CODEN:LUTEDX/(TEIE-7294)/1-33/(2023)

Industrial Electrical Engineering and Automation

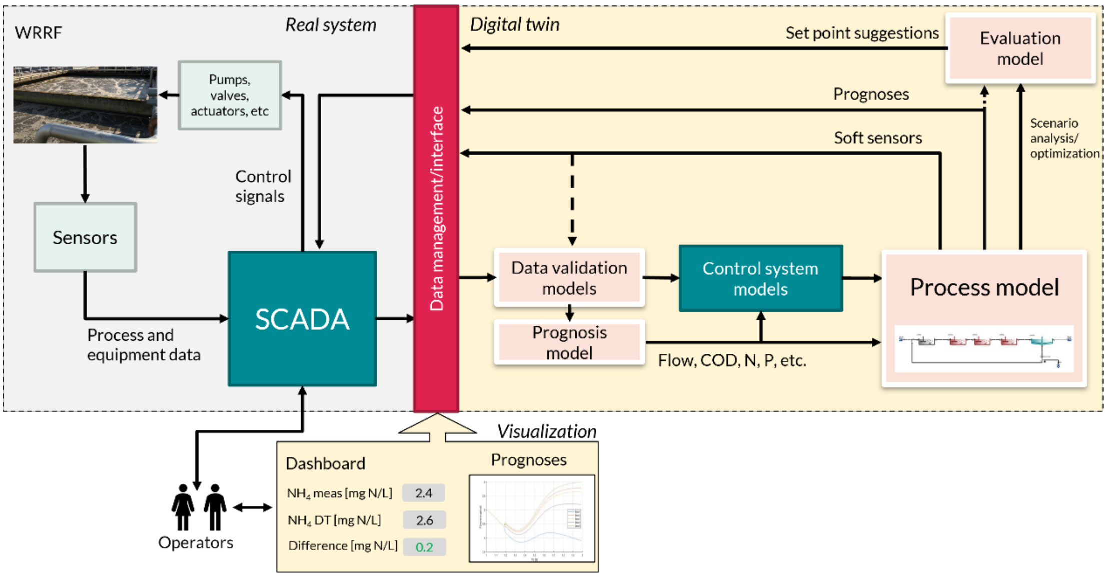
Operational Digital Twins for Water Resource Recovery Facilities –

Rationale, Components and Case Studies

Christoffer Wärff

Division of Industrial Electrical Engineering and Automation
Faculty of Engineering, Lund University

BUILT ENVIRONMENT
CIRCULAR WASTEWATER
SYSTEMS



Operational digital twins for water resource recovery facilities – Rationale, components, and case studies

Christoffer Wärff

RISE Report 2023:105

Operational digital twins for water resource recovery facilities – Rationale, components, and case studies

Christoffer Wärff

Key words: wastewater treatment plant; water resource recovery facilities; digital twin; digitalization; process modelling; real time simulation; automated simulation

RISE Research Institutes of Sweden AB

RISE Report 2023:105
ISBN: 978-91-89821-85-9
2023

Content

Preface

Digital twins for water resource recovery facilities/wastewater treatment plants are an emerging technology with large potential benefits to plant operations. This report is written as part of the project *Digital twin for sustainable and resource efficient operation of wastewater treatment plants* (Formas 2020-00222), with the aim to summarize the rationale for using digital twins, describe the different components that will be important for implementation of a digital twin, and describe available case studies. The report is written as part of the authors PhD studies at the department of Industrial Electrical Engineering and Automation (IEA), Lund University, Sweden.

Summary

Digital twins (DT) for water resource recovery facilities (WRRF) are different from regular process models. They require 1) a physical plant twin; 2) automatic data exchange with the real plant; 3) possibility to dynamically update models when or if required. Their use has the potential to improve understanding of plant behaviour and unmeasured variables; move towards proactive decision making at the plants when including influent forecasts; improve data quality control when comparing simulation results to measured values; and be used for predictive maintenance.

The model used in a DT can be mechanistic (i.e., describing underlying mechanisms/physics), data driven (empirical, based on observed relationships between variables) or a combination of both (hybrid model). Most of the commercially available (mechanistic) wastewater process simulators include the option to use them in (near) real time as digital twins.

Fault detection is important for DTs to avoid use of faulty input data. Methods range from dimensional reduction techniques to process models and statistical control charts. Automated methods for gap filling and corrections of sensor values based on laboratory measurements can be used to correct faulty data.

Forecasts of influent flow rate and concentration of pollutants can be useful for optimization and “what if”-scenarios. Forecast models can be data driven (e.g., many examples with time series models and artificial neural networks available in the literature) or detailed mechanistic models. Common for most examples is that weather forecasts (temperature and precipitation) are used, and the model accuracy of course depend on the quality of the forecast.

Automatic calibration can be used for both data driven/hybrid models (i.e., re-training) and mechanistic models. For mechanistic models, examples in the literature include simple changing of measured influent fractions or settler solids separation efficiency to global optimization of multiple variables over a plant-wide model. Automatic calibration can be done at fixed intervals or based on performance evaluation.

Model predictive control (MPC) has been widely studied in simulated settings, with few real examples for WRRFs. For digital twins, the possibility to combine a mechanistic model with influent forecasts and numerical optimisation for, e.g., setpoints over a future time interval to achieve a certain goal is promising. The faster control applications can then be handled using regular PID-controllers.

Few examples of implemented digital twins for WRRFs have so far been published in the literature. Here, one example of a digital twin is presented. It includes automatic data transfer, automatic calibration, and forecasts, but is (at the time of writing based on the available literature) only used as an advisory tool and not for direct control.

Digital twins of water resource recovery facilities are complex with many different parts and models that work together. They can be used for fault detection, predictions, and optimization/control. This report summarizes some of the components that can be used to build digital twins, which ones to include of course depends on the scope and goals of the specific project. In all cases, the flow of data from collection to use must be well designed to avoid unnecessary interruptions in operation.

1 Introduction

This document is intended to provide an overview of operational *digital twins* of water resource recovery facilities (WRRF), including a description of the different components that can comprise a digital twin, methods that can be used and examples of how these have been implemented in the literature.

1.1 What is a digital twin?

The term “digital twin” is not universally defined and can mean different things in different sectors, with the broad definition being a digital representation of a real object (Trauer et al., 2020). This has caused confusion about the term when used in the context of WRRFs and other parts of the water sector where it is sometimes used for conventional process models which have been used for several decades. Torfs et al. (2022) mention the risks associated with re-branding of established terms and defines the following conditions for the definition of digital twin of a WRRF:

1. There must be a physical counterpart.
2. There is an automated data feed from the physical to the digital twin.
3. The twin is updated dynamically as new information is available.

This means that there must exist a real plant from which data is automatically fed to the digital twin. This distinguishes it from a process model, which is usually not connected to automated data feeds, and which can be used to evaluate operations and designs before a real plant is built. For the context of WRRFs in this report, the focus is on digital twins used as an **operational tool for proactive decision making**, i.e., operational digital twins (Valverde-Pérez et al., 2021). A schematic representation of such a digital twin is shown in Figure 1.

