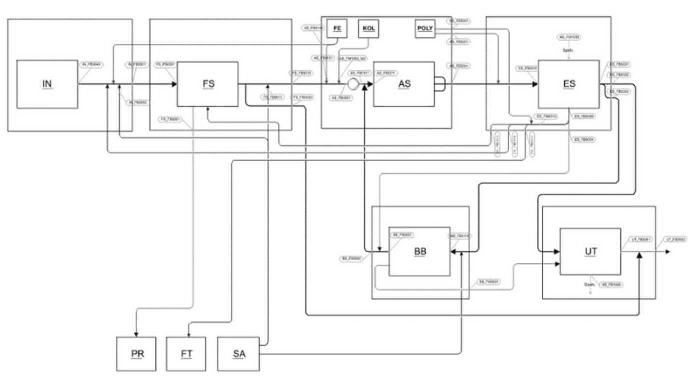

Figure 31. Plant-wide integrated flow model. (Äijälä & Lumley, 2006)

5.3.2.2. Process data analysis platform

Location	Käppala WWTP, Stockholm, Sweden
Developer	Gemit Solutions AB
Type	Descriptive analysis for supervisory control
Application	aCurve application enables visualization and analysis of process data collected from the DCS, LIMS, maintenance software and other databases. Filters and rules can be applied to improve data quality. In addition to basic statistics, derivative and integral functions, aCurve can be used for soft sensors to identify gradual changes in the process or equipment, *e.g.* gradual change in the ratio between pump production and energy consumption.
Benefits	At Käppala WWTP, work hours saved using aCurve instead of previous reporting systems were estimated. In total, about 700 hours are said to be annually saved in general reporting and monitoring tasks, equaling to 50 % of a full-time process engineer. When the value of saved time was estimated as relative to cost savings achieved with process optimization, described as 100 000 – 200 000 € annually, the payback time was less than one year.
Pre-requisites	Integrations are available for all common automation, laboratory and maintenance software at WWTPs. aCurve may be used for general reporting and monitoring work by all personnel with varying process and IT expertise.
Feasibility for Finnish WWTPs	aCurve application is readily applicable to Finnish WWTPs. It may be beneficial for those utilities with challenges in data integration, reporting and process monitoring using current systems. Software is available in Finnish.

Figure 32. Example from a dashboard view with different KPIs and visualizations.

5.3.2.3 Process data visualization and dashboards

Location	Langmatt WWTP, Switzerland
Developer	Rittmeyer Group
Type	Descriptive analysis for supervisory control
Application	At Langmatt WWTP, various visualizations and dashboards have been developed for the use of plant personnel, external specialists and general public. Visual dashboards of RITUNE software at control room screen provide quick overview to process state and potential disturbances. For example, visualizations include heatmaps (Figure 33), combining 42 signals of seven different variables in two basins, and a calendar view combining daily, weekly, monthly and annual averages for different process and energy efficiency KPIs. In addition, software modules for optimization of the precipitant dosage for phosphorus removal are activated.
Benefits	The dashboards are described to facilitate process monitoring and fault detection and promote cooperation between operating and planning personnel. They are also seen to function as positive feedback and an incentive for daily operation. Module for optimization of precipitant dosage has reduced precipitant consumption by over 15%.
Pre-requisites	The visualizations have been implemented with readily available modules in RITUNE process data analysis software. The software can be used from cloud or in-premise, runs in web-browser and may be integrated to most of the common automation systems.
Feasibility for Finnish WWTPs	RITUNE software is available in English and may be directly applied at most Finnish WWTPs as a complementary software on top of automation systems. The software has in total over 100 different modules for the purposes of process optimization, visualization and supervision.

Figure 33. Control room dashboards for process heatmaps (left) and biogas production (right), including a visual CO_2 savings calculator (© Rittmeyer Group).

5.3.2.4 Integrated network and plant control

Location	BlueKolding WWTP, Kolding, Denmark
Developer	Veolia AQUAVISTA^{TM} (former Kruger STAR)
Type	Proactive analysis for system-wide control
Application	Process control platform Aquavista has been used at BlueKolding WWTP since 2006 to optimize the plant control, equalize the influent and reduce combined sewer overflows (CSO). Aquavista defines set-points for PLCs to implement local control through OPC UA, see Figure 34. In addition to influent control, 12 other control modules are used in *e.g.* air supply, sludge circulation and phosphorus precipitation, to maximize the treatment capacity and to prevent the sludge from escaping from the clarifier. In SewerFlex module, settings of weirs, gates and pumps in the sewage network are adjusted based on the real time network data and rain forecasts. Setpoints are generated with a continuous dynamic overflow risk assessment algorithm, that takes into account actual basin levels, predicted flows based on rain forecast, relative cost of overflows in each location, hydraulic emptying capacity and hydraulic capacity at receiving WWTP.
Benefits	Aquavista has enabled to operate the plant largely remotely with no need for night shifts or alarms even in extreme flow conditions. Flow equalization enables more stable plant control. Occurrence of CSOs is minimized and avoided especially at most vulnerable locations. When increase of storage volume was planned, optimal operation of network pumping stations reduced the additional storage need by 50 %.
Pre-requisites	Reliability of online sensors is highly important. In Kolding, their accuracy is monitored with regular comparison to lab samples. In the network, conventional flow and level measurements are sufficient.
Feasibility for Finnish WWTPs	Some Finnish WWTPs have basins that can be utilized for flow equalization. Even without a basin, some transfer tunnels might provide means for equalization with automatic control. Other control modules could be tested for *e.g.* optimal management of wet weather flows.

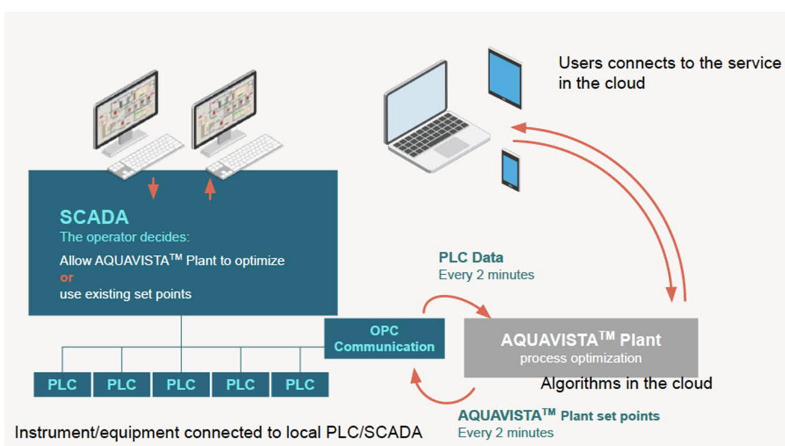

Figure 34. AQUAVISTA^{TM} Plant configuration in plant control system (© Veolia).

5.3.2.5 Aeration control performance monitoring

Location	Grand Rapids WWTP, Michigan, United States
Developer	inCTRL Solutions Inc.
Type	Diagnostic analysis for regulatory control
Application	opsCTRL is a software used on top of current DCS for data integration, clean-up and augmentation, sensor quality control and controller performance monitoring. Additional functionalities of opsCTRL include closed loop autotuning tools and a digital twin setup, *i.e.* running a process model with a simulator in parallel to the real plant. In Grand Rapids, also inCTRL's patented ammonia-based aeration control and sludge retention time controller (ABAC-SRT) has been implemented. In the software, overview to aeration control can indicate faults and bottlenecks in controller, actuator and sensor performance (Figure 35). In addition, model-based autotuning allows simulation and selection of optimal control parameters. If *e.g.* control valves are identified to continuously hit the lower bound, air pressure can be adjusted accordingly.
Benefits	Better data quality enables testing and implementation of more advanced control strategies. Monitoring of controller, actuator and sensor performance enables to identify the limiting components in different flow and process scenarios. Consequently, control bottlenecks can be resolved with controller tuning or re-dimensioning of equipment accordingly.
Pre-requisites	Software is set up as a local installation to DCS using an OPC connection, after which it can be used from a browser through VPN connection.
Feasibility for Finnish WWTPs	WWTPs with challenges in tuning and optimizing the aeration process control could benefit from monitoring the controller performance and its bottlenecks. Some operator training would be required for implementation.

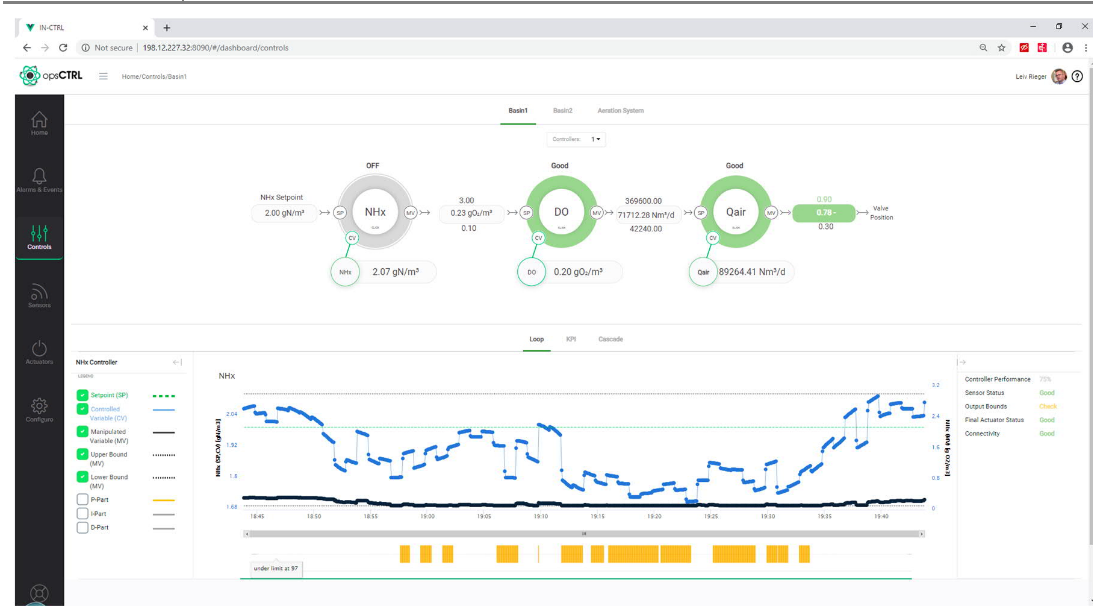

Figure 35. Example of aeration control loop performance dashboard.

