CODEN:LUTEDX/(TEIE-7276)/1-20/(2021)

Industrial Electrical Engineering and Automation

Implementation of digital twins at water resource recovery facilities

Hanna Molin

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

Implementation of digital twins at water resource recovery facilities

Identification and description of challenges and enablers for the use of real-time process models for decision making within the water and wastewater sector and other relevant industries

Hanna Molin

2021-03-02

Table of contents

1. Introduction

Models to describe the processes occurring in water resource recovery facilities (WRRFs) have been developed and used over the last decades (Jeppsson, 1999; Gernaey et al., 2004). In many applications the process models are used for design and optimization but are usually run separate from the plant itself. The increased knowledge and understanding of the value of process modelling in recent years, together with development in IT and data access are driving forces that pushes development towards extending process models. One extension is to have the process model run simultaneously with the plant, and to feed it with real-time data from the plant – i.e. to create a 'digital twin'.

The purpose of this study is to describe how digital twins have been implemented previously, both in WRRFs and in other relevant industries, and to identify key enabling factors and technologies. Literature reviews on digital twins have been made by many authors before (see e.g. Negri et al., 2017; Barricelli et al., 2019; Kritzinger et al., 2018). Although this study is more of a literature review than anything else, it is in no way complete. It is rather a first scan of the field to provide insight for further research. Instead of conducting yet another extensive review the aim has been to find case-studies on the topic, and the focus has been on the *implementation* and *application* of Digital Twins. Related topics within water and wastewater resource management, such as sewer flow models and soft sensors, are also discussed.

2. The definition of a digital twin

Although the term *digital twin* has been given a lot of attention both in research and in media in recent years, an unambiguous definition of the concept is yet to be found. The United States National Aeronautics and Space Administration (NASA) presented an early definition (Glaesson & Stargel, 2012):

> *"A digital twin is an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin."*

This definition is quite specific to the aerospace sector but is often referred to in literature. Many authors have since made their contributions in the field and have presented broader definitions (see for example Negri et al., 2017; Eyre et al., 2018; Zhao et al., 2020). To summarize, a digital twin is a virtual representation of the real entity that interact with a virtual space through a communication network. The digital twin includes both static and dynamic information and can have 'intelligence' (Schröder et al., 2016). Digital twins are a real-time reflection of a physical asset, achieved through continuous communication and synchronization between the real entity and the digital twin states so that they can interact with current and historical data of the real asset. This makes it possible for the digital twin to continuously improve performance and directly compare and analyze predicted and measured values of the physical asset (Martinez et al., 2018).

In this work, a digital twin for smart operation of WRRFs is defined as a combination of different models that are used in real-time for collection, control, validation, simulation and visualization of data to aid operators in decision making. The model parameters are continuously updated and the models are automatically calibrated.

3. Components in the digital twin system

Simplified, the digital twin system consists of three main parts: 1) physical entities in the physical or real space, 2) virtual entities in virtual space, and 3) the communication between the two where data and information can be seamlessly shared between the physical entity and its virtual counterpart (Grieves, 2015). Outside of this system is also the user space, which in some applications is regarded to belong to the digital twin system (Bevilacqua et al., 2020).

The physical entity can be a product, a process, or a system. Within manufacturing, the physical entity often is a product but might as well be the production plant or the complete supply chain (Grieves, 2015; Söderberg et al., 2017; Alcáer & Cruz-Machado, 2019). For applications in WRRFs, the physical entity is the plant itself with biological, chemical and physical processes, as well as equipment such as pumps etc., and the SCADA (Supervisory Control And Data Acquisition) system. The data from the physical entity is collected using sensors, laboratory data or data from actuators (Wanasinghe et al., 2020).

The virtual entity are models of the physical entity. Within manufacturing it is often 3D models that can be used for geometry assurance, but applications where the production plant or supply chain is modelled are also common. The models can be physical-based (such as 3D models) or data-driven where historical data are used for forecasting, often made possible by the use of machine learning or artificial intelligence algorithms (Kritzinger et al., 2018). For WRRFs the models include process models, hydraulic models, control system models amongst others.

3.1. Communication and computation

Connected to the three main building blocks (physical space, virtual space, and communication between them) of the digital twin system, there are mainly three types of data that must be coordinated: data from the physical entity (measurements), data from the virtual entity (models, descriptions of the system's configuration and construction), and data from simulations and analyses of the system (Cameron et al., 2018). The communication system between the physical and virtual entity is thereby essential to have a functional digital twin. Companies like ABB and Siemens nowadays offer software and hardware for implementation of digital twins with communication protocols that seamlessly integrate the physical and virtual entities (ABB, 2021; Siemens, 2021). Standardized communication platforms and protocols such as FMI (Functional Mock-up Interface) or OPC UA (Open Platform Communications Unified Architecture) are thought to be key enablers for digitalization of the industry. FMI is a standard that is used to exchange dynamic models into a single file which speed up communication and facilitates communication between different software or protocols (Blochwitz, 2012). OPC UA is a secure communication protocol for machine-to-machine communication. It is platform-independent which means it can operate on several hardware platforms and operating systems. This feature makes it suitable for communication between different machines as well as with operators (OPC Foundation, 2021; Övern, 2018).

Edge, cloud, and fog computing are also thought to be enablers for digitalization. In cloud computing, the data is processed and stored on multiple servers and can be accessed online from any device which offers benefits compared to traditional storage on local hard drives. Fog computing is a mix of cloud computing and more traditional data storage systems, where computing is made locally on decentralized servers. Edge computing is performed on so called edge devices away from centralized storage which keeps the information on local parts of the network. The data does not have to be sent to the centralized cloud but is rather processed directly on the edge device. These three technologies have different areas of use, but all facilitate data storage and processing. Fog and edge computing decrease latency and offers greater security compared to cloud computing which can enable the use of process models in real time – one of the core concepts of the digital process twin.