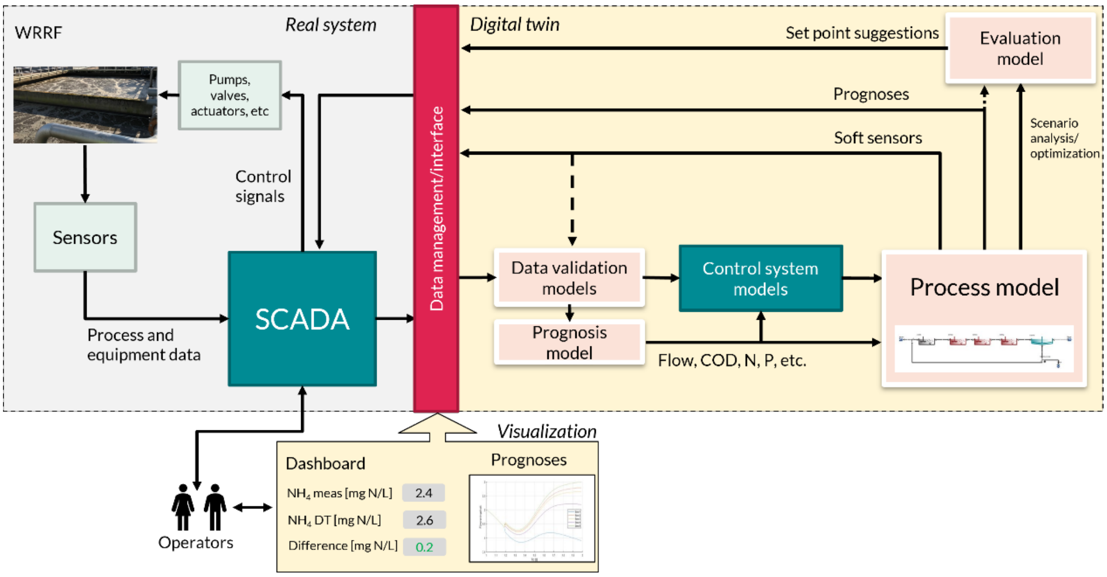


Figure 1. Components and structure of an operational digital twin of a water resource recovery facility.

For this type of digital twin, the physical system consists of a WRRF with its different process units, sensors and valves/actuators, all connected to a Supervisory Control And Data Acquisition (SCADA) system. A data management interface is used to collect data from the SCADA system historian (and potentially other data sources) and transfer it to the digital twin in the right format and at the right time. The digital twin consists of a collection of models:

- a data validation model for making sure the data used for analyses is feasible;
- a forecast model to predict the future influent (flow and pollutant loads) to the WRRF;
- a model of the real control system;
- a WRRF process model for prediction of plant performance;
- an evaluation model for scenario analyses and optimization.

The digital twin can produce WRRF operational performance prognoses, optimized set points (which can be used as suggestions or implemented as automatic control) and soft sensor values which is fed back to the data interface. From there, the data and predictions are visualized for the operators in dashboards with information such as Key Process Indicators (KPI), operational prognoses, uncertainty, etc., depending on the needs at the plant. The operators can interact with the digital twin through the interface and perform scenario analyses. In all applications, the flow of data from collection to use must be well designed to avoid unnecessary interruptions in operation (Therrien et al., 2020).

1.2 Potential benefits

Some of the potential benefits of using digital twins include:

- **Increased insight and process understanding:** components that are difficult/impossible/expensive to measure in reality can be approximated by the digital twin (so called *soft sensors*). For example: NH_4-N might be measured at the end of a treatment train while the digital twin gives information about the concentration in every zone; the concentration of nitrifiers in sludge is (nearly) impossible to measure accurately in reality but the digital twin provides this information. The potential use of this type if data for process understanding, control and fault detection is intriguing.
- **Operational forecasts for proactive decision making:** the combination of influent forecasting tools and digital twins of the WRRF can provide operational forecasts for the plant, such as the effluent quality for the coming days. This can in turn be used for predictive control or manual decisions by operators to achieve an optimal outcome based on, e.g., chemicals dosing, effluent quality, operational costs, or greenhouse gas emissions.
- **Automated data quality control:** data quality control is something that is both required before raw data is used in the digital twin as well as something the digital twin can be used for. Data driven/statistical methods can be used for raw data checks while comparison of prediction values from the digital twin with sensor data can help detect abnormalities.

- **Predictive maintenance:** models of equipment can be included in the digital twin such that predictions of deterioration of data quality (e.g., fouling) can be used for smarter maintenance planning. This will not be further discussed in this report.

1.3 Components of digital twins at water resource recovery facilities

The type of components that are included in a digital twin depends on the objective of the twin and is therefore case specific. This report will focus on:

- The type of process models that are suitable for digital twins.
- Methods for data analysis, fault/event detection and data correction.
- Model predictive control.
- Data transfer.
- Automatic calibration.
- Forecasting tools.
- Case studies.

2 Process models for use in digital twins

When deciding on the model to use in a project, good practice is to use the simplest one that achieves the objectives. This gives many options to use more or less detailed models to describe the required process(es). Mainly three types of mathematical models can be distinguished:

- **Mechanistic models** describe the underlying phenomena which affects observable behaviour. Sometimes also referred to as first principles models or white box models. For example: nitrification is described by modelling the growth of nitrifiers who consumes NH_4-N in the growth process (we can observe the depletion of NH_4-N).
- **Data driven/empirical models** describes behaviour based on observable relationships with other variables. Sometimes also referred to as black box models. For example: can range from a simple linear regression model to more complex machine learning methods such as artificial neural networks.
- **Hybrid data driven/mechanistic models** are a combination of mechanistic and data driven components. Sometimes also referred to as grey box models.

The different types of models have their advantages and disadvantages, but all three model types above can be suitable for use in digital twins of WRRFs. The model types are described below.

2.1 Mechanistic models

The well renown *activated sludge models* (ASM): ASM1, ASM2, ASM2d and ASM3 (Henze et al., 2000), have been widely used since the publication of ASM1 in the mid-eighties. Many extensions to the models have been developed to include new processes and variables, many of which are available in commercial simulators. The Benchmark Simulation Model No 1 and 2 (Gernaey et al., 2014), based on the ASM models, were developed for benchmarking of control strategies at WRRFs and have been widely used to test control and fault detection strategies at WRRFs. In commercial simulators the ASM models or further developments based on the ASM models are often available. Mechanistic models are usually calibrated when used for real plants, meaning that some model parameters are adjusted until a good match between model predictions and measured data is achieved. However, usually only a few parameters are required to be adjusted for calibration due to well validated default parameter values (based on long experience of using the models). This means that the models can be used to extrapolate outside of the operating conditions that the model has been calibrated on for each case study (although care is of course recommended with use in such cases). Some parameters though, might require re-calibration between different time periods. The mechanistic models are well suited to interpret results as the model structure and equations are (often) available for the user.

These types of models have the potential to also be used in digital twin applications, although one of the largest bottlenecks could be the computational demand (Johnson et al., 2021) depending on how it is used. Another benefit of using mechanistic models is that they generally require less data compared to data driven models as much knowledge is included in the equations from the beginning.

2.2 Data driven models

Data driven models have the benefit of (in general) lower computational demand than mechanistic models while drawbacks are their need of large amounts of data and bad extrapolation capabilities. When developing data driven models, the fitting of the model to data is usually called "training" the model. The model usually becomes sensitive to the training data used, meaning that bad data and extrapolation of conditions beyond the training data set can cause large errors in predictions. If the conditions that the model was trained on changes, the model must thus be recalibrated. Another drawback of data driven models are that for advanced models the internal calculations are difficult to access, severely limiting interpretability of the results.

Depending on the objective, data driven models with a wide range of complexity can be used for digital twin applications. Often some kind of Artificial Neural Network (ANN) is used. For example, Han et al. (2012) used a radial basis function neural network model to predict dissolved oxygen concentration in a plant for predictive control purposes. Su et al. (1992) used a recurrent neural network to predict concentration values of ammonia, nitrate and phosphate in different locations at a wastewater treatment plant with good results.

Due to the lower computational demand, data driven models are promising for use in model predictive control (Bernardelli et al., 2020; Stentoft et al., 2021) (see Section 6) and forecasting (Kim et al., 2006; Vezzaro et al., 2020) (see Section 4.1) applications.

2.3 Hybrid models

Hybrid mechanistic/data driven models have been especially mentioned in recent research as a promising way to achieve models with high predictive power for use in digital twins due to: 1) their lower computational demand than mechanistic models, while preserving some of the knowledge base of those models; and 2) lower data requirements than data driven models while having the ability to capture dynamics not included in the mechanistic model (Schneider et al., 2022).

