2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14247

ROADMAP TOWARDS SMART WASTEWATER TREATMENT FACILITIES

Daniel Aguado[1], Henri Haimi[2], Michela Mulas[3] and Francesco Corona[4]

[1] CALAGUA – Unidad Mixta UV-UPV, Institut Universitari d'Investigació d'Enginyeria de l'Aigua i Medi Ambient – IIAMA, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain.
[2] FCG Finnish Consulting Group Ltd, Osmontie 34, P.O. Box 950, FI-00601 Helsinki, Finland
[3] Department of Teleinformatics Engineering, Federal University of Ceará, Brazil
[4] School of Chemical Engineering, Aalto University, Finland.

[1] daaggar@hma.upv.es, [2] Henri.Haimi@fcg.fi, [3] michela.mulas@ufc.br, [4] francesco.corona@aalto.fi.

Abstract

To protect human health and natural ecosystems, wastewater treatment plants (WWTPs) have been traditionally designed to remove pollutants from wastewater. With remarkable success, WWTPs continuously adapt to increasingly stringent discharge limits. Nowadays, municipal wastewater treatment facilities are facing a double transition and new challenges: On the one hand, the transition towards a sustainable and circular water economy, in which resource recovery from wastewater (water, energy, and nutrient recovery) plays a fundamental role for its effective implementation. On the other hand, the digital transition, which aims at making the operation of these facilities smart, will undoubtedly have a synergistic effect together with the paradigm shift towards the effective implementation of a circular water economy.

To make our current facilities smart, there is a growing interest in finding the way to convert the collected process data into intelligent actions for improving their operation. This is not an easy task for many reasons:

- the harsh environment in which the instrumentation must work (corrosive, sludgy, biofilm formation with biological activity…),
- almost complete absence of metadata that would make it easy the interpretation of the process data that it is being collected and that would enable its future use,
- the almost complete absence of automated data quality assurance, required to avoid "garbage in – garbage out"
- the ever-increasing number of available process sensors (data overload), that must be properly processed and made easily available for further use to make them useful
- large amounts of data are collected and stored in databases but not wisely used, thus, resulting in data graveyards,
- the excessive cost of nutrient and organic matter sensors/analysers which moreover are labour maintenance intensive, fact that restrict their availability to the range of large facilities, thus, they are not usually available for small size facilities (which are the vast majority),
- the intelligent sensors and data-driven models must be maintainable by the plant workers (not by Data scientists),
- the lack of process expertise in the development of the artificial intelligent tools,
- plant operators are often accustomed to their operational routines and, therefore, cultural change is needed in the organization for successful digital transition and adopting new intelligent tools.

The progress in computing capabilities together with the large amount of collected process data in WWTPs have created the perfect storm for the machine learning boom we are observing, but all the aforementioned issues can make the incredible digital transition opportunity that exists today completely lost. In an attempt to avoid this disaster, this paper tries to shed light on the path towards increasing the value of the large amount of data that nowadays are being collected in WWTPs and WWRFs. Thus, digital transition could be safely embraced and the enormous potential of data analytics fully exploited, enabling it to play an essential role in the future automation and operation of our municipal facilities.

Keywords
Control, data analytics, digitalization, metadata, wastewater treatment plant, water resource recovery facility.

1 INTRODUCTION

The presence of organic compounds, nutrients, solids, pathogens, and other pollutants in wastewater made it to be traditionally considered an undesired waste. Due to the impact of these pollutants on the environment, wastewater must undergo energy-intensive treatment to remove them, prior to its discharge into natural aquatic environments. To protect human health and natural ecosystems, Wastewater Treatment Plants (WWTPs) have been designed to remove the pollutants contained in wastewater (Metcalf & Eddy, 2013).

With remarkable success WWTPs have fulfilled the tasks and, over the years, they relentlessly evolved to adapt to the increasingly stringent discharge limits. In the last decade, the transition towards a sustainable and circular water economy (CE) has stimulated a paradigm shift that is transforming the perception of sewage- and wastewater from an undesirable waste into a product that is rich in valuable resources to be recovered (Guest *et al.*, 2009), such as reusable water itself, nutrients and energy. From this perspective, a CE process aims at improving productivity of resources by keeping products, materials and infrastructure in use for longer than in the traditional linear 'take-make-consume-waste' economic model. CE has been promoted by policymakers (e.g., European Commission 2020) and adopted by many industries (Mhatre *et al.*, 2021). To reflect the increased focus on resource recovery in wastewater treatment (MacDonald & Crawford 2017) many WWTPs have been rebranded as Water Resource Recovery Facilities (WRRFs) as long as they incorporate some type of resource recovery process. However, the introduction of new process units for recovery and additional operational goals resulting from the shift from WWTPs to WRRFs renders the operation of the facilities more challenging.

The structural changes and new goals emphasize the need of efficient process monitoring and control tools, as well as multi-objective process optimization strategies (Arnell *et al.*, 2017; Solon *et al.*, 2019). At the same time, these technologies are key enablers of a second transition that municipal WWTPs are facing today: the digital transition. This transition aims at making WWTPs and WRRFs smart and their operations intelligent and efficient. According to Ingildsen and Olsson (2016) a smart water utility operates according to an optimal decision-making management, deployed at all process levels. Primary enablers of smart operations are online water quality and quantity sensors and process actuators and control levers. The use of these devices and their inclusion in computational solutions for designing operational strategies that account for the full water cycle, from water intake to water effluent, must be systematic and pervasive. The ultimate goal is a system-level management of the operations that is autonomously able of ensuring adequate water quality and quantity, with a minimum consumption of energy and materials, and minimum environmental impact.

Yet, today, most of wastewater treatment facilities operate using only basic sensor arrangements and the coupling between sensors and actuators is limited to simplified control schemes, if not ad hoc rules. Already at this level, the potential benefits offered by instrumentation, control, and automation (ICA) technologies remains largely under-utilized. Ingildsen and Olsson (2016) estimate that the use and exploitation of ICA technologies could, however, improve the capacity of a biological nutrient removal WWTP by 10–30% in the short term, and by 20–50% in the mid- and long-term (10–20 years from now). Moreover, the opportunities resulting from the availability of modern sensor technologies and more instrumented facilities are still to be discovered and used to support plant operation and management. Efficient statistical and data-driven models can be used to explore and model the wealth of information available in process data. This could lead to the discovery of new phenomena and would enable the development of process models and control strategies.