4. Applications of digital twins

The digital twin can be designed and used for monitoring, simulations, optimization, and verification of different activities in the real-life entity/process (Martinez et al., 2018). Most implementations are presently found in the manufacturing industries. Modelling and simulation tools have been used for many years in process industries (such as chemical, mineral processing, pulp and paper industries etc.) but even though the models are run close to real-time, the term digital twin is rarely mentioned within this field. Case studies from the two different fields are presented in the following chapters, separated since the manufacturing and production industries in general are simulated based on discrete events, whereas the process industries operate continuously. In that sense, WRRF:s are more alike process industries, but knowledge and experience from the manufacturing and production industries might still be valuable.

4.1. Manufacturing and production

The digital twins within manufacturing are still in its infancy, but benefits and applications of digital twins within this field are thought to be many, and an enabler of the ongoing industrial revolution Industry 4.0 (Zheng et al., 2019; Uhlemann et al., 2017). With applications such as production planning and control, maintenance, and layout planning, a smarter and more efficient production can be achieved by using digital twins (Kritzinger et al., 2018). Digital twins can be used for product design, smart manufacturing, and for usage monitoring to simulate the product's life cycle. A digital twin of a product enables an iterative and optimized product design. Product functions, behaviour, structures, and manufacturability can be verified in the virtual space before going into production. If coupled to a digital factory twin, the resources and capacities can be allocated to set up an optimal production plan on beforehand, which facilitates the manufacturing of the product (Qi & Tao, 2018). Digital twins give opportunities to support operators in the day to day operation, but also to plan services and maintenance, and make forecasts by using the digital twin for simulations (Rosen et al., 2015).

Vachálek et al. (2017) created a digital twin system consisting of the physical production line and a digital copy, with an interface through which the data was exchanged. The digital model was created in a simulation tool (*Plant Simulation* developed by Siemens). The data was transferred via an OPC data server. The simulation set up created by Vachálek et al. (2017) can be used for multiple purposes, including analyzing effects of changed parameters, or to discover errors/malfunctions in the production.

Zheng et al. (2019) built a digital twin system of a welding production line including a welding machine for weld assembly, a transporter for delivery, and a multi-layered warehouse for storage. The digital twin consisted of 3D geometric models to describe the shapes and sizes of devices, physical models to define the mechanical and thermodynamic states of devices, and kinematic models that receives kinematic data used to update all components positions in the devices in the real plant. An OPC UA server was used for communication. The digital twin system is said to be able to ensure operation efficiency of equipment in the production line and provide data to enhance welding quality.

Söderberg et al. (2017) used a digital twin for geometry assurance. The digital twin was used in the design phase where it was fed with variation data for parts and fixtures, and later used as a real-time controller for the assembly system in the production phase. The produced parts are scanned, and the data is stored as deviations from a predetermined nominal (defined during the design phase). The parts are sorted accordingly. In this particular case, the produced parts are adjusted and matched pairwise for further assembly in order to minimize deviation after welding. The digital twin technology used creates new possibilities to adjust the manufacturing process to compensate for geometrical deviations.

Bottani et al. (2017) created a digital twin of an automated guided vehicle (AGV) used in a production system. The production system consisted of four processing stations between which the AGV could move along a predefined circuit, while the digital twin AGV made decisions based on orders. Decisions were based on a self-adaptive algorithm that used data from each processing station. A number of optimization problems were tested, such as minimizing the logistic time, and the processing time. The study showed that it is possible to connect cyber physical systems, digital twins and Big Data to achieve decision-making autonomy and self-adaption of machines.

4.2. Process industries

Modelling and simulation tools have been used for many years in process industries, mainly to optimize the process to make it more efficient and thereby profitable. These tools can also be used to learn how to tackle and to minimize the occurrence of shutdowns/breakdowns. Simulation-based applications, of which some might be regarded as a sort of digital twin, are already in use in many process industries. The applications include plant monitoring, plant forecasting, open-loop decision support, and closed-loop control and automation (Martinez et al. 2018; Pantelides & Renfro, 2013).

Digital twins can be used to reduce fluctuations in industrial processes, as shown by He et al. (2019). They implemented a data-driven digital twin system by integrating process monitoring, diagnosis, and optimized control and managed to reduce severe fluctuations in a chemical process plant. The study included sensor faults, actuator faults, and process disturbances and developed a method used to accommodate different faults. The digital twin system was developed to shorten the downtime of a factory by extending the operation time of any faulty process plant with an acceptable performance. This is achieved by timely monitoring, diagnosis, and tolerance control.

The most anticipated applications areas for digital twins in the oil and gas industry is asset monitoring, asset maintenance, project planning, and lifecycle management. Other important applications are collaboration and virtual learning and training (Wanasinghe et al., 2020). Maximizing profit and plan production are also important applications. Makarov et al. (2019) created a simulation model to optimize the profit of an oil production enterprise. Artificial neural networks where used to estimate input characteristics to predict the average daily production rate. This method made it possible to update the simulation model in real time based on the physical characteristics of the oil field.

The Swedish companies Calejo Industrial Intelligence and SCA recently implemented a digital process twin at a pulp and paper plant in Obbola, Sweden. It was integrated with a 3D model of the plant itself, creating a versatile tool that mainly will be used to reduce the need for fossil fuel. Turbines used in the production process are run on steam or oil and by forecasting the need for steam in a 50h horizon it is possible to plan operations and reduce the need for oil which has both environmental and economic benefits (CIO Sweden, 2020).