Schneider et al. (2022) define three types of architectures for hybrid models: serial; parallel; and surrogate (Figure 2). In the serial case, a mechanistic model and data driven model are used in series: the output of one model is transferred to the next. In the parallel case, the mechanistic and data driven model are run in parallel, fed with the same input. The data driven part can then be used to predict the error of the mechanistic model and correct for this in the output. In the surrogate case a data driven model is trained on the output from a mechanistic model, e.g., to allow for faster simulations. Another method is embedded hybrid models, where the differential equations include a data driven component, such as neural differential equations (Chen et al., 2018).

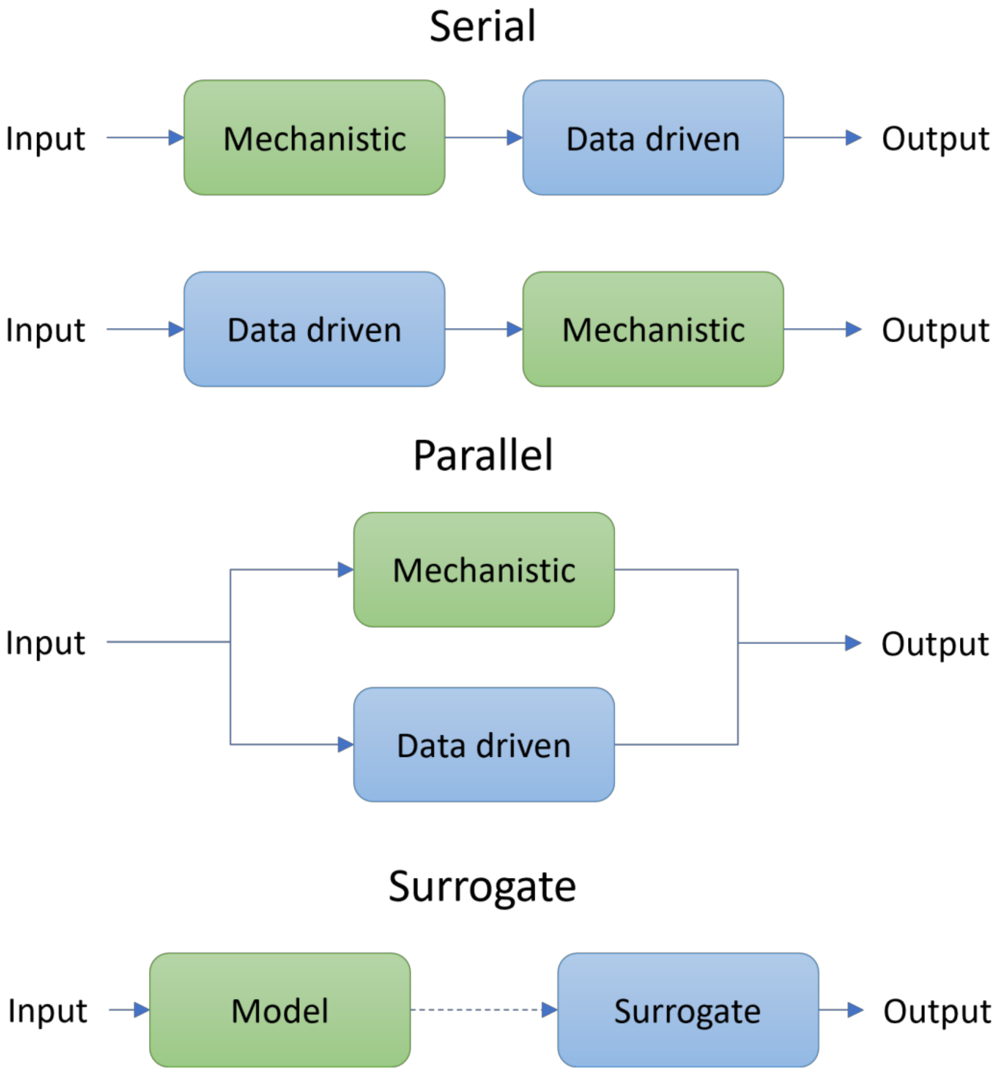


Figure 2. Different hybrid model structures. Adapted from Schneider et al. (2022).

Construction of hybrid models can be done in various ways. Lee et al. (2002) compared a pure neural network model with a serial hybrid model, where a neural network was used to estimate kinetics of the mechanistic (modified ASM1) model, and a parallel hybrid model where a neural network was used to correct the difference between mechanistic model output and measured data. They evaluated the models with data from an industrial wastewater treatment plant and found that the parallel hybrid model performed best, also during process upsets such as toxic shocks. Lee et al. (2005) compared different data driven components for parallel hybrid modelling and found that a neural network partial least squares model performed best when considering predictive power, model construction and interpretability. Hvala & Kocijan (2020) combined a Gaussian process (GP) model with ASM2d in a parallel configuration to predict the residuals from the mechanistic model predictions. According to the authors, GP models are as predictive as neural networks while they are probabilistic and thus includes information about uncertainty in the predictions. Quaghebeur et al. (2022) modified a mechanistic model (BSM1, Alex et al. (2008)) with neural differential equations, i.e. the Ordinary Differential Equations (ODE) of the model was extended with a neural network term. They tested the predictive power and extrapolation capabilities under realistic conditions of incomplete knowledge in the model structure (simplified hydraulic model). The model was compared to the mechanistic model and a pure neural network model for conditions with input data unseen during model training (wet weather flow). The hybrid model outperformed both the other models for this test. The authors highlight the benefit of this type of hybrid model structure as compared to other serial/parallel structures as the model learns the missing information of the mechanistic model (e.g. kinetic rates) rather than just correcting the errors. De Jaegher et al. (2021) also used a neural ODE structure to predict fouling in ion exchange membranes combined with a mechanistic model for prediction of flux. Nielsen et al. (2020) used a neural network model to compute kinetic rates in particle processes based on data from image analysis. The computed kinetic rates were fed to a mechanistic model and performed as well as established models, even with limited data.

2.4 Real time use process models

The real time component of a digital twin can be organized in different ways, and most of the major wastewater simulation software providers include an option for use with digital twins. Depending on the purpose of the digital twin, the data connection can be in essentially real time (as soon as a sensor value is updated, it is transferred to the twin). Another other option is to use a scheduled approach, where simulations are run in fixed intervals (i.e., every 10 minutes, every hour, or once per day). The first approach can be useful if the twin is mostly used for monitoring (e.g., fault detection of sensors) or direct feedback control based on modelled (soft sensor) variables. The second approach is more useful for model-based optimization control and forecasts. Both approaches can of course be combined as well.

In any case, the digital infrastructure at the plant must be organized to allow for transfer of data from the plant to use in a digital twin. The requirements and rules at each plant decides on how this must be set up. An example is if cloud services are allowed or if everything must be installed on dedicated servers on the premises. This can be both challenging and time consuming and must be taken into consideration for any digital twin project. For the case presented by Johnson et al. (2021), a database is updated with data from both the SCADA system as well as laboratory data. Python scripts are used for all automation of the data cleaning, calculations and simulations. All simulations are run on dedicated on premises servers as no cloud computations were allowed by the utility.

Brentan et al. (2017) also demonstrate the requirement to update the offline model structure for online applications, such as using a hybrid model. Although they used a data driven model (support vector machine) for the base (offline) model, an extension based on adaptive Fourier series was required to correctly predict the short-term dynamics.

3 Data management and analysis

Access to data of sufficient quality is essential in all types of modelling projects. In the context of digital twins, which are automatically fed with data from several sources, the data analysis and validation must also be performed in an automated fashion. In essence, three targets must be fulfilled for digital twins:

- Data faults and errors must be detected and identified.
- Faulty or missing data must be replaced with correct data (or a best estimate) to avoid data gaps. Filtering of data can be required to remove noise or outliers.
- Data must be efficiently transferred between storage locations and the digital twin.

Details about these targets will be presented below.

3.1 Fault and event detection

Fault detection usually rely on different types of machine learning models. The literature review is here organized under sub-topics with groups of machine learning models. Some of the most widely cited methods are presented here and more information can be found in Newhart et al. (2019).