Although recommended and beneficial, it is essential to understand that having many installed sensors is not sufficient for a WWTPs and WRRFs to be considered smart facilities. To benefit from process data, raw measurements must be firstly processed and then made easily available for further use at different levels of granularity by process operators, engineers, and plant managers. The data collected from sensors must be (a) digested into process insight, the resulting knowledge is then (b) used to develop predictive models that can help characterise the state of the plant and its units, before it is eventually (c) embedded in automatic control structures where it is transformed into optimal control actions aiming at driving plant operations: The technology involved along this workflow can be understood as a combination of statistics, optimisation, and control theory or, as the general public oftentimes denotes it, machine learning and artificial intelligence.

Designing, developing, and then effectively using these technologies for operating wastewater treatment facilities is not a straightforward task. To start with, it is well-known that the quality and reliability of sensors' signals is affected by the harsh environment in which the instruments operate (corrosive and sludgy environment, as well as biofilm formation with biological activity are commonplace). In addition, the biochemical processes and the hydraulics occurring in wastewater treatment plants are complicated, highly nonlinear, and only partially understood from a mechanistic point of view. The portrait is completed when adding the challenges of controlling monolithic facilities perpetually operated in transient conditions. With such complex systems, when large amounts of data are collected, but simply stored in databases and not modelled to achieve specific process monitoring and control tasks, the risk of forming data graveyards is not negligible. This is a risk that must be confronted, mitigating actions must be taken to minimise it, and a clear roadmap to harness the full potential made available by modern data-based technologies laid down.

Our starting point is the recognition that, regretfully, in most wastewater treatment facilities, data quality and data analysis procedures are still rudimental, if existent at all.

To reverse this situation, several limitations need to be overcome and counteractions taken:

- the almost complete absence of automated data quality assurance, required to avoid "garbage in – garbage out"
- the almost complete absence of metadata collection that would make it easy the interpretation of the process data that it is being collected and that would enable its future use,
- the excessive cost of nutrient and organic matter sensors/analysers which moreover are labour maintenance intensive, fact that restrict their availability to the range of large facilities,
- the inclusion of process expertise in the development of the artificial intelligent tools and make them understandable by WWTPs' personnel to favour their short-term adoption.

Importantly, it is believed that a successful transition in the direction of digitalization and the adoption of new computational tools must be achieved through a cultural change in the organization and the management of the facilities and the resources available for their operation. This has been recognized by the many water utilities that have accepted the challenge and developed explicit roadmaps to be implemented step by step. In this regard, the human factor plays a central role in catalysing the digital transition. Because the primary users of smart applications are frontline staff responsible for daily operational decisions, it is important to involve them in developing these tools and to ensure that new information is presented in a user-friendly format and is actionable for their needs (Torfs *et al.*, 2022).

Adequate training of end-users is also a necessity to enable the successful adoption of new tools. Eerikäinen *et al.* (2020) found that employees of WWTPs are expecting next generation of digital tools for process data analysis. They emphasize that those tools should combine competences of both automation providers and wastewater process experts with a thorough understanding of treatment phenomena. The challenge for this is that the number of experts who have adequate skills and experience in both data techniques and treatment processes is limited. Therefore, it is expected that universities also enable and encourage learning this kind of mix of technologies in their curricula. That would promote new business opportunities in the form of, for instance, machine learning applications tailored for WWTPs, as well as services that facilitate the introduction of new tools with in-house data methodology understanding of new generation of workforce members. In fact, it also is of primary importance that universities provide students with sufficient mathematical education as that is the backbone required for learning advanced data analysis methods.

Even though the majority of smart water products that have been traditionally available on the market are targeted for water and wastewater network operations, some companies have started offering data quality management solutions and advanced modelling services which are specifically tailored for WWTPs (Corominas *et al.*, 2018). In addition, a large number of the computational tools that could be used to support the operations of WWTPs have been designed and developed in the academia. While many of these tools have mainly kept their research-oriented nature (Haimi *et al.*, 2013; Corominas *et al.*, 2018; Newhart *et al.*, 2019), it is important to note that successful smart water companies have often strong connections with university and routinely adopt ideas from novel academic research to add functionalities to their products.

Academia also plays a highly important role in showcasing the benefits of advanced data mining, modelling, and control systems to WWTP decision-makers and other stakeholders. To reach the multiple operational goals set when operating a modern facility, it is crucial that end-users are offered the opportunity to clearly appreciate the potential of advanced monitoring and control systems from an economic and safety of people, the environment as well as the equipment: As a driver of the cultural change towards efficient utilization of measured data for improved operation, that would actuate inclusion, for instance, of advanced monitoring and control systems in public procurements of water utilities when upgrading plants. In addition, procurements of smart systems for WWTPs are challenging to master, for example, because individual application solutions should be integrated into existing and future software platform solutions (Müller-Czygan 2020). Nevertheless, procuring advanced systems would also act as a driver for an increased competence of automation companies and consultants providing services for water utilities.

To avoid that the extraordinary digital transition opportunity that exists today be completely lost due to the aforementioned issues, this paper tries to shed light on the path towards increasing the value of the large amount of data that are being collected in current WWTPs and WRRFs making it possible to leverage the machine learning boom. It serves also as a roadmap to ease the intelligent automation of these facilities, thus paving the way to their digitization. Thus, digital transition could be safely embraced and the enormous potential of data analytics fully exploited, enabling it to play an essential role in the future automation and operation of our municipal facilities.

2 MAKING WASTEWATER TREATMENT FACILITIES INTELLIGENT

2.1 The digital transition in WWTPs

The digital transition aims at making the operation of wastewater facilities intelligent, being on-line monitoring, real-time process control and automation essential parts of the digitalization of wastewater facilities. Thus, on-line measurement data from the process are the foundation of a smart facility. To be profitable and allow process control and decision-making, the frequency of measurement of a variable (i.e., the temporal resolution of the sensor) should make it possible to capture its dynamics, i.e., the phenomena of interest is captured by the sensor (e.g., from lower to higher frequency needed: suspended solids in the reactor, influent flow rate, dissolved oxygen concentration in the aerobic reactor).