5.3.2.6 UV254-based ozone control

Location	Neugut WWTP, Wallisellen, Switzerland
Developer	ARA Neugut
Type	Proactive analysis for regulatory control
Application	At Neugut WWTP, ozonation process is utilized to achieve 80% removal of micropollutants (Schachtler & Hubaux, 2016). Removal efficiency is monitored as an average of 12 indicator substances. A new control strategy for ozone dosing was developed with UV254 measurements. Absorbance of UV at 254 nm correlates well with overall removal efficiency of micropollutants. For this purpose, a plant-specific correlation coefficient was defined. UV254 sensors are placed at both inlet and outlet of the ozonation tank. Sensors were purchased from three different technology providers. Ozone is dosed to both first and third chamber to minimize production of oxidation byproducts. Different control modules are used for flow scenarios of dry, normal and wet flow. Control setup is presented in Figure 36.
Benefits	As a result, treatment efficiency of 82% ± 2 % has been reached in a stable manner. Control performance has been studied to result into 20% less ozone consumption as compared to conventional control based on incoming absorbance or flow.
Pre-requisites	Stable control result requires accurate UV measurements, *i.e.* reliable sensors, sufficient sensor maintenance and quality assurance methods.
Feasibility for Finnish WWTPs	For ozonation, similar control strategy could be tested and evaluated also at Finnish WWTPs. Suitability of UV254 measurement for monitoring of micropollutant removal could be studied also in other unit processes.

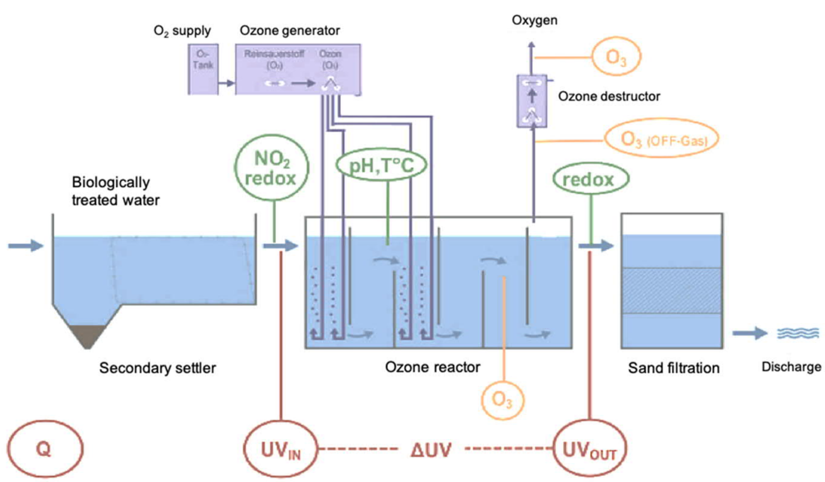

Figure 36. Control setup and measurements in the ozone reactor (translated, Schachtler & Hubaux, 2016).

5.3.3 Process industries

5.3.3.1 Process diagnostics and root cause analysis

Location	Forestry, mining, food processing and chemical industries, Finland
Developer	Trimble Wedge
Type	Diagnostic analysis for supervisory control
Application	Wedge is an online process diagnostic system originally developed for Finnish forestry industry. It enables root cause analysis for multivariable processes with multiple statistical methods, *e.g.* PCA, correlation, waveform and fluctuation analysis. The analyses can be automatically run with data from different sources. For example, a food processing plant experienced variation in the end product quality as seen in Figure 37. A fluctuation analysis was done to identify periodic fluctuations and their quantity, frequency and magnitude. From the analysis, a variation in steam temperature of an earlier process step was identified and concluded as the root cause.
Benefits	Wedge enables process diagnostics for operators, engineers and managers with no need for extensive knowledge of statistical methods. Process anomalies previously investigated manually can be automatically scanned and diagnosed. Process data can be automatically cleansed, or only the relevant data period or range visually selected.
Pre-requisites	Most benefits are achieved in complex dynamic processes with multiple couplings and interactions of process variables. Poor data quality could degrade some benefits but can be improved with automatic data cleansing.
Feasibility for Finnish WWTPs	Wedge is technically readily applicable to WWTPs. Potential use cases and achievable benefits could be studied for root causes of variations in *e.g.* effluent quality, activated sludge process or biogas production.

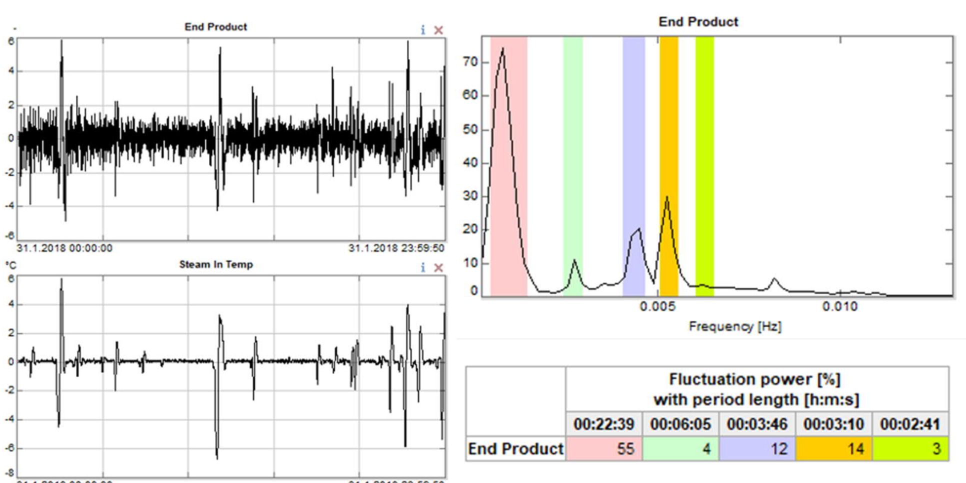

Figure 37. Fluctuation analysis indicating a fluctuating phenomenon (right) and its source (lower left), which causes a variation in the end product (upper left).

5.3.3.2 Model predictive control of paste thickener

Location	Yara apatite mine, Siilinjärvi, Finland
Developer	Outotec Oy
Type	Proactive analysis for regulatory control
Application	Yara Siilinjärvi mine produces apatite concentrate, which generates a side stream of approx. 10 million tons of tailings. Tailings are treated at the paste thickener plant. From the thickener, the paste is pumped to a piling area through a kilometer-long pipeline. For this purpose, a steady paste quality with 68–70 % solids content is necessary. Because of the large variance in raw paste quality and volume, steady treatment result is difficult to reach with conventional control methods. MPC algorithm is applied to optimize the thickening process and underflow. MPC control system, demonstrated in Figure 38, utilizes five process variables, from which underflow density and overflow solids are the priority ones. The control logic is located in a separate server, which is installed on top of the customer DCS.
Benefits	As a result, the process performance is more accurate and requires less manual work from the operator. New control strategy has enabled remote operation of the plant from the concentrator control room, 6 km from the paste thickener plant. Steady quality of the paste enables better piling result, which further increases the lifetime of the current piling area.
Pre-requisites	MPC algorithm requires high quality data and reliable sensors to result in a good control performance. Control performance is revised by the technology supplier about 2–3 times a year. In addition to control design, implementation and update, lot of resources has been put to train the operators properly.
Feasibility for Finnish WWTPs	Paste thickener process can be considered analogic to pre-sedimentation, post-sedimentation or sludge thickening in wastewater treatment. Automatic control requires accurate measurements for underflow solids, overflows solids and sludge blanket level.

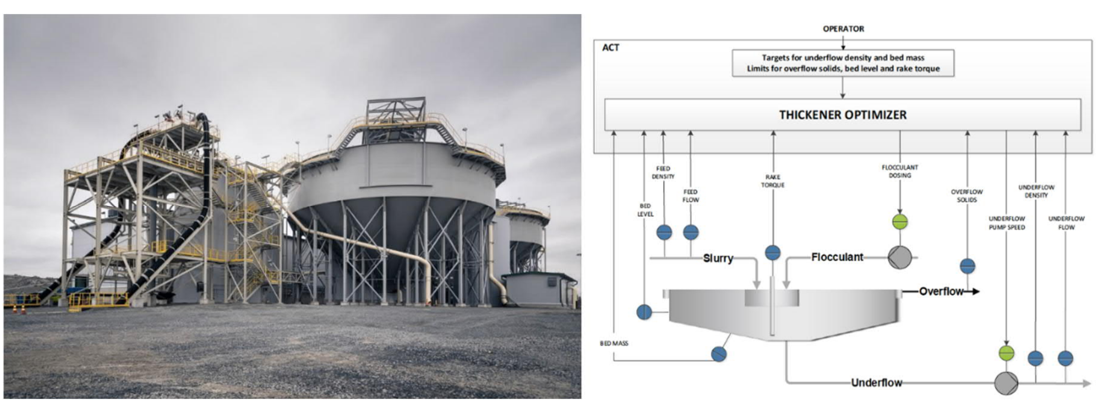

Figure 38. Yara paste thickener plant and control system (Ruhanen *et al.,* 2018).