3.1.1 Dimensional reduction techniques

Since many monitored variables at WRRFs often are highly correlated, dimensional reduction techniques can be useful for creating new variables which can be more easily monitored to detect unusual process conditions. This makes them popular for fault detection. Examples include *principle component analysis* (PCA) and *partial least squares* (PLS). In PCA, new variables (called principal components or latent variables) are created from the chosen variables with the aim that the new variables explain as much as possible of the variance of the desired process variable (Rosen & Olsson, 1998). In PLS on the other hand, the latent variables are constructed to maximize the correlation between input and output variables (Rosen & Olsson, 1998). PLS can also be made non-linear for non-linear processes. Another technique used is *independent component analysis* (ICA). In ICA, the extracted latent variables are independent from one another (as compared to PCA where the variables are uncorrelated, meaning only linearly independent) (Lee et al., 2004c).

Rosen & Olsson (1998) provide an example of the use of PCA for process monitoring at a WRRF, where influent temperature, conductivity, ammonia, pH and flow rate are used to create two principal components which explain 75 % of the variance of these variables. They show how an abnormal change in flow rate is the primary cause that affects the process, something that would be more difficult to show in normal time series plots. Yoo et al. (2002) and Lee et al. (2004a) used a dynamic PCA to better monitor non-stationary behaviour as a static PCA proved inadequate for dynamic WRRF monitoring. Many authors also use adaptive PCA, where the PCA is updated to reflect long term trends in the process (Aguado & Rosen, 2008; Baggiani & Marsili-Libelli, 2009; Haimi et al., 2016; Lee et al., 2005; Lee & Vanrolleghem, 2004; Lennox & Rosen, 2002). Lee et al. (2004b) used kernel PCA to capture nonlinearities in data and found the method superior to linear PCA. Liu et al. (2014) used variational Bayesian PCA to detect disturbances in the BSM1 and a real plant process and found the method more effective than PCA when dealing with missing data.

PLS has also been used at WRRFs by several authors. Ferrer et al. (2008) used PLS for soft sensor construction and process diagnosis and highlights the method's benefits of low risk of overfitting, efficient handling of missing data and outlier detection. Choi & Lee (2005) used multiblock PLS for fault detection, monitoring and modelling of a chemical process. Rosen & Olsson (1998) expanded on the previous example used with PCA also for PLS and related the influent variables to the effluent turbidity. With the model they were able to detect disturbances in the effluent turbidity 10 minutes before it was observed in the turbidity sensor.

Lee et al. (2004c) used ICA for process monitoring on the BSM1 plant and highlights that the method is more sophisticated than PCA while effectively identifying and isolating disturbances. Lee & Qin (2006) used a modified ICA, also on BSM1, and found it more efficient than PCA in detecting faults. An improvement of ICA, so called kernel ICA, was developed by Wang & Shi (2010) and evaluated on BSM1.

3.1.2 Process models

A process model can be used to predict the states of measured variables at different parts of a plant. Therefore, if the model is correct, it can be used to detect discrepancies between predicted and measured variables, i.e. for fault/event detection. Both mechanistic and data driven process models can be used for this purpose.

Time series models are mainly used to predict future variable values based on previous values of the same and/or other variables. They can therefore be used for forecasting or as process models. Due to their predictive capabilities, they have also been used to detect faults. For example, Sánchez-Fernández et al. (2018) used time series models to model a plant and used statistical techniques on model/observation residuals to detect faults.

Artificial neural network is a data driven method that can be used both as time series models or for classification purposes, e.g. for fault detection. Zumoffen et al. (2008) used PCA and adaptive PCA for fault detection, then used an ANN to determine the magnitude of the fault. Caccevale et al. (2010) used ANN to predict future NH4-N and NO3-N concentration values, then combined it with a time series model to detect faults.

3.1.3 Statistical control charts

Statistical control charts are often used for fault detection, with several methods available in the literature. They work by comparing values over time to lower and upper limits, where a fault is considered detected if the value crosses the limits.

One of the early examples of control charts are the Shewhart control charts (Shewhart, 1938). Shewhart control charts were developed for quality control in manufacturing and works by dividing samples into small groups. The mean for each group is plotted over time, with limits calculated from, e.g., the mean standard deviation of the groups. Any plotted point below or above the limits are considered deviating.

A later development resulted in Cumulative sum (CUSUM) charts (Page, 1954), which (compared to Shewhart charts) are better at detecting small changes in mean values over time since the cumulative sum of residual between measurement and mean is plotted.

Later, Roberts (1958) developed a chart based on the exponentially weighted moving average (EWMA) of the process variable, designed to also detect small changes which may not be detected with the Shewhart chart.

Marais et al. (2022) compared the performance of Stewhart, CUSUM and EWMA charts on a wastewater treatment plant NO_x-N sensor in the BSM1 model, testing for detection of both drift and bias faults in an offline and online setting. Discrepancies in the performance between offline and online settings were found due to compensation of sensor failure by the feedback controllers used. For the online setting, the CUSUM and EWMA both performed well for bias faults, while the EWMA performed better for drift faults. The Stewhart chart did not perform well in the online setting. The sensitivity of the methods for the parameters used was also pointed out, where careful calibration of the charts is required to correctly detect faults.

3.2 Data corrections

For digital twin predictions to be robust, data gaps must be avoided. At WRRFs however, data gaps will inevitably occur at times due to for example signal loss or faulty or uncalibrated sensors. After the first step (to detect the fault, as described in Section 3.1) the next step is to replace faulty data with correct (reconciled) data or a best estimate.

3.2.1 Data reconciliation

Historical datasets are usually used for model calibration or training in WRRF modelling. Measured data always contains errors due to measurement uncertainty, noise and sometimes gross errors. This means that calculations of mass balances over process units never close (i.e. mass flow in = mass flow out). Mass balance-based data reconciliation techniques have been developed to correct data so that relevant mass balances can be closed (Rieger et al., 2010). Although these techniques usually are used for low frequency steady state data (Le et al., 2018), examples exist from industrial applications where dynamic data reconciliation is used for high frequency data, for example to remove measurement noise to improve controller performance (Hu et al., 2021; Zhu et al., 2021).

3.2.2 Automatic gap filling

De Mulder et al. (2018) developed a software to systematically and semi-automatically fill high frequency data gaps in (influent) time series data for WRRFs. The algorithm detects values tagged as not-a-number (NaN), constant signals, unrealistically fast changes (noise) and outliers and filters them by either interpolation (short gaps), correlation with other variables, daily average profile values, values from the day before or from an influent generator model. They tested the reliability of the methods by introducing artificial gaps in data and computing the deviation of the filled data from the original and found that the influent generator model provided the lowest deviation (6 % for influent NH_4-N).

3.2.3 Sensor corrections based on laboratory measurements

High frequency (online) influent data is beneficial for use in digital twins for improved predictions, but this type of measurements (e.g., chemical oxygen demand (COD), NH_4-N, PO_4-P) are rarely found at WRRFs in Sweden (Andersson et al., 2019). One likely reason could be that the information historically has not been providing enough value to motivate the investment and increased maintenance in sensors, but another issue is that it is difficult to measure accurately in the influent. Automatic analysers for, e.g., NH_4-N require the samples to be filtered, and due to the large amounts of suspended solids, fats and grease in the influent these are prone to clogging quickly. Another option is *ion sensitive electrodes* (ISE), but these have the downside of drifting and can easily be wrongly calibrated, making them mainly useful for observing trends and less for obtaining accurate values. The benefit of ISEs is that they are much lower investment costs compared to automatic analysers.

Nivert et al. (2009) used automatic corrections of optical and ISE sensors based on grab samples at the Rya wastewater treatment plant in Sweden. They used so-called break point curves to adjust the raw signal to better conform to the measured grab samples, which was implemented in an automated fashion. A follow up study (Lumley et al., 2013) showed results from both the Rya plant as well as the Viikinmäki plant in Finland where similar methods are used. The authors report savings of 10-40 hours per sensor and year in maintenance since the automatic adjustments were installed.