A wide range of physical and chemical parameters relevant to operation of WWTPs can be measured continuously or semi-continuously with commonly used sensors and analysers (Ingildsen & Olsson 2016). Development of novel instruments is still taking place, for instance, for measurement of organic carbon, metals, and emerging contaminants in wastewaters, for more affordable and improved nutrient sensors (Zhang *et al.*, 2020), and for enabling continuous monitoring of some parameters relevant to optimization of anaerobic digestion processes (e.g., measuring individual volatile fatty acid species) (Jimenez *et al.*, 2015).

Figure 1 shows a scheme of the water line and the sludge line including the typical sensors that could usually be deployed within a standard wastewater facility. Please note that mainly process variables (also known as secondary variables) are recorded which are easy-to-measure with relatively cheap sensors. Less frequently available (in small and medium-size WWTPs) quality variables (also known as primary variables) - like nutrients and organic matter which are measured with expensive sensors/analyzers.

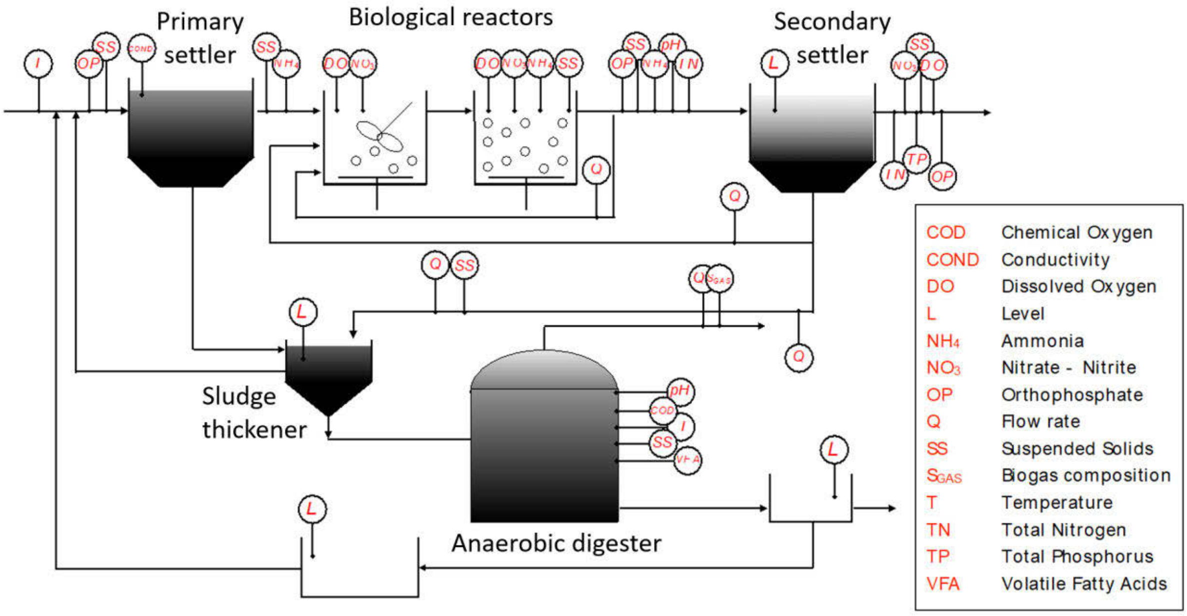

Figure 1. Typical sensors found in the water and sludge lines of a conventional wastewater treatment facility.

Although there is a wide variety of sensors available, practical challenges still exist: operators of treatment facilities find, for instance, quality of measured data and laborious maintenance needs of instrumentation as barriers for efficient use (Eerikäinen *et al.*, 2020). Particularly, there are simpler sensors (e.g. dissolved oxygen, pH, flow, level) that have been proven to be robust, sufficiently accurate and need minimum maintenance, but challenges concern more advanced sensors (e.g. ammonium, nitrate, nitrite and phosphate) that have been found to be less reliable, preventing a wider application of the advanced control algorithms that are dependent on these sensors (Yuan *et al.*, 2019). Even there are a number of undesired sensor states: excessive drift, shift, fixed value, complete failure, wrong gain and isolated fault (a single incorrect value) (Therrien *et al.*, 2020).

Recent instrumentation surveys in Swedish and Danish WWTPs (Åmand *et al.*, 2017; Nilsson & Andersson 2018, respectively) indicate that less than 70 % of investigated facilities in both countries had written instructions for quality control of at least part of the instruments. However, because large amounts of measured data are generated in modern WWTPs, in addition to conventional sensor maintenance efforts, incorporation of techniques for fast detection and diagnosis of faults are needed to guarantee sufficient data quality (Corominas *et al.*, 2018). Another challenge that is also emphasized by increasing amounts of measured data in WWTPs, is that stored sensor data is often not augmented with adequate meta-data i.e. descriptive information, which hinders the use of historical data to address future problems (IWA 2021a).

In our experience, efforts for keeping instrumentation operational varies quite much between facilities even if their size and personnel numbers would be similar. Also other reasons than laborious maintenance and incomplete technology have a crucial impact on this: if operators do not understand the value of certain measurements for process operation, they easily loose interest in those sensors (Olsson & Ingildsen 2018). Therefore, implementation of each sensor should be justifiable, and their purpose and received benefits need to be thoroughly explained to employees who operate the process. This is of key importance for successful instrumentation and control design projects. Moreover, adequately maintained sensors are vital for smart wastewater treatment facilities.