5.3.3.3 Neural MPC operator support system for crude oil distillation

Location	Oil refinery, Finland
Developer	CuriousAI and NAPCON
Type	Predictive analysis for plant-wide control
Application	Neural MPC utilizes a non-linear deep neural network for estimating the parameters for MPC. For complex process dynamics and varying raw material in crude oil distillation, it enables a more accurate process model for optimizing the control. Neural MPC will be utilized for instructing the operator on optimal actions during state transitions in an oil refinery. Process performance is predicted for 20 hours ahead and optimal control actions defined accordingly. Control objective is set with a cost function that accounts for 236 defined process constraints. Instead of constant values, setpoints of six controlled variables change in time to enable optimal state transition.
Benefits	Neural MPC has proven to predict process performance within confidence limits and yield a better control strategy compared to the operator-led control. When altering the process state, it can reach a steady state much faster as demonstrated in Figure 39. In crude oil distillation, this translates directly into higher profit from the end products.
Pre-requisites	The algorithm may be applied to any automation system with OPC UA or similar communication interface. Measurement accuracy of parameters used especially in the cost function, is essential for good control performance. In general, neural networks tolerate measurement noise and faults better than conventional control.
Feasibility for Finnish WWTPs	Applicability of Neural MPC could be studied *e.g.* as an operator support system for the control of activated sludge process, posing complex non-linear process dynamics.

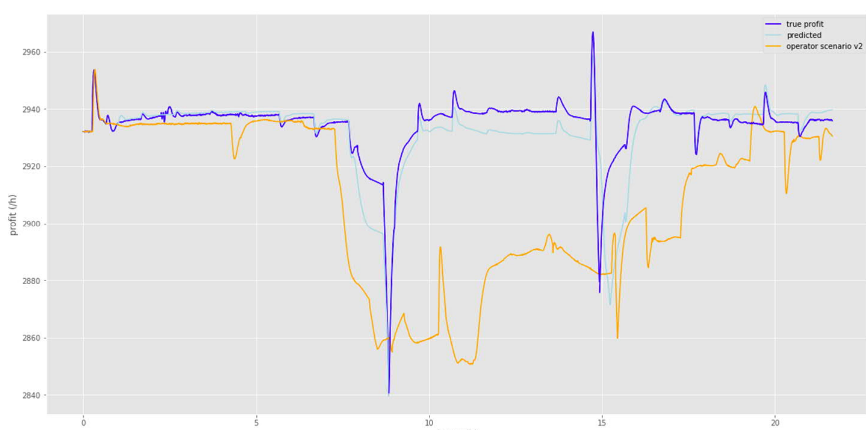

Figure 39. Simulated control performance with neural MPC (blue) compared to operator scenario (yellow) when altering process state in crude oil distillation (NAPCON, 2019).

5.3.3.4 Control loop performance monitoring

Location	Pulp and paper industries, Finland
Developer	Valmet Automation Oy
Type	Diagnostic analysis for regulatory control
Application	Controller and actuator performance can be automatically or periodically monitored. Performance monitoring tools enable automatic detection of poorly dimensioned or poorly performing actuators, equipment, unit processes and production units. Proper tuning of control loops can fix undesired process and control behavior, such as unstable control loops, large deviations and actuator malfunctions. Identified issues may be resolved by *e.g.* changing the operating point, tuning PID parameters and replacing faulty actuators.
Benefits	Automatic controller performance monitoring enables quality assurance and proactive maintenance of the control system. As a result, production capacity is increased and end-product quality improved. Incorrect dimensioning and faults in *e.g.* valve operation are acknowledged faster. Most critical poorly performing loops can be prioritized for maintenance and tuning.
Pre-requisites	Control performance tools can be utilized as an additional service for ValmetDNA DCS. Also, remote tuning services are available.
Feasibility for Finnish WWTPs	Control loop performance monitoring could reveal major control loop faults also in WWTPs. Systematic study could be done to a few different types and sizes of plants to assess the prevalence and impact of controller tuning and operation on overall plant performance.

Loop Monitoring
TOP-listaus

Osasto: Kamyr1
PVM: 08.08.2016

P = Performance = osuus ajasta, jonka suorituskykyindeksit ovat olleet "hyviä"
A = Automation = osuus ajasta, jonka säädin on ollut automaatilla
S = Saturation = osuus ajasta, jonka säätimen ohjaus on ollut fyysisten rajojen välissä
F = Fault = osuus ajasta, jolloin tiedon keruu on pelannut oikein

100 = hyvä
0 = huono

Valmet

Positio	Kuvaus	Tila	Diagnoosi	P	A	S	F	Käsin tehdyt ohjausmuutokset (kpl)	Auto-manual vaihdot (kpl)	Ranking suorituskyky	Ranking autotila	Ranking saturaatio	Ranking käsin tehdyt muutokset	Total ranking (1= Huonoin)	Kommentit
TIC-098	PUSKUPUTKI LÄMPÖTILA	BAD	Controller not operational	0	0	100	100			1	1	17		2,4	
LIC-104	PAISUNTASÄILIÖ 1	BAD	Controller not operational	0	0	100	100			2	2	23		3,7	Pintamittaus rikki. Mittaus tulee vaihtaa seisokissa.
TIC-111.1	VESI LX6 JÄLKEEN LÄMPÖTILA	BAD	Controller not operational	0	0	100	100			3	3	26		4,7	
TIC-111.2	VESI LX6 JÄLKEEN LÄMPÖTILA	BAD	Controller not operational	0	0	100	100			4	4	27		5,5	
LRC-080B	KEITTIMEN LIPEAPINTA	BAD	Controller not operational	0	0	100	100			5	5	32		6,7	
TIC-163	MUSTALIPEÄ LX8 , LÄMPÖTILA	BAD	Controller not operational	0	0	100	100			6	6	40		8,2	Lämpötilasäädin ei ole käytössä koska lämmönvaihdin on rikki. Jjotta säädin saataisiin toimimaan tulisi lämmönvaihdin korjata
PIC-6404.2	LAIM HAJUK PAINE SUORAKÄYT	BAD	Controller not operational	0	0	100	100			7	7	43		9,2	
LRC-065	KEITTIMEN HAKEPINTA	BAD	Small or slow oscillations	2	100	100	100			10	17	18		9,5	
FIQC-147	PASUTUSASTIA KAASAUS	BAD	Control saturated	41	100	33	100			11	28	7		10,1	
TIC-6406	LAIM. HAJUK. LÄMPÖTILA	BAD	Control saturated	52	100	39	100			12	22	8		10,2	
PIC-063	KEITT.HÖYRYTILA	BAD	Controller not operational	0	0	100	100			8	8	47		10,3	
LIC-154	HAKESIILO ,PINTA	OK	OK / Control error and variability small	82	100	100	100			14	13	14		11,1	
LIC-105	PAISUNTASÄILIÖ 2	OK	OK / Good control quality	80	99	100	100			13	11	22		11,1	
LRC-080.1	KEITTIMEN KOK. PINTA	BAD	Controller not operational	0	0	100	100			9	9	49		11,2	
FICQ-027	VALKOLIPEÄ IM.TORNI	OK	OK / Good control quality	84	100	100	100			15	23	20		13,3	
LIC-112	S.LAUHDES SL19 ALAP	OK	OK / Good control quality	93	100	10	100			19	19	6		13,9	

Figure 40. Example overview of control loop performance report.

5.3.3.5 Predictive maintenance with artificial neural network

Location	Hydro, combined heat and waste-to-energy plants, Finland
Developer	MPIntelligence® (part of Caverion Group)
Type	Predictive analysis for plant-wide control
Application	MPIntelligence® is a predictive maintenance software that can learn complex system behavior and identify anomalies in non-linear processes. Among others, it is used for process and equipment condition diagnostics in hydro, combined heat and waste-to-energy power plants of Fortum. An artificial neural network algorithm continuously receives monitoring data from tens of thousands of data points at each plant and detects potential disturbances before they impact the process. The software allows identification of anomalies from multivariate signals, which do not raise an alarm in the automation system. When the operator issues the alarms as true or false, the algorithm updates and becomes more accurate over time.
Benefits	In a combined heat power plant, the software identified a periodic variation in bearing temperature of a flue gas fan in the main boiler. Physical inspection revealed a bearing fault, which was then repaired, and potential costly shutdown was avoided. In a hydro power plant, an axle vibration in a generator was identified when operated in a lower power level. The vibration remained below the alarm limit but shortened the generator lifetime. As a result, operating practices were changed, and a premature equipment breakdown prevented.
Pre-requisites	Data may include process measurements, equipment and actuator data, such as bearing temperatures and valve positions. The software is mostly used from browser as cloud-based, but available also as on-premise installation.
Feasibility for Finnish WWTPs	Software could be applied for monitoring the process equipment also at WWTPs. Potential savings from avoided disturbances are assumed to be lower compared to power generation industry. Therefore, economic feasibility of the solution could be studied with a pilot.