Pedersen et al. (2020) installed an ISE NH_4-N sensor at the influent to a WRRF and evaluated its performance. They concluded that despite maintenance protocols faulty data such as drift and corrupt calibration occurred. They developed an automatic data correction algorithm which considers the daily composite sample values for influent NH_4-N and dynamic flow measurements and corrects the ISE data to fulfil the daily average NH_4-N values. This algorithm corrects both slow signal drift and sudden large changes from sensor calibration and improved data for 87 % of the days in the data set where composite samples were available.

4 Influent forecasting tools

A substantial potential benefit of digital twins is to use them for proactive decision making, as opposed to the reactive decision making often practiced in plants today with feedback control. To achieve this, the future conditions of the plant must be taken into account to either analyse effects of decisions or to optimize the plant performance. Therefore, influent forecasting tools are essential to estimate future influx of water flows and pollutant loads to the plant.

4.1 Data driven forecasting models

Stentoft et al. (2019) used an adaptive stochastic model to predict influent NH_4-N concentration and load to a plant. The benefit of a stochastic model structure is that uncertainty of the predictions is included in the output, which can then be further used to assess the uncertainty propagation to downstream models. The forecasted load was used as input to a model predictive controller for optimization of the plant and compared to a situation without a forecast. The system showed improved performance with the forecast.

Vezzaro et al. (2020) used a simple approach where the daily ammonium load is described by a Fourier series approach. The parameters for describing the load variation for the next day is based on the measured load parameters for n previous dry weather days. A benefit from this approach is that the variations in load in the previous days are used to describe the uncertainty of the forecast.

Different types of time series models can be used for forecasting influent loads and flow rates. Kim et al. (2006) used an autoregressive integrated moving average (ARIMA) model to predict future daily values of NH_4-N, PO_4-P, temperature, and flow with good results for one day ahead forecasts. Li & Vanrolleghem (2022a) used an ANN with multivariate regression to model influent flow rate and pollutant concentrations based on time dependent patterns (obtained from a Fourier transformation) and weather data (temperature and precipitation). In a subsequent paper, Li & Vanrolleghem (2022b) used a long short-term memory (LSMT) model to predict influent water flow and temperature at a treatment plant while also including data on snowmelt. The LSMT prediction was compared to a phenomenological model and was shown to produce better results. The models by Li & Vanrolleghem were not used specifically as forecast models in the papers, but similar forecast models could be build by including the relevant weather forecasts.

Sokolova et al. (2022) compared several data driven models (exponential smoothing; ARIMA; naïve baseline; least absolute shrinkage and selection operator (LASSO) regression; random forest; vector autoregression; and a tree-based model) for predicting E. coli concentration in a river at a drinking water production plant intake. They used past values of lab samples of E. coli concentrations, water temperature, turbidity, precipitation and water flow for future predictions. Multivariate predictions models were shown to work better than univariate ones.

4.2 Mechanistic forecasting models

The project Future City Flow (see e.g., Valverde-Pérez et al. (2021)) in Sweden developed a mechanistic model for flow predictions in the sewer system. The model uses weather forecasts in combination with data from sensors in the sewers (e.g., level and flow sensors) to produce forecasts of the influent flow rate to the treatment plant. The models are used in real time to optimize the system to minimize combined sewer overflows and reduce flow variations to the plant.

Since the predictions of influent flow rate are dependent on the weather forecasts, the accuracy of those forecasts will determine the accuracy of the influent flow predictions when the hydraulic model has been sufficiently calibrated. This poses difficulties since the weather forecasts can sometimes change substantially in a short period of time, meaning that the influent forecast for a given day and time can change quickly if there is unstable weather. This is shown by Valverde-Pérez et al. (2021) where the forecasts made 30 minutes before the given time have considerably lower peaks than the forecast made for the same time interval 4.5 hours earlier.

5 Automatic calibration

The ability to automatically adjust/update model parameters is one of the aspects defining a digital twin according to Torfs et al. (2022), which is relevant regardless of the type of model that is used (mechanistic, data driven or hybrid).

5.1 Data driven models

For data driven/hybrid models, the automatic calibration can mean that the model is retrained in regular intervals (e.g., in a rolling time window fashion) or when deemed necessary (e.g. after a certain deviation between measured and simulated values occur). New/changed dynamics are then automatically incorporated in the model structure. Besides changes in for example environmental conditions at the plant (such as rainfall patterns or changing influent load), other factors such as changing data patterns at the facility due to for example new hardware or changing user patterns can also affect model accuracy over time (Kidane et al., 2022). While retraining is required to retain accuracy over time by always including the latest data in training, it also means that computer intensive tasks such as communication with databases and the model training itself must be done at regular intervals (Schule et al., 2021). This must be taken into account for digital twin applications to make sure the relevant computer power is available.

Choosing the right sample size for retraining is important to achieve optimal performance. Brentan et al. (2017) solved this problem for retraining of a water demand model by calculating a training efficiency, considering the sample size, time for training and resulting root mean square error (RMSE) of the model predictions. When plotting the efficiency for a range of sample sizes the optimum sample size could be determined.

Deciding when to retrain the model is also of importance to avoid unnecessary computational demand. Wei et al. (2022) included this in a real time model application by analysing the input data. If new input data was deemed to contain new information relative to the data previously used for training, the model was retrained using the new data.

5.2 Mechanistic models

For mechanistic models, the automatic calibration can be done as a simple approach to update e.g. TSS removal efficiencies in primary clarifiers or update influent fractions, based on new laboratory measurements (Johnson et al., 2021). In more advanced examples, a number of model parameters (e.g. biological parameters, influent fractions, settling parameters) can be adjusted simultaneously in a global optimization. The latter approach was used by Gómez et al. (2023), where 29 parameters for the BSM2 model was automatically calibrated using an ANN with the neuroevolution of augmenting topologies (NEAT) algorithm. Three calibration scenarios were tested: 1) the probability density function (PDF) of each calibration parameter is known a priori; 2) the correct parameter value is the mean of the PDF; 3) both 1) and 2) simultaneously. It was found that with expert knowledge of the system (i.e. known PDFs of all calibration parameters), adequate calibration (close to the real values) could be achieved efficiently.

6 Model predictive control

Model predictive control (MPC) uses a model that can predict the value of a controlled variable over a prediction horizon and optimize manipulated variables to achieve a desired setpoint. Historically, few examples of MPC utilized for control at a real WRRF are available in the literature. This is possibly due to the large computational demand associated with this type of control in combination with the strongly non-linear biological processes normally occurring at WRRFs. Many studies instead focused on theoretical, simulated cases (Han et al., 2012; Shen et al., 2008, 2009). With integrated (sewer and WRRF) MPC, system wide optimizations could be very beneficial as it will take the total loads to receiving water bodies into consideration (Rauch & Harremoës, 1999). More recently, simplified, data driven model structures have been proposed (Stentoft et al., 2021) to allow MPC at intermittently aerated WRRFs allowing taking e.g. electricity cost and greenhouse gas emissions into consideration as additional constraints beside effluent quality.

Using digital twins for control would likely result in a different structure than conventional MPC. Some examples have been presented (but are still not scientifically published, e.g., Sparks et al. (2023)) where simplified plant models are combined with influent forecasts (see Section 4) and optimization algorithms to produce process setpoints. The regular PID controllers at the plant can then be used to achieve these setpoints. This can be a good approach for processes such as aeration, where it can be challenging to use a full mechanistic model for fast control actions. This type of digital twin-based optimization control is shown in Figure 3.