2.2 Stages to transform raw data into actionable insight

2.2.1 From raw data to quality checked data fitted for purpose

As can be seen in Figure 2, the first tough challenge for ICA in a wastewater facility is the harsh environment in which the instrumentation has to work (corrosive, sludgy, biofilm formation with biological activity…), which directly impacts the quality and reliability of a sensors' signal. Surprisingly, despite this is evident and known, data quality and data analysis are essentially non-existent in most wastewater facilities around the world. Therefore, nowadays with numerous sensors installed in many wastewater facilities, a huge amount of data is being collected that is neither analysed nor utilized, resulting in data graveyards (Corominas *et al.*, 2018). To make the situation even worse, the almost complete absence of metadata (that would enable the correct interpretation of the collected data) prevents its future use.

(a) **(b)** **(c)**

Figure 2. Pictures of a suspended solids probe installed in the buffer tank of a WWTP: (a) new probe (b) freshly removed from the buffer tank (c) after cleaning with water jet. This illustrates the harsh environment in which the instrumentation has to work (corrosive, sludgy, biofilm formation with biological activity…).

Since the on-line measurements are the basis of a smart facility, appropriate data quality check is vital to make it possible the exploitation of the ICA. It is evident and of paramount importance for process operation that every on-line measurement should have to be quality checked prior to be used in a control loop.

The raw collected data must be properly processed and made easily available for further use to make them truly useful. However, there is no general standard on how the on-line quality check should be done (Ingildsen & Olsson 2016). Figure 3, shows different pre-processing steps that can be applied to the collected data (raw data) to check its quality and improve its information content. Depending on each particular case, some steps or others can be applied. For example, for automatic control purposes human intervention should be not required, while for process modelling purposes it will be a quite important component: starting from the initial visualization of the data by the wastewater treatment expert to gain an overview and feeling of the plotted data to the application of mass balances to detect inconsistencies in the collected data.

A relatively easy to implement automated data quality assurance workflow would include the following stages: **raw data -> sanity checks -> outlier detection -> pre-processing (missing data imputation, scaling and filtering) -> Data fit for purpose**. As can be seen in Figure 3, the end of the pre-processing steps results in “data fit for purpose”, which is quality-checked data that has been reshaped into a better form for further analysis by the methods shown in Figure 4.

The pre-processed on-line measurement can be used in a control loop (i.e., used as input of the control algorithm), either directly (e.g., the oxygen concentration that is used to regulate the amount of air supplied by the blowers to the aerobic reactor) or indirectly via the on-line estimation of another variable using for example a data-driven model (e.g. a soft-sensor based on an artificial neural network or on a support vector machine or on a partial least squares model....) or simply multiplied by another variable. There are several signals of interest in the wastewater treatment context that are obtained by a simple combination of multiple signals (e.g., the organic mass flow (Q x COD), the solids mass flow (SS x Q),...).

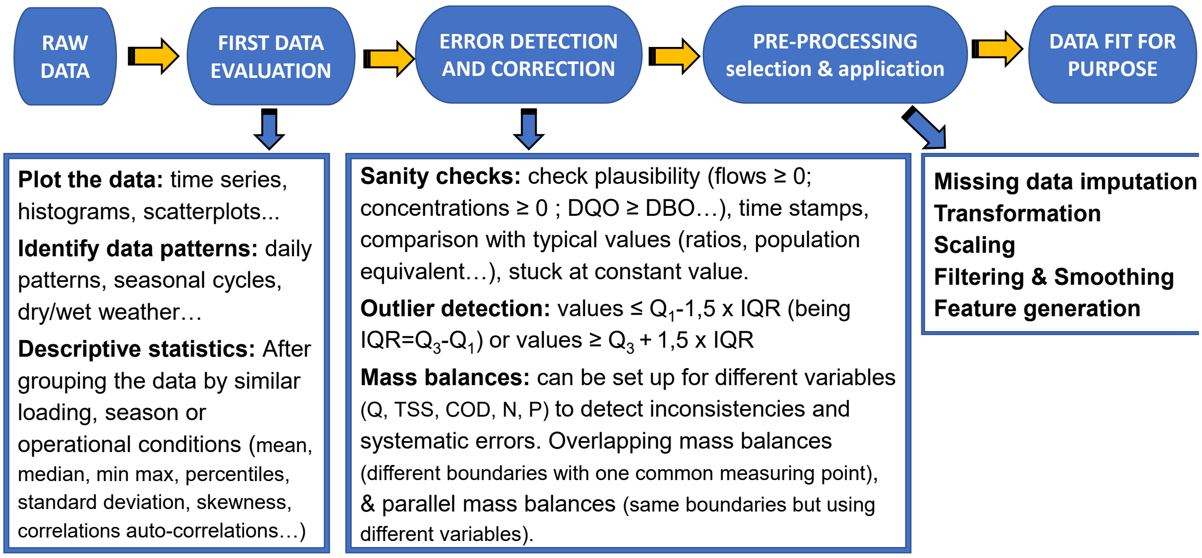

Figure 3. Pre-processing steps that can be applied on raw data to check its quality and improve its information content. Selecting appropriate steps, a data pipeline can be developed for each case.

2.2.2 From good quality data to data-driven models

From the perspective of data processing, the challenge is to effectively use the data that can be acquired for a modern wastewater treatment plant for developing models useful for achieving reuse, resource and energy recovery, and minimal carbon and greenhouse gas footprint. Within machine learning and artificial intelligence, the task is approached mainly as supervised and unsupervised learning.

In a typical supervised learning scenario, the output, called response or dependent variable, is a quantitative (such as "difficult-to-measure-variables") or categorical (such as effluent quality or process indicator) that we wish to predict. The output type leads to a naming convention for the prediction tasks: regression (to predict quantitative outputs) and classification (to predict qualitative outputs). In both situations, a set of measurements data from the sensors and the laboratory/experimental analysis is available. These are the features, also called predictors, inputs variables or independent variables. We have a training set of data, in which we observe the outcome and feature measurements for a set of samples. Using these data a prediction model is built which will enable to predict the outcome for new unseen test samples. Unsupervised learning, often performed as part of the exploratory data analysis, refers to a situation in which for every feature, we observe a vector of measurements but no associated response. The goal is to directly infer some interesting properties of the process from the available features without the help of an associated response variable providing correct answers or degree-of-error for each observation (Hastie *et al.*, 2017).