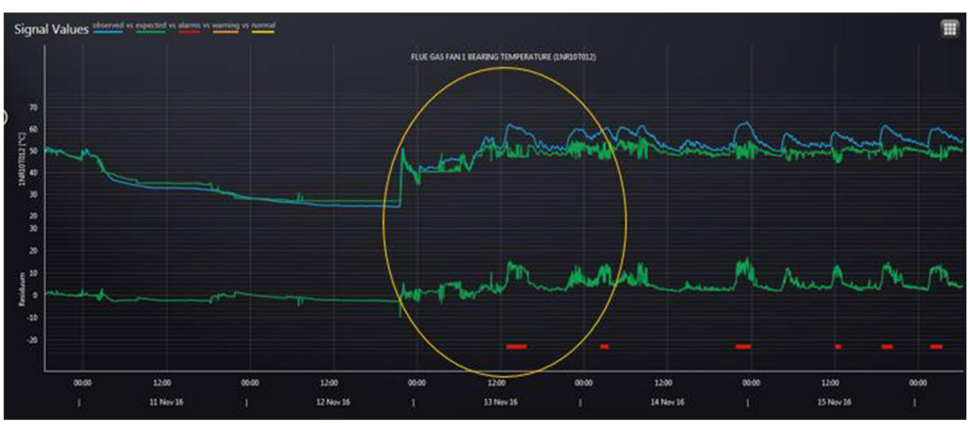

Figure 41. Anomalies in the variation of flue gas fan temperature, indicating a fault, were automatically identified with the software (© MaintPartner).

6 Discussion

6.1 Incentives for developing ICA with data analytics

The results of this study indicate that the development of ICA systems with data analytics is hindered by several barriers in both physical and digital systems, control strategies and operating environment. Equipment might be poorly dimensioned, instrument data untrustworthy, controllers lack proper tuning and staff incentives be disregarded. Barriers and their root causes slightly vary between plant types and sizes, with smaller plants facing the largest challenges with physical systems.

It should be noted that in addition to size, a range of other factors distinguish the interviewed WWTPs. Year of construction and possible renovations greatly influence the technical performance and maturity of the plant. As discussed, several different plant configurations and process types were included, some of them in major towns and some in *e.g.* holiday resort areas with significantly higher influent variation. In addition, economic status of the municipality may influence the available resources for water and sanitation services. Most of the barriers, such as instrumentation maintenance, were shared by all types of WWTPs, implying that some deficits may lie in the general practices in the sector.

For especially the smaller plants, keeping the plant running and satisfying the effluent permit requirements were of primary concern. If achieving them is not fully granted, minimizing operating expenses was not considered relevant. These findings are well in line with control objectives defined by Olsson *et al.* (2014). In plants where permit requirements are presumably always met, key incentives for process optimization were mentioned as energy efficiency and improved treatment result, both much more often than cost efficiency. This can be interpreted to reflect the emphasis of WWTPs on environmental protection and sustainability. Then again, low attention for operational expenses, in comparison to investments, might originate from budgeting practices in municipality-led organizations.

It is clear from the interviews that optimizing the process by compromising the robustness is strongly opposed. This can be seen to reflect the general work culture of the water sector with naturally high emphasis on operational reliability and safety. Acknowledging the risks involved in *e.g.* new control strategies is indeed necessary: quite often, seemingly most optimal solutions often come with the downside of reduced safety margins of the process (Rieger & Olsson, 2012). One example for this is given from ammonia-based control of aeration. While reduction of aeration intensity improves energy efficiency, it increases the risk of violating the ammonia limit, if a system failure or unanticipated disturbance occurs.

Yet, prudence towards new practices can be observed to reach beyond its original purpose. With operational reliability valued high, one could assume that a range of quality assurance measures throughout the process would be easily adopted. Yet, interviews indicated that in addition to conventional alarms indicating that a clear fault already took place, very few verification methods for process, actuator and sensor performance are in use. Even though the personnel often mentioned many development ideas, there rarely seemed to be strong incentives to put them into practice. Quite the opposite, since many interviewees mentioned the high risk of negative media attention with long-term damage for reputation. Although cautiousness in experimentations is necessary, Rieger and Olsson (2012) discuss that this type of “blame culture” is a common reason for sub-optimal WWTP performance.

Considering the rapidly changing operating environment, sticking to the familiar might soon no longer work. Not only is the aging infrastructure degrading the performance and posing more risks of failure (Silfverberg, 2017), also other changes are taking place. ICT systems in water utilities have also become targets for denial of service attacks and other cyber security threats (Rautiainen, 2019). Climate-change induced variation in influent flow rate and temperature will further increase the flooding risk and need for bypassing control. The global pandemic in Spring 2020 provoked new risk scenarios and pre-cautions for the personnel to secure operational reliability of water and wastewater services (World Health Organization, 2020). Effective use of data analytics in ICA systems can contribute to overcoming all these challenges.

6.2 Development barriers and needs

6.2.1 Stakeholder perceptions

Division of perceived barriers and future development was assessed by interviewee groups and sub-themes. In semi-structured interviews, some deviation on mentioned issues can originate simply from the different paths the discussion has taken. It is however clear, that same open questions raised differing opinions in different interviewees. The background and professional experience of interviewees clearly guided their answers. This suggests that more active interaction between different professions might already fill some gaps and allow more interdisciplinary solutions.

Perhaps most notable variation is the emphasis of WWTP personnel on the performance of physical systems, *i.e.* their daily work tasks, tools and problems, which was not highlighted in a similar manner by interviewed companies. In practice, this was reflected for example in ammonia-based aeration control, which had been initially planned for several WWTPs and naturally, assumed by the designers to be also used. However, problems in sensor maintenance and control stability had eventually led the plant personnel to apply conventional DO cascade control instead. Potential bias in views of different stakeholders could be acknowledged and mitigated by closer communication of plant operators, process designers and technology providers already in the design phase. Initiative for this should probably come from the WWTP as the client that sets the criteria and requirements for all collaborators.

Key development barriers include laborious instrumentation maintenance, difficult data processing, large flow variation and lack of personnel resources. These barriers are often inter-linked: with little personnel resources, allocating a lot of time for proper instrumentation maintenance is not possible. Consequently, poor or unreliable quality of process data makes the processing and analyzing of data seem useless for the interviewees. Without accurate data of process performance, more efficient (and robust) control strategies for varying flow conditions cannot be developed. Most of these barriers could be at least partly solved by simply increasing the personnel resources at the plant. Yet, with current and expectedly increasing economic constraints in several municipalities, all potential solutions should be evaluated.

In the state-of-the-art survey of Haimi *et al.* (2009), some key development and research areas were identified as prediction of wastewater load and characteristics, fault modelling, diagnosis and isolation and operator support systems. This study indicates many similar but also additional needs, possibly arising from distinct interviewee group and shift in perspectives over time. It could be argued that when including the views from consulting, automation and technology companies, demand for quality assurance is highlighted more.

User perspectives were studied also in power generation industry in the US to find out why advanced process control (APC) tools had not been adopted at the same pace with other industries (Smuts & Hussey, 2011). Identified barriers included *e.g.* lack of skilled personnel to operate the systems, perceived uncertainty of cost-benefits and poor regulatory incentives. Findings of this study are rather similar though barriers from personnel time over expertise and organizational incentives over regulatory ones are highlighted. All these barriers are probably widely encountered in adoption and user acceptance of any new technologies.

Digitalization strategy of Finnish water utilities outlines six themes with stepwise goals for digitalization, themes being i) asset management, ii) information management, iii) customer service and communication, iv) digital platforms, tools and sensors, v) knowhow development and vi) digital security. Similar to this study, equipment condition monitoring with data analysis, utilizing *e.g.* pump energy consumption data, is discussed (Ikäheimo, 2020). Other than that, concrete measures for WWTPs to utilize the process data more effectively are not given in the strategy. This study aims to complement also the recommendations of the strategy with further emphasis on plant processes and real-life examples from other sectors.

6.2.2 Quality assurance and data analysis

During the course of this study, focus was gradually shifted from the conventional outlook on optimization – aiming for the most optimal operating point with evermore accurate, customized and complex tools – to simple quality assurance. After the interviews, it seemed evident that for more advanced control strategies or tools to be feasible in practice, all layers in the process control system should function in a predictable, transparent and trustworthy manner, which the operator can monitor and verify at all times. If this is not the case, the plant personnel were seen to – understandably – be too busy, cautious or sceptic to attempt any sort of fine-tuning of the systems.

Currently, dosing of different chemicals seems to be mostly based on flow measurements. Faulty flow measurements may lead to excess dosing that increases the operating costs and potentially disturbs the process. Excess chemical use will also unnecessarily increase the ecological footprint of the plant. In the interviewed plants, some flow measurements were known to be poor. For the majority of flow sensors, no quality assurance methods were in place and signal was assumed to be mostly correct.

A lot of quality assurance for process and control performance can be done with already existing tools or only minor investments. Data already collected from the process, instrumentation, equipment and automation system can be utilized for fault detection and diagnosis – and in some cases, even fault prediction and prevention. Real-life examples of quality assurance tools and benefits were identified from both Finnish and international WWTPs and process industries, see *e.g.* 5.3.1.2 (Auto-adjustment of online instruments), 5.3.2.1 (Plant-wide flow model) and 5.3.5 (Predictive maintenance with artificial neural network).