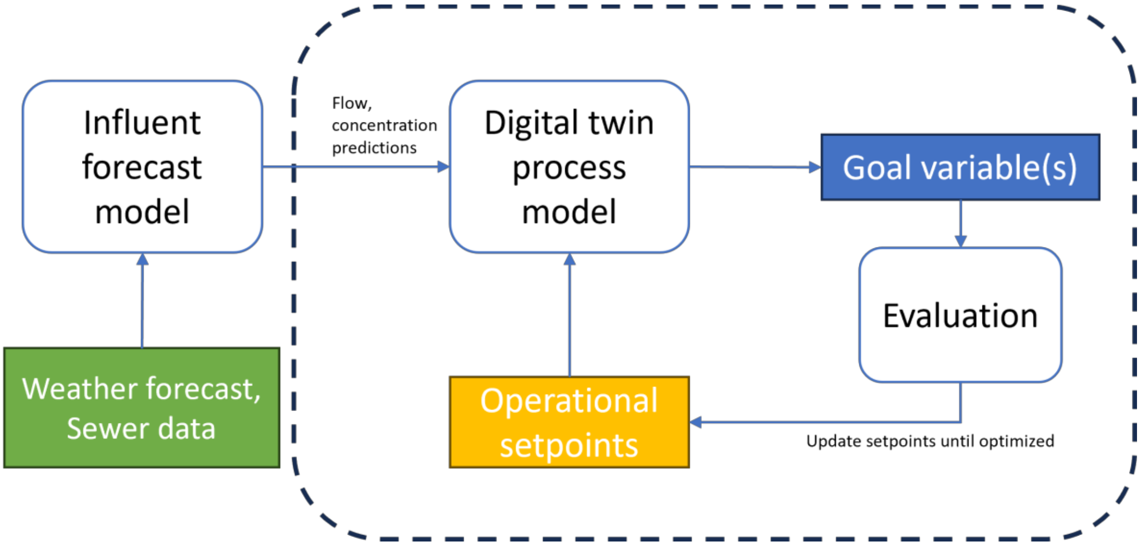

Influent forecast model
Flow, concentration predictions
Digital twin process model
Goal variable(s)
Evaluation
Weather forecast, Sewer data
Operational setpoints
Update setpoints until optimized

Figure 3. Scheme for digital twin-based optimization.

7 Case studies

Very few examples exist in the literature of implemented digital twins at water resource recovery facilities, although many more are known to be implemented and currently under development within Europe and North America. Some cases also exist with digital twins of sewer networks (i.e., Future City Flow, which was mentioned earlier in this report).

The first (according to the author's knowledge) implemented WRRF digital twin was developed for Changi Water Reclamation Plant in Singapore, and has been presented by e.g., Valverde-Peréz et al. (2021) and Johnson et al. (2021). The DT is constructed from a mechanistic process model, built in the simulator Sumo (Dynamita, France). The twin also features a separate more detailed model for the hydraulics and controls of the plant. Forecasts include five days into the future performance of the plant, with uncertainty included by Monte Carlo analysis for feasible variations in influent characteristics. The influent characteristics are derived from measurements of the primary effluent ammonia concentration. A model of the plant upstream the measurement point is used to find the required influent ammonia concentrations to produce the observed primary effluent concentrations. The remaining variables are calculated through observed relationships such as NH_4-N/TKN (total Kjeldahl nitrogen) and TKN/COD ratios. Automatic calibration of the model is performed weekly by adjusting the soluble COD/total COD ratio in the influent as well as solids removal in the primary clarifier, based on laboratory measurements. Several machine learning methods are used for data treatment and gap filling, including: isolation forests; interquartile range; K means clustering; sequential least squares programming; and ARIMA. The model is (currently) only used as an advisory tool, not for direct control purposes.

8 Conclusions

Digital twins of water resource recovery facilities are complex with many different parts and models that work together. They can be used for fault detection, predictions, and optimization/control. This report summarizes some of the components that can be used to build digital twins, which ones to include of course depends on the scope and goals of the specific project. In all cases, the flow of data from collection to use must be well designed to avoid unnecessary interruptions in operation.

Much work is ongoing on the topic of digital twins around the world. Likely, many articles and other publications will be published in the next few years highlighting methods and case studies, and above all the value of using digital twins in the daily operations of plants. This will lead to more use cases and application areas and will hopefully lead to the more proactive plant decision making that it has the potential to.

Acknowledgements

This work has been performed during the project *Digital twin for sustainable and resource efficient operation of wastewater treatment plants*. The project is funded by the Swedish Research Council Formas (2020-00222), the utility organization Nordvästra Skånes Vatten och Avlopp (NSVA) AB and the Swedish Water and Wastewater Association (SVU) (through the research cluster VA-kluster Mälardalen). In-kind funding is provided by the project partners RISE, Lund University, NSVA and Gemit Solutions AB.

References

Aguado, D., & Rosen, C. (2008). Multivariate statistical monitoring of continuous wastewater treatment plants. *Engineering Applications of Artificial Intelligence*, *21*(7), 1080–1091. https://doi.org/10.1016/j.engappai.2007.08.004

Alex, J., Benedetti, L., Copp, J., Gernaey, K. V, Jeppsson, U., Nopens, I., Pons, M.-N., Rieger, L., Rosen, C., Steyer, J. P., Vanrolleghem, P., Winkler, S., Magdeburg, I. E. V, & Benedetti, G. L. (2008). *Benchmark Simulation Model no. 1 (BSM1)*. Report by the IWA Taskgroup on benchmarking of control strategies for WWTPs.

Andersson, S. L., Åmand, L., Samuelsson, O., & Nilsson, S. (2019). *Correct instrumentation at wastewater treatment plants [Instrumentera rätt på avloppsreningsverk, in Swedish]*. Stockholm, Sweden: Svenskt Vatten Utveckling.

Baggiani, F., & Marsili-Libelli, S. (2009). Real-time fault detection and isolation in biological wastewater treatment plants. *Water Science and Technology*, *60*(11), 2949–2961. https://doi.org/10.2166/wst.2009.723

Bernardelli, A., Marsili-Libelli, S., Manzini, A., Stancari, S., Tardini, G., Montanari, D., Anceschi, G., Gelli, P., & Venier, S. (2020). Real-time model predictive control of a wastewater treatment plant based on machine learning. *Water Science and Technology*, *81*(11), 2391–2400. https://doi.org/https://doi.org/10.2166/wst.2020.298

Brentan, B. M., Luvizotto, E., Herrera, M., Izquierdo, J., & Pérez-García, R. (2017). Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics*, *309*, 532–541. https://doi.org/10.1016/j.cam.2016.02.009

Caccavale, F., Digiulio, P., Iamarino, M., Masi, S., & Pierri, F. (2010). A neural network approach for on-line fault detection of nitrogen sensors in alternated active sludge treatment plants. *Water Science and Technology*, *62*(12), 2760–2768. https://doi.org/10.2166/wst.2010.025

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural Ordinary Differential Equations. *Proceedings from 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), 3-8 December 2018, Montréal, Canada*, 6572–6583.

Choi, S. W., & Lee, I. B. (2005). Multiblock PLS-based localized process diagnosis. *Journal of Process Control*, *15*(3), 295–306. https://doi.org/10.1016/j.jprocont.2004.06.010

De Jaegher, B., De Schepper, W., Verliefde, A., & Nopens, I. (2021). Enhancing mechanistic models with neural differential equations to predict electrodialysis fouling. *Separation and Purification Technology*, *259*, 118028. https://doi.org/10.1016/j.seppur.2020.118028

De Mulder, C., Flameling, T., Weijers, S., Amerlinck, Y., & Nopens, I. (2018). An open software package for data reconciliation and gap filling in preparation of water and resource recovery facility modeling. *Environmental Modelling and Software*, *107*, 186–198. https://doi.org/10.1016/j.envsoft.2018.05.015

Ferrer, A., Aguado, D., Vidal-Puig, S., Prats, J. M., & Zarzo, M. (2008). PLS: A versatile tool for industrial process improvement and optimization. *Applied Stochastic Models in Business and Industry*, *24*(6), 551–567. https://doi.org/https://doi.org/10.1002/asmb.716

Gernaey, K. V., Jeppsson, U., Vanrolleghem, P. A., & Copp, J. B. (2014). *Benchmarking of control strategies for wastewater treatment plants*. London, UK: IWA Publishing.