Vattenfall developed a digital twin of parts of its nuclear power plant Ringhals. The digital twin is used for monitoring and fault detection. Models of the turbines predict a theoretical power production and compares it with the measured production. The turbines are cooled with sea water, causing a risk for biological growth on the condenser which reduces its capacity. As the difference between the modelled and measured data increases, an indication that cleaning is required is sent to the operators (Ny Teknik, 2018).

Sun et al. (2020) have proposed a multilayer cyber-physical system (CPS) based supervision management framework for real-time control of an urban water cycle. The multilayer CPS-based supervision management framework consists of four layers: (1) Supervision and control layer, (2) Scheduling layer, (3) Digital twin layer, and (4) Water users, and Environment layer. Each layer is interconnected with the physical world. The supervision and control layer is connected to the real control system. Advanced control algorithms and conceptual models of the urban water cycle are used to solve control problems with goals defined in the fourth layer. The scheduling layer is used to compute the supervisory control into set-points for actuators and to send sensor data to the Supervision and control layer. The digital twin layer consists of high-fidelity digital models that replicate the assets in the urban water cycle (treatment plants, sewer network etc.). The digital twin can be designed and used for various purposes, e.g. soft sensors. Finally, in the Water users and Environment layer, the impact of the control actions in the cyber physical system are evaluated.

5. Related work

A related topic to digital twin systems is using soft sensors for monitoring of treatment plants. Äijälä and Lumley (2006) integrated a soft sensor model for flow control at the Swedish WRRF Ryaverket in Gothenburg. The concept was quite simple. They used a few measured flows to estimate flows at various points in the WRRF where flow measurements were hard to accomplish. Instead of using a traditional hydrodynamic model, that could be beneficial but difficult to implement and maintain, they created flow balance models based on available on-line instrumentation. The control model was found to be effective in monitoring flows in the WRRF. It also functions as a fault detector for physical flow meters as it can detect when the modelled value and the measured value differs from each other (Äijälä and Lumley, 2006).

In the city of Copenhagen, a couple of digitalization projects within water management are currently being implemented. Using machine-learning algorithms they created a tool for forecasting flow in the sewer network and inflow to the WRRF. It uses real-time data (water level and flow) from the sewer system and combines it with rain gauge data and weather radar observations, nowcasts and forecasts. The toolbox is currently deployed in Copenhagen covering a catchment of ca 55 km^2. The toolbox is to be complemented with a decision tool for operation of the integrated sewer network and WRRF. The aim is to create a tool that can be used to optimize treatment capacities, pumping strategies and in-sewer flow allocation by combining level and flow sensors in the sewer network, WRRF operation data and the sewer flow forecast toolbox (Digital water city, n.d.). A similar project, *Furture City Flow*, is currently conducted in Sweden and Norway. The goal is to develop a decision support system to facilitate the information flow between different stakeholders within urban water management. Pilot trials are undertaken in Helsingborg, Gothenburg, Trollhättan, Malmö/Lund (Sweden) and Oslo (Norway) (Sweden Water Research, 2020).

6. Implementation of digital twins in WRRFs

Digitalization in the water and wastewater sector is ongoing. Implementing digital technologies in WRRFs, digital twins among others, are expected to offer many benefits, not only from an operational point of view. Applications of digital twins within WRRFs include, but are not limited to, optimization of the treatment process, predictive maintenance and fault detection, model predictive control, soft sensor implementation, forecasting and planning, and operators' training. In a broader perspective, this may lead to lower environmental impact (both with regard to energy usage and effluent discharge) and thereby better regulatory compliance. Financial benefits include reduced operational and maintenance costs and increased revenue (Sarni et al., 2019). Challenges to full-scale implementation of digital twins at WRRFs, key enabling factors and technologies to mitigate the challenges, and an outlook for applications of digital twins in WRRFs are discussed in the following sections.

6.1. Challenges for full-scale implementation of digital twins

A general challenge for full-scale implementation of digital twins is the lack of a clear definition and standardization of the concept (Kritzinger et al., 2018; Fuller et al., 2020). Efforts have been made to create frameworks for development and implementation of digital twins, although none have yet gained public acceptance (see e.g. Schleich et al., 2018; Shao & Kibira, 2019; Autiosalo et al., 2020; Eyre et al., 2020). Furthermore, the systems are often heterogenous which can make them difficult to model and monitor, but it also impedes the use of standardized solutions (Fuller et al., 2020).

The challenges affiliated to implementation of digital twins can be divided into three categories: (1) data management (2), technical, and (3) operational challenges. The challenges related to data management ranges from the acquisition to storage and ownership. Data is acquired from sensors, equipment and laboratory test. Sensors must have high accuracy and reliability and are in need of regular maintenance and calibration (Eerikäinen et al., 2020). The amount of data is extensive and comes with noise and missing data.

Therefore, data must often be preprocessed. Algorithms that clean raw data, detect outliers, fill missing data and present it in an ordered, simplified form is essential (Lu et al., 2020; Rasheed et al., 2020).

Data quality is key if the digital twin is to produce interpretable and reliable results. Models require tuning which is normally done by minimizing the difference between the acquired data from the process and the modelled data. Although, the differences between the two may stem from faulty sensors, uncertainty in data, missing components in the model or malfunction at the plant which give rise to the need of algorithms for data validation and fault detection (Rasheed et al., 2020; Wanasinghe et al., 2020). Maintaining the models is also needed to prevent the model from drifting.

Large amount of data will be transmitted between sensors, actuators and other equipment in the physical entity, and the digital twin, and vice versa. The digital twin should be able to interact with both historical and real-time data and thus data storage is key (Lu et al., 2020; Wanasinghe et al., 2020). Access to data is complicated with different stakeholders in place. The ownership and sharing principles must be established (Rasheed et al., 2020; Wanasinghe et al., 2020).