Currently, many service providers saw insufficient business potential in data analytics services and tools for WWTP process optimization. If process data is indeed utilized more in the future as stated by the interviewees, demand for tools and services can be expected to increase. Essential criteria for the tools can be assumed to include *e.g.* flexible integration with other data management systems, easy cleaning and filtering of data, modular structure to serve various sizes and types of plants, and readily available custom dashboards with relevant information for various user types.

6.2.3 Influent flow variation

Limited personnel resources can be seen as one root-cause barrier, and influent flow and quality variation as another. Influent variation induces and accelerates challenges in the further process steps, in *e.g.* sensor behavior, controller tuning and equipment dimensioning, and by also increasing the required safety margins that further limit the process efficiency. Although variation originating from water demand patterns cannot be fully eliminated, some equalization could be potentially done by reducing the network infiltration and optimizing the operation of network pumping stations and inlet pumps, such as the case study 5.3.2.4 (Integrated network and plant operation).

Small emphasis of WWTP personnel on network issues probably originates from the separate divisions (or even separate companies) operating the network and the plant. For example, reliable prediction of influent flow rate could already be utilized for optimizing the hydraulic capacity and treatment efficiency of the plant. Integration of network and plant operation towards more system-wide control would benefit especially the plant performance and operational reliability. For now, expanding the control horizon seems to attract relatively little interest from the practitioners.

6.2.4 Process equipment

To not only maintain but to improve process equipment performance, the limiting element in each control loop should be first identified. This was deemed to be at times difficult also in the larger WWTPs. In case of more complex control loops, such as aeration system, use of control loop performance monitoring discussed in case study 5.3.3.4 (Control loop performance monitoring) could be beneficial.

Similar to findings of Olsson (2012), challenges in controllability of actuating devices were seen to limit optimal operation in several utilities. Once bottlenecks of the control loop are identified, payback time or potential performance improvement of replacing the equipment before scheduled renewal could be calculated. If the behavior of the whole process control system in various flow conditions would be studied already in the design phase, with *e.g.* detailed process model as mentioned in case study 5.3.2.5 (Aeration control performance monitoring), equipment dimensioning could be more accurate and yield significant savings and improvements for operational reliability from the very beginning.

Maintenance of process equipment in interviewed WWTPs was mostly predictive and time-based. One WWTP mentioned that also condition-based maintenance of pumps had been tried but resulted into too long periods without maintenance and deteriorated the operational reliability. Although condition-based maintenance has been found as the most effective strategy for most process equipment (Hashemian, 2010), integration of new data to current maintenance system can be difficult in practice. This was also noted by *e.g.* Myhre *et al.* (2014) when the use of wireless vibration monitoring for condition-based maintenance of network pumping stations was piloted. Technical performance was proven good, but further challenges were found in how to actually utilize the condition data in the maintenance scheduling and decision-making.

6.2.5 Instrumentation

Future expectations of interviewed WWTPs include also more reliable sensors. Technology providers argued that better-quality sensors with *e.g.* self-calibration and self-diagnosis functionalities already exist, but application to wastewater utilities is limited by the tendering criteria. In the literature review, maintenance was found to be one of the largest group of expenses in any industrial plant. Evaluation of life-cycle costs from instrumentation maintenance would enable investment to better performing instrumentation in the long run, although initial investment might be higher. While importance of life-cycle costs in public procurement criteria has been increased in the EU legislation, generally accepted guidance for practical implementation seems to be lacking (von Deimling *et al.,* 2016).

Challenges in data quality and sensor maintenance could be overcome also by outsourcing the whole data collection, *i.e.* the equipment purchase, maintenance and data processing. This approach called Data as a Service (DaaS) has been recently discussed for wastewater networks (Cahn, 2020). During the interviews of this study, some interest for applying DaaS operation model also in plant processes was indicated by the instrumentation providers. Cost-efficiency of DaaS model for specific parameters could be evaluated from not only direct savings in human resources, but also potential indirect savings from the process optimization achieved with *e.g.* reliable ammonia measurement used for aeration control. This would be well justified, since several WWTPs were not utilizing the control strategy because of the sensor maintenance and reliability issues.

6.2.6 Future operation model

Operation of WWTPs is widely considered to move towards more centralized operation model. Along with increased automation, new operation model is expected to reformulate the work duties of WWTP personnel from daily operation and maintenance work to more knowledge-intensive process optimization work, attracting personnel with different skill set. This development can be assumed to increase the need for centralized data management.

While the organizational change takes place, well-defined development targets and corresponding incentive structures should be considered. At the moment, various interviewed WWTPs indicated poor or even negative incentives for development efforts. Olsson & Rieger (2012) describe similar findings of misleading incentive structures, but also successful examples for various management levels. At national level, effluent taxes are described to function well as a continuous incentive for improvement. At plant level, clearly communicated incentives with defined goals and regular feedback are mentioned.

New operation model is also expected to increase outsourcing of different plant operations. Outsourcing of maintenance could centralize both the expertise and collected data to a few specialized organizations, which might lead to higher efficiency. Even without outsourcing, reduced personnel resources in the future will probably serve as an incentive for not only automation but also condition-based maintenance. Optimal planning of maintenance might become especially useful in organizations that operate several plants with long distances in between.

Several consulting and automation companies mentioned the need to integrate the network and plant control more closely. In overall, plant-wide control was not directly highlighted by the WWTPs. However, many interviewees saw need for more predictive control and higher resources recovery in the future. For them to advance, plant- and system-wide approach can be assumed to become increasingly relevant. Optimizing the wastewater treatment is only possible if the complex interactions between influent flow, biological process, sludge treatment, energy consumption and effluent quality are taken into account. For this purpose, advanced data analysis and control tools will be evermore necessary.

6.2.7 Research

Comparing the findings to WWTP research, some distinguishing factors can be identified. The review of Corominas *et al*. (2018) discussed the quantity of peer-reviewed publications of data-based control and information extraction techniques studied for WWTPs and classified solutions by their primary objective and data analysis method. Some of the most studied subjects were plant performance prediction with ANN algorithms and process control with fuzzy logic were, with 49 and 35 related papers respectively. In addition, process control with qualitative methods and fault detection with PCA had more than 20 publications each.

Findings of this study indicate that potential benefits of predicting the plant performance have been partly identified also by the practitioners. It seems that automatic control applications dominate the research, while not receiving high interest for implementation. For fault detection, *i.e.* quality assurance, only PCA has seemed to gain some traction with more than 10 papers identified. Some interviewees stated that research on WWTP process control, including modelling, should consider more the performance and maintenance requirements – and limitations – in all layers of the process control system. For example, several studied solutions were considered economically unfeasible, if the cost of instrumentation maintenance would be taken into account.

6.3 Case studies

Based on the two distinct types of classification for the case studies, coverage of different data analysis methods and process control layers discussed in the examples is considered good. Further research efforts could be directed to a more complete review of real-life applications concerning one specific process control layer or analysis method of interest, *e.g.* system-wide or predictive control. With a more comprehensive review, example applications could be further compared to develop more detailed understanding of the practical prerequisites for successful implementation.

Feasibility of the case study applications for Finnish WWTPs has been briefly discussed in the description and should be further evaluated by means of research, piloting or implementation. Case studies of 5.3.1.2 (Auto-adjustment of online instruments) and 5.3.2.1 (Plant-wide flow model) are examples of simple quality assurance tools the WWTPs can already add to their current automation systems without major investments or external expertise. More information is easily available in the referred studies and can be further provided by the plant persons involved.

Case studies of 5.3.1.3 (Sludge dewatering optimization), 5.3.1.1 (Energy efficiency monitoring), 5.3.2.2. (Process data analysis platform) and 5.3.2.3 (Process data visualization and dashboards) present commercially viable products that have proven extensive benefits and contributed to resolving some of the challenges in WWTP operation and control identified also in this study. Despite the proven results, their applicability to the current control system and needs of a specific WWTP naturally has to be evaluated in more detail. Piloting could serve as a low-effort method to assess the benefits from these applications.

Based on the experiences in case study 5.3.3.2 Model predictive control of paste thickener), MPC applications require highly accurate process data and regular maintenance. Similar solutions have been studied in full-scale also for *e.g.* a Finnish WWTP in Kotka (Mulas *et al.,* 2016). With current resources for instrumentation and maintenance, operational reliability of MPC applications might often turn out to be insufficient. Case study of 5.3.3.3 (Neural MPC operator support system for crude oil distillation) presents an MPC application further refined with deep neural network, which is described to tolerate measurement noise and uncertainty better than a conventional MPC. Alike, Forbes *et al.* (2015) describe that in chemical engineering, research focus of advanced control strategies has shifted from further optimization efforts towards ease of installation, operation and maintenance. This might indicate that also the barriers from poor data quality might become negligible, as the data processing algorithms are further developed to suit the application needs.

Case study 5.3.3.4 (Control loop performance monitoring) presents a potential tool for major development leap in operation practices. Potential of applying this type of controller monitoring and tuning tools could be piloted in various different types of WWTPs to investigate the prevalence, common causes and achievable benefits from control loop performance monitoring. The example of 5.3.2.5 (Aeration control performance monitoring) indicates that benefits can be foreseen in not only process industries, but also in WWTPs.

Case study 5.3.3.5 (Predictive maintenance with artificial neural network) is most probably technically applicable also to WWTP process and equipment data. With such high emphasis on fault prevention, cost-benefits can be difficult to assess with a simple pilot. Economic and technical feasibility of the software could perhaps be analyzed also with historical data or simulated process model, which include examples of fault scenarios that would otherwise go unnoticed.