Figure 4 schematically summarises the methodologies applied to the pre-processed data within the wastewater treatment applications. Most studies focused on regression problems to predict and monitor the output variables based on a given number of historical observations, as summarized by Haimi *et al.* (2013) and lately by Ching *et al.* (2021). Recent machine learning algorithms for WWTP classification problems include random forest, tree-based algorithms, support vector machine and the comparison of different methods in various applications (as for instance in Guo *et al.*, 2015, Nourani *et al.*, 2018, Wang *et al.*, 2021 and 2022). Unsupervised methods have been utilized for determining changes in process variables and for anomaly detection (Corominas *et al.*, 2018). A recent example of clustering application is Xu *et al.* (2021) for optimizing the processes configuration of full-scale WWTP predesign through an integrated strategy consisting of t-distributed stochastic neighbour embedding (t-SNE) and deep neural networks (DNNs).

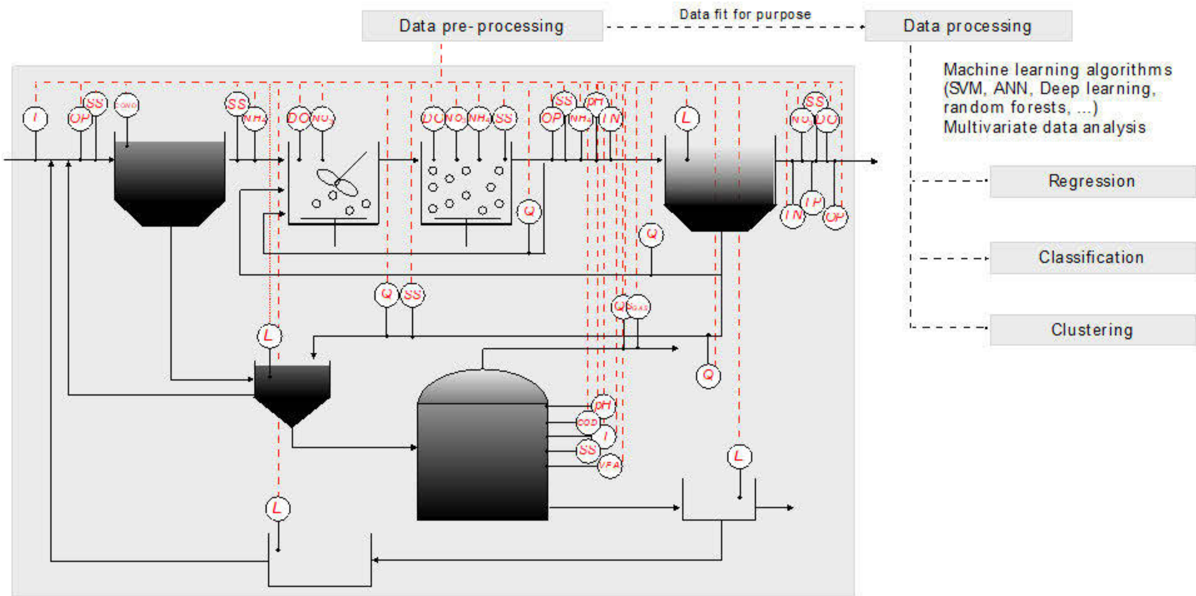

Figure 4. Data-driven techniques that can be applied to the quality-checked data that is fit for purpose.

2.2.3 From good quality data to WWTP process control

The development and implementation of data-driven models and automatic process controllers will make the collected data truly useful. Real-time process control and automation can significantly contribute to the optimization of different processes (chemical dosing, pumping, aeration, energy consumption) and can take care of the repetitive low-level tasks necessary to keep the facility running. It should be highlighted that sound control actions require good quality data. The increasingly complex control strategies (required to efficiently operate the increasingly complex WWTPs) will fail more often the lower the quality of the measured process data. There are control loops based on variables that are directly measured from the process (e.g., dissolved oxygen concentration) or on the output of a data-driven model (e.g., soft-sensor, or a data-driven predictive model).

From the perspective of automatic control, the challenge is a system-level technology to optimally operate WWTPs in a management of wastewater that includes reuse, resource and energy recovery, and minimal carbon and greenhouse gas footprint. Ideally, facilities and recipients of recovered resources must be operated as interacting entities, to satisfy operational objectives that aim at matching demand and resources, while always ensuring safety of people and equipment, environmental permits, and sustainability boundaries.

Nowadays, there is no standardised control solution for WWTPs that aim at being operated as WRRFs. General solutions for control algorithms are useless because they are case-specific, and the algorithms require tunning on a regular basis. To tune them, skilled personnel are needed. It is still rare to encounter WWTPs in which downstream operations are accounted for to define the planning over the recovery of energy and materials: When the operational goal is dictated only by disposal permits that are unaware of the fate of recoverable resources, plant management cannot be expected run WWTPs as bio-refineries within the water chain. Yet, there exists practical evidence that wastewater treatment facilities often times have the technological capability and flexibility to be operated towards these objectives.

In a less heuristic approach, the goals of resource recovery, subject to neutral- or positive-energy constraints and minimal air emission footprint, must be formulated as explicit control objectives to be achieved by manipulating material and energy fluxes across WWTPs, in response to downstream needs and upstream conditions. By coordinating the right synergy between treatment plants on the one hand, and the receivers of recovered resources on the other, flexibility between treatment, reuse, and resource recovery can be largely achieved by making the best use of existing facilities, with little-to-none capital investments for upgrading or retrofitting. In this framework, plant-wide planning must be designed around controllers that must be capable to (a) deal with complex and uncertain unit- and plant-wide dynamics; (b) satisfy the constraints given by current and forthcoming permits; (c) manage the production of reuse water and recovered resources; (d) enforce energetic neutrality, if not positivity; and (e) minimise the environmental impact of the plant. Moreover, controllers must be capable to determine an optimal equilibrium when conflicting objectives are at stake.