On the technical side of it, the main challenge is to implement the digital twin in such a way that the communication between the digital twin and the physical asset can run near real-time. There are three main communication pathways that have to be established: (1) between the digital twin and the real system, (2) between the digital twin and possible other digital twins in the system, and (3) between the digital twin and the operators. This requires good computational capacity, sufficient bandwidth, and a standardized information and communication platform or system (Schleich et al., 2017; Barricelli et al., 2019).

High fidelity and accurate models of the WRRF is needed to produce reliable information. The models can either be physical-based, mechanistical models, like WRRFs have traditionally been modelled, or data-driven. Data-driven models are easy to develop but as they are based on historical data obtained at the WRRF the model will not be able to forecast events or operational states outside of the dataset used to develop the model. Another drawback is that the output may be hard to interpret since it is based solely on data and not any physical, known relations between process parameters (Martinez et al., 2018). The use of hybrid models can make use of the benefits of both methods (Gernaey et al., 2004). The models are to be run at, or close to, real-time which put high demands on both the computational and communication capacities.

Introducing digital twins in an existing company might require a change in both the business model and the current work practice. The user acceptance for new tools is generally low and the operators tend to stick to familiar work practice (Eerikäinen et al., 2020). It is important that the digital twin is useable, functional and easy to maintain in order to have the operators use the digital twin. The twin must be trusted, i.e. the results must be accurate, and updated as the process or WRRF is changed (Fuller et al., 2020; Wanasinghe et al., 2020). Transparent and interpretable digital twins is important and should be considered when choosing models (Rasheed et al., 2020; Martinez et al., 2018). Understanding the operators and their requirements are essential to achieve this (Wanashinghe et al., 2020). It is also important to clearly state the scope and focus of the digital twin, define applications, and thereby set reasonable expectations on what the digital twin can achieve (Cameron et al., 2018; Fuller et al., 2020).

Other issues to attend include economical costs, such as the cost of development of the digital twin, investments in new equipment (sensors, IT systems etc.) and maintenance costs (Barricelli et al., 2019). From the operator or plant owner's perspective, the benefits of introducing new digital tools are not clear enough to overcome the costs for development and implementation (Eerkäinen et al., 2020). Whereas the economic benefits are easy to evaluate and visualize in industries, WRRFs have less incentives to optimize the treatment plant, and thus the economic benefits are less obvious (Eerikäinen et al., 2020). Another aspect is that WRRFs in general are in a quite early stage of digitalization. A general hinder for full-scale implementation of digital twins is that in-house implementation of digital techniques is insufficient (Uhlemann et al., 2017).

To have functional digital twins, there are many different digital systems that must be connected or developed. In each stage, cyber security must be addressed. The systems must be robust and secure from hacking and viruses (Baricelli et al., 2019; Wanasinghe et al., 2020; Fuller et al., 2020).

6.2. Key enabling factors and technologies

The core of developing and using digital twins is data. However, data alone will not be sufficient. One must go from *data*, to *information*, to *knowledge* to '*intelligence'*. While *data* is retrieved by quantitative or qualitative measurements from sensors or equipment in the WRRF, *information* is what can be obtained from the data through data analysis. Piecing together different information and insight to the mechanisms and processes in the WRRF gives *knowledge*. When the knowledge is used to develop new ideas and perspectives, it is commonly said to be '*intelligent'*, albeit that the term intelligent can be somewhat misleading (Therrien et al., 2020).

Data is collected from sensors, actuators and other process equipment, as well as from laboratory tests. As the data sources may have different interfaces and supported protocols, means to gather the data is needed to increase the connectivity (Tao et al., 2018). Internet of Things (IoT), or rather Industrial IoT (IIoT), are important technologies to achieve this. IIoT is regarded as a key enabler for digitalization within the industry as it enables the connectivity between devices and thereby facilitating data collection, exchange and analysis (Malakuti & Gruener, 2018).

The gathered data is typically noisy and must be pre-processed to detract outliers or detect anomalies in the data set before using it (Wanasinghe et al., 2020). The collected data require storage and the data (both historical and real-time data) must be well organized and easily accessible to ensure fast communication. IT systems, servers and communication protocols are needed. The large amount of real-time communication data gives rise to a need of data compression, data fusion, data encapsulation, data mapping technologies and communication technologies (Rasheed et al., 2020; Tao et al., 2018, Wanasinghe et al., 2020). Simple and open communication protocols and formats facilitate implementation in existing systems (Barthelemey et al., 2019).

Data must be stored in a secure way, protected from cyber-attacks. Unauthorized data access and modifications must be prevented (Wanasinghe et al., 2020). Edge, cloud, or fog computing might facilitate communication and storage. With cloud computing, files and data can be performed on the Internet instead of on physical servers. Edge computing might be more ideal when it comes to pre-processing the collected data as the computation would be made where the data is gathered and thereby lowering latency and required bandwidth (Al-Ali et al., 2020).

The data is used to monitor the WRRF, detect faults or anomalies in the WRRF, as well as to develop models and keep them up to date. Models can be physical-based, data-driven or hybrid models. Besides the process and plant models, models for fault detection, predictive maintenance, evaluation and validation is needed within the digital twin depending on pre-defined applications. Different techniques and research fields are needed depending on what sort of models are to be implemented. Simplified models of lower order or data-driven models may reduce the computational demands but with the drawback of not being as easy to interpret as traditional physical-based models (Martniez et al., 2018). Having a digital twin capable of predicting the future and realize the current state of the plant demands high-fidelity models and simulators with accurate input data (Rasheed et al., 2020). High-performance and large-scale computation is needed. The computational infrastructure must be considered. The use of cloud, fog and edge computing might be feasible in some cases (Tao et al., 2018; Rasheed et al., 2020).