In addition to WWTP operation, case study 5.3.3.1 (Process diagnostics and root cause analysis) could be applied also for general research purposes with need to understand different patterns and interactions in various unit processes. In fact, the software has been used for educational purposes a different department in Aalto university. Similar to other WWTP process control research, most benefit would be achieved by analyzing both process and control equipment data of a real-life plant.

Some of the presented technologies have potential also for other application areas within the water sector. For example, the algorithm presented in case study 5.3.3.3 (Neural MPC operator support system for crude oil distillation) was mentioned to be successfully piloted also for operation of district heat networks. Similarly, it could be studied for optimizing the operation of water or sewage networks. Due to long distances and location underground, condition monitoring and maintenance of assets can be generally perceived more challenging in distribution networks than at the plant. Correspondingly, the demand for data-based predictive maintenance, such as the case study 5.3.3.5 (Predictive maintenance with artificial neural network), can be expected to be even higher.

6.4 Reliability of the study

In this study, interviews were conducted only for a limited group and type of professionals. Including a larger variety of organizations, or people in various positions in the same organizations, could have provided additional insights. For example, mechanical engineers responsible for detailed design and equipment specifications might identify alternative needs and barriers in process control and operation. Interviews with *e.g.* representatives of research, municipalities or industrial wastewater treatment sector could further complement these findings.

An interviewer with more background knowledge on the discussed topics could have come up with more specific issues and findings. Nevertheless, conducting the interviews as an outsider can be also seen to reduce conflicts of interest and yield more information especially on the personal perspectives and opinions of the interviewees.

Thematic analysis enabled assessment of also non-technical development needs and barriers, which indeed have a major influence on the process performance in practice. A major part of the material is based on personal perceptions and experience, which might not always be fully justified or correct. Yet, thematic analysis enabled generalizing these individual perceptions to in a sense sector-wide views, which determine the pace and direction of development regardless of the validity.

Information for case studies was collected from limited sources in a limited time. Most of the case study descriptions have been revised by the users or developers involved in the implementation to ensure they are accurate and updated. One case study, 5.3.2.6 (UV254-based ozone control), was included and described without consulting the implementing parties, since already the vast amount of published material was found to be of sufficient coverage. This example demonstrates how new measurements and control strategies can be effectively applied also for emerging pollutants.

Data collection, *e.g.* planned site visits, was largely affected by the meeting and travel restrictions that took place in the midst of the work. Most case studies were described based on phone interviews and written publications, which could lack some aspects that a site visit would have revealed. In this type of short review, described performance and benefits of these applications could not be separately verified from *e.g.* actual process data. However, most solutions have demonstrated measurable results confirmed by the users. Based on the initial information, most promising solutions can be studied further.

7 Conclusions

In this study, barriers, needs and incentives for developing ICA systems in WWTPs with data analytics were assessed with interviews, thematic analysis and case studies. Several identified needs and solutions were related to improving the operational reliability, but also process optimization for improving the treatment and resource efficiency.

Instrumentation, control and automation in WWTPs were reviewed from the perspective of process control system structure, components and factors affecting their performance. In addition, common control algorithms and a range of data analysis methods and example applications were reviewed.

19 interviews with personnel of WWTPs, consulting, automation and technology companies were conducted. This sample gave a good overview of municipal wastewater sector in Finland, which could be further complemented by perspectives from *e.g.* research, mechanical designers or municipalities. Perceptions of different interviewees were assessed with thematic analysis.

Objective of the interviews was to identify what needs and barriers for developing ICA with data analytics are seen by the practitioners. Results of the analysis were concluded to four themes: physical systems, digital systems, process control and operating environment. A range of development needs and barriers were identified and discussed for each theme. Thematic analysis enabled generalization of personal views and consideration of various non-technical barriers.

Key development barriers include laborious instrumentation maintenance, difficult data processing, large flow variation and lack of personnel resources. Quality assurance in all process control layers was considered often necessary, before further optimization actions can take place.

Currently, use of online nutrient sensors for control is not a widespread practice. Control strategies are most often based on flow measurements, indicating even higher relevance for their quality assurance. Higher emphasis on performance and maintenance requirements was considered necessary also for the research conducted in the field.

Energy efficiency was identified as an important incentive for developing ICA systems with data analytics. Tools and methods currently used for data processing and analysis are perceived old-fashioned and inconvenient for the purpose. It is expected that once operation model of WWTPs is shifted towards more centralized organizations and outsourced activities, the demand for more systematic data processing, higher level of automation and predictive maintenance is increased.

Another objective of the study was to review if potential solutions for Finnish WWTPs have been taken into use in other countries and process industries. For this purpose, material was identified and collected from both users and developers of the solutions to describe the potential benefits, pre-requisites and limitations in practice. In total, 14 case studies of real-life applications in WWTPs and process industries were identified and described. The case studies include also three example practices already used in Finnish WWTPs.

Feasibility of these solutions for Finnish WWTPs was initially reviewed with the support of the project steering committee. Case studies present both simple quality assurance tools and advanced control strategies the example utilities and plants are using to ensure both maximum operational reliability and optimal process performance.

Some of the case study solutions could be directly applied at Finnish WWTPs. For majority of the examples, piloting could be done to ensure feasibility of the solution for the needs of the specific WWTP at hand. Further research could be focused on assessing the cost-benefit of the described case studies and factors influencing it.

In the future, demand for predictive control and plant-wide approach will further increase the need for more advanced data analysis and control tools. Plant-wide and system-wide control will be essential for especially resource efficiency and recovery, but also operational reliability. Changes in the operating environment arising from climate change and aging infrastructure, among other changes in the society, will eventually push the development also in the wastewater sector.

References

Agachi, P.S., Cristea, M.V. (2014). Basic Process Engineering Control. Berlin, Germany: Walter De Gruyter, Inc.

Äijälä, G., Lumley, D. (2006). Integrated soft sensor model for flow control. Water Science and Technology, 53(4-5), pp. 473-482.

Al-Khalifah, M. and McMillan, G.K. (2012). Control valve versus variable-speed drive for flow control. ISA Automation Week, 60(4), pp. 42-46.

Åmand, L., Olsson, G. and Carlsson, B. (2013). Aeration control – a review. Water Science and Technology, 67(11), pp. 2374-2398.

Anderson, N.A. (1997). Instrumentation for Process Measurement and Control. 3rd edition. Boca Raton, FL: CRC Press.

Antsaklis, P.J. (1997). Intelligent Control. Encyclopedia of Electrical and Electronics Engineering, 10, pp. 493-503. Notre Dame, IN: John Wiley & Sons, Inc.

Bailey, D. and Wright, E. (2003). Practical SCADA for Industry. Burlington, MA: IDC Technologies.

Banerjee, A., Bandyopadhyay, T. and Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential? Vikalpa, 38(4), pp. 1-12.

Bauer, M., Horch, A., Xie, L., Jelali, M. and Thornhill, N. (2016). The current state of control loop performance monitoring – A survey of application in industry. Journal of Process Control, 38, pp. 1-10.

Beall, J. (2010). Improving control valve performance. Chemical Engineering, 117(10), pp. 41-45.

Bernardelli, A., Gelli, P., Manzini, A., Marsili-Libelli, S., Stancari, S. and Venier, S. (2020). Real-time Model Predictive Control of a Wastewater Treatment Plant based on Machine Learning. Water Science and Technology, in press.

Bloch, H.P. (2011) Pump Wisdom: Problem Solving for Operators and Specialists., Hoboken, NJ: John Wiley & Sons, Inc.

Boger, Z. (1992). Application of neural networks to water and wastewater treatment plant operation. ISA transactions, 31(1), pp. 25-33.

Bolton, W. (2015). Programmable Logic Controllers, 6th Edition. Oxford, Great Britain: Newnes.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), pp. 77-101.

Buckbee, G. (2009). Best practices of controller tuning. International Society of Automation. Available at: https://www.isa.org/standards-and-publications/isa-publications/intech-magazine/white-papers/ (Accessed: 29.5.2020).

Carrasco, E.F., Rodríguez, J., Punal, A., Roca, E. and Lema, J.M. (2004). Diagnosis of acidification states in an anaerobic wastewater treatment plant using a fuzzy-based expert system. Control Engineering Practice, 12(1), pp. 59-64.

Chism, N.V.N., Douglas, E. and Hilson Jr, W.J. (2008). Qualitative research basics: A guide for engineering educators. Rigorous Research in Engineering Education NSF DUE-0341127.

Corominas, Ll., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U. and Poch, M. (2018). Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. Environmental modelling & software, 106, pp. 89-103.

DeepAI. Narrow AI definition. Available at: https://deepai.org/machine-learning-glossary-and-terms/narrow-ai (Accessed: 1.6.2020).

Eerikäinen, S., Haimi, H., Mikola, A. and Vahala, R. (2020). Data analytics in control and operation of municipal wastewater treatment plants: qualitative analysis of needs and barriers. Water Science and Technology, in press.

Forbes, M.G., Patwardhan, R.S., Hamadah, H. and Gopaluni, R.B. (2015). Model predictive control in industry: challenges and opportunities. IFAC 9th International Symposium on Advanced Control of Chemical Processes, 48(8), pp. 531-538.

Friese, S., Soratto, J. and Pires, D. (2018). Carrying out a computer-aided thematic content analysis with ATLAS.ti. MMG Working Paper, 18-02.

Garcia-Alvarez, D. (2009) Fault detection using principal component analysis (PCA) in a wastewater treatment plant (WWTP). Proceedings of the International Student's Scientific Conference.