The importance of optimising the operations of a WWTP using automatic control is largely recognised, both from an environmental and an economical viewpoint (Ingildsen & Olsson 2016). Efforts made to reduce energy use by replacing specific devices (Daw *et al.*, 2012) focused on the control of aerobic processes (Mulas *et al.*, 2015, Stentoff, 2020], but also on structural changes in process configuration (Sarpong and Gude, 2020). Recently, Neto et al. (2022) have shown that conventional activated sludge processes can be optimised with respect to non-conventional objectives, like quality and quantity of released water. While these results are rooted on the availability of dynamic models (Henze 2020) that enable the definition of advanced control strategies, progress remains to be made to integrate the mechanistic models in the determination of the technological margins for controlling WWTPs as WRRFs, and on how to safely complement them with empirical counterparts. There exist critical knowledge gaps and open questions also regarding the feasibility of full energy and nutrient recovery, and on the thresholds as of when these practices are safe. It is also unclear what are the conditions for environmental neutrality. These questions are at the core of the research.

A key pathway towards a smart management of wastewater treatment systems and the foundation for a sustainable water management is built upon the development of information and decision-support systems. In a model-based approach, state-space process models derived from first-principles, either mechanistic or statistical, are learn using process measurements, then analysed and finally embedded in receding-horizon controllers for optimal decision-making and planning. The integration of explicit operational policies for supervising smart operations, on a high-level, and the deployment of regulatory actions with low-level controllers can be developed according to a general control architecture consisting, in its basic formulation, of a dynamic process model, a state estimator, and a predictive controller.

In Figure 5, the integration of an optimal controller is illustrated on a conventional activated sludge process consisting of a certain number of actuators (in this case, 13), operated using low-level PI controllers whose set-points are dynamically determined using a predictive controller. The controller, in turn, determines the set-points as the decision variables that optimize a user-defined operational objective over a fixed time-horizon, subject to the dynamics of the plant (a process model) and to a number of technological and operational constraints. As the controller requires knowledge of the current state of the process, this information is reconstructed for sensor measurements (here, 14) by a state estimator, again based on a dynamical model of the process. In the example, the process model used to represent the plant is the BSM1.

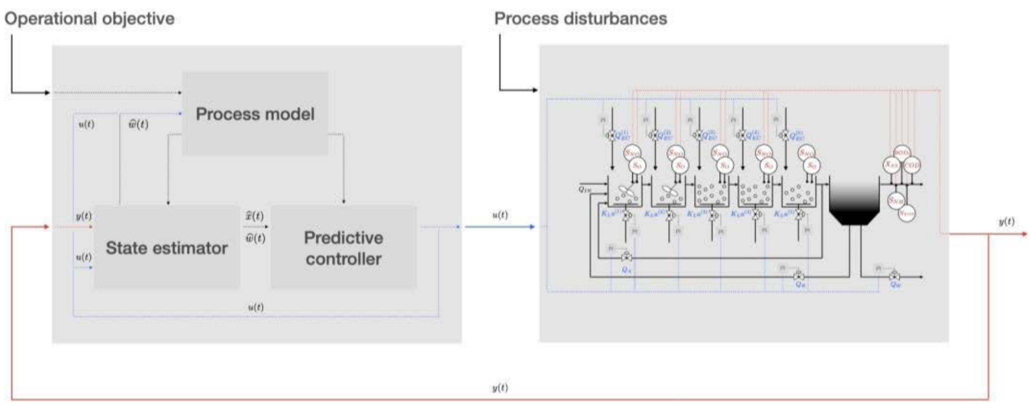

Figure 5. A schematic of a model-based predictive controller for a conventional activated sludge process. The actions of the actuators (in blue) are defined using PI controllers whose setpoints are determined using a predictive controller. Sensor measurements (in red) are used by a state estimator to determine the current state of the process. Both the controller and the state estimated are based on a mechanistic process model.

In Figure 6, the architecture is instantiated to use a model-predictive controller (MPC) and a moving-horizon estimator (MHE). The process model is explicitly described by a state-space formulation, with a set of dynamic models (the 145 differential equations in the BSM1) and measurement models (14 algebraic equations defining the measurements as function of the process state, again based on the BSM1). From a process perspective the actuators allow for the control of aeration and the addition of external carbon sources to the biological reactors, and the control of sludge recycles and removal in the secondary settler.

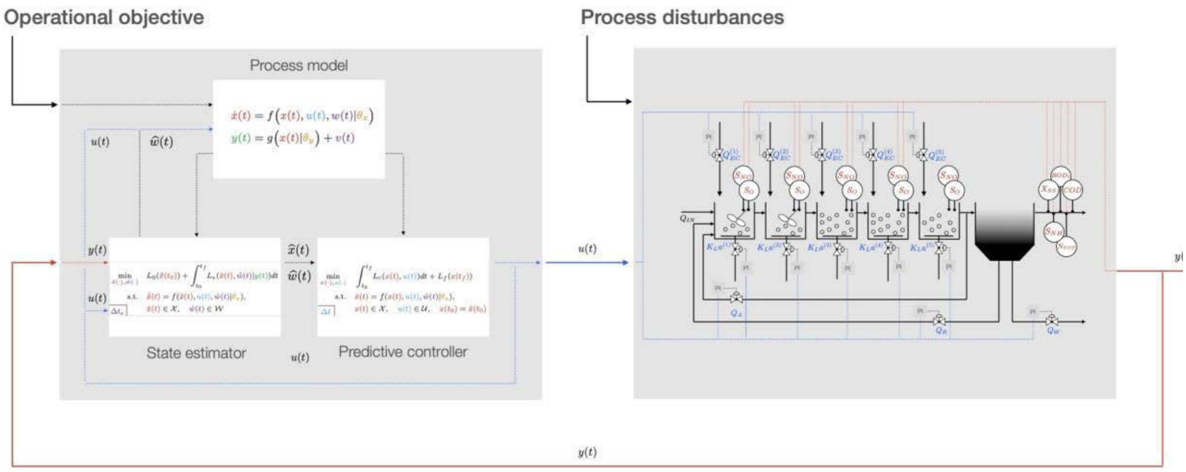

Figure 6. General structure of a model-based predictive control (MPC) and moving-horizon state estimator (MHE). Both the MPC and the MHC solve a nonlinear optimization problem, subject to the process dynamics, encoded by the process model, and a number of technological and operational constraints after

When compared to simpler error-feedback control strategies that require an ad-hoc pairing between controlled and manipulated variables, a model-based control architecture defines the control actions over the entire set of actuators comprehensively, at the plant-level, according to a state-feedback principle. It is also important to note that additional modules can be integrated into this general structure. Typically, it is expected to include a fault-analysis module that certifies the normality of the operations and health status of the equipment: This module is necessary to authorize the deployment of the supervisory actions to the regulatory layer. Moreover, it is often beneficial to add a module that support planning by predicting the future evolution of the disturbances. More advanced modules with data analysis capabilities can be included to support monitoring, from high-level KPIs to instrumentation, at different time scales and process levels.

To simulate real-world situations in a more realistic way allowing to test control actions and the staff to experience what different situations would be like in real life (without the associated costs and dangers), there is a current trend to develop digital twins (DT). DTs are virtual representations that serve (near) real-time counterparts of physical objects (twins). The core of DTs of wastewater treatment processes are process models that often are mechanistic models (for instance, ASM and ADM models), but hybrid models have also been found promising approaches (Torfs *et al.,* 2022). According to these authors, three key features that separate DTs of wastewater treatment processes from off-line models are that (a) a physical counterpart for model must exist, (b) there is an automated data connection to the physical twin, and (c) there needs to be means to continuously update the process model according to evolution of the physical process over time. Because an essential property of a DT is use of near live data, appropriate automated data management is crucial for successful implementations. DTs have been used, for instance, for evaluating current process status and for performing automated scenario analysis in the Singapore PUB Changi WRRF (Johnson *et al.*, 2021), for process monitoring and operational advice with focus on improving resource recovery and reducing energy footprint in Egå WRRF (Denmark) (Polesel *et al.,* 2021) and for predictive control of influent flow in Kolding WRRF (Denmark) (Stentoft *et al.,* 2020). Other potential applications of DTs in treatment plants include operator training, failure analysis, and asset management and predictive maintenance (Torfs *et al.*, 2022).

3 CONCLUSIONS

Collected data in WWTPs and WRRFs can be utilized for many useful purposes: process monitoring, real-time process control and automation, increase process knowledge, process optimization, data-driven modelling, data mining, better understanding of process state, fault prediction, decision-making. Currently a lot of data is collected in WWTPs and WRRFs, but most of them is not used and the almost complete absence of metadata prevents its future use not having the necessary knowledge of its context to allow their interpretation. In this paper it has been shown how to increase the value of the large amount of data that nowadays are being collected in WWTPs and WWRFs. The main take-home messages are the following:

- Nowadays, many facilities around the world are not exploiting the value of the available data, being data rich yet information poor a frequent situation that results in data graveyards.
- It has been shown that it is possible to take advantage of the data that nowadays it has been collected with the already deployed instrumentation in each facility, by assuring its quality (an automatic quality workflow has been proposed) and storing its context and related information (metadata) to enable its interpretation and future use.
- The deployment of real-time automatic process control algorithms and data-driven models can avoid some repetitive low-level tasks to process operators while keeping the WWTP running in a cost-effective way, thus making the collected data truly valuable and useful.
- The human factor is vital for a successful digital transition: the staff should be involved in the development of the artificial intelligent tools (e.g., taking advantage of their process expertise and know-how), as well as be trained on these tools (so the staff can understand the new smart-tools and some of them even update or tuning them).
- More information from the process can be obtained deploying new sensors and quality probes, which will allow the development of more complex control strategies, as well as a more detailed and in-depth knowledge of the monitored processes, which is always valuable for an informed decision-making.

4 REFERENCES

[1] Arnell M., Rahmberg M., Oliveira F, Jeppsson U. (2017): Multi-objective performance assessment of wastewater treatment plants combining plant-wide process models and life cycle assessment. Journal of Water and Climate Change, 8(4), pp. 715-729.

[2] Ching P.M.L, So R.H.Y., Morck T. (2021) Advances in soft sensors for wastewater treatment plants: A systematic review. Journal of Water Process Engineering, 44.

[3] Corominas Ll., Garrido-Baserba M., Villez K., Olsson G., Cortes U., Poch, M. (2018). Transforming data into knowledge for improved wastewater treatment operation: a critical review of techniques. Environ. Model. Software, 106, 2018, 89–103.

[4] Daw J., Hallet K., DeWoldfe J., Venner I. (2012) Energy efficiency strategies for municipal wastewater treatment facilities. Technical Report. Department of Energy Office of Scientific and Technical Information.

[5] Eerikäinen S., Haimi H., Mikola, A., Vahala R. (2020) Data analytics in control and operation of municipal wastewater treatment plants: qualitative analysis of needs and barriers. Water Sci. Technol. 82(12), 2681–2690.

[6] Elsayed A., Siam A., El-Dakhakhni W. (2022) Machine learning classification algorithms for inadequate wastewater treatment risk mitigation, Process Safety and Environmental Protection, 159, pp.1224-1235.

[7] European Commission (2020) A new circular economy action plan – for a cleaner and more competitive Europe. Brussels, Belgium.

[8] Guest J.S., Skerlos S.J., Barnard J.L., Beck M.B., Daigger G.T., Hilger H., Jackson S.J., Karvazy K., Kelly L., Macpherson L., Mihelcic J.R., Pramanik A., Raskin L., van Loosdrecht M.C.M., Yeh D. and Love N.G. (2009) A New Planning and Design Paradigm to Achieve Sustainable Resource Recovery from Wastewater, Environmental Science and Technology, 43, 6126-6130

[9] Guo H., Jeong K., Lim J., Jo J., Kim Y. M., Park J., Kim J.H., Cho K. H. (2015) Prediction of effluent concentration in a wastewater treatment plant using machine learning models, Journal of Environmental Sciences, 32, pp. 90-101.

[10] Haimi H., Mulas M., Corona F., Vahala R. (2013) Data-derived soft-sensors for biological wastewater treatment plants: An overview. Environ. Model. Softw. 47, 88–107.

[11] Hastie T., Tibshirani R., Friedman J. (2017) The Elements of Statistical Learning - Data Mining, Inference, and Prediction. Second Edition, Springer.

[12] Henze (2000). Activated sludge models ASM1, ASM2 and ASM3 IWA Publishing. ISBN:9781900222242

[13] Ingildsen P., Olsson G. (2016) Smart Water Utilities: Complexity made simple. IWA Publishing, London.

[14] IWA (2021a) Digital Water: The Value of Meta-Data for Water Resource Recovery Facilities. White Paper. London, UK, p. 15.

[15] Jimenez J., Latrille E., Harmand J., Robles A., Ferrer J., Gaida D., Wolf C., Mairet F., Bernard O., Alcaraz-Gonzalez V., Mendez-Acosta H., Zitomer D., Totzke D., Spanjers H., Jacobi F., Guwy A., Dinsdale R., Premier G., Mazhegrane S., Ruiz-Filippi G., Seco A., Ribeiro T., Pauss A., Steyer J.-P. (2015) Instrumentation and control of anaerobic digestion processes: a review and some research challenges. Rev. Environ. Sci. Bio/Technology 14, 615–648.

[16] Johnson B.R., Kadiyala R., Owens G., Mak Y.P., Grace P., Newbery C., Sing S., Saxena A., Green J. (2021) Water reuse and recovery facility connected digital twin case study: Singapore PUB's Changi WRP process, control, and hydraulics digital twin. Proceedings of WEFTEC 2021, Chicago, USA.

[17] McDonald G., Crawford J. (2017). Energy neutral WWTPs - A dream or a reality for New Zealand? The Water New Zealand conference & Expo 2017.

[18] Mhatre P., Panchal R., Singh A., Bibyan S. (2021): A systematic literature review on the circular economy initiatives in the European Union. Sustainable Production and Consumption, 26, pp. 187-202

[19] Metcalf & Eddy (2013). Wastewater Engineering: Treatment and Resource Recovery. McGraw Hill; 5th edition; ISBN-13 : 978-0073401188.

[20] Mulas M., Tronci S., Corona F., Haimi H., Lindell P., Heinonen M., Vahala R., Baratti R. (2015). Predictive control of an activated sludge process: An application to the Viikinmäki wastewater treatment plant. Journal of Process Control, 35:89-100

[21] Müller-Czygan, G. (2020) Smart Water–How to Master the Future Challenges of Water Management; Chandrasekaran P.T., Javaid M.S., Sadiq A., Eds.; Resources of Water; IntechOpen: London, UK, 19–33.

[22] Neto O.B.L., Mulas M., Corona F. (2022). About the classical and structural controllability and observability of a common class of activated sludge plants. Journal of Process Control, 11, 8-26.

[23] Newhart K.B., Holloway R.W., Hering A.S., Cath T.Y. (2019) Data-driven performance analyses of wastewater treatment plants: A review. Water Research, 157(6), pp. 498-513

[24] Nilsson S., Andersson S. (2018) Summary of instrumentation survey Denmark. IVL Swedish Environmental Research Institute Ltd., Stockholm, Sweden.

[25] Nourani V., Elkiran G., Abba, S. I. (2018) Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach, Water Science and Technology, 78 (10), pp. 2064-2076.

[26] Olsson G ., Ingildsen P. (2018) Making water operations smarter. In Advances in Wastewater Treatment (Mannina, G., Ekama, G., Ødegaard, H. and Olsson, G., eds). IWA Publishing, London.

[27] Polesel, F., Remigi, E.U., Sanford, R.K.E., Merrild, M.H.B., Nissen, Thage, R.K., Lynggaard-Jensen, Ramin, A.P., Drejer, L.H., Andreasen, P., Sørensen, H.R. and Dalkvist, T. (2021) Digital Twinning of Water Resource Recovery Facilities – From concept to implementation. Proceedings of 7th IWA Water Resource Recovery Modelling Seminar WRRMod2021, virtual conference.

[28] Sarpong G., Gude V.G. (2020). Near future energy self-sufficient wastewater treatment schemes. Int. Journal of Environmental Research, 14:479-488.

[29] Solon K., Volcke E.I.P., Sperandio M., van Loosdrecht M.C.M. (2019) Resource recovery and wastewater treatment modelling. Environmental Science: Water Research and Technology, Royal Society of Chemistry, 2019, 5 (4), pp.631-642.

[30] Stentoff P.A. (2020). Stochastic modelling and predictive control of wastewater treatment processes. PhD thesis, Technical University of Denmark.

[31] Stentoft P.A., Vezzaro L., Mikkelsen P.S., Grum M., Munk-Nielsen T., Tychsen P., Madsen H., Halvgaard R. (2020) Integrated model predictive control of water resource recovery facilities and sewer systems in a smart grid: example of full-scale implementation in Kolding. Water Science and Technology, 81, 1766–1777.

[32] Therrien J.-D., Nicolaï N. Vanrolleghem P.A. (2020) A critical review of the data pipeline: How wastewater system operation flows from data to intelligence. Water Sci. Technol. 82 (12), 2613–2634.

[33] Torfs E., Nicolaï N., Daneshgar S., Copp J.B., Haimi H., Ikumi D., Johnson B., Plosz B.B., Snowling S., Townley L.R., Valverde-Pérez B., Vanrolleghem P.A., Vezzaro L. Nopens I. (2022). The transition of WRRF models to Digital Twin applications. Water Science & Technology, accepted for publication.

[34] Xu R., Cao J., Fang F., Feng Q., Yang E., Luo J. (2021) Integrated data-driven strategy to optimize the processes configuration for full-scale wastewater treatment plant predesign, Science of The Total Environment, 785, 147356,

[35] Yuan Z., Olsson G., Cardell-Oliver R., van Schagen K., Marchi A., Deletic A., Urich C., Rauch W., Liu Y. Jiang G. (2019) Sweating the assets – the role of instrumentation, control, and automation in urban water systems. Water Res., 155 (2019), pp. 381-402

[36] Zhang W., Tooker N.B., Mueller A.V. (2020) Enabling wastewater treatment process automation: leveraging innovations in real-time sensing, data analysis, and online controls. Environ. Sci. Water Res. Technol., 6, pp. 2973-2992