Model development and continuous updating (or recalibration) can be facilitated by the use of computer and data science technologies, specifically Artificial Intelligence (AI) and its subsets like Machine learning, Reinforcement learning and Neural networks (Shao & Kibira, 2019; Rasheed et al., 2020). Some may argue that the use of AI, IoT and IIoT technologies are key enablers for digital twins (Fuller et al., 2020). Creating an autonomous digital twin that can make decisions on its own, continuously monitor and analyze the condition of the physical entity, predict maintenance, and detect faults requires AI (Autiosalo et al., 2020). While the use of AI may facilitate modelling and computation, it is important that the output is interpretable by the operators (Rasheed et al., 2020). Model development and interpretation of the digital twin and its performance require good understanding of the processes, and knowledge in control and automation engineering.

To have a functional and interpretable digital twin, which is easy to use and of value to the operators, the human machine interface (HMI) and good ways of visualizing results from the digital twin is needed (Autiosalo et al., 2020). Digital platforms for data management, privacy and security, and quality is important as well (Rasheed et al., 2020).

Key enablers

- Communication platforms and protocols
- IT infrastructure
- Cyber security
- IoT, IIoT
- Big data technologies
- Storage capacity
- Cloud, edge or fog computing
- Modelling and simulation tools/platforms
- Data science
- Computer science
- Artificial intelligence (Machine learning, Deep learning)
- Visualization and graphical user interface
- Transparency and interpretability

6.3. Possible applications, value and usability of digital twins in WRRFs

There are probably endless possibilities to what can be achieved with digital twins. A digital twin can monitor, control and optimize processes and functions in the physical object and continuously predict future scenarios (Barricelli et al., 2019). Applications such as real-time monitoring, asset management, predictive maintenance, fault detection, state monitoring, and virtual testing, as well as traditional WRRF process model applications like optimization, control, planning, evaluation, and automation are all thought to be achievable by the use of digital twins, all of which contribute to a robust, resource effective and optimized operation of WRRFs (Liu et al., 2020; He et al., 2019). However, to create a comprehensive and usable digital twin, easy to maintain and that produces reliable and interpretable results, the scope and focus of the digital twin must be well-defined. The digital twin should be made as simple as possible without compromising on quality (Cameron et al., 2018).

6.3.1. Fault detection

Sensor faults, actuator faults, and process disturbances commonly occurs at WRRFs. As faults or unexpected events can cause process irregularities in WRRFs and thereby influencing the treatment process or equipment it is important to incorporate fault detection to achieve effective operation and to shorten downtime (Newhart et al., 2019; Samuelsson et al., 2017). Methods for fault detection include traditional statistical methods, such as principal component analysis (PCA) and partial least squares, as well as more complex machine learning algorithms such as deep learning or reinforcement learning (Zhang et al., 2018; Newhart et al., 2019; Mamandipoor et al., 2020).

6.3.2. Predictive maintenance

The digital twin makes it possible to retrieve accurate and up to date estimations of the status of the process and the equipment. This opens up possibilities to use the digital twin for predictive maintenance (Avialiotis et al., 2019). Predictive maintenance ranges from detecting anomalies (fault detection) to prognosis of the lifetime of the plant and its equipment (Barthelemey et al., 2019). The use of artificial intelligence (AI) algorithms is a facilitator for this application but requires lots of labelled data. While faults or anomalies are possible to detect from the data with data analysis, more sophisticated methods are required to determine the cause and effect of the anomaly (Barthelemey et al., 2019).

6.3.3. Soft sensors

Parameters are not always easy to measure in a reliable way either due to physical constraints of the WRRF or the lack of reliable sensor data. By the use of data and data science algorithms, such as machine learning or deep learning algorithms, it is possible to create soft sensors that facilitate monitoring of parameters that are hard or impossible to measure (Haimi et al., 2013; Wu et al., 2019, Yan et al., 2017).

6.3.4. Prediction, forecasting and planning

The digital twin can be used to predict performance to either improve the operation of the treatment plant or to inform the operator about expected poor performance (Gardner et al., 2020). Making the digital twin self-adapting, using AI algorithms or Real-Time

Optimization, so that it learns from the data and the process over time, it can adapt to unforeseen events and make decisions (Autiosalo et al., 2020; Gardner et al., 2020; Matias & Jäschke, 2020). The digital twin can either make decisions on its own or, send alarms, stop operation, and provide insight for the operators to aid them in decision making (Autiosalo et al., 2020). It is possible to couple the plant digital twin to prediction models of the sewer network to predict incoming flows and thereby give early warnings to the operators so that either the digital twin or the operator actively can control the plant based on predicted events (Zhang et al., 2019). It is also possible to use the digital twin in a traditional process modelling way to simulate and analyze future scenarios.

6.3.5. Control and Optimization

Optimization, control, and automation are traditional applications of and areas where WRRF process model are used. Model predictive control (MPC) have been used since the 1980s and is still an important use of process models (Stentoft et al., 2019). Inclusion of a control system model in the digital twin makes it possible to develop and evaluate control strategies on online data before implementation in the real system (Wanasinghe et al., 2020). In a recent study by Lindblom and Samuelsson (2020) a digital twin was used to conduct a virtual acceptance test of new control system code for the process under construction at Henriksdal WRRF in Stockholm. The results are promising, and the digital twin was also judged to be feasible for training of operational staff.

6.3.6. Operator's training

The digital twin can increase the operators' knowledge of the treatment process and be used by operators to train or test future scenarios. Possible applications also include risk assessments and emergency response training (Bevilaqua et al., 2020; Wanasinghe et al., 2020).