Geissdoerfer M., Savaget, P., Bocken, N.M. and Hultink, E.J. (2017). The Circular Economy – a new sustainability paradigm? Journal of Cleaner Production, 143, pp. 757-768.

Groves, K.P., Daigger, G.T., Simpkin, T.J., Redmon, D.T. and Ewing, L. (1992). Evaluation of oxygen transfer efficiency and alpha factor on a variety of diffused aeration systems. Water Environment Research, 64(5), pp. 691-698.

Haimi, H., Mulas, M., Corona, F. and Vahala, R. (2013). Data-derived soft-sensors for biological wastewater treatment plants: An overview. Environmental Modelling & Software, 47, pp. 88-107.

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

Harju, T. and Marttinen, A. (2000). Säätötekniikan koulutusmateriaali. Suomen Automaatioseura ry. Available at: https://www.automaatioseura.fi/site/assets/files/1367/pid_kirja_1-1.pdf (Accessed: 12.6.2020).

Hashemian, H.M. (2010). State-of-the-art predictive maintenance techniques. IEEE Transactions on Instrumentation and measurement, 60(1), pp. 226-236.

Hegg, B.A., Rakness, K.L. and Schultz, J.R. (1979). Evaluation of operation and maintenance factors limiting municipal wastewater treatment plant performance. Municipal Environmental Research Laboratory, U.S. Environmental Protection Agency.

IDC. (2011). The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. Available at: https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm (Accessed: 8.6.2020).

Ikäheimo A. (2020). Vesihuoltolaitosten digistrategia – portaat digitalisaation hyödyntämiseen. Vesilaitosyhdistyksen monistesarja no 59. Helsinki, Finland. Available at: https://www.vvy.fi/site/assets/files/3211/vvy_digitalisaatiostrategia_loppuraportti.pdf (Accessed: 11.6.2020).

Ingildsen, P. and Olsson, G. (2016). Smart water utilities: Complexity made simple. London, UK: IWA Publishing.

Jämsä-Jounela, S. (2011). Fault diagnosis methods and their applications in the process industry.

Jämsä-Jounela, S. (2007). Future trends in process automation. Annual Reviews in Control, 31(2), pp. 211-220.

Jokelainen, M. (2011). Jätevedenpuhdistamojen on-line analysaattorien autokalibrointi (Master's Thesis). Aalto University, Finland.

Kadlec, P., Gabrys, B. and Strandt, S. (2009). Data-driven soft sensors in the process industry. Computers & Chemical Engineering, 33(4), pp. 795-814.

Kaliman, A., Rosso, D., Leu, S. and Stenstrom, M.K. (2008). Fine-pore aeration diffusers: accelerated membrane ageing studies. Water Research, 42(1-2), pp. 467-475.

Ketonen, M. and Marttinen, A. (2001). Real-time control performance monitoring as a tool for remote maintenance service. IFAC Proceedings, 34(9), pp. 41-46.

Lumley, D. (2002). On-line instrument confirmation: how can we check that our instruments are working? Water Science and Technology, 45(4-5), pp. 469-476.

Mashford, J., Marlow, D., Tran, D. and May, R. (2011). Prediction of sewer condition grade using support vector machines. Journal of Computing in Civil Engineering, 25(4), pp. 283-290.

McHugh, M.L. (2013). The chi-square test of independence. Biochemia medica, 23(2), pp. 143-149.

McMillan, G.K. (2014). Tuning and control loop performance. 4th edition. New York, NY: Momentum Press.

Metcalf, L., Eddy, H.P. and Tchobanoglous, G. (1979). Wastewater engineering: treatment, disposal, and reuse. New York, NY: McGraw-Hill.

Mjalli, F.S., Al-Asheh, S. and Alfadala, H.E. (2007). Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. Journal of environmental management, 83(3), pp. 329-338.

Modrák, V. and Mandulák, J. (2009). Mapping Development of MES Functionalities. ICINCO 2009 - 6th International Conference on Informatics in Control, Automation and Robotics, Proceedings. 3, pp. 244-247.

Monteith, H.D., Sahely, H.R., MacLean, H.L. and Bagley, D.M. (2005). A rational procedure for estimation of greenhouse gas emissions from municipal wastewater treatment plants. Water Environment Research, 77(4), pp. 390-403.

Motiva. (2019). Energiatehokas ilmastus. Available at: https://www.motiva.fi/julkinen_sektori/vesihuoltolaitos/jateveden_puhdistus/energiatehokas_ilmastus (Accessed: 25.6.2020).

Mulas, M., Corona, F., Sirviö, J., Hyvönen, S. and Vahala, R. (2016). Full-scale implementation of an advanced control system on a biological wastewater treatment plant. 11th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems. 49(7), pp. 1163-1168.

Myhre, B., Petersen, S. and Ugarelli, R. (2014). Using wireless vibration monitoring to enable condition-based maintenance of rotating machinery in the water and wastewater industries. Procedia Engineering, 89, pp. 1397-1403.

NAPCON. (2019). White Paper: Machine Learning in NAPCON Advisor. Available at: https://www.napconsuite.com/wp-content/uploads/Machine-Learning-in-NAPCON-Advisor-White-Paper-2019-10-23.pdf (Accessed: 12.5.2020).

Nasr, M.S., Moustafa, M.A., Seif, H.A. and El Kobrosy, G. (2012). Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT. Alexandria Engineering Journal, 51(1), pp. 37-43.

Olsson, G. (2012). ICA and me – a subjective review. Water Research, 46(6), pp. 1585-1624.

Olsson, G., Carlsson, B., Comas, J., Copp, J., Gernaey, K.V., Ingildsen, P., Jeppsson, U., Kim, C., Rieger, L. and Rodriguez-Roda, I. (2014). Instrumentation, control and automation in wastewater – from London 1973 to Narbonne 2013. Water Science and Technology, 69(7), pp. 1373-1385.

Olsson, G. and Jeppsson, U. (2006). Plant-wide control: dream, necessity or reality? Water Science and Technology, 53(3), pp. 121-129.

Olsson, G., Nielsen, M., Yuan, Z., Lynggaard-Jensen, A. and Steyer, J. (2005). Instrumentation, control and automation in wastewater systems. London, UK: IWA publishing.

Oxford University Press. (2019). Definition of Big Data. Available at: https://www.lexico.com/definition/big_data (Accessed: 9.6.2020).

Patel, K.K. and Patel, S.M. (2016). Internet of things (IoT): definition, characteristics, architecture, enabling technologies, application & future challenges. International Journal of Engineering Science and Computing, 6(5), pp. 6122-6131.

Prado, T., Silva, D.M., Guilayn, W.C., Rose, T.L., Gaspar, A.M.C. and Miagostovich, M.P. (2011). Quantification and molecular characterization of enteric viruses detected in effluents from two hospital wastewater treatment plants. Water Research, 45(3), pp. 1287-1297.

Rautiainen, M. (2019). Vesihuolto heräsi kyberuhkiin – ”Uhkakuva ei ole stabiili”. Tekniikka & Talous, 29.1.2019. Available at: https://www.tekniikkatalous.fi/uutiset/vesihuolto-herasi-kyberuhkiin-uhkakuva-ei-ole-stabiili/1458f3da-a877-3496-9813-0ae21b7bc7c0 (Accessed: 12.3.2020).

Rieger, L. and Olsson, G. (2012). Why many control systems fail. Water Environment and Technology, 24(6), pp. 42-45.

Ringnér, M. (2008). What is principal component analysis? Nature biotechnology, 26(3), pp. 303-304.

Rosén, C., Rieger, L., Jeppsson, U. and Vanrolleghem, P.A. (2008). Adding realism to simulated sensors and actuators. Water Science and Technology, 57(3), pp. 337-344.

Rosso, D., Larson, L.E. and Stenstrom, M.K. (2008). Aeration of large-scale municipal wastewater treatment plants: state of the art. Water Science and Technology, 57(7), pp. 973-978.

Ruhanen, E., Kosonen, M., Kauvosaari, S. and Henriksson, B. (2018). Optimization of paste thickening at the Yara Siilinjärvi plant. Proceedings of the 21st International Seminar on Paste and Thickened Tailings, pp. 75-88. Perth, Australia: Australian Centre for Geomechanics.

Runeson, P. and Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, 14(2), pp. 131.

Runkler, T.A. (2012). Data Analytics. München, Germany: Springer Fachmedien Wiesbaden GmbH.

Santín, I., Pedret, C. and Vilanova, R. (2017). Control and Decision Strategies in Wastewater Treatment Plants for Operation Improvement. Cham, Switzerland: Springer International Publishing.

Schachtler, M. and Hubaux, N. (2016). BEAR: INNOVATIVE REGEL-STRATEGIE DER OZONUNG. Aqua & Gas, 5, pp. 84-92.

Seborg, D.E., Mellichamp, D.A., Edgar, T.F. and Doyle, F.J. (2014). Process dynamics and control. 3rd edition. New York, NY: John Wiley & Sons, Inc.

Silfverberg, P. (2017). Vesihuollon suuntaviivat 2020-luvulle. Publication series of Finnish Water Utilities Association. (44). Helsinki, Finland: Finnish Water Utilities Association.

Solon, K., Volcke, E.I., Spérandio, M. and van Loosdrecht, M.C. (2019). Resource recovery and wastewater treatment modelling. Environmental Science: Water Research & Technology, 5(4), pp. 631-642.