Gómez, C., Solon, K., Nopens, I., & Torfs, E. (2023). Automatic plant-wide calibration of a water resource recovery facility. *Proceedings from the 8th Water Resource Recovery Modelling Seminar (WRRmod 2022+), 18-22 January 2023, Stellenbosch, South Africa*.

Haimi, H., Mulas, M., Corona, F., Marsili-Libelli, S., Lindell, P., Heinonen, M., & Vahala, R. (2016). Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant. *Engineering Applications of Artificial Intelligence*, *52*, 65–80. https://doi.org/10.1016/j.engappai.2016.02.003

Han, H. G., Qiao, J. F., & Chen, Q. L. (2012). Model predictive control of dissolved oxygen concentration based on a self-organizing RBF neural network. *Control Engineering Practice*, *20*(4), 465–476. https://doi.org/10.1016/j.conengprac.2012.01.001

Henze, M., Gujer, W., Mino, T., & van Loosdrecht, M. C. M. (2000). *Activated sludge models ASM1, ASM2, ASM2d and ASM3*. IWA Scientific and Technical Report No. 9. London, UK: IWA Publishing.

Hu, G., Zhang, Z., Chen, J., Zhang, Z., Armaou, A., & Yan, Z. (2021). Elman neural networks combined with extended Kalman filters for data-driven dynamic data reconciliation in nonlinear dynamic process systems. *Industrial and Engineering Chemistry Research*, *60*(42), 15219–15235. https://doi.org/10.1021/acs.iecr.1c02916

Hvala, N., & Kocijan, J. (2020). Design of a hybrid mechanistic/Gaussian process model to predict full-scale wastewater treatment plant effluent. *Computers and Chemical Engineering*, *140*, 1–12. https://doi.org/10.1016/j.compchemeng.2020.106934

Johnson, B. R., Kadiyala, R., Owens, G., Ping, M. Y., Grace, P., 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, 16-20 October 2021, Chicago, USA*.

Kidane, L., Townend, P., Metsch, T., & Elmroth, E. (2022). When and how to retrain machine learning-based cloud management systems. *Proceedings - 2022 IEEE 36th International Parallel and Distributed Processing Symposium Workshops, IPDPSW, 30 May - 3 June 2022, Lyon, France*, 688–698. https://doi.org/10.1109/IPDPSW55747.2022.00120

Kim, J. R., Ko, J. H., Im, J. H., Lee, S. H., Kim, S. H., Kim, C. W., & Park, T. J. (2006). Forecasting influent flow rate and composition with occasional data for supervisory management system by time series model. *Water Science and Technology*, *53*(4–5), 185–192. https://doi.org/10.2166/wst.2006.123

Le, Q. H., Verheijen, P. J. T., van Loosdrecht, M. C. M., & Volcke, E. I. P. (2018). Experimental design for evaluating WWTP data by linear mass balances. *Water Research*, *142*, 415–425. https://doi.org/10.1016/j.watres.2018.05.026

Lee, C., Choi, S. W., & Lee, I. B. (2004a). Sensor fault identification based on time-lagged PCA in dynamic processes. *Chemometrics and Intelligent Laboratory Systems*, *70*(2), 165–178. https://doi.org/10.1016/j.chemolab.2003.10.011

Lee, D. A. E. S., Jeon, C. O., Park, J. M., & Chang, K. S. (2002). Hybrid neural network modeling of a full-scale industrial wastewater treatment process. *Biotechnology and Bioengineering*, *78*(6), 670–682. https://doi.org/10.1002/bit.10247

Lee, D. A. E. S., Park, J. M., & Vanrolleghem, P. A. (2005). Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor. *Journal of Biotechnology*, *116*(2), 195–210. https://doi.org/10.1016/j.jbiotec.2004.10.012

Lee, D. A. E. S., & Vanrolleghem, P. A. (2004). Adaptive concensus principal component analysis for on-line batch process monitoring. *Environmental Monitoring and Assessment*, *92*, 119–135. https://doi.org/https://doi.org/10.1023/b:emas.0000014498.72455.18

Lee, D. A. E. S., Vanrolleghem, P. A., & Jong, M. P. (2005). Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. *Journal of Biotechnology*, *115*(3), 317–328. https://doi.org/10.1016/j.jbiotec.2004.09.001

Lee, J. M., & Qin, S. J. (2006). Fault detection and diagnosis based on modified independent component analysis. *AIChE Journal*, *52*(10), 3501–3514. https://doi.org/https://doi.org/10.1002/aic.10978

Lee, J. M., Yoo, C. K., Choi, S. W., Vanrolleghem, P. A., & Lee, I. B. (2004b). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, *59*(1), 223–234. https://doi.org/10.1016/j.ces.2003.09.012

Lee, J. M., Yoo, C. K., & Lee, I. B. (2004c). Statistical process monitoring with independent component analysis. *Journal of Process Control*, *14*(5), 467–485. https://doi.org/10.1016/j.jprocont.2003.09.004

Lennox, J., & Rosen, C. (2002). Adaptive multiscale principal components analysis for online monitoring of wastewater treatment. *Water Science and Technology*, *45*(4–5), 227–235. https://doi.org/10.2166/wst.2002.0593

Li, F., & Vanrolleghem, P. A. (2022a). An essential tool for WRRF modelling: A realistic and complete influent generator for flow rate and water quality based on data-driven methods. *Water Science and Technology*, *85*(9), 2722–2736. https://doi.org/10.2166/wst.2022.095

Li, F., & Vanrolleghem, P. A. (2022b). Including snowmelt in influent generation for cold climate WRRFs: Comparison of data-driven and phenomenological approaches. *Environmental Science: Water Research and Technology*, *8*(10), 2087–2098. https://doi.org/10.1039/d1ew00646k

Liu, Y., Pan, Y., Sun, Z., & Huang, D. (2014). Statistical monitoring of wastewater treatment plants using variational Bayesian PCA. *Industrial and Engineering Chemistry Research*, *53*(8), 3272–3282. https://doi.org/10.1021/ie403788v

Lumley, D., Lindell, P., Lindqvist, P., & Heinonen, M. (2013). Experience using auto-adjustment for improving sensor signal robustness at two large wastewater treatment plants. *Proceedings of the 11th IWA Conference on Instrumentation, Control and Automation, 18-20 September 2013, Narbonne, France.*

Marais, H. L., Zaccaria, V., & Odlare, M. (2022). Comparing statistical process control charts for fault detection in wastewater treatment. *Water Science and Technology*, *85*(4), 1250–1262. https://doi.org/10.2166/wst.2022.037

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*, 498–513. https://doi.org/10.1016/j.watres.2019.03.030

Nielsen, R. F., Nazemzadeh, N., Sillesen, L. W., Andersson, M. P., Gernaey, K. V., & Mansouri, S. S. (2020). Hybrid machine learning assisted modelling framework for particle processes. *Computers and Chemical Engineering*, *140*. https://doi.org/10.1016/j.compchemeng.2020.106916

Nivert, G., Lindqvist, P., Lumley, D., Rosenqvist, F., & Gunnarsson, J. (2009). Implementing auto-adjustment and auto-validation of on-line instrument signals. *Proceedings of the 10th IWA Conference on Instrumentation, Control and Automation, 14-17 June 2009, Cairns, Australia.*

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, *41*(1), 100–115. https://about.jstor.org/terms

Pedersen, J. W., Larsen, L. H., Thirsing, C., & Vezzaro, L. (2020). Reconstruction of corrupted datasets from ammonium-ISE sensors at WRRFs through merging with daily composite samples. *Water Research*, *185*, 116227. https://doi.org/10.1016/j.watres.2020.116227

Quaghebeur, W., Torfs, E., De Baets, B., & Nopens, I. (2022). Hybrid differential equations: Integrating mechanistic and data-driven techniques for modelling of water systems. *Water Research*, *213*, 118166. https://doi.org/10.1016/j.watres.2022.118166