7. Conclusions

This work was done to describe how digital twins have been implemented previously, both in WRRFs and in other relevant industries, and to identify challenges as well as key enabling factors and technologies for full-scale implementation of digital twins. The concept 'digital twin' lacks an unambiguous definition which is limiting for the research area. A unifying definition seems hard to reach which is why the following definition, specifically for digital twins of WRRFs, was used:

> *A digital twin for smart operation of WRRFs is defined as a combination of different models that are used in real-time for collection, control, validation, simulation and visualization of data to aid operators in decision making. The model parameters are continuously updated and the models are automatically calibrated.*

The use of digital twins has great potential and may well be the future of process plants, including WRRFs. Applications like real-time monitoring, predictive maintenance, fault detection, and virtual testing amongst others offers benefits for the operators. As an end result, the use of digital twins is expected to increase efficiency, reduce cost and energy consumption, and lower the overall environmental impact, as well as help operators in their day to day work. Challenges for full-scale implementation of digital twins should be addressed early to mitigate them. Set aside technical challenges such as IT infrastructure, cyber security, data storage etc., a main concern is to develop a digital twin that is comprehensive and usable. To make the digital twin the powerful tool it has the potential to be, and a tool that adds value for the operators, the scope of it and the applications to develop it for should be determined together or in close collaboration with the end user.

8. References

ABB (2021). *Digital twin applications*. https://new.abb.com/control-systems/features/digital-twin-applications [2021-02-23]

Al-Ali, A. R., Gupta, R., Batool, T. Z., Landolsi, T., Aloul, F. & Al Nabulsi, A. (2020). Digital twin conceptual model within the context of Internet of Things. *Future Internet*, vol. 12, p. 163. doi: 10.3390/fi12100163

Alcáer, V. & Cruz-Machado, V. (2019). Scanning the industry 4.0: a literature review on technologies for manufacturing systems. *Engineering Science and Technology, an International Journal*, vol. 22, pp. 899-919. doi: 10.1016/j.jestch.2019.01.006

Autiosalo, J., Vepsäläinen, J., Viitala, R. & Tammi, K. (2020). A feature-based framework for structuing industrial digital twins. *IEEE Access*, vol. 8, pp. 1193-1208. doi: 10.1109/ACCESS.2019.2950507

Avialiotis, P., Georgoulias, K., Arkouli, Z. & Makris, S. (2019). Methodology for enabling digital twin using advanced physics-based modelling in predictive maintenance. *Procedia CIRP*, vol. 81, pp. 417-422. doi: 10.1016/j.procir.2019.03.072

Barthelemey, A., Lee, E., Hana, R. & Deuse, J. (2019). Dynamic digital twin for predictive maintenance in flexible production systems. *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, Lisbon, Portugal, 2019, pp. 4209-4214, doi: 10.1109/IECON.2019.8927397

Barricelli, B. R., Cashiraghi, E. & Fogli, D. (2019). A survey on digital twin: definitions, characteristics, applications, and design implications. *IEEE Access*, vol. 7, pp. 167653-167671. doi: 10.1109/ACCESS.2019.2953499.

Bevilacqua M., Bottani, E., Ciarapica F.E., Costantino, F., Di Donato, L., Ferraro, A., Mazzuto, G., Monteriù, A., Nardini, G., Ortenzi, M., Paroncini, M., Pirozzi, M., Prist, M., Quatrini, E., Tronci, M., and Vignali, G. (2020). Digital twin reference model development to prevent operators' risk in process plants. *Sustainability*, vol. 12, pp. 1088. doi: 10.3390/su12031088

Blochwitz, T., Otter, M., Åkesson, J., Arnold, M., Clauss, C., Elmqvist, H., Friedrich, M., Junghanns, A., Mauss, J., Neumerkel, D., Olsson, H. & Viel, A. (2012). Functional Mockup Interface 2.0: the standard for tool independent exchange of simulation models. In *Proceedings of the 9th International Modelica Conference* (pp. 173-184). The Modelica Association. https://doi.org/10.3384/ecp12076173

Bottani, E., Cammardella, A., Murino, T. & Vespoli, S. (2017). From the cyber-physical system to the digital twin: the process development for behavior modelling of a cyber guided vehicle in M2M logic. *XXII Summer School "Francesco Turco" – Industrial Systems Engineering*. September 13 2017, Palermo, Italy.

Cameron, D. B., Waaler, A. & Komulainen, T. M. (2018). Oil and gas digital twins after twenty years. How can they be made sustainable, maintainable and useful?. *Linköping Electronic Conference Proceedings*, vol. 153, pp. 9-16. *Proceedings of The 59th Conference* *on Simulation and Modelling (SIMS 59)*, September 26-28 2018, Oslo Metropolitan University, Oslo, Norway. doi: 10.3384/ecp181539

CIO Sweden (2020). *Digital tvilling optimerar ångproduktionen i pappersbruket.* https://cio.idg.se/2.1782/1.742405/digital-tvilling-pappersbruk-sca [2020-11-24]

Digital Water City (n.d.). *Sewer flow forecast toolbox.* https://www.digital-water.city/city/copenhagen/ [2020-09-02]

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 Science and Technology*, vol. 82(12), pp. 2681–2690. doi: 10.2166/wst.2020.311

Eyre, J. M., Dodd, T. J., Freeman, C., Lanyon-Hogg, R., Lockwood., A. J. & Scott, R. W. (2018). Demonstration of an industrial framework for an implementation of a process digital twin. *Proceedings of the ASME 2018 International Mechanical Engineering Congress and Exposition.* November 9-15 2018, Pittsburgh, PA, USA.

Fuller, A., Fan, Z., Day, C. & Barlow, C. (2020). Digital twin: enabling technologies, challenges and open research. *IEEE Access*, vol. 8, pp. 108952-108971. doi: 10.1109/ACCESS.2020.2998358

Gardner, P., Dal Borgo, M., Ruffini, V., Hughes, A. J., Zhu, Y. & Wagg, D. J. (2020). Towards the development of an operational digital twin. *Vibration*, vol. 3, pp. 235-265. doi: 10.3390/vibration3030018

Gernaey, K. V., van Loodsrecht, M. C. M., Henze, M., Lind, M. & Jörgensen, B. (2004). Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling & Software*, vol. 19, pp. 763-783. doi: 10.1016/j.envsoft.2003.03.005

Glaesson, E. H. & Stargel, D. S. (2012). The digital twin paradigm for future nasa and u.s. Air force vehicles. *Proceedings of the 53rd Structures, Structural Dynamics and Materials Conference*, 23–26 April 2012, Honolulu, HI, USA, pp. 1–14. doi: 10.2514/6.2012-1818

Grieves, M. (2015). Digital twin: manufacturing excellence through virtual factory replication. White paper, available online: https://www.researchgate.net/publication/275211047_Digital_Twin_Manufacturing_Excellence_through_Virtual_Factory_Replication [2021-02-23]

Haimi, H., Mulas, M., Corona, F. & Vahala, R. (2013). Data-derived soft-sensors for biological wastewater treatment plants: an overview. *Environmental Modelling & Software*, vol. 47, pp. 88-107. doi: 10.1016/j.envsoft.2013.05.009

He, R., Chen, G., Dong, C., Sun, S. & Shen, X. (2019). Data-driven digital twin technology for optimized control in process systems. *ISA Transactions*, vol. 95, pp. 221-234. doi: 10.1016/j.isatra.2019.05.011

Kritzinger , W., Karner, M., Traar, G., Henjes, J. & Sihn, W. (2018). Digital twin in manufacturing: a categorical literature review and classification. *IFAC PapersOnLine*, vol. 51(11), pp. 1016-1022. doi: 10.1016/j.ifacol.2018.08.474

Lindblom, E. & Samuelsson, O. (2020). *Virtuell driftsättning avstyrsystem på reningsverk*. (B 2399). Stockholm: IVL Swedish Environmental Institute. ISBN 978-91-7883-230-9.

Liu, M., Fang, S., Dong, H. & Xu, C. (2020). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*. doi: 10.1016/j.jmsy.2020.06.017

Lu, Y., Liu, C., Wang, K., Huang, H., Xu, X. (2020). Digital twin-driven smart manufacturing: connotation, reference model, applications and research issues. *Robotics and Computer Integrated Manufacturing*, vol. 61, pp. 101837. doi: 10.1016/j.rcim.2019.101837

Makarov, V. L., Bakhtizin, A. R. & Beklaryan, G. L. (2019). Developing digital twins for production enterprises. *Business Informatics*, vol. 13(4), pp. 7-16. doi: 10.17323/1998-0663.2019.4.7.16

Malakuti, S. & Gruener, S. (2018). Architectural aspects of digital twins in iiot systems. In *Proceedings of the 12th European Conference on Software Architecture: Companion Proceedings (ECSA '18)*, September 24-28 2018, Madrid, Spain. doi: 10.1145/3241403.3241417

Mamandipoor, B., Majd, M., Sheikhalishahi, S., Modena, C. & Osmani, V. (2020). Monitoring and detecting faults in wastewater treatment plants using deep learning. *Environment Monitoring and Assessment*, vol. 192(2), pp. 148-159. doi: 10.1007/s10661-020-8064-1

Martinez, G. S., Sierla, S., Karhela, T. & Vyatkin, V. (2018). Automatic generation of a simulation-based digital twin of an industrial process plant. *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, 2018, Washington, DC, USA. pp. 3084-3089, doi: 10.1109/IECON.2018.8591464.

Matias, J. & Jäschke, J. (2020). Online model maintenance in real-time optimization methods. *Computers and Chemical Engineering*, vol. 145 p. 107141. doi: 10.1016/j.compchemeng.2020.107141

Negri E., Fumagalli, L. & Macchi, M. (2017). A review of the roles of Digital Twin in CPS-based production systems. *Procedia Manufacturing, 11. 27th International Conference on Flexible Automation and Intelligent Manufacturing.* June 27-30 2017, Modena, Italy. pp. 939-948

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*, vol. 157, pp. 498-513. doi: 10.1016/j.watres.2019.03.030

Ny teknik (2018). *Den digitala tvillingen avslöjade läckaget i kärnkraftverket.* https://www.nyteknik.se/energi/den-digitala-tvillingen-avslojade-lackaget-i-karnkraftverket-6898789 [2020-11-24]

OPC Foundation (2021). *Unified Architecture.* https://opcfoundation.org/about/opc-technologies/opc-ua/ [2021-02-23]

Pantelides, C. C. & Renfro, J. G. (2013). The online use of first-principles models in process operations: Review, current status and future needs. *Computers and Chemical Engineering*, vol. 50, pp. 136-148. doi: 10.1016/j.compchemeng.2012.07.008

Qi, Q. & Tao, F. (2018). Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. *IEEE Access,* vol. 6, pp. 3585-3593. doi: 10.1109/ACCESS.2018.2793265.

Rasheed, A., San, O. & Kvamsdal, T. (2020). Digital twin: values, challenges and enablers from a modeling perspective. *IEEE Access*, vol. 8, pp. 21980-22012. doi: 10.1109/ACCESS.2020.2970143

Rosen, R., von Wichert, G., Lo, G. & Bettenhausen, K. D. (2015). About the importance of automomy and digital twins for the future of manufacturing. *IFAC PapersOnLine*, vol. 48(3), pp. 567-572. doi: 10.1016/j.ifacol.2015.06.141

Samuelsson, O., Björk, A., Zambrano, J. & Carlsson, B. (2017). Gaussian process regression for monitoring and fault detection of wastewater treatment processes. *Water Science and Technology*, vol. 75(12), pp. 2952-2963. doi: 10.2166/wst.2017.162

Sarni, W., White, C., Webb, R., Cross, K. & Glotzbach, R. (2019). *Digital Water – Industry leaders chart the transformation journey*. (White paper). London: International Water Association. Accessible online: https://iwa-network.org/wp-content/uploads/2019/06/IWA_2019_Digital_Water_Report.pdf [2020-11-10]

Schleich, B., Anwer, N., Mathieu, L. & Wartzack, S. (2018). Shaping the digital twin for design and production engineering. *CIRP Annals – Manufacturing Technology*, vol. 66, pp. 141-144. doi: 10.1016/j.cirp.2017.04.040

Schröder, G. N., Steinmetz, C., Pereira, C. E. & Espindola, D. B. (2016). Digital twin data modeling with automationml and a communication methodology for data exchange. *IFAC PapersOnLine*, vol. 49(30), pp. 12-17. doi: 10.1016/j.ifacol.2016.11.115

Shao, G. & Kibira, D. (2019). Digital manufacturing: requirements and challenges implementing digital surrogates: Rabe, M., Juan, A.A., Mustafee, N., Skoogh, A., Jain, S. & Johansson, B. (eds.). *Proceedings of the 2018 Winter Simulation Conference 2018*, December 9-12 2018, Gothenburg, Sweden, pp. 1226-1237, doi: 10.1109/WSC.2018.8632242.

Siemens (2021). *Digital twins*. https://new.siemens.com/global/en/company/stories/research-technologies/digitaltwin/digital-twin.html [2020-02-23]

Stentoft, P. A., Guericke, D., Munk-Nielsen, T., Mikkelsen, P. S., Madsen, H., Vezzaro, L. & Möller, J.K. (2019). Model predictive control of stochastic wastewater treatment process for smart power, cost-effective aeration. *IFAC-PapersOnLine*, vol. 52(1), pp. 622-627. doi.: 10.1016/j.ifacol.2019.06.132

Sun, C., Puig, V. & Cembreno, G. (2020). Real-time control of urban water cycle under cyber-physical systems framework. *Water*, vol. 12, p. 406. doi: 10.3390/w12020406

Sweden Water Research (2020). *Future city flow*. https://www.swedenwaterresearch.se/en/projekt/future-city-flow-3/ [2020-09-02]

Söderberg, R., Wärmefjord, K., Carlsson, J. S. & Lindkvist, L. (2017). Toward a digital twin for real-time geometry assurance in individualized production. *CIRP Annals – Manufacturing Technology*, vol. 66, pp. 137-140. doi: 10.1016/j.cirp.2017.04.038

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H. & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, vol. 94, pp. 3563–3576. doi: 10.1007/s00170-017-0233-1

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 & Technology*, wst2020393. doi: 10.2166/wst.2020.393

Uhlemann, T. H.-J., Lehmann, C. & Steinhilper, R. (2017). The digital twin: realizing the cyber-physical production system for industry 4.0. *Procedia CIRP*, vol. 61, pp. 335-340. doi: 10.1016/j.procir.2016.11.152

Vachálek, J., Bartalský, L., Rovný, O., Šišmišová, D, Morhac, M. & Loksik, M. (2017). The digital twin of an industrial production line within the industry 4.0 concept. M. Fikar & M. Kvasnica (eds.) *2017 21st International Conference on Process Control (PC)*. June 6-9 2017, Štrbské Pleso, Slovakia, pp. 258-262.

Wanasinghe, T. R., Wroblewski, L., Petersen, B., Gosine, R. G., James, L. A., de Silva, O., Mann, G. K. I. & Warrian, P. J. (2020). Digital twin for the oil and gas industry: overview, research trends, opportunities, and challenges. *IEEA Access*, vol. 8, pp. 104175-104197. doi: 10.1109/ACCESS.2020.2998723

Wu. J., Cheng, H., Lio, Y., Liu, B. & Huang, D. (2019). Modeling of adaptive multi-output soft-sensors with applications in wastewater treatments. *IEEE Access*, vol. 7, pp. 161887-161898. doi: 10.1109/ACCESS.2019.2950034

Yan, W., Tang, D. & Lin, Y. (2017). A data-driven soft sensor modeling method based on deep learning and its application. *IEEE Transactions on Industrial Electronics*, vol. 64(5), pp. 4237-4245. doi: 10.1109/TIE.2016.2622668

Zhang, D., Lin, Z. & Gao, Z. (2018). A novel fault detection with minimizing the noise-signal ratio using reinforcement learning. *Sensors*, vol. 18, pp. 3087-3107. doi: 10.3390/s18093087

Zhang, Q., Li, Z., Snowling, S., Siam, A. & El-Dakhakhni, W. (2019). Predictive models for wastewater flow forecasting based on time series analysis and artificial neural network. *Water Science and Technology*, vol. 80(2), pp. 243-253. doi: 10.2166/wst.2019.263

Zhao, P., Liu, J., Tang, M., Sheng, S., Zhou, H. & Lio, X. (2020). The modeling and using strategy for the digital twin in process planning. *IEEE Access*, vol. 8, pp. 41229-41245. doi: 10.1109/ACCESS.2020.2974241.

Zheng, Y., Yang, S. & Cheng, H. (2019). An application framework of digital twin and its case study. *Journal of Ambient Intelligence and Humaized Computing*, vol. 10, pp. 1141-1153. doi: 10.1007/s12652-018-0911-3

Äijälä, G. & Lumley, D. (2006). Integrated soft sensor model for flow control. *Water Science and Technology*, vol. 4(5), pp. 473-482. doi: 10.2166/wst.2006.152

Övern, A. (2018). *Industry 4.0 – Digital Twins and OPC UA*. (Master thesis). Norwegian University of Science and Technology. Department of Mechanical and Industrial Engineering. http://hdl.handle.net/11250/2561319