Stamboliska, Z., Rusiński, E. and Moczko, P. (2014). Proactive Condition Monitoring of Low-Speed Machines. Proactive condition monitoring of low-speed machines. Cham, Switzerland: Springer International Publishing.

Tchobanoglous, G., Burton, F.L. and Stensel, H.D. (1991). Wastewater engineering, treatment and reuse. 4th edition. New York, NY: McGraw-Hill.

Tolvanen, J. (2007). Life cycle energy cost savings through careful system design and pump selection. World Pumps, 2007(490), pp. 34-37.

Traore, A., Grieu, S., Puig, S., Corominas, L., Thiéry, F., Polit, M. and Colprim, J. (2005). Fuzzy control of dissolved oxygen in a sequencing batch reactor pilot plant. Chemical Engineering Journal, 111(1), pp. 13-19.

Vanrolleghem, P.A. (2003). Models in advanced wastewater treatment plant control. Proceedings Colloque Automatique et Agronomie. Montpellier, France.

Vienonen, S., Rintala, J., Orvomaa, M., Santala, E. and Maunula, M. (2012). Ilmastonmuutoksen vaikutukset ja sopeutumistarpeet vesihuollossa. Suomen ympäristö, 24/2012.

Von Sperling, M. (2007). Wastewater characteristics, treatment and disposal. London, UK: IWA publishing.

Creaco, E., Campisano, A., Fontana, N., Marini, G., Page, P.R. and Walski, T., (2019). Real time control of water distribution networks: a state-of-the-art review. Water Research, 161, pp. 517-530.

World Health Organization. (2020). Water, sanitation, hygiene and waste management for COVID-19: technical brief. Available at: https://apps.who.int/iris/bitstream/handle/10665/331305/WHO-2019-NcOV-IPC_WASH-2020.1-eng.pdf (Accessed: 24.5.2020)

Ye, N. and Chen, Q. (2001). An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. Quality and Reliability Engineering International, 17(2), pp. 105-112.

Yetilmezsoy, K., Turkdogan, F.I., Temizel, I. and Gunay, A. (2013). Development of ann-based models to predict biogas and methane productions in anaerobic treatment of molasses wastewater. International journal of green energy, 10(9), pp. 885-907.

Appendices

Appendix 1. Thematic analysis codings and their groundedness

Theme 1: Physical systems

Now		Needs		Barriers		Future	
Condition monitoring with manual parameter comparison	7	Using existing data for condition-based monitoring	4	Instrumentation maintenance need too high	12	Sensor self-diagnostics	4
Schedule-based maintenance	6	Easy condition diagnostics with current systems	3	Poor dimensioning of equipment	9	Equipment fault prediction from big data	4
Manual sensor adjustment with lab-results	5	Better planning of maintenance	3	Sensor accuracy & reliability	9	Sensors become more reliable	3
Maintenance software coming into use	3	Dimensioning of equipment with historical data	2	Sensor calibration problems	4	Sensors with more functionalities	3
Condition monitoring from alarms & observations	2	Reliable sludge blanket level measurement	2	Faulty or poorly installed actuators	4	Instrumentation will be added	2
Maintenance outsourced	2	Predictive alarms for maintenance	2	Condition monitoring data exists, but is not used	3		
Maintenance not outsourced	2			Maintenance providers protect their own business	2		
Maintenance partly outsourced	2			User expertise of instrumentation	2		

Theme 2: Digital systems

Now		Needs		Barriers		Future	
Data processed and stored in excels	4	Better understanding of action impacts	4	Difficult & laborious data processing	9	Focus on energy efficiency	8
A lot of data is collected, but not utilized	3	Better understanding of process state	3	Data coverage & quality not sufficient	9	Data analysis will be used more	8
Energy efficiency monitored at utility level	3	Data quality assurance	3	ICT systems not designed with process expertise	8	Data management is centralized	7
Trends followed in real-time	2	Quick/visual access to relevant info	3	Integration of software	6	Condition monitoring	4
Cloud access to automation	2	Energy efficiency monitoring	3	User acceptance	4	More demand for flexibility & modularity	4
Data transferred from excels to QlikView	2	Gradual changes, leakages, malfunctions	2	Insufficient business potential	3	Alarm monitoring	3
				Clear targets missing	2	Data quality assurance	3
				Common KPI’s		Competed markets	2

Theme 3: Process control

Needs		Barriers		Future	
Controller tuning	5	Large flows & variations	5	Chemical & energy consumption optimization	10
Inlet pumping optimization	2	Instable control loops	4	Predictive control	9
Optimizing air pressure level	1	User & owner expertise	4	Increased automation	5
Decrease of personnel work	1	Instrumentation malfunctioning	3	Advanced Process Control	4
Separate project, tender criteria for expertise	1	Old automation system	3	Outsourcing will increase the role of fieldbus	2
N2O optimization with NH4 control	1	Slow response time	2	Remote control, occasional site visits	2

Theme 4: Operating environment

Barriers		Future	
Lack of time and resources of utility personnel	15	Satellite model will push development	9
Lack of client demand	13	New staff generation with new skills & requirements for digital tools	7
Purchasing & tendering criteria	6	More outsourcing	4
Cost-benefit of new tools not sufficient	5	Higher treatment requirements	3
No incentive for cost & energy efficiency from the owner	5	CO2 & LCA requirements extend to production chain e.g. chemicals	1
Tech providers not included early enough in the planning	1		
Sticking to familiar technologies & methods	1		

Appendix 2. Interview questions for WWTPs (in Finnish)

Jätevedenpuhdistamot

Prosessinohjaus & monitorointi

- Käytetyt prosessimittaukset
- Käytetyt ajotavat
- Ohjausnäkymät & avaintunnusluvut

- Mitä manuaalista ohjausta tyypillisesti tehdään?
- Mitkä prosessivaiheet vaativat eniten huomiota/työtä operaattorilta?
- Mitä kehitettävää näette prosessinohjauksessa ja prosessin tilan monitoroinnissa?

Sensorit ja analysaattorit

- Miten seuraatte eri mittalaitteiden kuntoa?
- Onko teillä haasteita mittalaitteiden toimivuudessa?
- Onko teillä haasteita mittalaitteiden kunnossapidossa & kalibroinnissa?

- Mitä kehitettävää näette instrumentoinnin hallinnassa?

Toimilaitteet

- Miten seuraatte eri toimilaitteiden kuntoa?
- Onko teillä haasteita toimilaitteiden toimivuudessa (toiminta-alue & säätö)?
- Onko teillä haasteita toimilaitteiden kunnossapidossa?

- Mitä kehitettävää näette toimilaitteiden hallinnassa?

Mittausdatan laatu & käsittely

- Miten seuraatte (instrumentaation) mittausdatan laatua?
- Mihin mittausdata tallennetaan?
- Kerääkö joku muu toimija (esim. laitetoimittajat) myös jotain prosessi/laitedataa?
- Mitä raportointia mittausdatan pohjalta tehdään?
- Miten mittausdata valmistellaan raportteihin?

- Mitä kehitettävää näette mittausdatan käsittelyssä & laadunvarmistuksessa?

Muu tiedon jalostaminen

- Mihin muuhun prosessista kerättyä dataa hyödynnetään?
- Miten kerättyä dataa visualisoidaan?
- Miten energiankulutusta seurataan?

- Mitä kehitettävää näette mittausdatan hyödyntämisessä & raportoinnissa?

- Mitä haasteita uskotte uusien analytiikkamenetelmien ratkaisevan?

Kehityssuunnitelmat

- Onko teillä suunnitelmia kehittää prosessinohjausta tai ajotapoja?
- Onko teillä suunnitelmia kehittää/lisätä prosessi- tai toimilaitemittauksia?
- Onko teillä kehitysideoita tai -suunnitelmia liittyen kerätyn datan hyödyntämiseen?
- Onko teillä kehitysideoita tai -suunnitelmia liittyen kerätyn datan raportointiin?
- Onko teillä kehitysideoita tai -suunnitelmia liittyen kerätyn datan visualisointiin?

- Miten uskotte prosessiautomaation, monitoroinnin ja analytiikan kehittyvän laitoksellanne seuraavan 20 vuoden aikana? Entä seuraavan 5 vuoden aikana?

Appendix 3. Interview questions for companies (in Finnish)

Yritykset

- Mitä kehitettävää näette jätevedenpuhdistamoilla…

  … tulopumppauksen ohjauksessa?
  … ilmastuksen ohjauksessa?
  … muiden osaprosessien ohjauksessa?

  … instrumentoinnissa?
  … mittalaitteiden kunnonvalvonnassa?
  … toimilaitteiden kunnonvalvonnassa?

  … mittausdatan laadunvarmistuksessa?
  … mittausdatan hyödyntämisessä?
  … energiatehokkuudessa?

- Mitä pullonkauloja näette puhdistamojen prosessiautomaation kehittämisessä?
- Mitä pullonkauloja näette puhdistamojen operointityön kehittämisessä?
- Mitä pullonkauloja näette puhdistamojen kunnossapidon kehittämisessä?

- Mitä kannustimia näette prosessidatan tehokkaammalle hyödyntämiselle?
- Millaisia säätö-, monitorointi- tai analytiikkaratkaisuja olette selvittäneet tai kehittäneet?
- Millaisia säätö-, monitorointi- tai analytiikkaratkaisuja myytte muille teollisuudenaloille?

- Miten uskotte jätevedenpuhdistamojen prosessinohjauksen, monitoroinnin ja mittausdatan analytiikan kehittyvän seuraavan 20 vuoden aikana?
- Entä seuraavan 5 vuoden aikana?