Rauch, W., & Harremoës, P. (1999). Genetic algorithms in real time control applied to minimize transient pollution from urban wastewater systems. *Water Research*, *33*(5), 1265–1277. https://doi.org/10.1016/S0043-1354(98)00304-2

Rieger, L., Takács, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P. A., & Comeau, Y. (2010). Data reconciliation for wastewater treatment plant simulation studies-planning for high-quality data and typical sources of errors. *Water Environment Research*, *82*(5), 426–433. https://doi.org/10.2175/106143009x12529484815511

Roberts, S. W. (1958). Properties of control chart zone tests. *Bell System Technical Journal*, *37*(1), 83–114. https://doi.org/10.1002/j.1538-7305.1958.tb03870.x

Rosen, C., & Olsson, G. (1998). Disturbance detection in wastewater treatment plants. *Water Science and Technology*, *37*(12), 197–205. https://doi.org/https://doi.org/10.1016/S0273-1223(98)00372-2

Sánchez-Fernández, A., Baldán, F. J., Sainz-Palmero, G. I., Benítez, J. M., & Fuente, M. J. (2018). Fault detection based on time series modeling and multivariate statistical process control. *Chemometrics and Intelligent Laboratory Systems*, *182*, 57–69. https://doi.org/10.1016/j.chemolab.2018.08.003

Schneider, M. Y., Quaghebeur, W., Borzooei, S., Froemelt, A., Li, F., Saagi, R., Wade, M. J., Zhu, J.-J., & Torfs, E. (2022). Hybrid modelling of water resource recovery facilities: Status and opportunities. *Water Science and Technology*, *85*(9), 2503–2524. https://doi.org/10.2166/wst.2022.115

Schule, M., Lang, H., Springer, M., Kemper, A., Neumann, T., & Gunnemann, S. (2021). In-Database Machine Learning with SQL on GPUs. *SSDBM '21: Proceedings of the 33rd International Conference on Scientific and Statistical Database Management, 6-7 July 2021, Tampa, Florida, USA*, 25–36. https://doi.org/10.1145/3468791.3468840

Shen, W., Chen, X., & Corriou, J. P. (2008). Application of model predictive control to the BSM1 benchmark of wastewater treatment process. *Computers and Chemical Engineering*, *32*(12), 2849–2856. https://doi.org/10.1016/j.compchemeng.2008.01.009

Shen, W., Chen, X., Pons, M. N., & Corriou, J. P. (2009). Model predictive control for wastewater treatment process with feedforward compensation. *Chemical Engineering Journal*, *155*(1–2), 161–174. https://doi.org/10.1016/j.cej.2009.07.039

Shewhart, W. A. (1938). Application of statistical methods to manufacturing problems. *J. Franklin Inst.*, *226*(2), 163–186. https://doi.org/https://doi.org/10.1016/S0016-0032(38)90436-3

Sokolova, E., Ivarsson, O., Lillieström, A., Speicher, N. K., Rydberg, H., & Bondelind, M. (2022). Data-driven models for predicting microbial water quality in the drinking water source using E. coli monitoring and hydrometeorological data. *Science of the Total Environment*, *802*, 149798. https://doi.org/10.1016/j.scitotenv.2021.149798

Sparks, J., Vanrolleghem, P. A., & Bott, C. (2023). Advanced Ammonia Based Aeration Control (ABAC) using predictive modelling. In *Advanced control systems for nitrogen removal in full-scale water facilities*. Presentation in International Water Association (IWA) webinar, 26 July 2023. https://iwa-network.org/learn/nitrogen-removal/

Stentoft, P. A., Munk-Nielsen, T., Møller, J. K., Madsen, H., Valverde-Pérez, B., Mikkelsen, P. S., & Vezzaro, L. (2021). Prioritize effluent quality, operational costs or global warming? – Using predictive control of wastewater aeration for flexible management of objectives in WRRFs. *Water Research*, *196*, 116960.

Stentoft, P. A., Vezzaro, L., Courdent, V., Pedersen, J. W., Thomsen, H., Mikkelsen, P. S., Tisserand, B., & Amiel, C. (2019). Real time forecasting of flows and loads to WWTPs for enhanced hydraulic and biological capacity during stormwater events. *Proceedings of 10th Edition of the Novatech Conference, 1-5 July 2019, Lyon, France.*

Su, H. Te, McAvoy, T. J., & Werbos, P. (1992). Long-term predictions of chemical processes using recurrent neural networks: A parallel training approach. *Industrial and Engineering Chemistry Research*, *31*(5), 1338–1352. https://doi.org/10.1021/ie00005a014

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 Science and Technology*, *82*(12), 2613–2634. https://doi.org/10.2166/wst.2020.393

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 and Technology*, *85*(10), 2840–2853. https://doi.org/10.2166/wst.2022.107

Trauer, J., Schweigert-Recksiek, S., Engel, C., Spreitzer, K., & Zimmermann, M. (2020). What is a digital twin? Definitions and insights from an industrial case study in technical product development. *Proceedings of the Design Society: DESIGN Conference, Volume 1, May 2020*, 757–766. https://doi.org/10.1017/dsd.2020.15

Valverde-Pérez, B., Johnson, B., Wärff, C., Lumley, D., Torfs, E., Nopens, I., & Townley, L. (2021). *Operational digital twins in the urban water sector: case studies*. IWA Digital Water White Paper series. London, UK: International Water Association.

Vezzaro, L., Pedersen, J. W., Larsen, L. H., Thirsing, C., Duus, L. B., & Mikkelsen, P. S. (2020). Evaluating the performance of a simple phenomenological model for online forecasting of ammonium concentrations at WWTP inlets. *Water Science and Technology*, *81*(1), 109–120. https://doi.org/10.2166/wst.2020.085

Wang, L., & Shi, H. (2010). Multivariate statistical process monitoring using an improved independent component analysis. *Chemical Engineering Research and Design*, *88*(4), 403–414. https://doi.org/10.1016/j.cherd.2009.09.002

Wei, Y., Law, A. W. K., & Yang, C. (2022). Real-time data-processing framework with model updating for digital twins of water treatment facilities. *Water*, *14*(22), 3591–14. https://doi.org/10.3390/w14223591

Yoo, C. K., Choi, S. W., & Lee, I. B. (2002). Dynamic monitoring method for multiscale fault detection and diagnosis in MSPC. *Industrial and Engineering Chemistry Research*, *41*(17), 4303–4317. https://doi.org/10.1021/ie0105730

Zhu, W., Zhang, Z., Chen, J., Zhao, S., & Huang, S. (2021). Dynamic data reconciliation to enhance the performance of feedforward/feedback control systems with measurement noise. *Journal of Process Control*, *108*, 12–24. https://doi.org/10.1016/j.jprocont.2021.10.003

Zumoffen, D., Basualdo, M., & Molina, G. (2008). Improvements in fault tolerance characteristics for large chemical plants: 2. Pulp mill process with model predictive control. *Industrial and Engineering Chemistry Research*, *47*(15), 5482–5500. https://doi.org/10.1021/ie800100r

Through our international collaboration programmes with academia, industry, and the public sector, we ensure the competitiveness of the Swedish business community on an international level and contribute to a sustainable society. Our 2,800 employees support and promote all manner of innovative processes, and our roughly 100 testbeds and demonstration facilities are instrumental in developing the future-proofing of products, technologies, and services. RISE Research Institutes of Sweden is fully owned by the Swedish state.

I internationell samverkan med akademi, näringsliv och offentlig sektor bidrar vi till ett konkurrenskraftigt näringsliv och ett hållbart samhälle. RISE 2 800 medarbetare driver och stöder alla typer av innovationsprocesser. Vi erbjuder ett 100-tal test- och demonstrationsmiljöer för framtidssäkra produkter, tekniker och tjänster. RISE Research Institutes of Sweden ägs av svenska staten.

RISE Research Institutes of Sweden AB
Box 857, 501 15 Borås
Telephone: +46 10-516 50 00
E-mail: info@ri.se, Internet: www.ri.se

Circular wastewater systems
RISE Report 2023:105
ISBN: